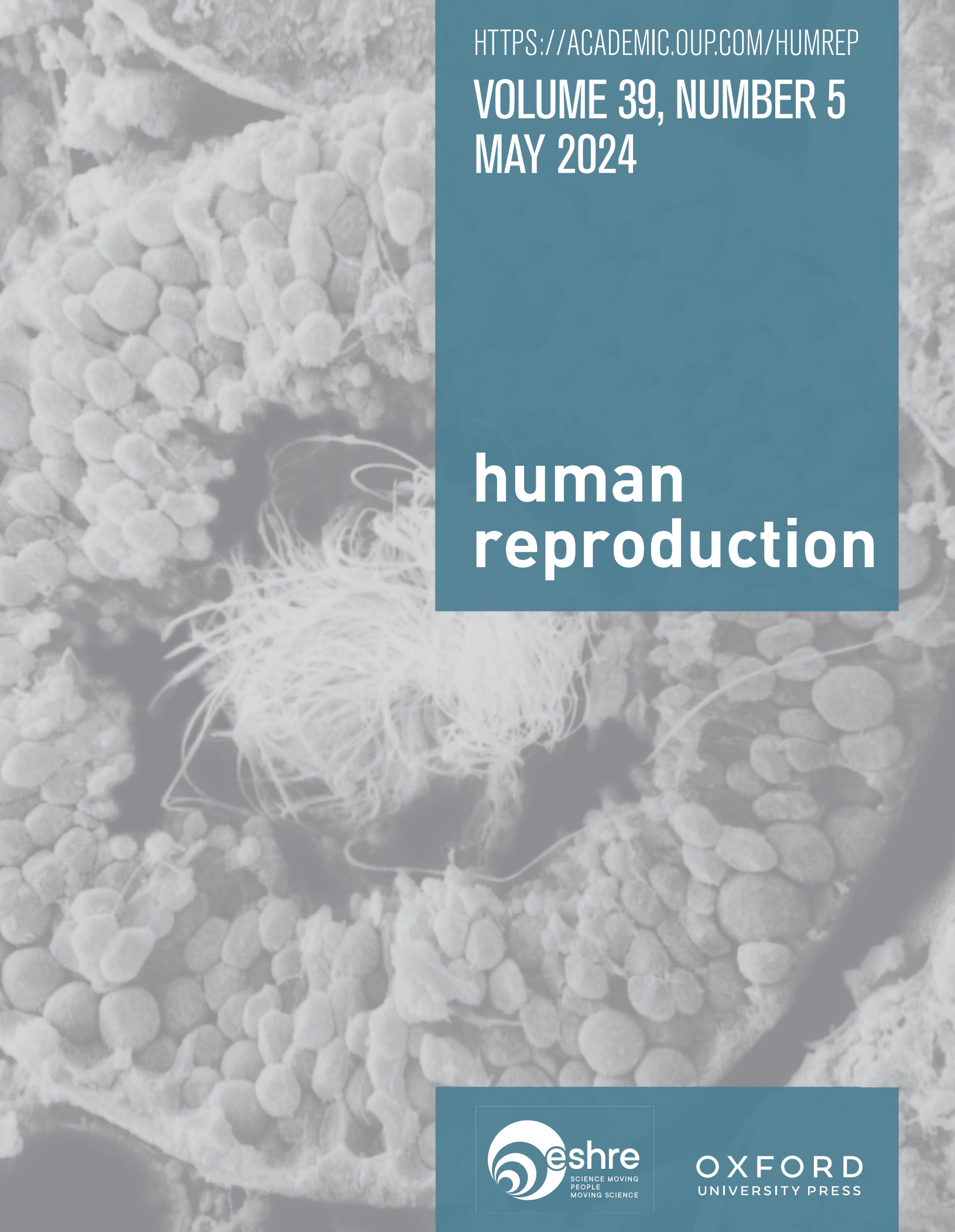




[HTTPS://ACADEMIC.OUP.COM/HUMREP](https://academic.oup.com/humrep)

VOLUME 39, NUMBER 5
MAY 2024

human reproduction



OXFORD
UNIVERSITY PRESS

human reproduction

Editor-in-Chief

Professor Cornelis (Nils) Lambalk

Deputy Editors

D. Sakkas R. Wang L. Wise

Commissioning Editor

K. Kirkegaard

Associate Editors

J. Acierno, *Switzerland & USA*
L. Aghajanova, *USA*
J. Aitken, *Australia*
G. Benagiano, *Italy*
A. Beristain, *Canada*
E. Bertone-Johnson, *USA*
A. Boggio, *USA*
L. Cacciottola, *Belgium*
A. Campbell, *UK*
A. Capalbo, *Italy*
E. Caroppo, *Italy*
D. Christopikou, *Greece*
D. Cimadomo, *Italy*
M.G. Da Broi, *Brazil*
M. De Vos, *Belgium*
A. Diez-Juan, *Spain*
A. Dolinko, *USA*
A. Domar, *USA*
G.L. Dunietz, *USA*
A. Dwyer, *USA*

S. Esteves, *Brazil*
F. Facchin, *Italy*
T. Freour, *France*
S. Gameiro, *UK*
J. Grosbois, *Belgium*
E. Hadar, *Israel*
B.P. Hermann, *USA*
L. Kahn, *USA*
K. Kim, *USA*
J. Knez, *Slovenia*
T. Kohn, *USA*
A. La Marca, *Italy*
T. Laisk-Podar, *Estonia*
M. Lambertini, *Italy*
G. Lambert-Messerlian, *USA*
T. Lanisnik Rizner, *Slovenia*
K. Lattes, *Spain*
S. Laurentino, *Germany*
S. Lensen, *Australia*
Y. Liu, *Australia*

Y. Louwers, *Netherlands*
M. Lydic, *USA*
R. Machtinger, *Israel*
M. Magnus, *Norway*
A. Mantzouratou, *UK*
W. Martins, *Brazil*
S. Martins da Silva, *UK*
H. Mbarek, *Netherlands*
V. Mijatovic, *Netherlands*
M.G. Minasi, *Italy*
D. Modi, *India*
M. Moravek, *USA*
S. Mumusoglu, *Turkey*
M. Munro, *USA*
N. Neuhaus, *Germany*
H. Niemann, *Germany*
C. O'Flaherty, *Canada*
L. Priskorn, *Denmark*
G. Quinn, *USA*
C. Racowsky, *France*

J. Ramalho-Santos, *Portugal*
R. Rey, *Argentina*
A. Romano, *Netherlands*
K. Ruth, *UK*
W. Salem, *USA*
S. Santos-Ribeiro, *Portugal*
I. Sfontouris, *Greece*
C. Simintras, *USA*
J. Subirá, *Spain*
A. Torre, *UK*
R. van Eekelen, *Netherlands*
D. Vaughan, *USA*
M. Vazquez-Levin, *Argentina*
P. Vercellini, *Italy*
R. Wang, *Australia*
A. Wesselink, *USA*
J. Woodward, *USA*

Statistical Advisory Board

O. Basso (Canada) S. Roberts (UK) S. Missmer (USA) C. Venetis (Greece)

Founding Editor

R. Edwards

Editors Emeriti

D.H. Barlow P. Crosignani C. De Geyter J.L.H. Evers R. Sharpe E. Somigliana A. Van Steirteghem M. van Wely

Managing Editor

A.C. Williams

Associate Managing Editors

J.M. Hastings and E.J. Andrew

Editorial Administrator

R. Madison

Editorial Office

E-mail: editorial@humanreproduction.co.uk

Published for the
European Society of Human Reproduction and Embryology
by Oxford University Press,
Oxford, UK

SUBSCRIPTIONS

A subscription to *Human Reproduction* comprises twelve issues. Human Reproduction Advance Articles contains papers that have been finalized, but have not yet been included within an issue.

Subscription and single issue pricing information is available at <https://academic.oup.com/humrep/subscribe>.

ESHRE members

Application forms for ESHRE membership can be obtained from ESHRE Central Office, Meerstraat 60, 1852 Grimbergen (Beigem), Belgium. Tel: +32-(0)2-263 64 63; Fax: +32-(0)2-269 56 00, E-mail: membership@eshre.eu. Webaddress: <http://www.eshre.eu/membership>. Please note that the ESHRE subscription rate is only available to ESHRE members.

Full prepayment, in the correct currency, is required for all orders. Orders are regarded as firm and payments are not refundable. Subscriptions are accepted and entered on a complete volume basis. Claims cannot be considered more than FOUR months after publication or date of order, whichever is later. All subscriptions in Canada are subject to GST. Subscriptions in the EU may be subject to European VAT. If registered, please supply details to avoid unnecessary charges. For subscriptions that include online versions, a proportion of the subscription price may be subject to UK VAT. Personal rate subscriptions are only available if payment is made by personal credit card and delivery is to a private address.

For further information, please contact: Journals Customer Service Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. E-mail: jnls.cust.serv@oup.com. Tel (and answerphone outside normal working hours): +44 (0) 1865 353907. **In the US, please contact:** Journals Customer Service Department, Oxford University Press, 4000 CentreGreen Way, Suite 310, Cary, NC 27513, USA. E-mail: jnlorders@oup.com. Tel (and answerphone outside normal working hours): 800 852 7323 (toll-free in USA/Canada). **In Japan, please contact:** Journals Customer Services, Oxford University Press, 4-5-10-8 F, Shiba, Minato-ku, Tokyo, 108-8386, Japan. E-mail: custserv.jp@oup.com. Tel: +81 (0)3 5444 5858.

Methods of payment. (i) Bank transfer to Barclays Bank Plc, Oxford Group Office, Oxford (bank sort code 20-65-18) (UK), overseas only Swift code BARC GB 22 (GB£ Sterling to account no. 70299332, IBAN GB89BARC20651870299332; US\$ Dollars to account no. 66014600, IBAN GB27BARC20651866014600; EU€ Euros to account no. 78923655, IBAN GB16BARC20651878923655).

(ii) Credit card (Mastercard, Visa, Switch or American Express).

Human Reproduction is published monthly (ISSN 1460-2350) by Oxford University Press, Oxford, UK. Annual subscription price is available at <https://academic.oup.com/humrep/subscribe>.

Instructions to Authors

For instructions on how to submit papers to *Human Reproduction*, please refer to the Journal website at: <https://academic.oup.com/humrep>.

Supplements, reprints and corporate sales

For requests from industry and companies regarding supplements, bulk article reprints, sponsored subscriptions, translation opportunities for previously published material, and corporate online opportunities, please email [special.sales@oup.com](mailto:sales@oup.com) or visit <http://oupmediainfo.com/#1/mediakit/human-reproduction>.

Permissions

For information on how to request permissions to reproduce article/information from this journal, please visit https://academic.oup.com/journals/pages/access_purchase/rights_and_permissions.

Advertising

Advertising, inserts and artwork enquiries should be addressed to Advertising and Special Sales, Oxford Journals, Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK. Tel: +44 (0) 1865 354767; Fax: +44 (0) 1865 353774; E-mail: jnlsadvertising@oup.com.

Material Disclaimer

The opinions expressed in *Human Reproduction* are those of the authors and contributors, and do not necessarily reflect those of the European Society of Human Reproduction and Embryology, the editors, the editorial board, Oxford University Press or the organization to which the authors are affiliated.

Drug Disclaimer

The mention of trade names, commercial products or organizations, and the inclusion of advertisements in the journals does not imply endorsement by the European Society of Human Reproduction and Embryology, the editors, the editorial board, Oxford University Press or the organization to which the authors are affiliated. The editors and publishers have taken all reasonable precautions to verify drug names and doses, the results of experimental work and clinical findings published in the journal. The ultimate responsibility for the use and dosage of drugs mentioned in the Journal and in interpretation of published material lies with the medical practitioner, and the editors and publishers cannot accept liability for damages arising from any errors or omissions in the journal. Please inform the editors of any errors.

© European Society of Human Reproduction and Embryology 2024

All rights reserved; no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior written permission of the Publishers, or a licence permitting restricted copying issued in the UK by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1P 9HE, or in the USA by the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For those in the USA or Canada not registered with CCC, articles can be obtained by fax in 48 hours by calling: WISE for Medicine™ 1-800-667-WISE.

Typeset by KnowledgeGlobal Works Limited, Chennai, India.



This journal is a member of the Committee on Publication Ethics (COPE).



human reproduction

Volume 39, Number 5 May 2024

<https://academic.oup.com/humrep>

OPINION

Estimating the impact of bias in causal epidemiological studies: the case of health outcomes following assisted reproduction	A.R. Walker, C.A. Venetis, S. Opdahl, G.M. Chambers, L.R. Jorm, and C.M. Vajdic	869

INVITED COMMENTARIES

Evolution of evidence-based diagnostic criteria in adolescents with polycystic ovary syndrome	H.J. Teede, A.C.H. Neven, and A. Pena	876
Progestin-primed ovarian stimulation and aneuploidy, innocent until proven guilty	B. Ata	878

ORIGINAL ARTICLES

Andrology

Deletion of <i>ACTRT1</i> is associated with male infertility as sperm acrosomal ultrastructural defects and fertilization failure in human	Q. Zhang, H. Jin, S. Long, X. Tang, J. Li, W. Liu, W. Han, H. Liao, T. Fu, G. Huang, S. Chen, and T. Lin	880
Identification of two hidden clinical subgroups among men with idiopathic cryptozoospermia	L.C. Schülke, J. Wistuba, V. Nordhoff, H.M. Behre, J.-F. Cremers, S. Kliesch, S. Di Persio, and N. Neuhaus	892
Microfluidic sperm sorting selects a subpopulation of high-quality sperm with a higher potential for fertilization	N. Sheibak, F. Amjadi, A. Shamloo, F. Zarei, and Z. Zandieh	902

Early pregnancy

First trimester maternal tryptophan metabolism and embryonic and fetal growth: the Rotterdam Periconceptional Cohort (Predict Study)	S.K.M. van Zundert, N.C.M. van Egmond, L. van Rossem, S.P. Willemsen, P.H. Griffioen, R.H.N. van Schaik, M. Mirzaian, and R.P.M. Steegers-Theunissen	912
Morphologic development of the first-trimester utero-placental vasculature is positively associated with embryonic and fetal growth: the Rotterdam Periconception Cohort	E.S. De Vos, A.G.M.G.J. Mulders, A.H.J. Koning, S.P. Willemsen, M. Rousian, B.B. Van Rijn, E.A.P. Steegers, and R.P.M. Steegers-Theunissen	923

Embryology

- Disposables used cumulatively in routine IVF procedures could display toxicity L. Delaroche, L. Besnard, V. Ouary, F. Bazin, and G. Cassuto 936
- Abnormal cleavage up to Day 3 does not compromise live birth and neonatal outcomes of embryos that have achieved full blastulation: a retrospective cohort study T. Lee, K. Peirce, J. Natalwala, V. Chapple, P.J. Mark, K. Sanders, and Y. Liu 955

Infertility

- Risk factors for poor oocyte yield and oocyte immaturity after GnRH agonist triggering S. Gambini, C. Sonigo, G. Robin, I. Cedrin-Durnerin, C. Vinolas, C. Sifer, Y. Boumerdassi, A. Mayeur, V. Gallot, M. Grynberg, and M. Peigné 963
- Does recurrent implantation failure exist? Prevalence and outcomes of five consecutive euploid blastocyst transfers in 123 987 patients P. Gill, B. Ata, A. Aranz, D. Cimadomo, A. Vaiarelli, H.M. Fatemi, F.M. Ubaldi, J.A. Garcia-Velasco, and E. Seli 974
- Economic evaluations of assisted reproductive technologies in high-income countries: a systematic review E. Olive, C. Bull, A. Gordon, M. Davies-Tuck, R. Wang, and E. Callander 981

Psychology and counselling

- An evaluation of sexual function and health-related quality of life following laparoscopic surgery in individuals living with endometriosis L. Van Niekerk, S. Pugh, A. Mikocka-Walus, C. Ng, R. O'Hara, M. Armour, M. Leonardi, and S. Evans 992

Puberty, aging and HRT

- Urinary gonadotropin assay on 24-h collections as a tool to detect early central puberty onset in girls: determination of predictive thresholds C. Janot, P. Perrin, V. Raverot, P. Bretones, R. Ecochard, S. Malburet-Testori, M. Nicolino, Z. Robert, F. Roucher-Boulez, C. Villanueva, K. Perge, and I. Plotton 1003
- Bone health in women with premature ovarian insufficiency/early menopause: a 23-year longitudinal analysis A.R. Jones, J. Enticott, P.R. Ebeling, G.D. Mishra, H.T. Teede, and A.J. Vincent 1013

Reproductive biology

- The impact of ovarian stimulation on the human endometrial microenvironment A. Chemerinski, M. Shen, N. Valero-Pacheco, Q. Zhao, T. Murphy, L. George, A. Lemenze, L. Sherman, D. Heller, X. Chen, T. Wu, P.G. McGovern, S.S. Morelli, R. Arora, A.M. Beaulieu, and N.C. Douglas 1023
- Deciphering the role of PGRMC2 in the human endometrium during the menstrual cycle and *in vitro* decidualization using an *in vitro* approach Y. Medina-Laver, R. Gonzalez-Martin, P. de Castro, I. Diaz-Hernandez, P. Alama, A. Quiñero, A. Palomar, and F. Dominguez 1042
- Targeting osteopontin alleviates endometriosis and inflammation by inhibiting the RhoA/ROS axis and achieves non-invasive *in vitro* detection via menstrual blood H. Wang, B. Wang, M. Wu, J. Lu, and P. Duan 1057

Reproductive endocrinology

- Adolescents diagnosed with polycystic ovary syndrome under the Rotterdam criteria but not meeting the diagnosis under the updated guideline J.J. Kim, K.R. Hwang, D. Lee, S. Kim, and Y.M. Choi 1072
- Resumption of ovulation in anovulatory women with PCOS and obesity is associated with reduction of 11 β -hydroxyandrostenedione concentrations Z. Wang, M. Van Faassen, H. Groen, A.E.P. Cantineau, A. Van Oers, A. Van der Veen, J.M. Hawley, B.G. Keevil, I.P. Kema, and A. Hoek 1078
- Natural proliferative phase frozen embryo transfer—a new approach which may facilitate scheduling without hindering pregnancy outcomes C. Mendes Godinho, S.R. Soares, S.G. Nunes, J.M.M. Martínez, and S. Santos-Ribeiro 1089
- Ovarian response and embryo ploidy following oral micronized progesterone-primed ovarian stimulation versus GnRH antagonist protocol. A prospective study with repeated ovarian stimulation cycles M. del Mar Vidal, F. Martínez, I. Rodríguez, and N.P. Polyzos 1098

Reproductive epidemiology

- A history of recurrent pregnancy loss is associated with increased perinatal complications, but not necessarily a longer birth interval: a population study spanning 18 years K. Cho, B. Fayek, Y.(D.) Liu, A. Albert, E. Wiesenthal, S. Dobrer, F.F. AbdelHafez, S. Lisonkova, and M.A. Bedaiwy 1105

Reproductive genetics

- Ready for polygenic risk scores? An analysis of regulation of preimplantation genetic testing in European countries M. Siermann, V. van der Schoot, E.M. Bunnik, and P. Borry 1117
- Systematic molecular analyses for 115 karyotypically normal men with isolated non-obstructive azoospermia Y. Muranishi, Y. Kobori, Y. Katoh-Fukui, S. Tamaoka, A. Hattori, A. Osaka, H. Okada, K. Nakabayashi, K. Hata, T. Kawai, H. Ogata-Kawata, T. Iwahata, K. Saito, M. Kon, N. Shinohara, and M. Fukami 1131
- Genomic alterations in ovarian endometriosis and subsequently diagnosed ovarian carcinoma A. Linder, S. Westbom-Fremer, C. Mateoiu, A. Olsson Widjaja, T. Österlund, S. Veerla, A. Ståhlberg, B. Ulfenborg, I. Hedenfalk, and K. Sundfeldt 1141

PEER PERSPECTIVES

- Understanding and addressing unexplained infertility: from diagnosis to treatment C. Massarotti, J.J. Fraire-Zamora, G. Liperis, J. Uraji, K. Sharma, M. Serdarogullari, O.F. Ammar, S. Makieva, Z.E. Ali, D. Romualdi, E. Somigliana, D. Sakkas, E. Dancet, and M. Mincheva 1155

CORRECTION

- Correction to: Identification of two hidden clinical subgroups among men with idiopathic cryptozoospermia L.C. Schülke, J. Wistuba, V. Nordhoff, H.M. Behre, J.-F. Cremers, S. Kliesch, S. Di Persio, and N. Neuhaus 1160

Estimating the impact of bias in causal epidemiological studies: the case of health outcomes following assisted reproduction

Adrian R. Walker ^{1,*}, Christos A. Venetis ^{1,2}, Signe Opdahl ^{1,3}, Georgina M. Chambers ⁴, Louisa R. Jorm ¹, and Claire M. Vajdic ⁵

¹Centre for Big Data Research in Health, Faculty of Medicine and Health, UNSW Sydney, Sydney, NSW, Australia

²Unit for Human Reproduction, 1st Department of Obstetrics and Gynaecology, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece

³Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway

⁴National Perinatal Epidemiology and Statistics Unit, Centre for Big Data Research in Health and School of Clinical Medicine, Faculty of Medicine and Health, UNSW Sydney, Sydney, NSW, Australia

⁵Kirby Institute, UNSW Sydney, Sydney, NSW, Australia

*Correspondence address. Centre for Big Data Research in Health, Faculty of Medicine and Health, UNSW Sydney, Sydney, NSW 2052, Australia.

Tel: +61-2-9385-1873; E-mail: adrian.walker@unsw.edu.au  <https://orcid.org/0000-0003-0801-2316>

ABSTRACT

Researchers interested in causal questions must deal with two sources of error: *random error* (random deviation from the true mean value of a distribution), and *bias* (systematic deviance from the true mean value due to extraneous factors). For some causal questions, randomization is not feasible, and observational studies are necessary. Bias poses a substantial threat to the validity of observational research and can have important consequences for health policy developed from the findings. The current piece describes bias and its sources, outlines proposed methods to estimate its impacts in an observational study, and demonstrates how these methods may be used to inform debate on the causal relationship between medically assisted reproduction (MAR) and health outcomes, using cancer as an example. In doing so, we aim to enlighten researchers who work with observational data, especially regarding the health effects of MAR and infertility, on the pitfalls of bias, and how to address them. We hope that, in combination with the provided example, we can convince readers that estimating the impact of bias in causal epidemiologic research is not only important but necessary to inform the development of robust health policy and clinical practice recommendations.

Keywords: statistics / epidemiology / assisted reproduction / ART / medically assisted reproduction / bias

Causal inference and quantifying bias

A central challenge for epidemiology is causal inference, that is how to determine if two events are causally related (Pearl, 2009). Generally, when trying to answer a causal question, researchers must deal with two sources of error: *random error* and *bias* (Hernan and Robins, 2020). Random error describes how, when sampling data from an underlying distribution, the samples will exhibit some random deviation from the true mean value of that distribution. If this deviation is too high, and the number of samples is too low, then true relationships in the data will not be discernible. Bias describes how the observed relationship between two variables differs systematically from the true value owing to factors other than the causal link between the two (Hernan and Robins, 2020). Bias can be split into three components (Lash et al., 2014):

- **Selection bias:** bias introduced through the tendency for some types of participants to be sampled more or less often than others relative to the exposure and outcome.

- **Confounding:** bias introduced through factors that influence both the exposure and the outcome ('confounders'), such that it produces a non-causal association, or masks a true causal association, between the exposure and outcome. Measured confounders are those known and measured, and their impact on the true causal association between exposure and outcome is controlled for in the research design and/or analysis. Unmeasured confounders are those either unknown to the researcher, or known but unmeasured (e.g. when data on a confounder is too difficult to collect).
- **Measurement error:** bias introduced through the incorrect measurement of the outcome, exposure, or any confounding or mediating variables.

Randomized controlled trials (RCTs), though not immune to selection bias, address confounding through randomization of participants to treatment/non-treatment conditions, and can attempt to minimize issues with measurement through methods such as blinding. However, practical and ethical concerns can

Received: October 03, 2023. Revised: February 19, 2024. Editorial decision: February 23, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

often restrict the use of RCTs in research. A prime example of this restriction is the study of medically assisted reproduction (MAR), such as ovulation induction and IVF, and health outcomes. That is, if one is interested in a health outcome that generally takes a long time after treatment to manifest (e.g. dementia), it is not only impractical, but potentially unethical, to ask women not to seek a treatment that increases the likelihood of conception in their childbearing years. When conducting an RCT is not possible, researchers must rely on observational data. As there is no inherent randomization, studies using these data can be subject to substantial bias (Lu, 2009).

Though both random error and bias are important to consider in the design of observational health studies, random error is generally measured through established statistical methods (such as the generation of confidence intervals) (Rothman et al., 2008; Barraza et al., 2019; Fox et al., 2022), while bias is rarely quantified (Petersen et al., 2021). Furthermore, the consequences of random and systematic error are not equivalent. Though both random error and bias can decrease researchers' ability to observe a true causal relationship between an exposure and an outcome, bias can also often lead to erroneously concluding an exposure/outcome relationship exists when it does not, and in extreme cases may reverse the direction of the true causal relationship (Barraza et al., 2019). Therefore, the assessment and, ideally, quantification of bias is an important step when assessing the internal validity of studies and the strength of evidence they produce.

Despite both the importance of understanding sources of bias in observational designs and the availability of various methods to assess such error, estimation of bias remains rare in the epidemiological literature (Petersen et al., 2021). This may be linked to a lack of awareness about when estimating bias is useful, or knowledge of how bias may affect research outcomes. However, even when researchers are aware of potential sources of bias, they may be unaware of the available methods to estimate it. To address this concern, in this article, we provide an overview of two common approaches to estimating the potential impact of bias: quantitative bias analysis; and E-values. We then show how these techniques can help assess the strength of evidence for health outcomes following MAR using cancer risk as an example. In doing so, we hope to convince the reader that such methods are an important step to consider when conducting observational research, particularly in the study of health outcomes following MAR.

Approaches to measuring bias

Substantial work has gone into developing ways to estimate the impact of bias. One approach, known as 'quantitative bias analysis', describes a set of approaches in which the researcher assesses how the relationship and strength of the causal relationship between an exposure variable and an outcome variable may change when one or more biases are explicitly assumed and modelled (Lash et al., 2014; Fox et al., 2022). These methods range from mathematically and computationally straightforward (e.g. simple bias-sensitivity analysis) to highly complex (e.g. multiple bias modelling) (Lash et al., 2014; Fox et al., 2022), and depending on the method chosen can be used to assess the principal types of bias (selection bias, unmeasured confounding, and measurement error) (Lash et al., 2014).

Another method of estimating the impact of bias caused by unmeasured confounding that has been proposed in causal questions is the 'E-value' (VanderWeele and Ding, 2017). Rather than

assessing how the strength of a causal association may change by adjusting for different assumed biases, the E-value provides an absolute measure of the minimum association (in the risk ratio scale) an unmeasured confounder would need to have on both the exposure and the outcome to potentially reduce any observed association to the null, thus 'explaining away' any observed effect. In comparison to quantitative bias analysis, the E-value is easily computed and can be reported for a range of study designs. However, it does not account for either selection bias or measurement error nor does it explicitly consider that there may be multiple interacting unmeasured confounders (Greenland, 2020).

The general equation for calculating an E-value (on the risk ratio scale) is given by

$$E = RR + \sqrt{RR(RR - 1)},$$

where E is the E-value and RR is the risk ratio of the exposure and outcome. To give an example of how the E-value is used, consider a scenario where the risk ratio of an association between an exposure and outcome is 1.50. Using the equation, we calculate our E-value to be 2.37. This E-value tells us that, for this observed association between the exposure and outcome of 1.50 to be reduced to the null, an unmeasured confounder would need to have a causal association on the risk ratio scale of strength at least 2.37 to the exposure and 2.37 to the outcome.

Case study: health outcomes following assisted reproduction and cancer

When the outcomes of research can affect policy and clinical decisions, researchers must be especially considerate of bias affecting their results (Lash et al., 2014). One example of where this research/policy interaction occurs is in the study of the relationship between MAR and later health outcomes. For example, observational studies have linked MAR to cardiovascular disorders, diabetes, thyroid disease, obesity, and cancer (Dayan et al., 2017; Vassard et al., 2018; Murugappan et al., 2021; Cullinane et al., 2022; Yiallourou et al., 2022; Magnus et al., 2023; Sachdev et al., 2023; Wei et al., 2023). Given that more than 1.9 million IVF cycles are performed each year globally, with a similar number of ovulation induction cycles estimated (Chambers et al., 2021), even a slight increase in the risk of poor health outcomes following such treatments may represent a significant burden of disease and concern for patients and clinicians.

To illustrate how bias may impact conclusions about MAR and later health outcomes, we provide an example using the relationship between MAR and cancer. There has been long-standing concern that the use of MAR may increase the risk of certain cancers, particularly hormone-sensitive ovarian, endometrial, and breast cancer (Fishel and Jackson, 1989).

Over the past 30 years, numerous studies in different countries have attempted to elucidate the relationship between MAR and cancer risk by analysing large-scale linked administrative health datasets, with meta-analyses published almost yearly summarizing the association (Gennari et al., 2015; Momenimovahed et al., 2019; Barcroft et al., 2021; Beebejaun et al., 2021; Cullinane et al., 2022). Though recent research has generally suggested limited to no association between hormone-related cancers and MAR (Barcroft et al., 2021), there is still substantial debate on the topic owing to the gravity of the concern for women seeking MAR treatment, the biological plausibility of associations (Fathalla, 1971; Auranen et al., 2005; Huhtaniemi, 2010; Liang and Shang, 2013; Brown and Hankinson, 2015), and the potential for random error

and systematic bias in studies reliant on secondary use of administrative datasets (Momenimovahed *et al.*, 2019).

The reported effect sizes from studies showing a positive association between MAR and cancer are generally small (Barcroft *et al.*, 2021), indicating that even a small amount of bias (such as confounding or selection bias) may affect the outcome of a given study. Furthermore, there are many potential confounders that can affect both exposure to MAR and the development of cancer in women. These can generally be split into four groups: health history (e.g. infertility, anovulation, endometriosis, cancer screening, smoking, excessive alcohol consumption, oral contraceptive use); reproductive history (e.g. parity, age at first/last birth); social factors (e.g. socioeconomic advantage, rurality); and familial history (e.g. genetic predisposition to cancer). Owing to the challenges associated with collecting, accessing, linking, and analysing population-based data, to our knowledge no existing studies have accounted for all known confounders in their analyses (Momenimovahed *et al.*, 2019; Bates, 2021). To illustrate how confounding can cause bias, consider the simple directed acyclic graph (DAG) in Fig. 1 showing the relationship between MAR, endometriosis, and cancer. If we conducted a cohort study, we might find an increased risk of cancer associated with MAR treatment. However, if we did not account for the impact of endometriosis, we could not be sure if this effect is driven by a causal effect of MAR on cancer, or by the fact that those with endometriosis are both more likely to receive MAR treatments and develop certain types of cancer, particularly ovarian cancer. An expanded DAG showing proposed relationships between all mentioned confounders is provided in Supplementary Fig. S1 and justifications for each causal link are provided in Supplementary Table S1.

As well as unmeasured confounding, accurately measuring factors associated with receiving MAR can be difficult. For example, it is likely that infertility is underreported in administrative medical records, as it is estimated that over half of all couples do not seek treatment for infertility (Boivin *et al.*, 2007). Similarly, there may be substantial underreporting of endometriosis in administrative medical records, with one study showing that over 40% of suspected endometriosis cases were not captured in such records (Rowlands *et al.*, 2021). Given the potential for bias, it is not surprising that many meta-analyses on MAR and cancer report a high proportion of total variability caused by between-

study heterogeneity (Gennari *et al.*, 2015; Del Pup *et al.*, 2018; Cullinane *et al.*, 2022).

Quantitative bias analysis

Quantitative bias analysis could be used to assess the impact of this bias by showing the precise change in the effect estimate assuming different strengths of confounder associations between MAR and cancer risk. For example, consider how the inability to control for endometriosis may bias an association between MAR and ovarian cancer in an Australian cohort. To test this association, we could take the following steps:

- Using either expert knowledge or prior literature, estimate the population prevalence of endometriosis (the confounder): we know around 10% of women of childbearing age have endometriosis in Australia (Australian Institute of Health and Welfare, 2023).
- Using prior literature, estimate the risk ratio of endometriosis to ovarian cancer: according to a systematic review-informed meta-analysis, women with endometriosis have a risk ratio of around 1.93 for ovarian cancer compared to women without endometriosis (Kvaskoff *et al.*, 2020).
- Using prior literature, estimate the risk ratio of endometriosis to receiving MAR: one prospective study reported the risk ratio of infertility for women with endometriosis to be around 2.12 (Prescott *et al.*, 2016).
- Examine how much the estimated effect of MAR on cancer would change if we had accounted for confounding between endometriosis, MAR, and ovarian cancer.

A simplified working example of this process with numerical estimates can be seen in Supplementary Data File S1. Taking these steps, we might find that any effect observed between MAR and ovarian cancer is attenuated when accounting for the higher rates of endometriosis in women who received MAR. Such a finding would lessen our confidence that we have observed a true causal effect between MAR and ovarian cancer, and not an effect observed caused by the confounding effect of endometriosis.

Quantitative bias analysis allows us to check many sources of bias in an analysis. For example, given not everyone who is infertile receives MAR, we might choose to test multiple plausible risk ratio values (say between 1.50 and 2.50) for the relationship between MAR and endometriosis and observe how the overall effect of MAR on the risk of ovarian cancer changes. If we had endometriosis in our dataset, but at a lower prevalence than the population level (perhaps because of underreporting), we could test for measurement error by simulating an increased proportion of endometriosis.

It is also possible to use quantitative bias analysis to test for more complicated types of confounding. For example, we could model the joint impact of two confounders (endometriosis and smoking), or model two confounders where one is assumed to affect both the exposure and the outcome, and the other confounder (e.g. age at menarche and endometriosis). Both these bias models can be seen in Fig. 2. In either case, the process of doing quantitative bias analysis is the same—use prior knowledge to assume the strength of the relationship between the confounders, the exposure, and the outcome, and then see how your estimate of the effect of exposure on the outcome changes when accounting for this confounding.

The main limitation of quantitative bias analysis is it tends to require a substantial amount of time and statistical expertise. Furthermore, though simple analyses can be conducted with

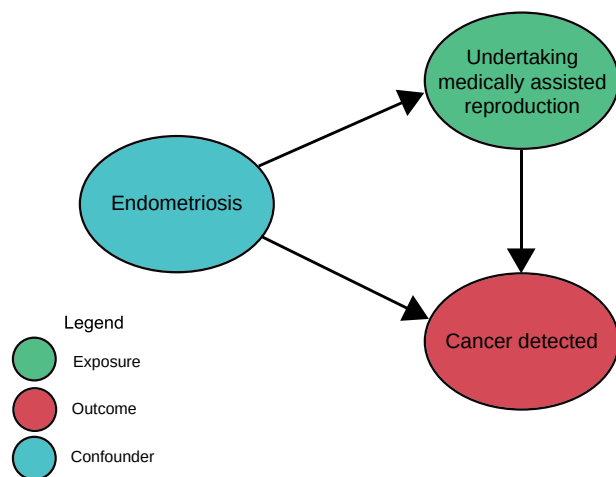


Figure 1. Example of the confounding relationship between endometriosis, MAR, and cancer. Causal direction is indicated by the arrows. MAR: medically assisted reproduction. Figures first made with the help of Dagitty (<https://dagitty.net>) (Textor *et al.*, 2016).

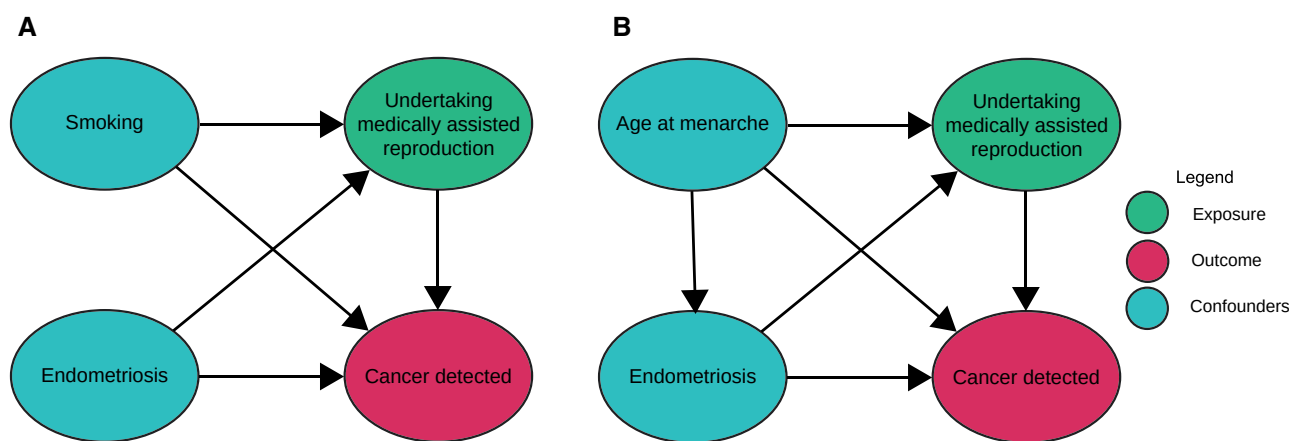


Figure 2. Example of confounding bias models that could be tested using quantitative bias analysis. (A) Two confounders affecting exposure and outcome. (B) Two confounders where both affect the exposure and the outcome and one confounder also affects the other confounder. Figures first made with the help of Dagitty (<https://dagitty.net>) (Textor et al., 2016).

summary data, more complex analyses (such as those involving person-time calculations) require access to person-level data (Fox et al., 2022). An alternative to quantitative bias analysis is to use E-values.

E-values

The E-value is simple and quick to compute using the formulae provided by VanderWeele and Ding (2017), even without person-level data. We calculate some E-values here for demonstration. Table 1 shows the E-values for selected results for four recent studies on women's cancer and MAR, based on the hazard ratios reported in the papers, and potential uncontrolled confounders for each study (Reigstad et al., 2017; Lundberg et al., 2019; Lindquist et al., 2022; Sandvei et al., 2023).

From Table 1, two points are evident. First, even in recent studies, there is variability in the potential confounders that have been controlled. Though most studies adjust for age and parity, they variably adjust for other key confounders like smoking, weight, and familial cancer history. Second, in most studies, it would only take one moderately strong uncontrolled confounder to nullify any observed association, particularly if the E-value is anchored to the lower bound of the confidence interval, as this represents the minimum level of confounding required to nullify the observed result. Though many studies list potential uncontrolled confounders in their limitations, E-values can be directly compared against existing literature. For example, endometriosis is not controlled for in any study in Table 1, despite being variably linked to both the development of cancers in Table 1 (Kvaskoff et al., 2020; Ye et al., 2022), and decreased fertility (Prescott et al., 2016; Thong et al., 2020; Mattsson et al., 2021).

If we consider the study on thyroid cancer by Lindquist et al. (2022), the lower-bound E-value of the relationship between MAR and thyroid cancer is 1.34. This E-value indicates that a relationship of 1.34 on the risk ratio scale between both endometriosis and MAR and endometriosis and thyroid cancer, would be sufficient to drive the observed association between MAR and thyroid cancer. We first note that the estimated risk ratio (compared to the general population) for infertility for women with endometriosis is 2.12 (Prescott et al., 2016), and the estimated risk ratio for thyroid cancer for women with endometriosis is 1.39 (Kvaskoff et al., 2020). Given that both risk ratios are above the lower-bound E-value for the relationship between endometriosis and thyroid cancer, and endometriosis and infertility, it is plausible that

endometriosis may be driving the observed relationship between MAR and thyroid cancer. By contrast, consider the relationship between MAR and ovarian cancer, where the reported lower-bound E-values are between 2.34 and 2.85. As the risk ratio for ovarian cancer for women with endometriosis is around 1.93 (Kvaskoff et al., 2020), it is unlikely that endometriosis is solely responsible for driving an observed association between MAR and ovarian cancer.

After estimating the potential for bias to affect the results through an E-value, the question remains of how our interpretation of the findings should change. It is important to state that efforts to measure the potential for bias should not serve to increase confidence in findings, given a major benefit of measuring bias is to counteract overconfidence in research findings (Lash et al., 2014). Furthermore, bias estimates are only interpretable in the face of domain-specific knowledge. That is, a high E-value for studies on ovarian cancer does not in itself mean the results are more likely to be true, particularly if there are plausible strong unmeasured confounders that may be affecting the results. Similarly, a low E-value should not necessarily decrease confidence in the results, especially if there are no plausible unmeasured confounders that could drive the effect. However, in cases where there are theoretically plausible unmeasured confounders, we should be particularly wary of low E-values. Based on our previous assessment of the association between endometriosis, MAR, and thyroid cancer, we may be more sceptical that a true causal relationship exists between MAR and thyroid cancer, or at the very least recommend that future studies attempt to control for endometriosis. Importantly, we may be more cautious about recommending against undertaking MAR treatments because of an increased risk of thyroid cancer.

Discussion and recommendations

Estimating and reporting the potential impact of bias is an important step to consider when conducting causal research on observational data, particularly when the conclusions may have implications for health policy and practice recommendations. Though overall the reporting of bias has been gaining traction, it has not been widely taken up in the epidemiological literature. Due to the risk of bias influencing the results of observational studies on MAR and fertility, it is imperative that researchers are aware of the potential pitfalls of bias, and subsequently the ways to estimate its impact on study findings. To promote the explicit

Table 1. Selected E-values calculated from recent studies on assisted reproductive therapy and cancer.

Study	Cancer type	Treatment type	Controlled confounders	Hazard ratio (95% CI)	E-value (95% CI)*
Sandvei et al. (2023)	Ovarian	IVF, ICSI (both fresh or frozen transfer)	Age, parity, age at first birth, BMI, smoking status, country	1.95 (1.49–2.55)	3.11 (2.34–4.54)
Lundberg et al. (2019)	Ovarian	IVF, ICSI (both fresh or frozen transfer)	Age, parity, calendar year, age at first birth, education level, country of birth, family history of breast or ovarian cancer	2.43 (1.73–3.42)	4.29 (2.85–6.30)
Lindquist et al. (2022)	Thyroid	Progesterone	Age, parity, calendar year of infertility diagnosis, education, obesity, thyroid disease, use of other fertility drugs	1.63 (1.07–2.48)	2.64 (1.34–4.40)
Reigstad et al. (2017)	Endometrial	Clomiphene citrate	Age, parity (only nulliparous women included), region of residence, exposure to GnRH/gonadotrophins	4.49 (2.66–7.60)	8.45 (4.76–14.68)

* E-value CI derived from values of hazard ratio CI.
 E-values calculated using the E-value calculator at: <https://www.evalue-calculator.com/> (VanderWeele and Ding, 2017; Mathur et al., 2018). All calculations assume the cancer outcome occurs in fewer than 15% of cases (to assume a reasonable approximation of risk ratios by hazard ratios).

consideration of bias in epidemiological studies, we outline two methods (quantitative bias analysis and E-values) researchers can use to estimate the potential impact of bias in their studies, and show how these might be used to enhance our understanding of the true relationship between MAR and later health outcomes (using cancer as an example).

Importantly, the point of estimating bias is not just to conduct better research but also to assist in the development of clinical practice guidelines. For example, reported links between MAR and certain health outcomes may impact both the health messaging doctors provide to prospective MAR patients and their decision to provide the treatment based on their patients' risk profile for those outcomes. This decision may have life-changing consequences for the patient, who may decide not to undergo treatment based on the advice they receive. Therefore, such advice should be given with careful consideration to the weight of evidence provided by research findings. Estimating bias is therefore an important step in assessing this weight of evidence for clinicians and policy-makers, who can then adjust their recommendations accordingly.

We must emphasize that estimating the impact of bias is not a silver bullet to improving study conclusions. Any methodological limitations that existed prior to considering bias continue to exist once that error is estimated. Furthermore, estimating the impact of bias should only ever serve to *decrease* our confidence in observed results (Lash et al., 2014), and estimates of bias should not be used to 'rule out' potential sources of bias. Though one unmeasured confounder may not appear strong enough to drive an effect alone, it may be part of a larger set of uncontrolled confounders that are consequential. However, estimating bias does provide a mechanism to better assess the strength of evidence provided, beyond looking simply at whether studies have a large population capture or high precision in coefficient estimates.

It is also important to stress that this piece is not a criticism of existing work on cancer and MAR. Rather it serves as an example of where quantifying bias may be particularly important for stakeholders. Especially in the case of unmeasured confounding, it is often impossible to accurately determine values of various confounders from administrative datasets (e.g. age of menarche or family history of cancer) that are relevant to questions of cancer and MAR. We stress that an inability to control for unmeasured confounding does not mean such studies are not useful.

However, the ongoing question of MAR's relationship to cancer does highlight that measuring bias may be particularly informative when effect sizes are generally small and implications for health outcomes large.

Though we have presented both quantitative bias analysis and E-values as methods to estimate bias, there are also other approaches. For example, the use of negative controls, where researchers model an outcome expected to have the same potential sources of bias as the outcome of interest, but no relationship to the exposure, is another method of estimating the impact of bias (Arnold et al., 2016; Shi et al., 2020). In general, simpler estimation approaches can be informative, but tend to be more limited in the scope of error they can account for. For example, E-values may give us information on unmeasured confounding, but they do not address complex causal structures, measurement errors, or selection bias. By contrast, more complicated approaches can be more informative, but also prohibitively difficult to run, requiring both specialist expertise and sometimes vast computational resources. It is worth noting proponents of quantitative bias analysis acknowledge the difficulties in implementation inherent in such methods and are making efforts to increase the accessibility of such analyses (such as through the distribution of readily modifiable implementation code) (Fox et al., 2023).

Researchers who choose to report E-values should be aware of their criticisms. E-values are fundamentally theory agnostic, incorporating only information about the observed risk ratio association between the exposure and outcome (Greenland, 2020). This fact can lead to both researchers being overly optimistic and pessimistic about their research findings (Sjölander and Greenland, 2022), as they may incorrectly assume a low E-value automatically indicates a problematic result (even when strong confounders have been controlled for) (Greenland, 2020), or a high E-value indicates a trustworthy result (even when there may be multiple strong uncontrolled confounders). It is therefore critical researchers do not use simpler methods like E-values as a proxy for theory development. Rather, they should be used in conjunction with careful consideration of background knowledge and research, as would be required if one were to perform quantitative bias analysis.

A major strength of quantitative bias analysis over E-values is that any uncertainty in the true relationships between confounders, exposure, and treatment can be explicitly modelled. Though

this strength means that quantitative bias analysis can provide a more comprehensive test of bias than E-values, there is no guarantee the researchers' model of these relationships is correct or useful, and there is no way to test this empirically (Lash et al., 2014). As a poorly specified bias model may lead to incorrect conclusions about the level of bias, researchers must be careful when designing and implementing these bias analyses. For example, in this commentary, we have used infertility owing to endometriosis as a proxy for receiving MAR treatment, but not everyone who experiences infertility will receive MAR treatment. As such, it may be important to include this pathway from infertility to MAR treatment in a bias model. If you were comparing only infertile women who received and did not receive MAR (an approach taken for the Dutch OMEGA cohort, for example) (van Leeuwen et al., 2011), this extra step may not be necessary. Despite the risk of poorly specified bias models, we believe the benefits outweigh the risk assuming researchers think carefully about the specific problem they are studying.

In conclusion, quantitative bias analysis and E-values are two methods that can be used to assess how bias may affect research conclusions in observational studies. We argue this assessment is important in questions where there are potentially many sources of bias (such as studies on health outcomes associated with MAR), particularly where the findings may have important implications for health policy and clinical practice. They are not a remedy for all methodological problems faced in observational research but can help to encourage researchers and readers to more carefully consider the strength of evidence provided to arguments around the presence or absence of causal associations.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

No new data were generated or analysed in support of this research.

Authors' roles

All authors made substantial contribution to the conception of the work. ARW was responsible for drafting the work, while CAV, SO, GMC, LRJ, and CMV were responsible for reviewing it critically for important intellectual content. All authors approved the final version to be published, and all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

This work was funded by the National Health and Medical Research Council (NHMRC: APP1164852).

Conflict of interest

ARW. declares that his involvement in this work was supported by employment at UNSW Sydney. CAV declares his membership in the Statistical Advisory Board of the journal *Human Reproduction*. SO declares that she has received payment to their institution from the National Health and Medical Research Council, Australia, and received salary payment from The

Norwegian University of Science and Technology. LRJ declares that they have received payment to their institution from the National Health and Medical Research Council, Australia. CMV declares that they have received payment to their institution from the National Health and Medical Research Council, Australia; they are a non-remunerated member of DSM on an investigator-initiated clinical trial (vitamin D and skin cancer risk).

References

- Arnold BF, Ercumen A, Benjamin-Chung J, Colford JM. Brief report: negative controls to detect selection bias and measurement bias in epidemiologic studies. *Epidemiology* 2016;**27**:637–41.
- Auranen A, Hietanen S, Salmi T, Grénman S. Hormonal treatments and epithelial ovarian cancer risk. *Int J Gynecol Cancer* 2005;**15**:692–700.
- Australian Institute of Health and Welfare. *Endometriosis*. Canberra: AIHW, 2023.
- Barcroft JF, Galazis N, Jones BP, Getreu N, Bracewell-Milnes T, Grewal KJ, Sorbi F, Yazbek J, Lathouras K, Smith JR et al. Fertility treatment and cancers—the eternal conundrum: a systematic review and meta-analysis. *Hum Reprod* 2021;**36**:1093–107.
- Barraza F, Arancibia M, Madrid E, Papuzinski C. General concepts in biostatistics and clinical epidemiology: random error and systematic error. *Medwave* 2019;**19**:e7687.
- Bates GW. Breast cancer risk clouds fertility care. *Fertil Steril* 2021;**116**:81–2.
- Beebejaun Y, Athithan A, Copeland TP, Kamath MS, Sarris I, Sunkara SK. Risk of breast cancer in women treated with ovarian stimulation drugs for infertility: a systematic review and meta-analysis. *Fertil Steril* 2021;**116**:198–207.
- Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod* 2007;**22**:1506–12.
- Brown SB, Hankinson SE. Endogenous estrogens and the risk of breast, endometrial, and ovarian cancers. *Steroids* 2015;**99**:8–10.
- Chambers GM, Dyer S, Zegers-Hochschild F, de Mouzon J, Ishihara O, Banker M, Mansour R, Kupka MS, Adamson GD. International committee for monitoring assisted reproductive technologies world report: assisted reproductive technology, 2014. *Hum Reprod* 2021;**36**:2921–34.
- Cullinane C, Gillan H, Geraghty J, Evoy D, Rothwell J, McCartan D, McDermott EW, Prichard RS. Fertility treatment and breast-cancer incidence: meta-analysis. *BJS Open* 2022;**6**:zrab149.
- Dayan N, Filion KB, Okano M, Kilmartin C, Reinblatt S, Landry T, Basso O, Udell JA. Cardiovascular risk following fertility therapy: systematic review and meta-analysis. *J Am Coll Cardiol* 2017;**70**:1203–13.
- Del Pup L, Peccatori F, Levi-Setti P, Codacci-Pisanelli G, Patrizio P. Risk of cancer after assisted reproduction: a review of the available evidences and guidance to fertility counselors. *Eur Rev Med Pharmacol Sci* 2018;**22**:8042–59.
- Fathalla MF. Incessant ovulation—a factor in ovarian neoplasia? *Lancet* 1971;**2**:163.
- Fishel S, Jackson P. Follicular stimulation for high tech pregnancies: are we playing it safe? *BMJ* 1989;**299**:309–11.
- Fox MP, MacLehose RF, Lash TL. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Springer, 2022.
- Fox MP, MacLehose RF, Lash TL. SAS and R code for probabilistic quantitative bias analysis for misclassified binary variables and binary unmeasured confounders. *Int J Epidemiol* 2023;**52**:1624–33.

- Gennari A, Costa M, Puntoni M, Paleari L, De Censi A, Sormani MP, Provinciali N, Bruzzi P. Breast cancer incidence after hormonal treatments for infertility: systematic review and meta-analysis of population-based studies. *Breast Cancer Res Treat* 2015; **150**:405–13.
- Greenland S. Commentary: an argument against E-values for assessing the plausibility that an association could be explained away by residual confounding. *Int J Epidemiol* 2020; **49**:1501–3.
- Hernan MA, Robins JM. *Causal Inference: What If*. Boca Raton, FL: Chapman & Hall/CRC, 2020.
- Huhtaniemi I. Are gonadotrophins tumorigenic—a critical review of clinical and experimental data. *Mol Cell Endocrinol* 2010; **329**:56–61.
- Kvaskoff M, Mahamat-Saleh Y, Farland LV, Shigesu N, Terry KL, Harris HR, Roman H, Becker CM, As-Sanie S, Zondervan KT et al. Endometriosis and cancer: a systematic review and meta-analysis. *Hum Reprod Update* 2020; **27**:393–420.
- Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol* 2014; **43**:1969–85.
- Liang J, Shang Y. Estrogen and cancer. *Annu Rev Physiol* 2013; **75**:225–40.
- Lindquist S, Kjær SK, Albieri V, Frederiksen K, Hansen T, Nøhr B, Jensen A. Fertility drugs and incidence of thyroid cancer in a Danish nationwide cohort of 146 024 infertile women. *Hum Reprod* 2022; **37**:838–47.
- Lu CY. Observational studies: a review of study designs, challenges and strategies to reduce confounding. *Int J Clin Pract* 2009; **63**:691–7.
- Lundberg FE, Johansson AL, Rodriguez-Wallberg K, Gemzell-Danielsson K, Iliadou AN. Assisted reproductive technology and risk of ovarian cancer and borderline tumors in parous women: a population-based cohort study. *Eur J Epidemiol* 2019; **34**:1093–101.
- Magnus MC, Fraser A, Håberg SE, Rönö K, Romundstad LB, Bergh C, Spangmose AL, Pinborg A, Gissler M, Wennerholm U-B et al. Maternal risk of cardiovascular disease after use of assisted reproductive technologies. *JAMA Cardiol* 2023; **8**:837–45.
- Mathur MB, Ding P, Riddell CA, VanderWeele TJ. Website and R package for computing E-values. *Epidemiology* 2018; **29**:e45–7.
- Mattsson K, Nilsson-Condori E, Elmerstig E, Vassard D, Schmidt L, Ziebe S, Jöud A. Fertility outcomes in women with pre-existing type 2 diabetes—a prospective cohort study. *Fertil Steril* 2021; **116**:505–13.
- Momenimovahed Z, Taheri S, Tiznobaik A, Salehiniya H. Do the fertility drugs increase the risk of cancer? A review study. *Front Endocrinol (Lausanne)* 2019; **10**:313.
- Murugappan G, Li S, Alvero RJ, Luke B, Eisenberg ML. Association between infertility and all-cause mortality: analysis of US claims data. *Am J Obstet Gynecol* 2021; **225**:57. e51–7. e11.
- Pearl J. Causal inference in statistics: an overview. *Statist Surveys* 2009; **3**:96–146.
- Petersen JM, Ranker LR, Barnard-Mayers R, MacLehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. *Int J Epidemiol* 2021; **50**:1708–30.
- Prescott J, Farland L, Tobias D, Gaskins A, Spiegelman D, Chavarro J, Rich-Edwards J, Barbieri R, Missmer S. A prospective cohort study of endometriosis and subsequent risk of infertility. *Hum Reprod* 2016; **31**:1475–82.
- Reigstad MM, Storeng R, Myklebust TÅ, Oldereid NB, Omland AK, Robsahm TE, Brinton LA, Vangen S, Furu K, Larsen IK. Cancer risk in women treated with fertility drugs according to parity status—a registry-based cohort study. *Cancer Epidemiol Biomarkers Prev* 2017; **26**:953–62.
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia, 2008.
- Rowlands I, Abbott J, Montgomery G, Hockey R, Rogers P, Mishra G. Prevalence and incidence of endometriosis in Australian women: a data linkage cohort study. *BJOG* 2021; **128**:657–65.
- Sachdev D, Yamada R, Lee R, Sauer MV, Ananth CV. Risk of stroke hospitalization after infertility treatment. *JAMA Netw Open* 2023; **6**:e2331470.
- Sandvei MS, Pinborg A, Gissler M, Bergh C, Romundstad LB, van Leeuwen FE, Spaan M, Tiitinen A, Wennerholm U-B, Henningsen A-K et al. Risk of ovarian cancer in women who give birth after assisted reproductive technology (ART)—a registry-based Nordic cohort study with follow-up from first pregnancy. *Br J Cancer* 2023; **128**:825–32.
- Shi X, Miao W, Tchetgen ET. A selective review of negative control methods in epidemiology. *Curr Epidemiol Rep* 2020; **7**:190–202.
- Sjölander A, Greenland S. Are E-values too optimistic or too pessimistic? Both and neither! *Int J Epidemiol* 2022; **51**:355–63.
- Textor J, Van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol* 2016; **45**:1887–94.
- Thong EP, Codner E, Laven JS, Teede H. Diabetes: a metabolic and reproductive disorder in women. *Lancet Diabetes Endocrinol* 2020; **8**:134–49.
- van Leeuwen FE, Klip H, Mooij TM, van de Swaluw AMG, Lambalk CB, Kortman M, Laven JSE, Jansen CAM, Helmerhorst FM, Cohlen BJ et al. Risk of borderline and invasive ovarian tumours after ovarian stimulation for in vitro fertilization in a large Dutch cohort. *Hum Reprod* 2011; **26**:3456–65.
- VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med* 2017; **167**:268–74.
- Vassard D, Schmidt L, Pinborg A, Petersen GL, Forman JL, Hageman I, Glazer CH, Kamper-Jørgensen M. Mortality in women treated with assisted reproductive technology—addressing the healthy patient effect. *Am J Epidemiol* 2018; **187**:1889–95.
- Wei SQ, Paradis G, Ayoub A, Lewin A, Auger N. Assisted reproductive technology and cardiovascular outcomes in parents and offspring. *Can J Cardiol* 2023; **40**:130–7.
- Ye J, Peng H, Huang X, Qi X. The association between endometriosis and risk of endometrial cancer and breast cancer: a meta-analysis. *BMC Womens Health* 2022; **22**:455.
- Yiallourou SR, Magliano D, Haregu TN, Carrington MJ, Rolnik DL, Rombauts L, Rodrigues A, Ball J, Bruinsma FJ, Da Silva Costa F. Long term all-cause and cardiovascular disease mortality among women who undergo fertility treatment. *Med J Aust* 2022; **217**:532–7.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Human Reproduction, 2024, 39, 869–875
<https://doi.org/10.1093/humrep/deae053>
Opinion

Evolution of evidence-based diagnostic criteria in adolescents with polycystic ovary syndrome

H.J. Teede ^{1,*}, A.C.H. Neven ¹, and A. Pena ²

¹Monash Centre for Health Research and Implementation, Monash University, Melbourne, VIC, Australia

²Discipline of Paediatrics, The University of Adelaide Robinson Research Institute, North Adelaide, SA, Australia

*Correspondence address. Monash Centre for Health Research and Implementation, Monash University, 268 Kanooka Grove Clayton, Melbourne, VIC 3168, Australia. E-mail: helena.teede@monash.edu  <https://orcid.org/0000-0001-7609-577X>

Polycystic ovarian syndrome (PCOS) is the most common endocrinopathy in reproductive age women and has significant clinical reproductive, cardiometabolic, dermatological and psychological features (Teede et al., 2010). Diagnostic criteria have evolved historically, resulting in variable global prevalence estimates. The key historical criteria include the 1990 National Institute of Health (NIH) which requires both hyperandrogenism (HA) (clinical and/or biochemical) and ovulatory dysfunction (OD) for diagnosis, without defining these individual criteria (Zawadzki and Dunaif, 1992). The 2003 Rotterdam consensus criteria led by the European Society for Human Reproduction and Embryology (ESHRE) and the American Society for Reproductive Medicine (ASRM), defined PCOS based on at least two out of three of HA, OD, and ultrasound characteristics of polycystic ovarian morphology (PCOM) after exclusion of other causes (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004).

In 2018, the field advanced considerably with ESHRE, ASRM and 38 other societies and patient groups partnering with the Australian Centre for Research Excellence Monash University, to develop rigorous, evidence-based International Guideline diagnostic criteria, applying best practice evidence synthesis and guideline methods (Teede et al., 2018). Optimizing evidence-based criteria were a key priority in the Guideline to limit under diagnosis or delayed diagnosis and reduce potential over diagnosis based on isolated PCOM features. The International Evidence-based PCOS Guideline built on consensus Rotterdam criteria, refined individual diagnostic components, and differentiated adult and adolescent criteria.

Adolescence is a period with critical changes in growth, puberty, and development. The normal physiological changes during this period including irregular menstrual cycles, acne, and PCOM, all of which overlap with adult diagnostic criteria for PCOS making the diagnosis of adolescent PCOS challenging and controversial (Pena and Metz, 2018). The 2018, and updated, expanded 2023, International Evidence-based PCOS Guidelines refined the diagnostic criteria during adolescence addressing the issue of overlap features with normal pubertal physiology (Peña et al., 2020; Teede et al., 2023). First, the Guideline provided a detailed definition of irregular menstrual cycles according to time post-menarche, limiting early physiological pubertal irregular

menstrual cycles from being labelled as abnormal and increasing capture of pathological irregular menstrual cycles during late post-menarcheal years. Second, it emphasizes that severe but not mild acne should be considered a sign of HA during adolescence, in addition to hirsutism evaluated by modified Ferriman Gallwey score. Third, the Guidelines highlight the importance of removing PCOM defined by ultrasound and not including anti-Müllerian hormone (AMH) in adolescent diagnostic criteria, due to lack of specificity at this life stage. Hence, the International Guidelines recommend diagnosis of PCOS during adolescence based on the presence of both OD defined according to time post-menarche and clinical/biochemical HA, after exclusion of other disorders. Importantly it was also recommended that adolescents with only one of these two features should be considered 'at risk' of PCOS and followed up (Teede et al., 2018, 2023; Peña et al., 2020). In terms of prevalence of PCOS by diagnostic criteria, existing literature and systematic reviews in this area are limited by heterogeneous source data, including medical records or registry studies with artificially low capture rates, highlighting the need for more research in well-defined adolescent populations (Mousa et al., 2023; Kim et al., 2024).

In this issue of *Human Reproduction*, a recent study by Kim et al., makes an important contribution to the field. This retrospective study explored a well characterized population of Korean adolescents presenting with clinical features of PCOS (Kim et al., 2024). They were diagnosed with PCOS on 2003 Rotterdam consensus criteria and separately on International Guideline criteria. Of the 315 girls with PCOS by Rotterdam criteria, 65%, 10%, 4%, and 21% were classified as having all three features, OD and HA, PCOM and HA, or PCOM and OD, respectively. Therefore, 236 adolescents with both OD and HA were diagnosed based on Guideline criteria and 79 were in the Guideline 'at risk' group, with only one of these two features. Another 415 were classified as controls with only hirsutism and cycle data, however no biochemistry or ultrasound data were available for controls. These results are consistent with another longitudinal study in a well characterized population followed from birth, that showed reduced prevalence of PCOS using International Guideline criteria, with the most severely affected being identified (Tay et al., 2020).

Received: February 18, 2024. Editorial decision: February 23, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved. For permissions, please email: journals.permissions@oup.com

The study by Kim *et al.* demonstrates that adolescents diagnosed on International Guideline criteria, were the most severely affected in terms of hirsutism and metabolic features. Those who only met Rotterdam criteria (requiring PCOM for diagnosis), thereby meeting the 'at risk' International Guideline criteria (Kim *et al.*, 2024), had comparatively milder clinical features, yet they had higher metabolic markers than controls. Here, International Guidelines would recommend symptomatic treatment and follow-up for those at risk, to clarify diagnosis, potentially including PCOM or AMH, once they reach adulthood.

Identification and follow-up of adolescents 'at risk' of PCOS is important to limit missed diagnosis, especially during the critical transition to adult care. Follow-up to capture those who clearly develop PCOS in early adulthood will also ensure appropriate metabolic and mental health screening and targeted treatment. Given the high prevalence of mental health challenges in PCOS, the recent evidence of an 8-fold increase in suicide in those with PCOS, together with patient dissatisfaction and adverse outcomes especially in fertility and the attainment of aspirational family size, such follow-up is important (Hsu *et al.*, 2024).

Importantly, the study by Kim *et al.* (2024) also highlights the need for life course studies on the long-term health outcomes and prevention opportunities in PCOS. For a condition that affects 140 million women globally, the lack of research funding, recognition, education, and clinical models of care, continues to present a major gap, emphasized in the Guidelines (Teede *et al.*, 2023). Key research priorities in adolescence identified during the broad consultation and Guideline development processes, include further refinement of individual components of the diagnostic criteria, investigation of natural history to ensure adolescent identification of those most at risk for acute and long-term manifestation, and the establishment of optimal treatments. Here, we also highlight the importance of an international individual patient data meta-analysis underway in the PCOS Phenotype in Unselected Population (PUPP) study to explore PCOS features across world regions (Kiconco *et al.*, 2021).

Overall, PCOS is a common endocrinopathy with diverse and significant clinical features. International Guideline criteria are now evidence-based and have clarified diagnosis in adolescents. They capture patients with both irregular cycles and HA, and designate an 'at risk' group with milder diagnostic features than those meeting full diagnostic criteria, but still with metabolic features who can still be treated for symptoms and should be followed up. For timely and appropriate diagnosis for adolescents, dissemination and implementation of International Guideline recommendations is vital. A multifaceted translation program aims to enhance health care practitioner confidence in diagnosing PCOS including at this life stage and to empower those with PCOS. It includes multilingual education outputs and evidence-based resources for consumers (such as the free [ASKPCOS](#) app which is used in 186 countries), alongside tools for healthcare professionals and for policy makers. Ultimately, more research is now needed on life course, natural history and timing for diagnosis, alongside ongoing education of healthcare providers, those affected and their guardians.

Authors' roles

All authors contributed to writing the manuscript and approved the final version.

Funding

H.J.T. is supported by the Australian National Health and Medical Research Council. A.C.H.N. is funded by a Monash International Tuition Scholarship (MITS) and Research Training Program Stipend (RTP stipend).

Conflict of interest

None declared.

References

- Hsu TW, Kao YC, Tsai SJ, Bai YM, Su TP, Chen TJ, Liang CS, Chen MH. Suicide attempts after a diagnosis of polycystic ovary syndrome: a cohort study. *Ann Intern Med* 2024. <https://doi.org/10.7326/M23-2240>.
- Kim JJ, Hwang KR, Lee D, Kim S, Choi YM. Adolescents diagnosed with polycystic ovary syndrome under the Rotterdam criteria but not meeting the diagnosis under the updated guideline. *Hum Reprod* 2024;**39**:1072–1077.
- Kiconco S, Mousa A, Azziz R, Enticott J, Suturina LV, Zhao X, Gambineri A, Tehrani FR, Yildiz BO, Kim JJ *et al.* PCOS Phenotype in Unselected Populations Study (P-PUP): protocol for a systematic review and defining PCOS diagnostic features with pooled individual participant data. *Diagnostics (Basel)* 2021;**11**:1953.
- Mousa A, Tay CT, Teede H. Technical report for the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Report. Monash University, 2023. <https://doi.org/10.26180/23625288.v1>.
- Pena AS, Metz M. What is adolescent polycystic ovary syndrome? *J Paediatr Child Health* 2018;**54**:351–355.
- Peña AS, Witchel SF, Hoeger KM, Oberfield SE, Vogiatzi MG, Misso M, Garad R, Dabadghao P, Teede H. Adolescent polycystic ovary syndrome according to the international evidence-based guideline. *BMC Med* 2020;**18**:72.
- Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod* 2004;**19**:41–47.
- Tay CT, Hart RJ, Hickey M, Moran LJ, Earnest A, Doherty DA, Teede HJ, Joham AE. Updated adolescent diagnostic criteria for polycystic ovary syndrome: impact on prevalence and longitudinal body mass index trajectories from birth to adulthood. *BMC Med* 2020;**18**:389.
- Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. *BMC Med* 2010;**8**:41.
- Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ; Network IP. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod* 2018;**33**:1602–1618.
- Teede HJ, Tay CT, Laven J, Dokras A, Moran LJ, Piltonen TT, Costello FM, Boivin J, Redman ML, Boyle AJ *et al.*; International PCOS Network. Recommendations from the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. *Hum Reprod* 2023;**38**:1655–1679.
- Zawadzki J, Dunaif A. Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. In: Dunaif A, Givens JR, Haseltine F (eds). *Polycystic Ovary Syndrome*. Boston: Blackwell Scientific, 1992, 377–384.

Progestin-primed ovarian stimulation and aneuploidy, innocent until proven guilty

Baris Ata ^{1,2,*}

¹ART Fertility Clinics, Dubai, United Arab Emirates

²Department of Obstetrics and Gynecology, Koç University School of Medicine, Istanbul, Türkiye

*Correspondence address. ART Fertility Clinics, Jumeirah Street Villa 869, Umm Suqeim 3, Dubai, United Arab Emirates. E-mail: baris.ata@artfertilityclinics.com

 <https://orcid.org/0000-0003-1106-3747>

In this issue of *Human Reproduction*, Vidal *et al.* contribute to the progestin-primed ovarian stimulation (PPOS) literature with a high-quality study in which they prospectively compared two consecutive cycles in the same patients who aimed to accumulate euploid blastocysts, the first with a GnRH antagonist and the second with a PPOS protocol that employed 200 mg/day micronized progesterone (MIP) in a single oral dose (Vidal *et al.*, 2024). Given the median age of 39.59 years with quartiles of 38.73 and 40.34, 75% of the 44 participants were over 38 years old, at which age 55–70% of blastocysts are expected to be aneuploid (Ata *et al.*, 2021b). This ensures an adequate event rate to compare euploidy rate between two cycles. As expected in this age group, the euploidy rates were 34.9% versus 29.5% in the GnRH antagonist and PPOS cycles, respectively (Vidal *et al.*, 2024). Since most of the blastocysts diagnosed as ‘mosaic’ should indeed be euploid, the inclusion of mosaics in the euploid count, further diminishes the statistically nonsignificant absolute difference between GnRH antagonists and PPOS in the study (Popovic *et al.*, 2024; Vidal *et al.*, 2024). Despite the non-randomized design, comparing two cycles in the same patient conducted within three months with a constant gonadotropin dosage ensures similar baseline characteristics and isolates the effect of pituitary suppression method, arguably even better than randomizing different individuals. Yet, the design could have been even better if the participants were randomized for the method of pituitary suppression in the first cycle, followed by crossing over to the other arm in the second cycle.

It is noteworthy that in addition to a similar percentage and number of euploid blastocysts, PPOS with 200 mg/day MIP yielded significantly higher number of oocytes and metaphase II oocytes than the GnRH antagonist. These findings are consistent with a prior study in which we also compared two cycles of the same oocyte donors, one with GnRH antagonist and the other with PPOS (Yildiz *et al.*, 2019). We also reported significantly higher numbers of oocytes and metaphase II oocytes with PPOS using 10 mg/day medroxyprogesterone acetate. Although pre-implantation genetic testing for aneuploidy was not applied in our study involving young oocyte donors, recipients had similar live birth or ongoing pregnancy rates with both protocols, suggesting similar embryonic potential (Yildiz *et al.*, 2019). It is intriguing to observe

higher oocyte yield in both studies comparing two cycles in the same women, which ensures similar baseline characteristics, in contrast to the pooled analyses of mostly retrospective cohort studies which report similar oocyte yield with PPOS and GnRH antagonist (Yildiz *et al.*, 2019; Ata *et al.*, 2021a; Vidal *et al.*, 2024). The higher oocyte yield could be due to a milder pituitary suppression by PPOS than GnRH antagonists, which is reflected by higher serum LH levels with the former (Vidal *et al.*, 2024). It would have been interesting to measure serum FSH levels during stimulation to see if it was higher in PPOS cycles, supporting the theory of milder pituitary suppression.

The study by Vidal *et al.* (2024), with its proper design and analyses, is an important contribution to the increasing number of studies demonstrating effectiveness and safety of PPOS from different clinics worldwide. The replication of favorable results by different teams enhances their authority. However, among these increasing numbers of publications, some suffer from serious methodological errors and report misleading negative conclusions. Some of such studies contribute to a wrong narrative about lower cumulative live birth rates (CLBR) and lower euploidy rates with PPOS as we discuss in detail in our recent review of PPOS (Ata and Kalafat, 2024). Among all the studies comparing blastocyst euploidy rates between PPOS and GnRH analogues, only one reported a significantly lower euploidy rate with PPOS, which was based on a subgroup analysis, with a limited sample of 19 PPOS cycles (Pai *et al.*, 2023). Vidal *et al.*’s prospective study is in agreement with all the other retrospective studies comparing euploidy rates between PPOS and GnRH analogues, and increases confidence in PPOS (Ata and Kalafat, 2024; Vidal *et al.*, 2024).

Unfortunately, Vidal *et al.* did not report live birth rates with the euploid blastocysts (Vidal *et al.*, 2024). Yet, embryo transfer outcomes of PPOS cycles seem similar with GnRH antagonists (Ata *et al.*, 2021a). While some early studies suggested lower cumulative live birth rates with PPOS, they suffer from serious flaws including arbitrary definitions of CLBR, arbitrary categorizations of participants, and inconsistencies in their results (Zhang *et al.*, 2021; Chen *et al.*, 2022; Li *et al.*, 2022; Zhou *et al.*, 2023a,b; Ata and Kalafat, 2024). In contrast, two studies, including a randomized controlled trial, reported similar CLBR with

Received: January 29, 2024. Revised: February 19, 2024. Editorial decision: February 23, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

PPOS and GnRH analogues using the correct definition of CLBR (Du *et al.*, 2021; Giles *et al.*, 2021). Given the high follicular fluid progesterone concentrations during the follicular phase of natural menstrual cycles, PPOS would be extremely unlikely to affect oocyte development and reproductive potential (Westergaard *et al.*, 1986).

Medicine and law are different, but the legal principle of the presumption of innocence seems applicable to medicine in the context of PPOS. Given the absence of a biological rationale and increasing evidence in favor of PPOS, it should be regarded as providing similar euploidy and CLBR with GnRH antagonists, unless high-quality studies prove otherwise.

Conflict of interest

There is no conflict of interest associated with this work.

References

- Ata B, Capuzzo M, Turkgeldi E, Yildiz S, La Marca A. Progestins for pituitary suppression during ovarian stimulation for ART: a comprehensive and systematic review including meta-analyses. *Hum Reprod Update* 2021a;**27**:48–66.
- Ata B, Kalafat E. Progesterin-primed ovarian stimulation: for whom, when and how? *Reprod Biomed Online* 2024;**48**:103639.
- Ata B, Kalafat E, Somigliana E. A new definition of recurrent implantation failure on the basis of anticipated blastocyst aneuploidy rates across female age. *Fertil Steril* 2021b;**116**:1320–1327.
- Chen H, Teng XM, Sun ZL, Yao D, Wang Z, Chen ZQ. Comparison of the cumulative live birth rates after 1 in vitro fertilization cycle in women using gonadotropin-releasing hormone antagonist protocol vs. progesterin-primed ovarian stimulation: a propensity score-matched study. *Fertil Steril* 2022;**118**:701–712.
- Du M, Zhang J, Li Z, Liu X, Li J, Liu W, Guan Y. Comparison of the cumulative live birth rates of progesterin-primed ovarian stimulation and flexible GnRH antagonist protocols in patients with low prognosis. *Front Endocrinol (Lausanne)* 2021;**12**:705264.
- Giles J, Alama P, Gamiz P, Vidal C, Badia P, Pellicer A, Bosch E. Medroxyprogesterone acetate is a useful alternative to a gonadotropin-releasing hormone antagonist in oocyte donation: a randomized, controlled trial. *Fertil Steril* 2021;**116**:404–412.
- Li Z, Jia R, Wang K, Zhang J, Ren B, Guan Y. Analysis of cumulative live birth rate and perinatal outcomes in young patients with low anti-müllerian hormone levels using two ovulation promotion protocols: a cohort study. *Front Endocrinol (Lausanne)* 2022;**13**:938500.
- Pai AH, Sung YJ, Li CJ, Lin CY, Chang CL. Progesterin primed ovarian stimulation (PPOS) protocol yields lower euploidy rate in older patients undergoing IVF. *Reprod Biol Endocrinol* 2023;**21**:72.
- Popovic M, Borot L, Lorenzon AR, Lopes ALRC, Sakkas D, Lledó B, Morales R, Ortiz JA, Polyzos NP, Parriego M *et al.* Implicit bias in diagnosing mosaicism amongst preimplantation genetic testing providers: results from a multicenter study of 36 395 blastocysts. *Hum Reprod* 2024;**39**:258–274.
- Vidal MDM, Martinez F, Rodriguez I, Polyzos N. Ovarian response and embryo ploidy following oral micronized progesterone primed ovarian stimulation vs. GnRH antagonist protocol. A prospective study with repeated ovarian stimulation cycles. *Hum Reprod* 2024;deae047. <https://doi.org/10.1093/humrep/deae047>.
- Westergaard L, Christensen IJ, McNatty KP. Steroid levels in ovarian follicular fluid related to follicle size and health status during the normal menstrual cycle in women. *Hum Reprod* 1986;**1**:227–232.
- Yildiz S, Turkgeldi E, Angun B, Eraslan A, Urman B, Ata B. Comparison of a novel flexible progesterin primed ovarian stimulation protocol and the flexible gonadotropin-releasing hormone antagonist protocol for assisted reproductive technology. *Fertil Steril* 2019;**112**:677–683.
- Zhang S, Yin Y, Li Q, Zhang C. Comparison of cumulative live birth rates between GnRH-A and PPOS in low-prognosis patients according to POSEIDON criteria: a cohort study. *Front Endocrinol (Lausanne)* 2021;**12**:644456.
- Zhou R, Dong M, Huang L, Wang S, Fan L, Liang X, Zhang X, Liu F. Comparison of cumulative live birth rates between progesterin-primed ovarian stimulation protocol and gonadotropin-releasing hormone antagonist protocol in different populations. *Front Endocrinol (Lausanne)* 2023a;**14**:1117513.
- Zhou R, Dong M, Huang L, Wang S, Wang Z, Xu L, Zhang X, Liu F. Comparison of cumulative live birth rates between progesterin and GnRH analogues in preimplantation genetic testing cycles. *J Clin Endocrinol Metab* 2023b;**109**:217–226.

Andrology

Deletion of ACTRT1 is associated with male infertility as sperm acrosomal ultrastructural defects and fertilization failure in human

Qi Zhang ^{1,2,†}, Huijuan Jin^{3,†}, Shunhua Long^{1,2}, Xiangrong Tang^{1,2}, Jiaxun Li^{1,2}, Weiwei Liu^{1,2}, Wei Han^{1,2}, Haiyuan Liao^{1,2}, Tao Fu^{1,2}, Guoning Huang ^{1,2,*}, Suren Chen ^{3,*}, and Tingting Lin ^{1,2,*}

¹Chongqing Key Laboratory of Human Embryo Engineering, Center for Reproductive Medicine, Women and Children's Hospital of Chongqing Medical University, Chongqing, China

²Chongqing Clinical Research Center for Reproductive Medicine, Chongqing Health Center for Women and Children, Chongqing, China

³Key Laboratory of Cell Proliferation and Regulation Biology, Ministry of Education, College of Life Sciences, Beijing Normal University, Beijing, China

*Correspondence address. Chongqing Key Laboratory of Human Embryo Engineering, Center for Reproductive Medicine, Women and Children's Hospital of Chongqing Medical University, No. 64, Jintang Street, Yuzhong District, Chongqing 400010, China. E-mail: gnhuang217@sina.com  <https://orcid.org/0000-0003-0196-3709> (G.H.); Department of Biology, College of Life Sciences, Beijing Normal University, No. 19, Xijiekouwai Street, Haidian District, Beijing 100875, China. E-mail: chensr@bnu.edu.cn  <https://orcid.org/0000-0002-9337-5412> (S.C.); Chongqing Key Laboratory of Human Embryo Engineering, Center for Reproductive Medicine, Women and Children's Hospital of Chongqing Medical University, No. 64, Jintang Street, Yuzhong District, Chongqing 400010, China. E-mail: yuting9263@163.com  <https://orcid.org/0000-0003-1559-5038> (T.L.)

[†]These authors contributed equally to this work.

ABSTRACT

STUDY QUESTION: Could actin-related protein T1 (ACTRT1) deficiency be a potential pathogenic factor of human male infertility?

SUMMARY ANSWER: A 110-kb microdeletion of the X chromosome, only including the ACTRT1 gene, was identified as responsible for infertility in two Chinese males with sperm showing acrosomal ultrastructural defects and fertilization failure.

WHAT IS KNOWN ALREADY: The actin-related proteins (e.g. ACTRT1, ACTRT2, ACTL7A, and ACTL9) interact with each other to form a multimeric complex in the subacrosomal region of spermatids, which is crucial for the acrosome-nucleus junction. Actrt1-knockout (KO) mice are severely subfertile owing to malformed sperm heads with detached acrosomes and partial fertilization failure. There are currently no reports on the association between ACTRT1 deletion and male infertility in humans.

STUDY DESIGN, SIZE, DURATION: We recruited a cohort of 120 infertile males with sperm head deformations at a large tertiary hospital from August 2019 to August 2023. Genomic DNA extracted from the affected individuals underwent whole exome sequencing (WES), and *in silico* analyses were performed to identify genetic variants. Morphological analysis, functional assays, and ART were performed in 2022 and 2023.

PARTICIPANTS/MATERIALS, SETTING, METHODS: The ACTRT1 deficiency was identified by WES and confirmed by whole genome sequencing, PCR, and quantitative PCR. Genomic DNA of all family members was collected to define the hereditary mode. Papanicolaou staining and electronic microscopy were performed to reveal sperm morphological changes. Western blotting and immunostaining were performed to explore the pathological mechanism of ACTRT1 deficiency. ICSI combined with artificial oocyte activation (AOA) was applied for one proband.

MAIN RESULTS AND THE ROLE OF CHANCE: We identified a whole-gene deletion variant of ACTRT1 in two infertile males, which was inherited from their mothers, respectively. The probands exhibited sperm head deformations owing to acrosomal detachment, which is consistent with our previous observations on Actrt1-KO mice. Decreased expression and ectopic distribution of ACTL7A and phospholipase C zeta were observed in sperm samples from the probands. ICSI combined with AOA effectively solved the fertilization problem in Actrt1-KO mice and in one of the two probands.

LIMITATIONS, REASONS FOR CAUTION: Additional cases are needed to further confirm the genetic contribution of ACTRT1 variants to male infertility.

WIDER IMPLICATIONS OF THE FINDINGS: Our results reveal a gene–disease relation between the ACTRT1 deletion described here and human male infertility owing to acrosomal detachment and fertilization failure. This report also describes a good reproductive outcome of ART with ICSI-AOA for a proband.

STUDY FUNDING/COMPETING INTEREST(S): This work was supported by the Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau, 2023MSXM008 and 2023MSXM054). There are no competing interests to declare.

Received: June 6, 2023. Revised: January 10, 2024. Editorial decision: February 5, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

TRIAL REGISTRATION NUMBER: N/A.

Keywords: male infertility / ACTRT1 deletion / actin-related protein T1 / acrosomal detachment / sperm head deformations / fertilization failure

Introduction

Male infertility is a pathological disease caused by multiple factors, affecting approximately 7% of the global male population (Vander Borgh and Wyns, 2018; Agarwal et al., 2021). Spermatogenesis is a complex and highly ordered process that involves the differentiation of spermatogonia into primary spermatocytes through mitosis, the formation of haploid round spermatids from spermatocytes via meiosis, and then the drastic transformation from round spermatids into spermatozoa, defined as spermiogenesis (Neto et al., 2016).

The perinuclear theca (PT) is a special cytoskeleton that surrounds the nucleus, consisting of two structurally continuous portions (the subacrosomal layer and the post-acrosomal region), which is crucial for the acrosome assembly and sperm head remodeling that occurs during spermiogenesis (Oko and Sutovsky, 2009). Although the structure, composition, and assembly of PT have been well studied, the physiological roles of PT-specific proteins remain largely unknown. Phospholipase C zeta (PLC ζ) is firstly identified as a sperm-borne oocyte activator located in the PT and many variants in phospholipase C zeta 1 (PLCZ1) have been associated with total fertilization failure (Heytens et al., 2009; Escoffier et al., 2016; Yuan et al., 2020). Our recent study on Ccin-knockout (KO) mice has shown that PT also has a physiological function of protecting nuclear structure from shaping forces during the nuclear condensation stage (Zhang et al., 2022b).

Recently, several actin-related proteins (ARPs) that share high amino acid identity and conserved folding domains with conventional actin have been identified as key components of sperm PT, such as actin-related protein T1 (ACTRT1) (Heid et al., 2002), ACTRT2 (Heid et al., 2002), ACTRT3 (Hara et al., 2008), actin-like 7A (ACTL7A) (Boëda et al., 2011; Xin et al., 2020), and ACTL9 (Dai et al., 2021). Increasing evidence indicated that the genetic defects of ARP-encoding genes are causative factors for sperm head deformations, fertilization failure, and male infertility. The destruction of ACTL7A or ACTL9 can lead to acrosomal ultrastructural defects and total fertilization failure in humans and mice (Xin et al., 2020; Dai et al., 2021). The expression and localization of PLC ζ are disrupted in the ACTL7A- or ACTL9-mutant spermatozoa (Xin et al., 2020; Dai et al., 2021; Wang et al., 2021). Our recent study has shown that Actrt1-KO mice are severely subfertile owing to malformed sperm heads with detached acrosomes and partial fertilization failure (Zhang et al., 2022c). ACTRT1 interacts with inner acrosomal membrane proteins and nuclear envelope proteins to mediate the acrosome-nucleus junction (Zhang et al., 2022c). Importantly, ACTRT1, ACTRT2, ACTL7A, and ACTL9 proteins can interact with each other to form a multimeric complex in the subacrosomal region of spermatids, highlighting the important roles of these sperm PT-specific ARPs (Zhang et al., 2022c).

In this study, we recruited 120 infertile males with sperm head deformations and identified a whole-gene deletion variant of ACTRT1 from two Chinese males. We also described the semen parameter characteristics, changes in sperm morphology and ultrastructure, as well as the down-regulated expression of ACTL7A and PLC ζ in the ACTRT1-deficient spermatozoa. ICSI combined with artificial oocyte activation (AOA) was effective for

rescuing the fertilization failure of ART in both Actrt1-KO mice and one ACTRT1-deficient proband. Our study describes for the first time the relation between ACTRT1 deficiency and human male infertility with sperm head deformation, providing information for appropriate genetic counseling and fertility guidance for men with ACTRT1 deficiency.

Materials and methods

Study participants and ethical approval

A total of 120 infertile Han Chinese males with teratospermia characterized by deformed sperm heads were recruited in the Center for Reproductive Medicine, Women and Children's Hospital of Chongqing Medical University (Chongqing, China) from August 2019 to August 2023. The detailed inclusion criteria were as follows: primary infertile; teratospermia characterized by deformed sperm heads, including tapered head, globozoospermia, small head, amorphous head, and pyriform head; normal chromosomal karyotypes (46, XY) without Y chromosome microdeletion; and no other identified pathogenic factors, such as absence of vas deferens, varicocele, viral infection. A written informed consent was signed before the collection of blood and semen samples. This study was approved by the Ethics Committee of Chongqing Health Center for Women and Children.

Whole exome sequencing

Whole exome sequencing (WES) was conducted according to a previously reported procedure (Long et al., 2023). Briefly, a QIAamp® DNA blood mini kit (51106, Qiagen, Hilden, Germany) was used to extract genomic DNA from peripheral blood samples. Then next-generation sequencing and data-analysis were conducted according to a standard operation procedure (Supplementary Fig. S1). Briefly, exomes were captured using the Agilent SureSelect Human All Exon V6 Kit (Agilent Technologies, Santa Clara, CA, USA) and sequenced on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA). The clean reads were filtered and aligned to the human reference genome (GRCh37/hg19) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Single-nucleotide variants (SNVs) and insertion-deletion variants (InDels) were called using Genome Analysis Toolkit (GATK) 4.0 (McKenna et al., 2010) and were functionally annotated with ANNOVAR (Wang et al., 2010). Copy number variants (CNVs) were called by DNA copy R package (Venkatraman and Olshen, 2007) and the genes included in the CNV interval were annotated by ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>), Decipher database (<https://www.deciphergenomics.org/>), and Online Mendelian Inheritance in Man database (OMIM, <https://www.omim.org/>). Moreover, PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>), Sorting Intolerant from Tolerant (SIFT, <http://sift.bii.a-star.edu.sg/>), MutationTaster (<https://www.mutationtaster.org/>), and Combined Annotation Dependent Depletion (CADD, <https://cadd.gs.washington.edu/>) were applied in the prediction of clinical significance of variants.

Whole genome sequencing

For whole genome sequencing (WGS), genomic DNA was extracted from peripheral blood samples and then broken with an ultrasonicator Q800R (Qsonica, Newtown, CT, USA) for random

fragmentation. Subsequently, the constructed library of genomic DNA was sequenced on the AmCareSeq-2000 platform (AmCare Genomics, Guangzhou, China) with a sequencing strategy of 200~500 bp insert size and 150 bp paired-ends. The raw data were cleaned by discarding reads with a base quality less than QC20 and aligned to the human reference genome (GRCh37/hg19) with BWA, as described in WES. The mean and SD of WGS depths for the two probands were 8.45 ± 2.23 and 6.18 ± 1.64 , respectively.

PCR, qPCR, and RT-PCR

The deletion of *ACTRT1* in the human genome was verified through PCR with primers covering the entire gene: 5'-TTCCTTT GCCACACTTTTC-3' (forward) and 5'-TTTGCTTTCCTTGGCTTT GT-3' (reverse). The predicted amplification products were about 1905 bp and analyzed by electrophoresis on 1% agarose gels. The copy number of *ACTRT1* gene in genomic DNA was detected through quantitative PCR (qPCR) with the following primers: forward 5'-CAGGCACACTGAAGGTCTCA-3' and reverse 5'-GGAGTA AAACCCAGCCAACA-3' for the coding sequencing (CDS); forward 5'-TTTCTTCTGGGCTTCAGCAT-3' and reverse 5'-CAGGCCATAG GAAAGAAGA-3' for the 3'-untranslated region (3'-UTR); forward 5'-CTGGTGGACTGAAGGTAGCC-3' and reverse 5'-CCTGGACTTC CCCCAAATAA-3' for the 5'-UTR. The copy number of *ACTRT1* was evaluated using TB Green Premix Ex Taq II (RR820A, Takara, Kyoto, Japan) on an ABI 7500 (Thermo Fisher, MA, USA) and analyzed using the $2^{-\Delta\Delta C_t}$ method by normalizing to that of beta-2-microglobulin gene as the inner control (forward 5'-TGCTGTCTC CATGTTTGATGTATCT-3' and reverse 5'-TCTCTGCTCCCCACCTC TAAGT-3').

The total RNA from mouse tissues was extracted with an RNA extraction kit (9767, Takara, Kyoto, Japan) according to the manufacturer's instructions. Then a RT-PCR kit (RR047A, Takara, Kyoto, Japan) was applied to obtain cDNA samples. The sequences of primers were as follows: *Actrt1*, 5'-GAACCCACAAGA GACCCGAG-3' (forward) and 5'-GAGTCGGGTGAGGTGTTCTG-3' (reverse); *Gapdh*, 5'-GGTGAAGGTCGGTGTGAACG-3' (forward) and 5'-CTCGCTCCTGGAAGATGGTG-3' (reverse).

Semen analysis and papanicolaou staining

As previously reported (Long et al., 2023; Tang et al., 2023), fresh semen samples were collected via masturbation after 2~7 days of abstinence and evaluated after liquefaction for 30 min at 37 °C following the World Health Organization (WHO) guidelines (Cooper et al., 2010; World Health Organization, 2021). A computer-aided sperm analysis system (CFT-9202, Jiangsu Rich Life Science Instrument Co., Ltd, Nanjing, China) was applied to evaluate the semen parameters, including sperm concentration and sperm motility. For Papanicolaou staining, the washed sperm samples were mounted on glass slides, air-dried, and fixed with 95.0% ethanol for 3 min. Then the slides were stained with modified Papanicolaou staining (RS-161, Cariad Medical Technology Co., Ltd, Zhuhai, China) and submerged in hematoxylin for 3 min, acidic ethanol for 5 s, and eosin and bright green for 3 min. More than 200 stained sperm were observed under a light microscope (TCS SP8, Leica, Wetzlar, Germany).

Evaluation of the acrosome reaction

Spontaneous and calcium ionophore-induced acrosome reactions were measured with *Pisum sativum* agglutinin conjugated to fluorescein (PSA-FITC) staining (GWDT01, Goodwe Biotechnology Taizhou Co., Ltd, Jiangsu, China), according to the manufacturer's instructions. Briefly, the washed sperm samples were treated with 10 μ M A23187 to induce the acrosome reaction for 15 min at 37 °C, while treatment with dimethylsulphoxide was

used as an inner control for a spontaneous acrosome reaction. Sperm samples were then stained with 0.5% PSA-FITC for 15 min under dark conditions. More than 5×10^3 sperm were analyzed with a flow cytometer (FACSCalibur, BD Biosciences) for the percentage of sperm with an induced but not a spontaneous acrosome reaction.

Electronic microscopy

Scanning electronic microscopy (SEM) and transmission electronic microscopy (TEM) were performed as previously reported (Long et al., 2023; Tang et al., 2023). Briefly, semen samples from normal controls and the probands were collected and fixed in 2.5% phosphate-buffered glutaraldehyde at 4 °C overnight. For the SEM assay, the samples were deposited on coverslips and dehydrated sequentially in 30%, 50%, 70%, 80%, 90%, 95%, and 100% ethanol. Subsequently, the prepared samples were dried with a CO₂ critical-point dryer (EikoHCP-2, Hitachi, Tokyo, Japan), sputter coated using an ionic sprayer meter (ACE200, Leica), and finally observed under the SEM (Nova NanoSEM 450, FEI, Hillsboro, OR, USA) at an accelerating voltage of 5 kV. For the TEM assay, the samples were post-fixed in 1% buffered osmium tetroxide, dehydrated in graded acetone solutions, and embedded in Epon 812 (90529-77-4, SPI, West Chester, PA, USA). Subsequently, ultrathin (70 nm) sections were double stained with lead citrate and uranyl acetate, and observed by a TEM (TECNAI-10, Philips, Eindhoven, The Netherlands) at an accelerating voltage of 80 kV.

Immunofluorescence

After permeabilization with 1% Triton X-100 for 30 min, the slides of human spermatozoa were blocked by 5% bovine serum albumin for 45 min. Samples were incubated with rabbit anti-ACTRT1 (1:200, PA5-31691, Invitrogen, CA, USA), anti-ACTRT2 (1:200, 16992-1-AP, Proteintech, Wuhan, China), anti-ACTL7A (1: 400, 17355-1-AP, Proteintech, Wuhan, China), or anti-PLC ζ (1:100, ab181816, Abcam, Cambridge, UK) antibodies overnight at 4 °C. After washing with phosphate buffer three times, slides were incubated with Alexa Fluor 647-labeled goat anti-rabbit IgG (H + L) (1:500, A0468, Beyotime, Shanghai, China) at room temperature for 1 h. Peanut agglutinin (PNA)-FITC (L7381, Sigma-Aldrich, MO, USA) was used to label the acrosome and the nuclei were further counterstained with DAPI (1:500, C1002, Beyotime, Shanghai, China). Photographs were taken using a Zeiss LSM780 confocal microscope (Carl Zeiss, Batenfuburg Prefecture, Germany).

Western blot

Total proteins from human sperm were extracted according to a previously reported procedure (Zhang et al., 2022c). Briefly, sperm were cracked with RIPA buffer (C1055, Applygen, Beijing, China) containing 1 mM phenylmethylsulphonyl fluoride and protease inhibitors on ice, and the supernatants were collected following centrifugation at 12 000g for 20 min at 4 °C. Proteins were electrophoresed in 10% sodium dodecyl sulphate-polyacrylamide gels and transferred to nitrocellulose membranes (GE Healthcare, WI, USA). The blots were blocked with 5% skim milk and incubated with the rabbit anti-ACTRT1, anti-ACTRT2, anti-ACTL7A, or anti-PLC ζ antibodies overnight at 4 °C, followed by incubation with anti-rabbit IgG H&L (HRP) (M212115, Abmart, Beijing, China) at a 1/10 000 dilution for 1 h. The signals were evaluated using the Super ECL Plus Western Blotting Substrate (P1050, Applygen, Beijing, China) and a chemiluminescence imaging system (Tanon, Shanghai, China).

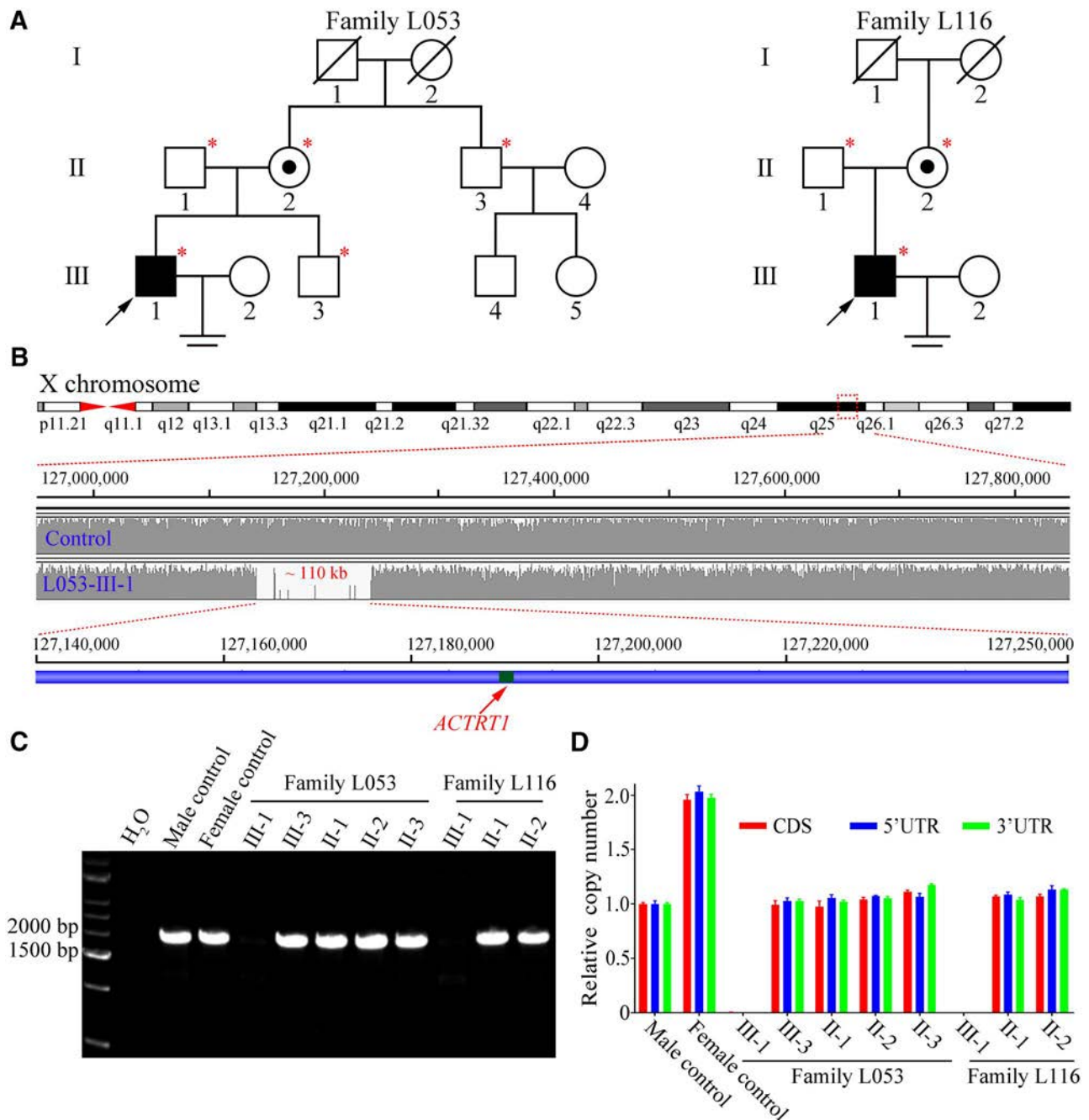


Figure 1. ACTRT1 gene located in X chromosome was deleted in the probands. **(A)** Pedigrees of the patients with male infertility. Male family members are denoted by squares and female family members are denoted by circles. Deceased family members are denoted by slashes. The infertile patients are denoted by black filled squares. Arrow indicates the probands (L053-III-1 and L116-III-1). The red asterisk indicates family members who have undergone validation for the presentation of actin-related protein T1 (ACTRT1) gene. **(B)** Whole genome sequencing (WGS) verification of the ~110-kb deletion in Chromosome X from the proband L053-III-1. WGS demonstrated the location and length of deleted variants in the proband, which only consisted of the ACTRT1 gene. **(C)** Image of agarose gel electrophoresis of the PCR products from the probands, their family members in Family L053 and L116, as well as H₂O, normal male and female controls. The full ACTRT1 gene is about 1.44 kb long and the amplicon covered the whole ACTRT1 gene. The absence of a specific band showed the deletion of ACTRT1 in the probands (L053-III-1 and L116-III-1). **(D)** Relative copy numbers of ACTRT1 gene in the probands, their family members, as well as normal male and female controls. CDS, coding sequence; UTR, untranslated region. The B2M (beta-2-microglobulin) gene was used as an internal control. As ACTRT1 is located at chromosome X, a normal male subject has only one copy while two copies are present in the normal female control. The zero copy confirmed the deletion of ACTRT1 in the probands. Only one copy in the probands' mothers demonstrated inheritance of the variant from the mothers (L053-II-2 and L116-II-2) to the probands (L053-III-1 and L116-III-1).

Mouse ICSI and AOA

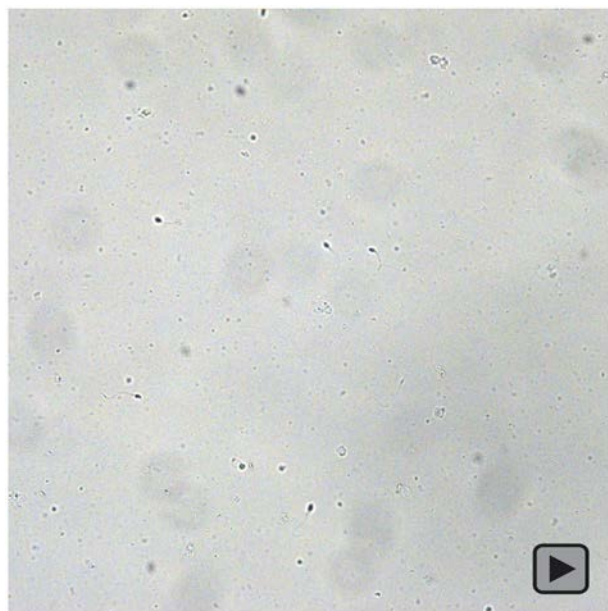
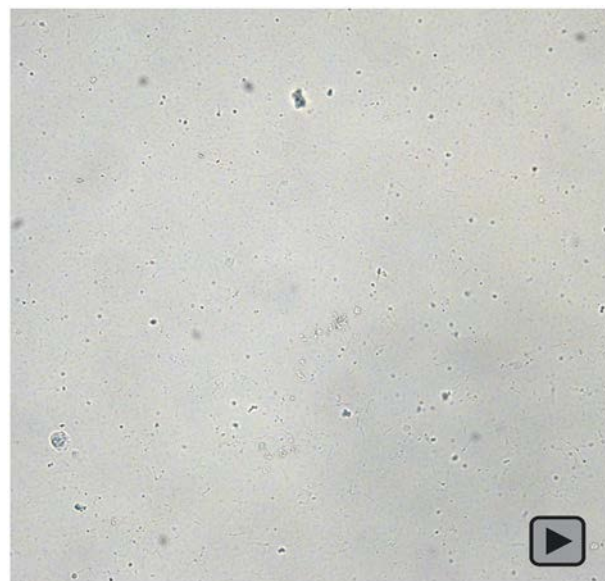
Animal experiments were approved by the Animal Care and Use Committee of the College of Life Sciences, Beijing Normal University. As in our previous protocol (Zhang et al., 2022b), 8-week-old C57BL/6J female mice were super-ovulated by injecting

5 IU pregnant mare's serum gonadotrophin combined with 5 IU hCG (48 h later). Oocytes were obtained from the ampulla of the uterine tube 14 h after hCG injection. Mouse sperm heads were separated from sperm tails and injected into mouse oocytes using a Piezo driven pipette (PrimeTech, Osaka, Japan). The injected

Table 1. Semen analysis using CASA in the probands with ACTRT1 deficiency.

Characteristics	L053 III-1 [#]	L116 III-1 [#]	Control*	Reference value
Age (mean ± SD, years)	25	36	32.2 ± 4.2	–
Duration of infertility (years)	1	7	–	–
Semen parameters				
Semen volume (mean ± SD, ml)	2.8 ± 0.9	2.0 ± 0.8	3.9 ± 1.3	≥1.4
Semen pH (mean ± SD)	7.3 ± 0.1	7.3 ± 0.1	7.3 ± 0.1	≥7.2
Sperm concentration (mean ± SD, 10 ⁶ /ml)	2.0 ± 0	1.0 ± 0	64.0 ± 39.6	≥16
Total sperm count (mean ± SD, 10 ⁶ /ejaculate)	6.0 ± 1.7	2.0 ± 1.4	237.9 ± 157.3	≥39
Sperm motility (mean ± SD, %)	27.0 ± 11.5	36.5 ± 5.0	60.5 ± 13.1	≥42
Progressive motility (mean ± SD, %)	18.0 ± 9.5	36.5 ± 5.0	52.5 ± 12.5	≥30
Immotility (mean ± SD, %)	73.0 ± 11.5	63.5 ± 5.0	39.6 ± 13.1	≥20
Normal sperm morphology ^{&} (mean ± SD, %)	2.33 ± 0.3	1.67 ± 0.8	21.0 ± 2.3	≥4
Sperm acrosome reaction (mean ± SD, %)	9.2	–	44.0 ± 25.9	–
Sperm locomotion parameters				
Curvilinear velocity (VCL) (mean ± SD, μm/s)	20.1 ± 9.0	27.0 ± 30.6	28.4 ± 7.2	–
Straight-line velocity (VSL) (mean ± SD, μm/s)	28.4 ± 14.9	29.3 ± 26.7	27.1 ± 7.4	–
Average path velocity (VAP) (mean ± SD, μm/s)	14.2 ± 6.1	13.0 ± 14.9	18.9 ± 5.5	–
Amplitude of lateral head displacement (ALH) (mean ± SD, μm)	1.9 ± 0.9	2.2 ± 2.6	2.7 ± 0.8	–
Linearity (mean ± SD)	1.4 ± 0.2	1.5 ± 0.6	1.0 ± 0.2	–
Wobble (mean ± SD, WOB, = VAP/VCL)	0.7 ± 0	0.5 ± 0	0.7 ± 0.1	–
Straightness (mean ± SD, STR, = VSL/VAP)	2.0 ± 0.2	3.2 ± 1.5	1.5 ± 0.2	–
Beat-cross frequency (mean ± SD, BCF, Hz)	4.4 ± 2.2	5.5 ± 7.8	9.9 ± 3.9	–

* 100 cases of normal males.

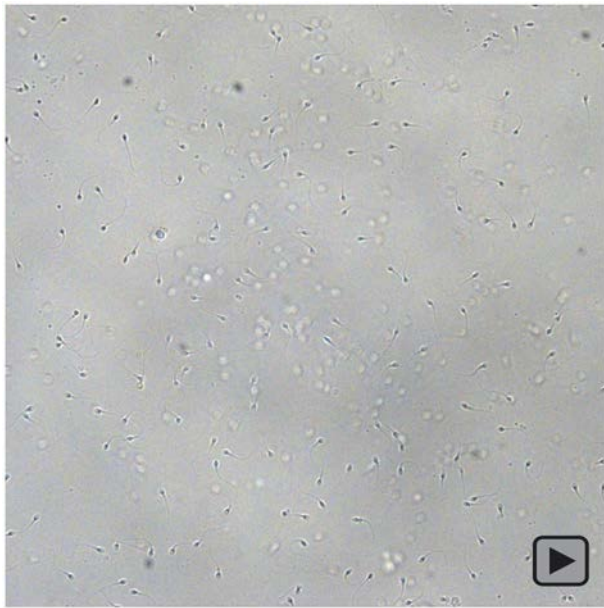
[#] Three different semen samples.[&] Three sets of sperm smears were collected for each group and a total of 200 sperm were counted in each sperm smear. –, no reference range; CASA, computer-assisted sperm analysis; ACTRT1, actin-related protein T1. The reference values refer to the World Health Organization laboratory manual for the examination and processing of human semen. 6th Edition.**Video 1. Semen parameters analysis of the proband (L053-III-1).** The video showed that the number of sperm with progressive motility from the proband is extremely small. The video was taken at 20× magnification under a light microscope.**Video 2. Semen parameters analysis of the proband (L116-III-1).** The video showed that the sperm concentration from the proband is extremely low. The video was taken after 20× magnification under a microscope.

oocytes were cultured in KSOM medium (M1430, Nanjing Aibei Biotechnology, Nanjing, China) at 37°C under 5% CO₂. For AOA, the injected oocytes from mice were exposed to 10 mM SrCl₂ (255521, Sigma, CA, USA) in KSOM medium for 6 h after ICSI. Subsequently, the oocytes were washed twice with KSOM medium and then cultured at 37°C and 5% CO₂. Two-cell embryos and blastocysts were counted 20 and 96 h later, respectively.

Ovarian stimulation and fertilization

A short-acting GnRH agonist (GnRH-a) long protocol was used for controlled ovarian stimulation. However, the wife of the second

proband (L116-III-1) was receiving treatment for chronic endometritis, and therefore ART was not undertaken. The patient (L053-III-2) received a daily s.c. injection of 0.1 mg short-acting GnRH-a (decaptyl, Ferring, Switzerland) from the midluteal phase of the preceding menstrual cycle for down-regulation of the pituitary gland. After 20 days of down-regulation, the levels of serum estradiol, LH, and progesterone were less than 30 pg/ml, 5 mIU/ml, and 0.9 ng/ml, respectively. Then the gonadotrophin (recombinant FSH) was administered with an initial dose of 150 IU, while monitoring serum hormone changes and follicle growth trends. Recombinant hCG (Merck Serono, Geneva, Switzerland) at a



Video 3. Semen parameters analysis of the normal control. The video showed that most of the sperm from the normal control were progressively swimming. The video was taken after 20× magnification under a microscope.

dose of 250 µg was administered to trigger the maturation of oocytes, and oocyte retrieval was performed at 34–36 h later. Consequently, 13 oocytes were obtained, including 10 metaphase II (MII) oocytes. After cumulus cell removal, ICSI were performed as previously described (Zhang *et al.*, 2022a). For AOA, the injected oocytes were transferred to G-IVF-plus medium (Vitrolife, Sweden) with 10 µM calcium ionophore (A23187, Sigma) and incubated for 10 min. After washing with G-IVF-plus medium, the oocytes were transferred to G1-plus medium (Vitrolife, Sweden) droplets for further culture. As previously reported (Lin *et al.*, 2023), a time-lapse monitoring system (Embryoscope Plus, Vitrolife, Sweden) was used for embryo culture and the embryo quality assessment was performed according to the previously reported criteria (Gardner and Lane, 1997; Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011).

Frozen embryo transfer

The patient (L053-III-2) was administered oral estradiol valerate (Progynova, Bayer, Germany) 4 mg daily for 14 days until the endometrial thickness reached 7 mm. Luteal phase support was started with 90 mg daily of vaginal progesterone (Crinone, Merck, Germany) 5 days before frozen embryo transfer. The patient underwent single blastocyst transfer and continued to be administered oral estradiol valerate plus vaginal and oral progesterone for intensive luteal phase support for 14 days.

Statistical analysis

The data are presented as the mean and SEM or the mean and SD, and Student's *t*-test (unpaired, two-tailed) was used for comparing the difference between groups. The differences were considered statistically significant at $P < 0.05$. The statistical analysis was performed with GraphPad Prism version 5.01 (Graph Pad Software, San Diego, CA, USA).

Results

Deletion of ACTRT1 was identified in two infertile men

To date, we have performed WES in a cohort of 120 affected individuals and screened out several reported genetic variants associated with sperm head deformations, including dpy-19 like 2 (DPY19L2) and centromere protein E (CENPE). We identified a novel genetic variant potentially related to sperm head deformations in two infertile males from two unrelated families through systematic bioinformatic analysis (Fig. 1A and Supplementary Fig. S1 and Table S1). Although we screened out five and seven candidate genes in the L053-III-1 and L116-III-1 patients, respectively (Supplementary Table S2), we only identified the microdeletion of X chromosome associated with head deformations in both probands after retrieving the OMIM database and literature. The two probands, L053-III-1 and L116-III-1, were 25 and 36 years of age, respectively, and both had normal hormone levels and testicular volume. Owing to the limitations of WES in detecting CNVs, the microdeletions were furtherly verified with WGS and defined from 127 140 001 to 127 250 001 in GRCh37 with a length of about 110 kb (Fig. 1B and Supplementary Fig. S2 and Table S3) in both probands. Intriguingly, the deleted fragment was located at X chromosome q25, and covered only one protein-coding gene, namely ACTRT1. The frequency of this CNV in the human Database of Genomic Variants is as low as 0.0016 and has not been described in the literature.

The ACTRT1 gene was mapped to 127 184 941–127 186 382 in GRCh37 and was intron-free with one coding exon (Heid *et al.*, 2002). RT-PCR of different mouse tissues showed that *Actrt1* was specifically and highly expressed in testis (Supplementary Fig. S3A). Within mouse testes, the expression of *Actrt1* begins from the third postnatal week, during which A1 spermatogonia transform into preleptotene spermatocytes, and reaches the peak in the seventh week, during which round spermatids transform into elongated spermatids (Supplementary Fig. S3A) (Kluin *et al.*, 1982; Griswold, 2016). Besides, according to the human protein atlas, ACTRT1 mRNA was specifically enriched in human testis tissues and mainly expressed in early/late spermatids (<https://www.proteinatlas.org/ENSG0000012316-5-ACTRT1>). We have recently reported that the *Actrt1*-KO male mice were severely subfertile owing to acrosomal detachment and partial fertilization failure (Zhang *et al.*, 2022c).

Males have an X chromosome corresponding to one copy of the ACTRT1 gene in genomic DNA while females have two copies of ACTRT1. PCR with specific primers covering the full-length of ACTRT1 showed the presence of ACTRT1 in the fathers of the probands (L053-II-1 and L116-II-1) and mothers (L053-II-2 and L116-II-2), with absence of ACTRT1 in the probands (L053-III-1 and L116-III-1) (Fig. 1C). The proband's brother (L053-III-3) and uncle (L053-II-3) also carried the ACTRT1 gene normally. To explain the origin of this deletion, qPCR with primers for coding sequence and untranslated regions was performed. The copy numbers of ACTRT1 in the probands (L053-III-1 and L116-III-1), their fathers (L053-II-1 and L116-II-1), and their mothers (L053-II-2 and L116-II-2) were 0, 1, and 1, respectively (Fig. 1D). This result indicated that their mothers probably carry an X chromosome with a deletion of ACTRT1. Thus, we speculated that absence of the maternally inherited ACTRT1 gene in the male progeny might cause teratospermia and infertility in humans.

Morphological abnormalities of the sperm heads in the probands

A detailed description of physical characteristics and semen parameters of the probands are presented in Table 1. According

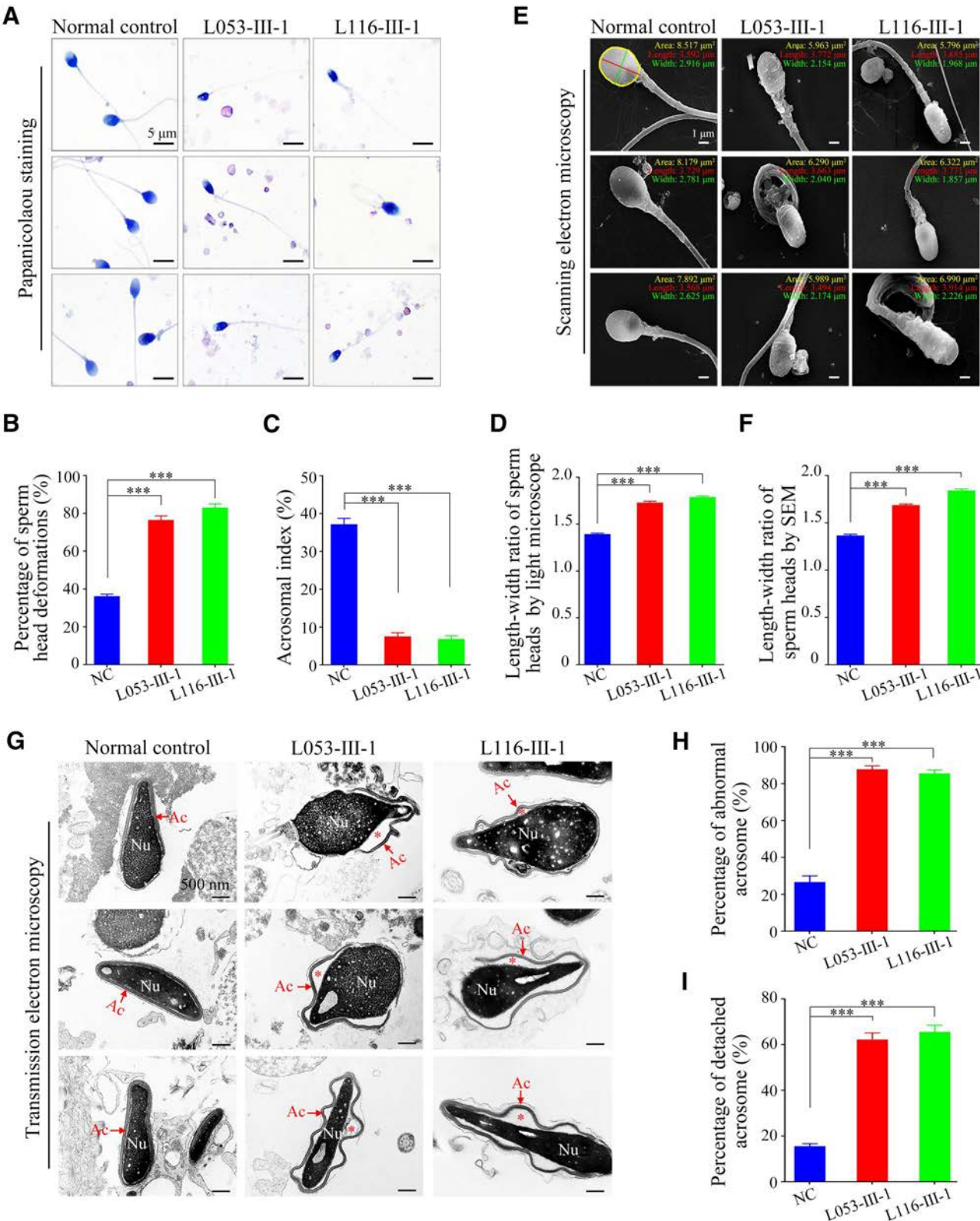


Figure 2. Sperm head are deformed in the probands owing to acrosomal detachment. (A) Representative images of sperm using a Papanicolaou staining from the probands (L053-III-1 and L116-III-1) and the healthy control. Staining experiments were performed for three times and representative images were provided. Scale bar, 5 μm . (B) The percentage of sperm with head deformations in the probands and the normal control. Three sets of semen samples were collected from each group and a total of 200 sperm were counted in each experiment. Data are mean $\pm$ SEM, Student's t-test, *** $P < 0.001$. (C) Acrosomal index in the probands and the normal control. acrosomal index (%) = (spermatozoa with normal acrosomes/total number of spermatozoa evaluated) $\times 100$. Three sets of semen samples were collected from each group and a total of 200 sperm were counted in each experiment. Data are mean $\pm$ SEM, Student's t-test, *** $P < 0.001$. (D) The length-width ratio of sperm heads by light microscope in the probands and the normal control. Three sets of semen samples were collected from each group and a total of 20 sperm were counted in each experiment. Data are mean $\pm$ SEM, Student's t-test, *** $P < 0.001$. (E) Representative images of sperm using scanning electronic microscopy (SEM) from the probands and the healthy control. The parameters of area, length, and width of sperm heads were analyzed by using the SEM measuring tool. SEM experiments were performed

(continued)

to the WHO guidelines, the values for sperm concentration and normal sperm morphology of probands (L053-III-1 and L116-III-1) were all below the minimum reference. In addition, the total sperm motility and progressive motility of proband (L053-III-1) was also below the minimum reference value (Table 1 and Videos 1, 2, and 3), defining the probands L053-III-1 and L116-III-1 as having oligoasthenoteratozoospermia (OAT) and oligoteratozoospermia, respectively. Besides, sperm with a normal acrosomal reaction accounted for 9.2% in proband L053-III-1, which was significantly lower than normal controls (44.0%) (Table 1).

To investigate the effect of ACTRT1 deletion on sperm morphology, Papanicolaou staining and SEM were performed. Sperm from the normal control showed oval heads with length-width ratios of 1.2~1.4:1 and ~45 μm -length flagella. The acrosome covered more than one-third of the head area (Fig. 2A). Compared with the typical morphology of normal sperm, sperm from the probands exhibited obvious head deformations, mainly showing long and thin heads, along with a smaller coverage area of acrosomes (Fig. 2A). No obvious morphological abnormalities were identified in flagella of the probands' sperm. Statistical analysis indicated a significant increase in the ratio of spermatozoa with head deformations, a noteworthy decrease in the acrosomal index, and a significant increase in the length-width ratio of sperm heads (assessed by light microscopy) in the probands compared with the control group (Fig. 2B–D). The parameters of area, length, and width of spermatozoa heads were further analyzed with the measure tool in SEM. The average head area of sperm from the normal control was $8.196 \pm 0.181 \mu\text{m}^2$ with an average length and width of $3.630 \pm 0.050 \mu\text{m}$ and $2.774 \pm 0.084 \mu\text{m}$, respectively (Fig. 2E). On the contrary, the average head area of sperm from the probands (L053-III-1 and L116-III-1) was $6.081 \pm 0.105 \mu\text{m}^2$ and $6.369 \pm 0.346 \mu\text{m}^2$, with an average length of $3.643 \pm 0.081 \mu\text{m}$ and $3.857 \pm 0.044 \mu\text{m}$, and an average width of $2.123 \pm 0.042 \mu\text{m}$ and $2.017 \pm 0.109 \mu\text{m}$, respectively. Statistical analysis indicated that the length-width ratios of sperm heads in the probands under SEM were significantly greater than that in the normal control (Fig. 2F). Collectively, these data suggest that ACTRT1 deletion is highly associated with sperm head deformations in humans.

Acrosomal detachment of the ACTRT1-deficient sperm

To explore the underlying cause of sperm head deformations in the probands, we further examined the ultrastructural abnormalities of sperm heads under TEM. In the normal control, acrosomes were closely attached to the nucleus; however, most of the probands' sperm exhibited acrosomal detachment (Fig. 2G). The percentage of sperm with acrosomal abnormalities was significantly higher in the probands compared with the normal control (Fig. 2H). The number of sperm with detached acrosomes in the probands was significantly more than that in the normal control (Fig. 2I). Taken together, acrosomal detachment is a major cause of head deformations identified in the proband's sperm,

which is consistent with our previous findings in *Actrt1*-KO mice (Zhang et al., 2022c).

Reduced expression and ectopic distribution of ACTL7A and PLC ζ

To further elucidate the effect of ACTRT1 deletion, immunofluorescence staining and western blotting analysis were performed with sperm samples of the probands and normal control. As expected, the specific localization of ACTRT1 at the post-acrosomal region of heads was absent in the probands' sperm (Fig. 3A). The distribution of ACTRT2 at the post-acrosomal region of sperm heads was similar in the probands and the normal control (Fig. 3A). In normal control sperm heads, ACTL7A was localized at both the subacrosomal layer and the post-acrosomal region while PLC ζ signals were distributed at the subacrosomal layer (Fig. 3A). However, their distributions were seriously disrupted in the proband sperm heads, showing ectopic expression as dots (Fig. 3A). As revealed by western blotting, ACTRT1 protein was absent in the probands' sperm samples and ACTRT2 level in sperm samples was similar in the probands and the normal control (Fig. 3B–D). The expression of ACTL7A and PLC ζ was significantly attenuated in the proband sperm samples compared with the normal control (Fig. 3B–D). These findings indicated that the absence of ACTRT1 resulted in the reduced expression and ectopic distribution of ACTL7A and PLC ζ in the probands' sperm, which suggested the possibility of fertilization failure in ART.

ICSI combined with AOA can overcome the fertilization failure

Our previous study indicated an extremely low rate of IVF using sperm from *Actrt1*-KO mice (Zhang et al., 2022c). Before the implementation of ART for the proband (L053-III-1), we first performed ICSI in *Actrt1*-KO mice to provide theoretical clues. The percentages of two-cell embryos and blastocysts using sperm from *Actrt1*-KO mice ($38.84 \pm 1.05\%$ and $35.18 \pm 8.07\%$, respectively) were significantly lower compared with those using wild-type sperm ($89.40 \pm 0.99\%$ and $70.20 \pm 3.53\%$, respectively) (Supplementary Fig. S3B and C). To solve the problem of low fertilization rate, we applied ICSI combined with AOA in *Actrt1*-KO mice. Outstandingly, 56 out of 71 oocytes were fertilized by sperm from *Actrt1*-KO mice and 40 out of these 56 embryos developed to the blastocyst stage. The percentages of two-cell embryos and blastocyst in ICSI-AOA group ($79.00 \pm 1.46\%$ and $71.67 \pm 3.70\%$, respectively) were significantly higher than that using conventional ICSI (Supplementary Fig. S3B and D). The study of ART in *Actrt1*-KO mice suggested that ICSI-AOA may be an effective strategy for the ACTRT1-deficient patients.

A detailed description of physical characteristics of the proband's wife (L053-III-2) and the clinical outcomes of ART are presented in Table 2. L053-III-2 was 27 years old, with normal ovarian reserves and BMI. Following ovarian stimulation with a short-acting GnRH-a long protocol, a total of 13 oocytes were retrieved during follicular puncture, of which 10 were in metaphase II (MII) stage. After ICSI manipulation for 30 min, the

Figure 2. Continued

three times and representative images are provided. Scale bar, 1 μm . (F) The length-width ratio of sperm heads in the probands and the normal control assessed by SEM. Three sets of semen samples were collected from each group and a total of 20 sperm were counted in each experiment. Data are mean $\pm$ SEM, Student's t-test, *** $P < 0.001$. (G) Transmission electronic microscope (TEM) analysis showing the detached acrosome in the probands' sperm. Arrows point to acrosomes and asterisks indicate the enlarged space between the acrosome and nucleus. Ac, acrosome; Nu, nucleus. Scale bars, 500 nm. (H) Percentage of sperm with acrosome abnormalities (total) in the probands and the normal control as revealed by TEM. Acrosome abnormalities included no acrosome, less acrosome, detached acrosome, deformed acrosome, and other types. Three sets of semen samples were collected from each group and a total of 30 sperm were counted in each experiment. Data are mean $\pm$ SEM, Student's t-test, *** $P < 0.001$. (I) Percentage of sperm with detached acrosomes as revealed by TEM. Three sets of semen samples were collected from each group and a total of 30 sperm were counted in each experiment. Data are mean $\pm$ SEM, Student's t-test, *** $P < 0.001$.

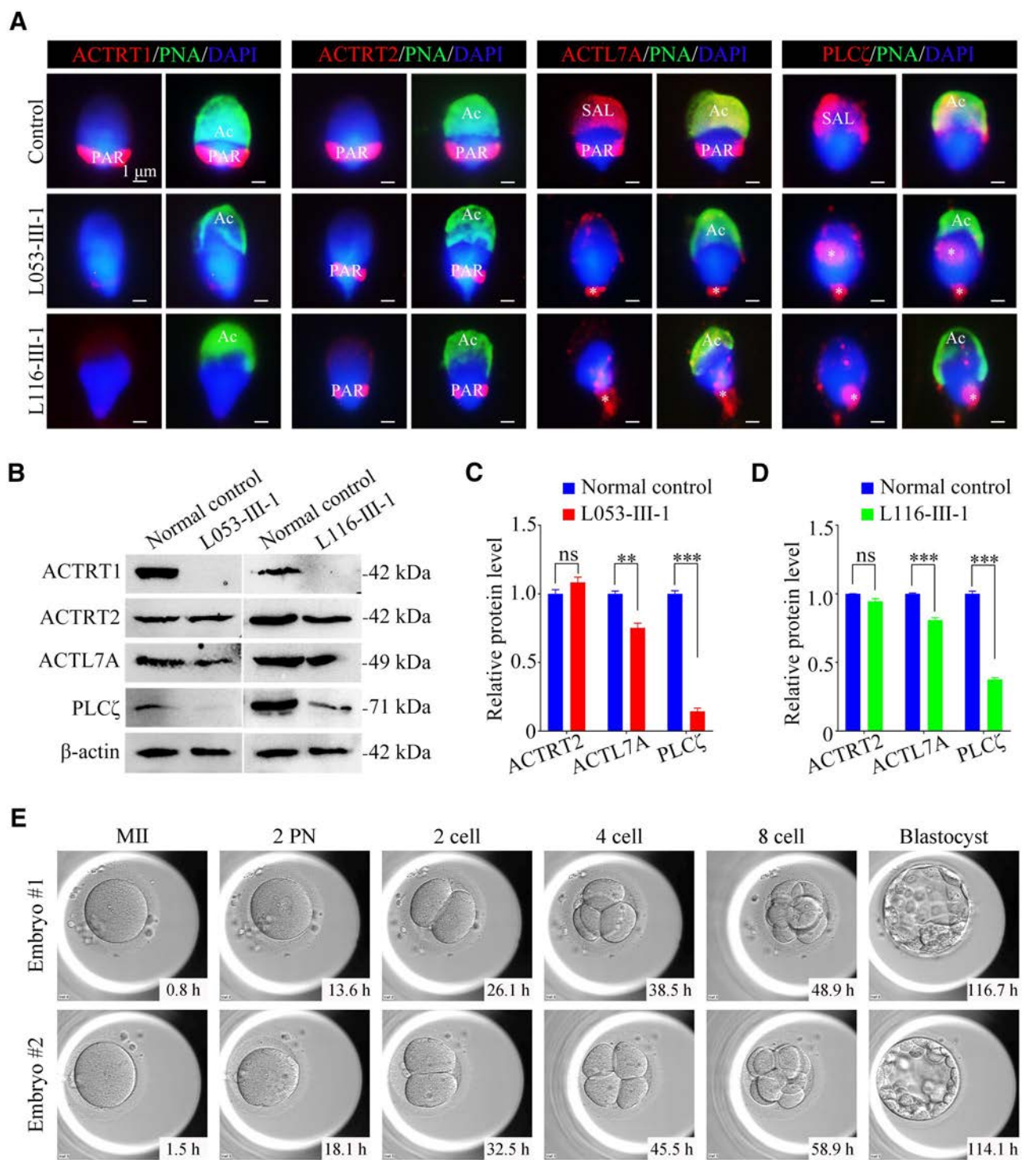


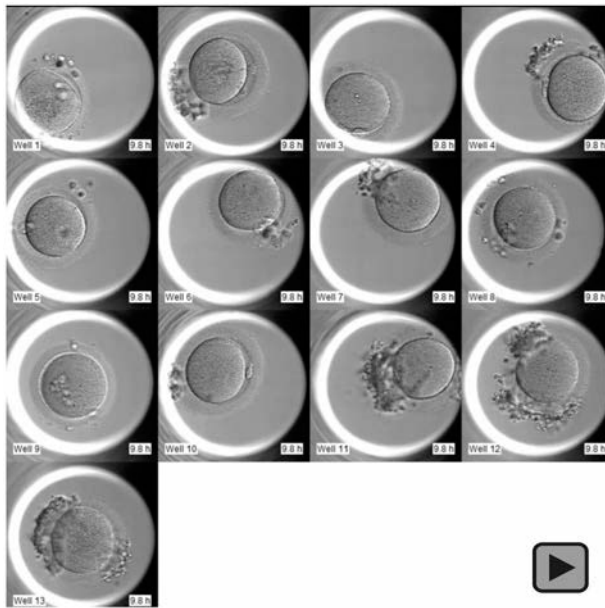
Figure 3. Analysis of sperm proteins and embryo development following ICSI combined with assisted oocyte activation using sperm from proband L053-III-1. (A) Immunofluorescence staining of sperm heads using anti-actin-related protein T1 (ACTRT1), ACTRT2, actin like 7A (ACTL7A), or phospholipase C zeta (PLC ζ) antibodies together with peanut agglutinin (PNA)-FITC (an acrosome fluorescence dye). The nucleus was counterstained with DAPI. Ac, acrosome; SAL, subacrosomal layer; PAR, post-acrosomal region. Asterisks indicate the ectopic distribution of dots of ACTL7A and PLC ζ . Scale bar, 1 μ m. (B) Western blotting analysis of ACTRT1, ACTRT2, ACTL7A, and PLC ζ expression in sperm samples of the probands and the normal control. Three sets of semen samples were collected from each group and representative images are provided. β -actin served as the internal control. (C) The relative protein levels of ACTRT1, ACTRT2, ACTL7A, and PLC ζ in sperm samples of the L053-III-1 proband and the normal control. Gray values of the protein bands were measured by Image J software. Data are mean $\pm$ SEM, Student's t-test, ** P < 0.01; *** P < 0.001; ns, not significant. (D) The relative protein levels of ACTRT1, ACTRT2, ACTL7A, and PLC ζ in sperm samples of the L116-III-1 proband and the normal control. Gray values of the protein bands were measured by Image J software. Data are mean $\pm$ SEM, Student's t-test, *** P < 0.001; ns, not significant. (E) ICSI combined with assisted oocyte activation (AOA) improves the fertilization of oocytes using the proband's sperm. After ICSI combined with AOA, 80% of metaphase II (MII) oocytes were successfully fertilized and six zygotes formed blastocysts. The figure shows the development process of two MII oocytes after fertilization.

Table 2. The clinical outcomes of ART in the wife of the affected man with the ACTRT1 deletion.

Characteristics	L053-III-2
Basal information	
Female age (years)	27
BMI (kg/m ²)	21.48
AMH (ng/ml)	5.65
FSH (mIU/ml)	2.37 [#]
LH (mIU/ml)	4.02 [#]
E2 (pg/ml)	142.08 [#]
Prog (ng/ml)	20.6 [#]
PRL (ng/ml)	22.21
Ovarian stimulation	
Protocol	Short-acting GnRH-agonist long protocol
E2 level on the trigger day (pg/ml)	3691
No. of follicles ≥14 mm on the trigger day	15
No. of follicles ≥18 mm on the trigger day	10
Laboratory indicators	
No. of oocytes retrieved	13
No. of MII oocytes (%)	10/13 (76.92)
Insemination method	ICSI combined with AOA
No. of 2PN (%)	8/10 (80.0)
No. of transferable embryos (%)	8/8 (100)
No. of blastocysts (%)	6/8 (75.0)
No. of embryo transfer	1

AMH, anti-Müllerian hormone; E2, estradiol; Prog, progesterone; PRL, prolactin; 2PN, two-pronuclei; MII, metaphase II; -, no reference range. AOA, artificial oocyte activation.

[#] Sex hormone levels in the luteal phase.

**Video 4.** Embryo cultures as revealed by a time-lapse monitoring system. In the wells numbered 1, 2, 3, 5, 6, and 8, the metaphase II (MII) oocytes eventually formed blastocysts. In the wells numbered 9 and 10, the MII oocytes did not form blastocysts. In the wells numbered 4 and 7, the MII oocytes failed to be fertilized. In the wells numbered 11, 12, and 13, the oocytes are metaphase I and germinal vesicle oocytes, which were not used for insemination.

injected oocytes were artificially activated by treatment with 10 μ M A23187 for 10 min. A time-lapse monitoring system showed that eight MII oocytes achieved successful fertilization with two pronuclei. Subsequently, all embryos successfully completed cleavage and developed into transferable embryos on Day 3. Finally, six blastocysts were obtained and subjected to

vitrification cryopreservation, owing to the risk of ovarian hyperstimulation syndrome (Fig. 3E and Video 4). Two months later, the couple completed the single blastocyst transfer, and the wife achieved a clinical pregnancy. These results indicate that ICSI combined with AOA may be an effective strategy for infertile males affected by ACTRT1 defects.

Discussion

In this study, we found that the deletion of ACTRT1 was associated with sperm head deformations, mainly characterized by acrosomal detachment. A 110-kb microdeletion of the X chromosome, only including the ACTRT1 gene, was identified in two infertile men (L053-III-1 and L116-III-1) from two unrelated families. Furthermore, abnormal levels and distributions of ACTL7A and PLC ζ proteins were identified in sperm from the two probands. Importantly, the infertility of *Actrt1*-KO mice and the proband (L053-III-1) could be overcome by applying a treatment of ICSI combined with AOA.

ACTRT1, ACTL7A, and ACTL9 proteins interact with each other to form sperm PT-specific ARPs., and KO mice all exhibit male infertility/subfertility with acrosomal ultrastructural defects and total/partial fertilization failure (Xin et al., 2020; Dai et al., 2021; Zhang et al., 2022c). Clinical studies have reported that genetic variants in ACTL7A and ACTL9 were associated with primary male infertility (Xin et al., 2020; Dai et al., 2021; Wang et al., 2021; Zhou et al., 2023). Reduced expression and an abnormal distribution of PLC ζ in sperm heads are observed in ACTL7A or ACTL9-mutant patients, explaining the fertilization failure (Xin et al., 2020; Dai et al., 2021; Wang et al., 2021). Both ACTL7A- and ACTL9-associated infertility could be successfully overcome with ICSI together with AOA (Xin et al., 2020; Dai et al., 2021; Wang et al., 2021; Zhang et al., 2022c). Our previous study discovered a phenotype of sperm acrosomal detachment in *Actrt1*-KO mice, and that protein levels of ACTL7A and PLC ζ were significantly lower in *Actrt1*-KO sperm than in wild-type sperm (Zhang et al., 2022c). Additionally, ACTRT1 interacts with the inner acrosomal membrane protein SPACA1 and the nuclear envelope proteins PARP11 and SPATA46 to mediate the acrosome-nucleus connection (Zhang et al., 2022c). IVF rate is extremely low in *Actrt1*-KO mice and attenuated expression of ACTL7A and PLC ζ is suggested to be a potential cause of the fertilization defects (Zhang et al., 2022c).

Taking into consideration the low fertilization rate of IVF and ICSI in *Actrt1*-KO mice, as well as the abnormal expression and distribution of ACTL7A and PLC ζ protein in the sperm of probands, ICSI combined with AOA was applied for proband L053-III-1 and a good outcome was achieved. In future, a large cohort is needed to further confirm the genetic contribution of ACTRT1 to male infertility. Moreover, the outcomes of ICSI+AOA or ICSI alone should be reported for more ACTRT1-mutant/deletion patients.

A recent study reported that ACTRT1 variants in infertile males were associated with acephalic spermatozoa syndrome (ASS) (Sha et al., 2021). However, we did not observe the ASS phenomenon in the sperm of probands or in the sperm of *Actrt1*-KO mice. Headless spermatozoa were originally mis-described as 'pin head spermatozoa' because a residual droplet of cytoplasm and misarranged mitochondria attached to the top of the flagella were usually considered to be shrunken head. Variants in Sad1 and UNC84 domain containing 5 (SUN5) (Zhang et al., 2021a), spermatogenesis associated 6 (SPATA6) (Yuan et al., 2015), polyamine modulated factor 1 binding protein 1 (PMFBP1) (Zhu et al.,

2018), testis specific 10 (TSGA10) (Sha et al., 2018), and the centlein (CNTLN) gene (Zhang et al., 2021b) have been reported to be related to ASS. Indeed, the residual cytoplasm and misarranged mitochondria can be observed at the top of flagella, and the axoneme and outer dense fibers of decapitated tails are also abnormal in affected patients or their corresponding gene-edited mice (Yuan et al., 2015; Zhu et al., 2018; Zhang et al., 2021a,b). However, these phenomena are not observed in *Actrt1*-KO sperm in the Sha et al., (2021) study. In addition, the head-tail integrity for sperm depends on the head-to-tail coupling apparatus (HTCA) and defective HTCA is suggested to be the primary cause of ASS; however, the relation between PT protein ACTRT1 and HTCA structure is far from clear. Besides, functional experiments are needed to verify the pathogenicity of identified missense variants in ACTRT1 in the Sha et al. (2021) study.

In conclusion, our study provides the first direct evidence that a deficiency in ACTRT1 causes male infertility with sperm head deformations. Acrosomal detachment and attenuated expression/ectopic distribution of PLC ζ and ACTL7A are observed in the sperm of two probands, consistent with our previous observations in *Actrt1*-KO mice. Importantly, ICSI combined with AOA could overcome the fertilization problem of one proband, which may guide clinical treatment strategies.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data in this article will be shared with upon reasonable request to the corresponding author.

Acknowledgements

The authors thank all the participants of this study.

Authors' roles

T.L., S.C., and G.H. conceived and designed the study. H.J. and S.L. performed the biochemistry experiments and bred the mice. Q.Z. and X.T. performed the genetic studies. S.L., X.T., and J.L. contributed to the recruitment and characterization of the patients. W.L. and W.H. performed the AOA. H.L. and T.F. performed the sperm mobility analysis and ICSI. Q.Z. and H.J. performed data analysis and drafted the manuscript. T.L., S.C., and G.H. revised the manuscript. All authors read and approved the final manuscript.

Funding

Chongqing Medical Scientific Research Project (Joint project of Chongqing Health Commission and Science and Technology Bureau, 2023MSXM008 and 2023MSXM054).

Conflict of interest

The authors declare that they have no competing interests.

References

- Agarwal A, Baskaran S, Parekh N, Cho CL, Henkel R, Vij S, Arafa M, Panner Selvam MK, Shah R. Male infertility. *Lancet* 2021; **397**:319–333.
- Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul Consensus Workshop on embryo assessment: proceedings of an expert meeting. *Hum Reprod* 2011; **26**:1270–1283.
- Boëda B, Knowles PP, Briggs DC, Murray-Rust J, Soriano E, Garvalov BK, McDonald NQ, Way M. Molecular recognition of the Tes LIM2-3 domains by the actin-related protein Arp7A. *J Biol Chem* 2011; **286**:11543–11554.
- Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, Haugen TB, Kruger T, Wang C, Mbizvo MT et al. World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 2010; **16**:231–245.
- Dai J, Zhang T, Guo J, Zhou Q, Gu Y, Zhang J, Hu L, Zong Y, Song J, Zhang S et al. Homozygous pathogenic variants in *ACTL9* cause fertilization failure and male infertility in humans and mice. *Am J Hum Genet* 2021; **108**:469–481.
- Escoffier J, Lee HC, Yassine S, Zouari R, Martinez G, Karaouzène T, Coutton C, Kherraf ZE, Halouani L, Triki C et al. Homozygous mutation of *PLCZ1* leads to defective human oocyte activation and infertility that is not rescued by the WW-binding protein PAWP. *Hum Mol Genet* 2016; **25**:878–891.
- Gardner DK, Lane M. Culture and selection of viable blastocysts: a feasible proposition for human IVF? *Hum Reprod Update* 1997; **3**:367–382.
- Griswold MD. Spermatogenesis: the commitment to meiosis. *Physiol Rev* 2016; **96**:1–17.
- Hara Y, Yamagata K, Oguchi K, Baba T. Nuclear localization of profilin III-ArpM1 complex in mouse spermiogenesis. *FEBS Lett* 2008; **582**:2998–3004.
- Heid H, Figge U, Winter S, Kuhn C, Zimbelmann R, Franke W. Novel actin-related proteins Arp-T1 and Arp-T2 as components of the cytoskeletal calyx of the mammalian sperm head. *Exp Cell Res* 2002; **279**:177–187.
- Heytens E, Parrington J, Coward K, Young C, Lambrecht S, Yoon SY, Fissore RA, Hamer R, Deane CM, Ruas M et al. Reduced amounts and abnormal forms of phospholipase C zeta (PLCzeta) in spermatozoa from infertile men. *Hum Reprod* 2009; **24**:2417–2428.
- Kluin PM, Kramer MF, de Rooij DG. Spermatogenesis in the immature mouse proceeds faster than in the adult. *Int J Androl* 1982; **5**:282–294.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**:1754–1760.
- Lin T, Liu W, Han W, Tong K, Xiang Y, Liao H, Chen K, He Y, Liu D, Huang G. Genetic screening and analysis of *TUBB8* variants in females seeking ART. *Reprod Biomed Online* 2023; **46**:244–254.
- Long S, Fu L, Ma J, Yu H, Tang X, Hu T, Han W, Liu W, Liao H, Fu T et al. Novel biallelic variants in *DNAH1* cause multiple morphological abnormalities of sperm flagella with favorable outcomes of fertility after ICSI in Han Chinese males. *Andrology* 2023; **38**:dead093.307.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**:1297–1303.
- Neto FT, Bach PV, Najari BB, Li PS, Goldstein M. Spermatogenesis in humans and its affecting factors. *Semin Cell Dev Biol* 2016; **59**:10–26.
- Oko R, Sutovsky P. Biogenesis of sperm perinuclear theca and its role in sperm functional competence and fertilization. *J Reprod Immunol* 2009; **83**:2–7.

- Sha Y, Liu W, Li L, Serafimovski M, Isachenko V, Li Y, Chen J, Zhao B, Wang Y, Wei X. Pathogenic variants in ACTRT1 cause acephalic spermatozoa syndrome. *Front Cell Dev Biol* 2021;**9**:676246.
- Sha YW, Sha YK, Ji ZY, Mei LB, Ding L, Zhang Q, Qiu PP, Lin SB, Wang X, Li P et al. TSGA10 is a novel candidate gene associated with acephalic spermatozoa. *Clin Genet* 2018;**93**:776–783.
- Tang X, Ma J, Wang X, Long S, Wan L, Yu H, Yang J, Huang G, Lin T. A novel variant in CFAP69 causes asthenoteratozoospermia with treatable ART outcomes and a literature review. *J Assist Reprod Genet* 2023;**40**:2175–2184.
- Vander Borgh M, Wyns C. Fertility and infertility: definition and epidemiology. *Clin Biochem* 2018;**62**:2–10.
- Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 2007;**23**:657–663.
- Wang J, Zhang J, Sun X, Lin Y, Cai L, Cui Y, Liu J, Liu M, Yang X. Novel bi-allelic variants in ACTL7A are associated with male infertility and total fertilization failure. *Hum Reprod* 2021;**36**:3161–3169.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164.
- World Health Organization (WHO). *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 6th edn. Geneva: World Health Organization, 2021.
- Xin A, Qu R, Chen G, Zhang L, Chen J, Tao C, Fu J, Tang J, Ru Y, Chen Y et al. Disruption in ACTL7A causes acrosomal ultrastructural defects in human and mouse sperm as a novel male factor inducing early embryonic arrest. *Sci Adv* 2020;**6**:eaaz4796.
- Yuan P, Yang C, Ren Y, Yan J, Nie Y, Yan L, Qiao J. A novel homozygous mutation of phospholipase C zeta leading to defective human oocyte activation and fertilization failure. *Hum Reprod* 2020;**35**:977–985.
- Yuan S, Stratton CJ, Bao J, Zheng H, Bhetwal BP, Yanagimachi R, Yan W. Spata6 is required for normal assembly of the sperm connecting piece and tight head-tail junction. *Proc Natl Acad Sci USA* 2015;**112**:E430–E439.
- Zhang D, Huang WJ, Chen GY, Dong LH, Tang Y, Zhang H, Li QQ, Mei XY, Wang ZH, Lan FH. Pathogenesis of acephalic spermatozoa syndrome caused by SUN5 variant. *Mol Hum Reprod* 2021a;**27**:gaab028.
- Zhang XD, Zhang Q, Han W, Liu WW, Shen XL, Yao GD, Shi SL, Hu LL, Wang SS, Wang JX et al. Comparison of embryo implantation potential between time-lapse incubators and standard incubators: a randomized controlled study. *Reprod Biomed Online* 2022a;**45**:858–866.
- Zhang XZ, Wei LL, Jin HJ, Zhang XH, Chen SR. The perinuclear theca protein Calicin helps shape the sperm head and maintain the nuclear structure in mice. *Cell Rep* 2022b;**40**:111049.
- Zhang XZ, Wei LL, Zhang XH, Jin HJ, Chen SR. Loss of perinuclear theca ACTRT1 causes acrosome detachment and severe male subfertility in mice. *Development* 2022c;**149**:dev200489.
- Zhang Y, Liu C, Wu B, Li L, Li W, Yuan L. The missing linker between SUN5 and PMFBP1 in sperm head-tail coupling apparatus. *Nat Commun* 2021b;**12**:4926.
- Zhou X, Xi Q, Jia W, Li Z, Liu Z, Luo G, Xing C, Zhang D, Hou M, Liu H et al. A novel homozygous mutation in ACTL7A leads to male infertility. *Mol Genet Genomics* 2023;**298**:353–360.
- Zhu F, Liu C, Wang F, Yang X, Zhang J, Wu H, Zhang Z, He X, Zhang Z, Zhou P et al. Mutations in PMFBP1 cause acephalic spermatozoa syndrome. *Am J Hum Genet* 2018;**103**:188–199.

Andrology

Identification of two hidden clinical subgroups among men with idiopathic cryptozoospermia

Lena Charlotte Schülke¹, Joachim Wistuba¹, Verena Nordhoff², Hermann M. Behre³, Jann-Frederik Cremers², Sabine Kliesch², Sara Di Persio^{1,†}, and Nina Neuhaus^{1,*} 

¹Centre of Reproductive Medicine and Andrology, Institute of Reproductive and Regenerative Biology, University of Münster, Münster, Germany

²Department of Clinical and Surgical Andrology, Centre of Reproductive Medicine and Andrology, University of Münster, Münster, Germany

³UKM Kinderwunschzentrum, Universitätsklinikum Münster, Münster, Germany

*Correspondence address. Centre of Reproductive Medicine and Andrology, University Hospital of Münster, Albert-Schweitzer-Campus 1, Building D11, 48149 Münster, Germany. E-mail: nina.neuhaus@ukmuenster.de  <https://orcid.org/0000-0003-0181-6194>

[†]These authors contributed equally to this work.

ABSTRACT

STUDY QUESTION: Are there subgroups among patients with cryptozoospermia pointing to distinct etiologies?

SUMMARY ANSWER: We reveal two distinct subgroups of cryptozoospermic (Crypto) patients based on testicular tissue composition, testicular volume, and FSH levels.

WHAT IS KNOWN ALREADY: Cryptozoospermic patients present with a sperm concentration below 0.1 million/ml. While the etiology of the severely impaired spermatogenesis remains largely unknown, alterations of the spermatogonial compartment have been reported including a reduction of the reserve stem cells in these patients.

STUDY DESIGN, SIZE, DURATION: To assess whether there are distinct subgroups among cryptozoospermic patients, we applied the statistical method of cluster analysis. For this, we retrospectively selected 132 cryptozoospermic patients from a clinical database who underwent a testicular biopsy in the frame of fertility treatment at a university hospital. As controls (Control), we selected 160 patients with obstructive azoospermia and full spermatogenesis. All 292 patients underwent routine evaluation for endocrine, semen, and histological parameters (i.e. the percentage of tubules with elongated spermatids). Moreover, outcome of medically assisted reproduction (MAR) was assessed for cryptozoospermic ($n = 73$) and Control patients ($n = 87$), respectively. For in-depth immunohistochemical and histomorphometrical analyses, representative tissue samples from cryptozoospermic ($n = 27$) and Control patients ($n = 12$) were selected based on cluster analysis results and histological parameters.

PARTICIPANTS/MATERIALS, SETTING, METHODS: This study included two parts: firstly using clinical parameters of the entire cohort of 292 patients, we performed principal component analysis (PCA) followed by hierarchical clustering on principal components (i.e. considering hormonal values, ejaculate parameters, and histological information). Secondly, for histological analyses seminiferous tubules were categorized according to the most advanced germ cell type present in sections stained with Periodic acid Schiff. On the selected cohort of 39 patients (12 Control, 27 cryptozoospermic), we performed immunohistochemistry for spermatogonial markers melanoma-associated antigen 4 (MAGEA4) and piwi like RNA-mediated gene silencing 4 (PIWIL4) followed by quantitative analyses. Moreover, the morphologically defined A_{dark} spermatogonia, which are considered to be the reserve stem cells, were quantified.

MAIN RESULTS AND THE ROLE OF CHANCE: The PCA and hierarchical clustering revealed three different clusters, one of them containing all Control samples. The main factors driving the sorting of patients to the clusters were the percentage of tubules with elongated spermatids (Cluster 1, all Control patients and two cryptozoospermic patients), the percentage of tubules with spermatocytes (Cluster 2, cryptozoospermic patients), and tubules showing a Sertoli cells only phenotype (Cluster 3, cryptozoospermic patients). Importantly, the percentage of tubules containing elongated spermatids was comparable between Clusters 2 and 3. Additional differences were higher FSH levels ($P < 0.001$) and lower testicular volumes ($P < 0.001$) in Cluster 3 compared to Cluster 2. In the spermatogonial compartment of both cryptozoospermic Clusters, we found lower numbers of MAGEA4⁺ and A_{dark} spermatogonia but higher proportions of PIWIL4⁺ spermatogonia, which were significantly correlated with a lower percentage of tubules containing elongated spermatids. In line with this common alteration, the outcome of MAR was comparable between Controls as well as both cryptozoospermic Clusters.

LIMITATIONS, REASONS FOR CAUTION: While we have uncovered the existence of subgroups within the cohort of cryptozoospermic patients, comprehensive genetic analyses remain to be performed to unravel potentially distinct etiologies.

WIDER IMPLICATIONS OF THE FINDINGS: The novel insight that cryptozoospermic patients can be divided into two subgroups will facilitate the strategic search for underlying genetic etiologies. Moreover, the shared alterations of the spermatogonial stem cell compartment between the two cryptozoospermic subgroups could represent a general response mechanism to the reduced output of sperm, which may be associated with a progressive phenotype. This study therefore offers novel approaches towards the understanding of the etiology underlying the reduced sperm formation in cryptozoospermic patients.

Received: July 12, 2023. Revised: December 20, 2023. Editorial decision: January 19, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

STUDY FUNDING/COMPETING INTEREST(S): German research foundation CRU 326 (grants to: SDP, NN). Moreover, we thank the Faculty of Medicine of the University of Münster for the financial support of Lena Charlotte Schülke through the MedK-program. We acknowledge support from the Open Access Publication Fund of the University of Münster. The authors have no potential conflicts of interest.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: cluster analysis / cryptozoospermia / piwi like RNA-mediated gene silencing 4 / male infertility / medically assisted reproduction / spermatogonial stem cells / meiotic arrest / Sertoli cell only phenotype / testis / spermatogenesis

Introduction

Men with cryptozoospermia have a sperm concentration of ≤ 0.1 million per milliliter in the ejaculate, rendering natural conception almost impossible. It is noteworthy that rare motile spermatozoa can usually only be detected after resuspension of semen sediments (World Health Organization, 2010), reflecting the severe impairment of spermatogenesis within testicular tissues.

First insights pointing to the stage of germ cell differentiation failure in cryptozoospermic testes were provided by single cell RNA-sequencing data (Di Persio et al., 2021; Di Persio and Neuhaus, 2023). These revealed that the proportions of spermatogonia and early meiotic stages were still comparable to the control situation. Only from pachytene stage onwards did cryptozoospermic samples show a reduction of germ cells (Di Persio et al., 2021). Moreover, these datasets highlighted alterations already in the spermatogonial stem cell compartment, which forms the basis of spermatogenesis. Unexpectedly, testicular tissues from cryptozoospermic patients showed increased proportions of piwi like RNA-mediated gene silencing 4 (PIWIL4⁺) spermatogonia (Di Persio et al., 2021), which are considered to be the origin of the spermatogonial differentiation process (Guo et al., 2018; Sohni et al., 2019; Shami et al., 2020; Di Persio et al., 2021). In contrast to this, the number of morphologically defined A_{dark} spermatogonia, which have been suggested to include the reserve stem cell population, was reduced (Clermont, 1963; van Alphen et al., 1988; Caldeira-Brant et al., 2020; Di Persio et al., 2021). Two possible scenarios can be considered as explanation for these findings. The increased proportion of PIWIL4⁺ spermatogonia may point to a stem cell defect prohibiting differentiation of the most undifferentiated spermatogonia and thereby leading to an accumulation of this cell type. Alternatively, the activation of the spermatogonial stem cell compartment may be a response mechanism to the reduced production of sperm. However, it seems likely that this over activation of the stem cell pool, which is also at the cost of reserve stem cell (A_{dark}) numbers, may lead to stem cell exhaustion and a progressive phenotype of cryptozoospermia. Analyses focusing also on the spermatogonial compartment are therefore essential to unravel the cellular changes leading to the severely reduced sperm production in cryptozoospermic men.

Provided that clinical parameters can be identified that distinguish subgroups among the highly heterogeneous group of cryptozoospermic patients, these may point to distinct etiologies. One example demonstrating the successful use of this approach is the genetic screening of men with meiotic arrest, which led to the identification of *meiosis 1 Associated Protein* (M1AP) (mutations causative for meiotic failure). Specifically, deleterious mutations in M1AP lead to complete meiotic arrest and azoospermia in some patients and to cryptozoospermia in others (Wyrwoll et al., 2020). An analysis of cryptozoospermic patients to unveil potential subgroups is therefore a necessity to reveal the underlying etiologies and improve counseling and, prospectively, treatment. To this end, the statistical approach of cluster analysis can be applied, which has been successfully used in other medical fields to

improve diagnostics in heterogeneous patient cohorts, building the basis for specialized treatments and a better prediction of the prognosis (Burgel et al., 2014; Zhu et al., 2019; Krenz et al., 2021).

This study therefore aimed to assess if subgroups of patients can be identified in a cohort of 132 cryptozoospermic men considering clinical/andrological parameters, and to evaluate the alterations in the spermatogonial compartment and outcome of medically assisted reproduction (MAR).

Materials and methods

Ethical approval

For the use of testicular tissue samples from patients with obstructive azoospermia and cryptozoospermia, ethical approval was obtained from the Ethics Committee of the Medical Faculty of Münster and the State Medical Board (No. 2008-090-f-S). Testicular biopsies for research were obtained during microsurgical testicular sperm extraction and histological analysis was carried out at the Department of Clinical and Surgical Andrology at the Centre for Reproductive Medicine and Andrology (CeRA), University Hospital Münster following written informed consent.

Selection of the patient cohort

We retrospectively selected all the 3539 patients from the database Androbase[®] (Tüttelmann et al., 2006) who underwent a testicular biopsy at the CeRA, Münster between 1991 and 2020. Only patients with a complete dataset including physical examination, hormone analysis (including LH, FSH, testosterone (T), and free testosterone), ultrasound of the testes, semen analysis according to the World Health Organization criteria (World Health Organization, 2010) and if indicated genetic analyses (including karyotype and azoospermia factor (AZF) deletions, cf transmembrane conductance regulator (CFTR) mutations, and single nucleotide polymorphism (SNP) FSHB-211 G>T genotyping; Busch et al., 2015) were included in the study cohorts.

To obtain a cohort of idiopathic cryptozoospermic men, patients with a sperm concentration ≤ 0.1 million per milliliter of ejaculate were selected. Rare motile spermatozoa can usually only be detected after resuspension of semen sediments (World Health Organization, 2010). Exclusion criteria were genetic diseases (i.e. Klinefelter syndrome, AZF deletions), hormonal treatments (i.e. hCG, rFSH, Tamoxifen, Anastrozol), and a history of cryptorchidism or oncological disease.

For the Control cohort, patients with obstructive azoospermia (full spermatogenesis, no sperm in the ejaculate) owing to congenital bilateral absence of the vas deferens (CBAVD), vasovasostomy, or a physical obstruction in the vas deferens were selected. Exclusion criteria were genetic diseases (i.e. Klinefelter syndrome, AZF deletions), hormonal treatments (i.e. hCG, rFSH, Tamoxifen, Anastrozol), and a history of cryptorchidism or oncological disease. Furthermore, patients who only had one testis or had a FSH concentration outside the reference range (1–7 U/l) and a Bergmann–Kliesch score below 7 were excluded (Bergmann and Kliesch, 2010). Applying these criteria, we selected a cohort

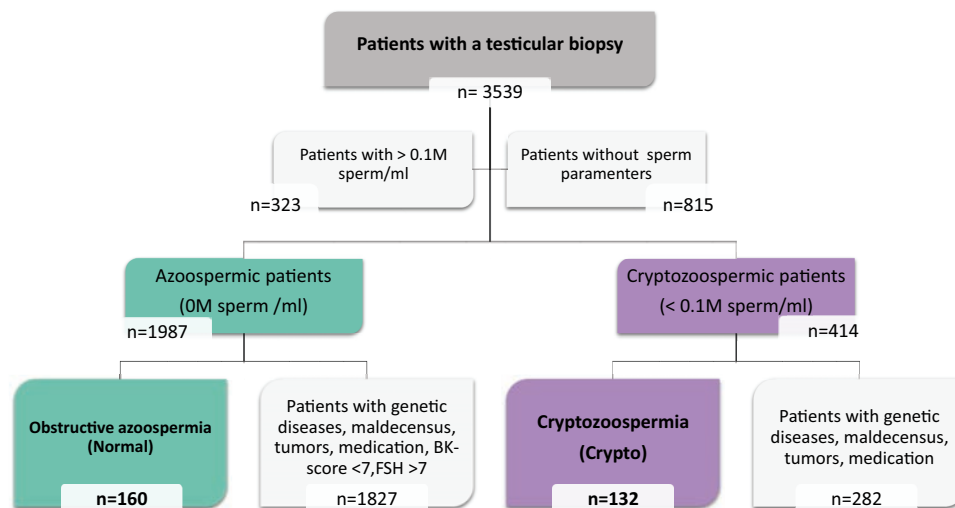


Figure 1. Selection of the study cohort. Flow chart depicting the selection of 160 Control (teal) and 132 cryptozoospermic (Crypto, purple) patients from the database Androbase. Exclusion criteria are provided in grey boxes. M, million; ml, milliliter; BK-Score, Bergmann–Kliesch Score.

of 160 Control and 132 cryptozoospermic patients (Fig. 1, Supplementary Table S1).

Hormone measurements

Hormone measurements, including gonadotrophins and T, were performed using blood samples taken in the morning. LH (Cat No. 02P40-25), FSH (Cat No. 07 K75-25), T (Cat No. 02P13-28), sex hormone-binding globulin (SHBG, Cat No. 08 K26-20), prolactin (Cat No. 07 K76-25), estradiol (Cat No. 07 K72-25), and prostate-specific antigen (PSA, Cat No. 07 K70-25) were measured using chemiluminescent microparticle immunoassays analyzed with the Architect i1000 (Abbott Diagnostics, Wiesbaden, Germany). Free testosterone was calculated considering the total testosterone and the fraction bound to SHBG using the published formula (<https://www.issam.ch/freetesto.htm>; Vermeulen et al., 1999). Moreover, ACTIVE® dihydrotestosterone (DHT) radioimmunoassay (DSL-9600 Beckmann-Coulter, Krefeld, Germany), was used to measure DHT. The assay shows less than 2% cross-reactivity with T and is calibrated against a standard of LCMS-MS with an accuracy of <15% within the range of 0.1 to 5 nmol/l. Intra-assay co-efficient of variation (CV) is 3.5%, mean inter-assay CV is 7% for the DHT assay.

Principal component analysis and hierarchical clustering

The principal component analysis (PCA) and clustering were performed using the hierarchical clustering on principal components (HCPC) method. R 4.0.5 and the R packages FactoMineR (Version 2.4) (Lê et al., 2008) and factoextra (Version 1.0.7) were used. For the cohort of 292 patients, the variables included in the analyses were: age of the patients at the time of surgery, LH, FSH, T, free testosterone, ejaculate volume, ejaculate pH, sperm concentration, total sperm count, and testicular volume (Busch et al., 2019; Krenz et al., 2021). In addition, the histological information (percentage of tubules with elongated spermatids, round spermatids, spermatocytes, spermatogonia as the most advanced germ cell types or displaying Sertoli cell only (SCO) phenotype or tubular shadows) was taken into consideration. An overview of all clinical parameters is provided in Supplementary Table S1. The PCA was performed to reduce the dimensionality of the data into few continuous variables using the function PCA() from the FactoMineR (Version 2.4) package. Prior to the analysis, the data

were scaled to avoid dominance by variables with large measurement units. Five dimensions were retained in the output. The results of the PCA were visualized using the factoextra (Version 1.0.7) package. The hierarchical clustering was performed using the Ward's criterion on the selected principal components using the function HCPC() from the FactoMineR (Version 2.4) package. The partition in different clusters was initially performed by cutting the hierarchical tree, and the K-means clustering was then used to improve the initial partition. The function fviz_cluster() in factoextra (Version 1.0.7) package was used to visualize individual clusters.

Analysis of ICSI results

Data reporting the outcome of MAR treatments and ICSI procedures performed at the UKM Kinderwunschzentrum and the Centre of Reproductive Medicine and Andrology were available for 160 of the 292 patients included in the study. The percentages of embryo transfer, clinical and biochemical pregnancies, and live birth were calculated for each Cluster based on the total number of patients for which ICSI data were available. The outcome was considered positive if a patient had at least one embryo transfer, a biochemical or clinical pregnancy, or a live birth, respectively, and results were not normalized by number of ICSI cycle.

Periodic acid-Schiff staining and histological evaluation of human testicular tissue sections

For routine histological evaluation, testicular tissues were processed and stained according to previously described protocols (Di Persio et al., 2021). To assess the histological status, periodic acid-Schiff-stained sections from two independent biopsies per testis were graded using the Bergmann–Kliesch (BK) score (Bergmann and Kliesch, 2010). A score from 0 to 10 was assigned to each patient based on the percentage of tubules containing elongated spermatids, with a score of 1 indicating 10% of the tubules with elongated spermatids. Additionally, the percentages of tubules with round spermatids, spermatocytes, and spermatogonia as the most advanced germ cell type as well as the percentage of SCO tubules and tubular shadows were evaluated.

Immunohistochemical staining of human testicular tissue sections

For immunohistochemical analyses, we randomly selected 13 samples from Cluster 1 (including 12 Control patients and one cryptozoospermic patient), 13 cryptozoospermic samples from Clusters 2 and 3. The samples from Clusters 2 and 3 were matched based on the BK score to ensure an equal representation of tubules with full spermatogenesis in the two groups. The immunohistochemical stainings were performed as previously published (Di Persio et al., 2021). Mouse monoclonal anti-melanoma-associated antigen 4 (MAGEA4) (provided by Prof. G.C. Spagnoli, University Hospital of Basel, Switzerland; dilution 1:20) and rabbit polyclonal anti-PIWIL4 (PIWIL4, Cat No. HPA036588, Atlas Antibodies, Bromma, Sweden; dilution 1:50) were used as primary antibodies. As negative controls, sections were incubated with species-specific immunoglobulin G (Mouse-IgG Control Antibody, unconjugated, Cat No. 15381, Sigma Aldrich, St Louis, MO, USA; dilution 1:100 and Rabbit-IgG Control Antibody, unconjugated, Cat No. 15006, Sigma Aldrich; dilution 1:100) or without primary antibody. After overnight incubation, the sections were incubated with a corresponding secondary antibody (Goat F(ab')₂ Anti-Mouse IgG – (Fab')₂ (Biotin), Cat No. ab5886, Abcam, Cambridge, UK; dilution 1:100 and Goat F(ab')₂ Anti-Rabbit IgG—H&L (Biotin), Cat No. ab6012, Abcam; dilution 1:100) and subsequently with streptavidin-peroxidase (Cat No. S5512, Sigma-Aldrich; dilution 1:500). 3,3'-diaminobenzidine tetrahydrochloride solution (Cat No. A0596.0001, Applichem, Darmstadt, Germany) was used to detect the peroxidase activity. Lastly, the nuclei were counterstained with Mayer's hematoxylin (Cat No. 1.092.490.500, Sigma-Aldrich). Stained sections were digitized employing the Precipoint M8 Microscope and Scanner (Precipoint, Freising, Germany), using a 60× objective and were analyzed using the Precipoint software Viewpoint (Precipoint, Freising, Germany).

Quantifications of spermatogonial subpopulations

The numbers of MAGEA4⁺ and PIWIL4⁺ spermatogonia per tubule were quantified in two independent testicular tissue sections and 20 round tubules per sample, which was the maximum number that could be reached in some cryptozoospermic samples owing to the limited number of tubules containing germ cells. Tubules were regarded as round when the ratio of the two diameters was between 1 and 1.5. To obtain the proportion of PIWIL4⁺ spermatogonia among the entire spermatogonial population, the number of PIWIL4⁺ spermatogonia was normalized to the number of MAGEA4⁺ spermatogonia. A_{dark} spermatogonia were identified based on morphological criteria (Clermont, 1966). The A_{dark} spermatogonia are characterized by an unstained rarefaction zone in the center of a nucleus with uniformly dark stained chromatin. The proportion of A_{dark} spermatogonia was quantified per 100 MAGEA4⁺ spermatogonia.

Statistical analyses and data presentation

For the clinical and histological data, normality (Shapiro–Wilk test), homogeneity of variances (Levene and Flinger tests), and homoscedasticity (Breusch–Pagan) tests were performed for all variables. For normally distributed data, the one-way ANOVA test followed by pairwise Student's t-test was performed. In case of data not normally distributed, the Kruskal–Wallis rank sum test was used to compare the values among the three patient groups, followed by pairwise Wilcoxon tests. Prior to correlation analysis, the data were checked for normality using the Shapiro–Wilk test. Based on the normality test results, the correlation

analyses were performed using the Spearman correlation method. Statistical analyses were executed using R 4.1.2 and specifically the package stats (Version 4.1.2) (*R a language and environment for statistical computing*, 2010). Details regarding the statistical analysis of each variable are provided in [Supplementary Tables S2 and S3](#).

Plots were generated using GraphPad Prism® Version 5.0 (GraphPad Software, Inc., San Diego, CA, USA). Box plots elements are defined as follows: center line represents the median; box limits depict upper and lower quartiles and whiskers the 1.5× interquartile range. Outliers are shown as individual points.

Results

PCA and hierarchical clustering based on clinical data of cryptozoospermic men

To assess if there are different subgroups of cryptozoospermic patients, we screened 3,539 patients who underwent a testicular biopsy and full clinical evaluation at the CeRA. We selected 160 patients with obstructive azoospermia (Control) and 132 cryptozoospermic patients, applying the criteria outlined in [Fig. 1](#).

The PCA based on clinical parameters showed a clear separation of the Control and the cryptozoospermic patient groups on the first principal component (PC1), which accounts for 30.3% of the total variation ([Supplementary Fig. S1A](#)). The variation on the PC1 was mostly due to the following variables: sperm concentration in the ejaculate, percentage of tubules with elongated spermatids, and FSH concentration ([Supplementary Fig. S1B](#)).

In order to evaluate whether the cryptozoospermic patients can be separated from the Controls and whether they can be divided into subgroups, we applied HCPC method and identified three different clusters. The first cluster contained all 160 Control patients and 2 cryptozoospermic patients (Cluster 1). The remaining cryptozoospermic patients were equally subdivided into Clusters 2 (n = 65) and 3 (n = 65), respectively ([Fig. 2](#)).

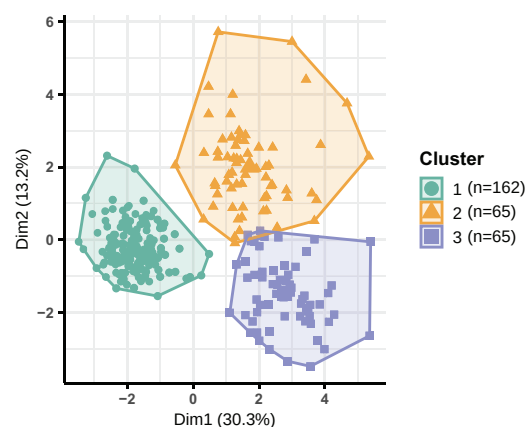


Figure 2. Hierarchical clustering on principal components revealed three distinctive clusters among the 292 included patients. The clustering is shown in a principal component analysis (PCA) space that was based on the clinical and histological parameters of each patient. The cluster analysis identified three different clusters. The first cluster contained all 160 Control patients and 2 cryptozoospermic patients (Cluster 1). The remaining cryptozoospermic patients were equally subdivided into Clusters 2 (n = 65) and 3 (n = 65), respectively. Each dot represents one patient and the lines that group the dots show the clustering. The percentages on the x- and y-axes show the variation of the data.

Sub-classification of the cryptozoospermic patient cohort

The main factors (leading factors) determining the three clusters were the percentage of tubules with elongated spermatids for Cluster 1, the percentage of tubules with spermatocytes for Cluster 2, and the percentage of tubules containing solely Sertoli cells for Cluster 3 (Fig. 3A, Table 1). The composition of testicular tissues in these clusters is therefore fundamentally different, with the majority of tubules showing an arrest at the spermatocyte level (median: 51.5%) in Cluster 2 compared to the majority of tubules devoid of germ cells (median SCO: 63.5%; median TS: 11%) in Cluster 3 (Fig. 3B, Table 1). Importantly, although these represent the predominant tubular phenotype in each cluster,

the percentage of tubules with full germ cell differentiation (up to the elongated spermatids) was comparable in cryptozoospermic Clusters 2 (median: 4%) and 3 (median: 3.5%), respectively (Fig. 3C, Table 1).

Comparison of clinical parameters between patient subgroups

In line with the comparable percentage of tubules containing elongated spermatids in Clusters 2 and 3, we found comparable sperm retrieval rates after testicular sperm extraction (TESE) between the two cryptozoospermic groups (Supplementary Fig. S2A). Additional clinical parameters, apart from the leading factors, that showed differences between the two cryptozoospermic

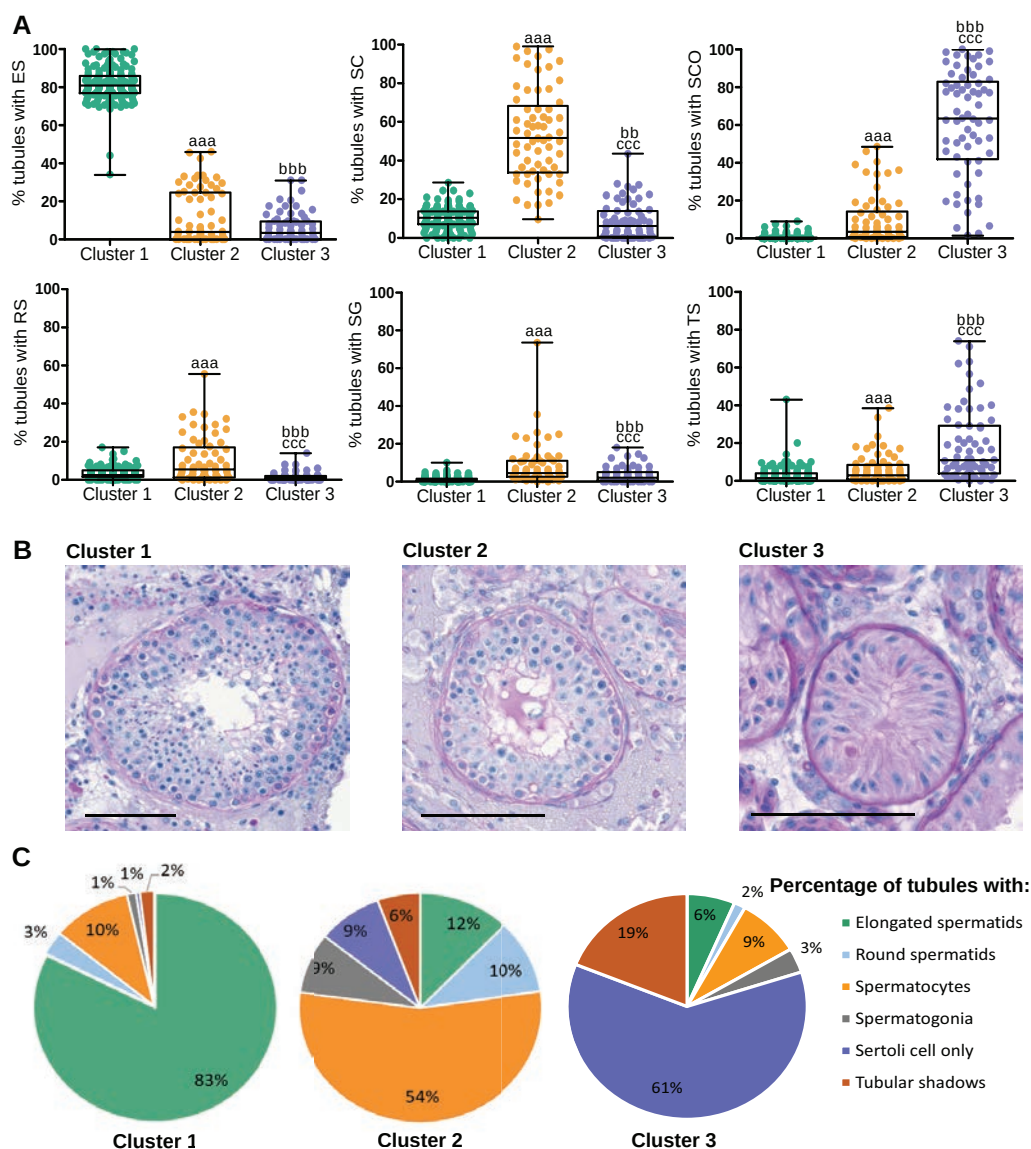


Figure 3. Comparison of the histological status of the testis tissues from patients in the three clusters. (A) Boxplots comparing the percentage of the most advanced germ cell types in the seminiferous tubules of each cluster. Cluster 1 (teal): $n = 162$ (160 Control and 2 cryptozoospermic patients), Cluster 2 (orange): $n = 65$ cryptozoospermic patients, Cluster 3 (purple): $n = 65$ cryptozoospermic patients. ES, elongated spermatids; RS, round spermatids; SC, spermatocytes; SG, spermatogonia; SCO, Sertoli cell only; TS, tubular shadow. Boxplot elements are defined as follows: center line: median; box limits: upper and lower quartiles; whiskers: $1.5 \times$ interquartile range; points: outliers. Statistically significant differences between the clusters are shown by letters: (a) Cluster 1 is different from Cluster 2, (b) Cluster 1 is different from Cluster 3, (c) Cluster 2 is different from Cluster 3. Significance levels are shown by the number of letters, e.g. $a = P < 0.05$, $aa = P < 0.01$, $aaa = P < 0.001$. (B) Representative images of periodic acid-Schiff-stained tubules with full spermatogenesis (Cluster 1), meiotic arrest at the spermatocyte stage (Cluster 2), and Sertoli cell only phenotype (Cluster 3). Scale bar: $100 \mu\text{m}$. (C) Pie charts showing the means of the percentage of the most advanced germ cell type in the seminiferous tubules of patients from each cluster.

Table 1. Hormonal and histological parameters of patients from the three clusters.*

Patient groups	Hormone parameters (normal range)			Histological parameters of tubules		
	Testicular volume (12–30 ml)	FSH (1–7 U/l)	LH (2–10 U/l)	% ES	% SC	% SCO
Cluster 1 (n = 162)	24.88 (20, 29.5)	3.35 (2.5, 4.38)	2.7 (1.9, 3.9)	81 (77, 86)	10.25 (7, 13.5)	0 (0, 0.5)
Cluster 2 (n = 65)	14.5 (11.5, 18.5)	6.4 (4.3, 13)	4.8 (3, 5.8)	4 (0, 24.5)	51.5 (34.5, 66.5)	3.5 (1, 13.5)
Cluster 3 (n = 65)	11.5 (8.5, 15)	20.8 (15.8, 24.1)	6.7 (4.8, 8.8)	3.5 (0, 9.5)	6 (0.5, 13.5)	63.5 (43, 82)

For each cluster (*Cluster 1: 160 Control and 2 cryptozoospermic patients, Cluster 2 and 3: cryptozoospermic patients) and parameter, the median is shown, as well as the 25th and 75th percentile in brackets (median, 25th, 75th percentile). ES, elongated spermatids; SC, spermatocytes; SCO, Sertoli cell only; ml, milliliter; l, liter; U, units.

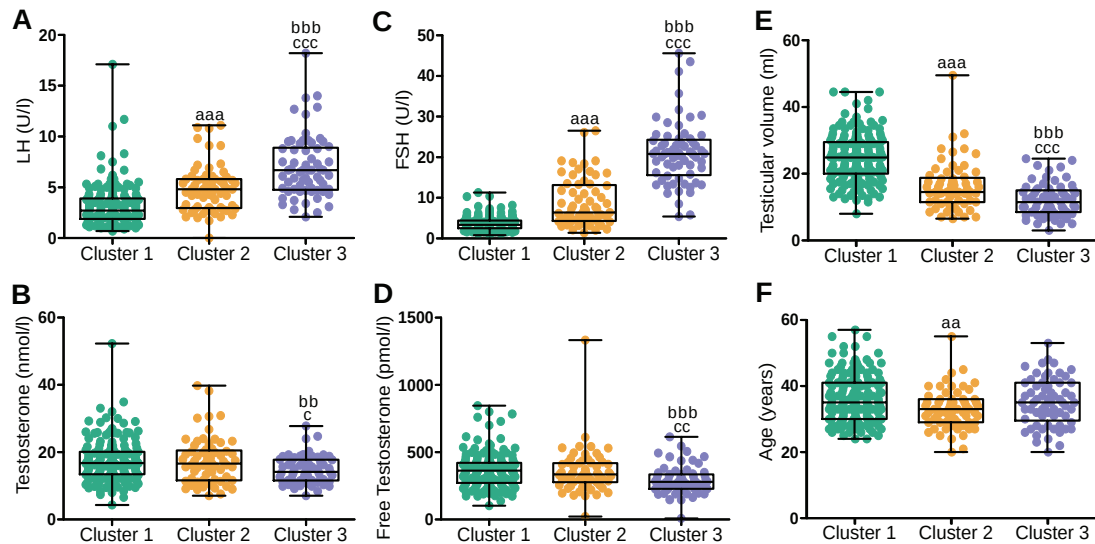


Figure 4. Comparison of clinical parameters among the three patient clusters. Boxplots comparing the (A) testicular volume, (B) LH, (C) FSH, (D) testosterone, (E) free testosterone values, and (F) age of the patients of the three clusters. Cluster 1 (teal): n = 162 (160 Control and 2 cryptozoospermic patients), Cluster 2 (orange): n = 65 cryptozoospermic patients, Cluster 3 (purple): n = 65 cryptozoospermic patients. Boxplot elements are defined as follows: center line: median; box limits: upper and lower quartiles; whiskers: 1.5 × interquartile range; points: outliers. Statistically significant differences between the clusters are shown by letters: (a) Cluster 1 is different from Cluster 2, (b) Cluster 1 is different from Cluster 3, (c) Cluster 2 is different from Cluster 3. Significance levels are shown by the number of letters, e.g. a = $P < 0.05$, aa = $P < 0.01$, aaa = $P < 0.001$.

groups are outlined below (Fig. 4, Table 1) and an overview of the results of all the parameters analyzed can be found in Supplementary Table S2.

Statistical analyses revealed significantly larger testicular volumes in the Cluster 2 patients compared to the Cluster 3 patients (Fig. 4A). Moreover, LH levels were significantly lower in Cluster 2 patients (median: 4.8 U/l) compared to Cluster 3 (median: 6.7 U/l, Fig. 4B). While the LH levels in both cryptozoospermic groups were still within the internal reference values, the FSH values in Cluster 3 patients (median: 20.8 U/l) were significantly higher compared to the reference. The FSH values in Cluster 2 were significantly lower (median: 6.4 U/l, Fig. 4C, Supplementary Table S2). The T and free testosterone levels were significantly higher in the Cluster 2 compared to Cluster 3 patients but still within the reference values (Fig. 4D and E, Supplementary Table S2). In contrast to the differences in hormonal values, there was no enrichment of specific FSHB-211 SNPs between the groups (Supplementary Fig. S2B). Interestingly, a significant difference could be found regarding the age of the patients with Cluster 2 patients being slightly younger compared to Control patients in Cluster 1 but not Cluster 3 (Fig. 4F).

In order to evaluate whether the three groups had different outcomes with regard to MAR treatment, we analyzed the ICSI results of 160 patients included in our study (n = 88 Cluster 1, n = 33 Cluster 2, n = 39 Cluster 3). Interestingly, the three groups showed comparable percentages of embryo transfers,

biochemical and clinical pregnancies, and live births (Fig. 5, Supplementary Table S4).

Histological analysis of the spermatogonial compartment

As alterations in the stem cell compartment of cryptozoospermic patients have been found, specifically a reduction of A_{dark} spermatogonia and an increased number of $PIWIL4^+$ cells (Di Persio et al., 2021), we evaluated whether these observed changes are shared between the two clusters of cryptozoospermic patients. We first examined the spermatogonial compartment by analyzing the pan-spermatogonial marker MAGEA4 (Fig. 6A and B, Supplementary Table S3). This analysis revealed a significant reduction of MAGEA4⁺ spermatogonia per tubule in both cryptozoospermic groups compared to the Control cluster but did not show a significant difference between the two cryptozoospermic clusters. Furthermore, Spearman correlation analysis revealed a significant positive correlation between the number of spermatogonia per tubule and the percentage of tubules containing elongated spermatids represented as BK-Score (Fig. 6C).

To assess changes in the population of reserve spermatogonial stem cells, we quantified the number of the morphologically defined A_{dark} spermatogonia (Fig. 6D, Supplementary Table S3). While we found significantly lower numbers of the A_{dark} spermatogonia in the cryptozoospermic clusters compared to the

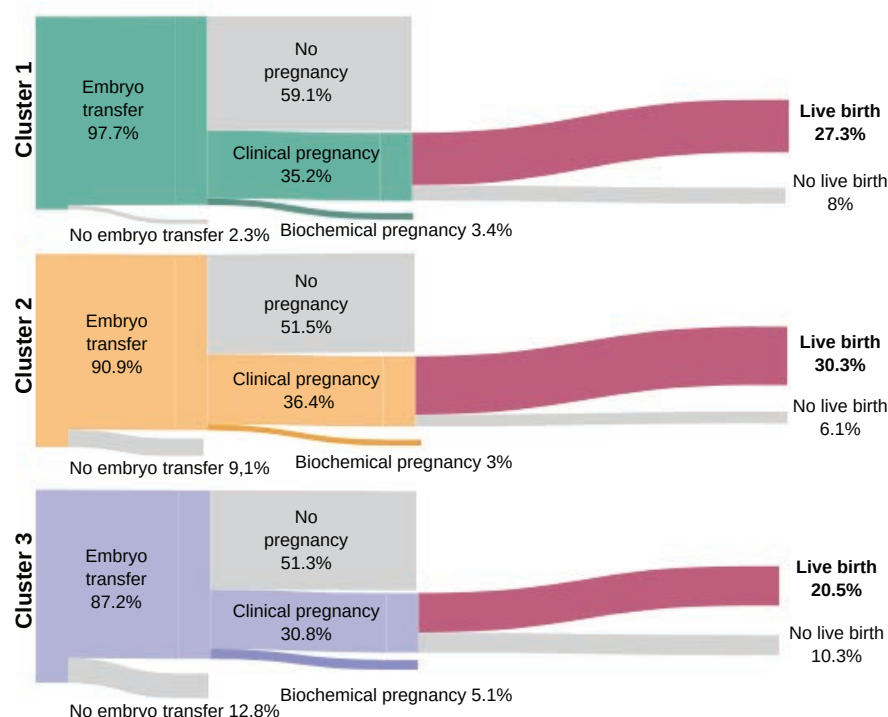


Figure 5. Sankey diagram summarizing the ICSI results in the patient cohort. From left to right the percentage of patients with or without embryo transfer, clinical and biochemical pregnancies, no pregnancy as well as live birth or no live birth are depicted for each cluster of patients (Cluster 1 (teal), Cluster 2 (orange), Cluster 3 (purple)). Numbers represent the percentage of patients in each category. Absolute numbers are shown in Supplementary Table S4.

Control samples (Fig. 6E), no significant difference between the two cryptozoospermic clusters was observed.

In order to assess if the reduction of reserve spermatogonial stem cells is linked to the spermatogenic state, we performed a correlation analysis. This showed that the number of A_{dark} spermatogonia is positively correlated with the percentage of tubules containing elongated spermatids (Fig. 6F).

Interestingly, the proportion of $PIWIL4^{+}$ cells, which are considered the most undifferentiated spermatogonial stem cell type (Guo et al., 2018; Sohni et al., 2019; Shami et al., 2020; Di Persio et al., 2021), among the $MAGEA4^{+}$ cells is increased in both cryptozoospermic groups (Fig. 6G and H, Supplementary Table S3). Moreover, the percentage of $PIWIL4^{+}$ spermatogonia was negatively correlated with the percentage of tubules containing elongated spermatids (Fig. 6I). Hence, both cryptozoospermic groups displayed shared alterations of their spermatogonial compartments with regard to the lower number of A_{dark} and higher number of $PIWIL4^{+}$ spermatogonia, respectively.

Discussion

Heterogeneity of clinical parameters—such as testicular tissue composition and endocrine levels—in cryptozoospermic men has hampered the identification of the etiology behind this infertility phenotype (Di Persio et al., 2021). The almost complete absence of sperm from the ejaculate often necessitates TESE to fulfill the wish for fatherhood in these patients, providing the unique opportunity to evaluate alterations in the spermatogenic process. First analyses point to changes already in the spermatogonial compartment indicating stem cell exhaustion and the existence of distinct severity phenotypes. Considering the heterogeneity of cryptozoospermic patients, we sought to assess if subgroups can

be identified by cluster analysis of clinical parameters, which is the prerequisite to unveil the etiology of cryptozoospermia.

Cluster analyses have been successfully used in other medical fields for the identification of clinical subgroups in heterogeneous patient cohorts, building the basis for specialized treatments and a better prediction of the prognosis (Burgel et al., 2014; Zhu et al., 2019). The PCA-based cluster analysis revealed two subgroups of cryptozoospermic patients characterized by a meiotic arrest- and SCO phenotype in the majority of testicular tissues, respectively. We consider the following scenario which could lead to the two clusters of cryptozoospermic patients. Cryptozoospermia may be a progressive disease, which develops from a meiotic arrest to a SCO phenotype, potentially caused by stem cell exhaustion over time. Also, these two histologically distinct phenotypes may be caused by distinct underlying genetic causes. It is of note that these two scenarios are not necessarily mutually exclusive. The findings supporting the first scenario are the lower testicular volume, higher FSH levels, and higher number of tubules displaying a SCO phenotype in Cluster 3 compared to Cluster 2. While we found no significant association between the patient age of cryptozoospermic patients and the cluster assignment, which may be expected if patients would transition from one phenotype to the other over time, the start of germ cell differentiation failure and the speed of its progression may be different among cryptozoospermic men compared to others. However, a study performed on men who underwent multiple TESE owing to obstructive and non-obstructive azoospermia reported that the sperm retrieval rate was comparable in all the attempts independent of the patient diagnosis, suggesting that the spermatogenic rate in these men did not change over time (Vernaev et al., 2006). For further validation of this hypothesis close monitoring of sperm parameters in individual oligozoospermic and cryptozoospermic patients over time is necessary. If this

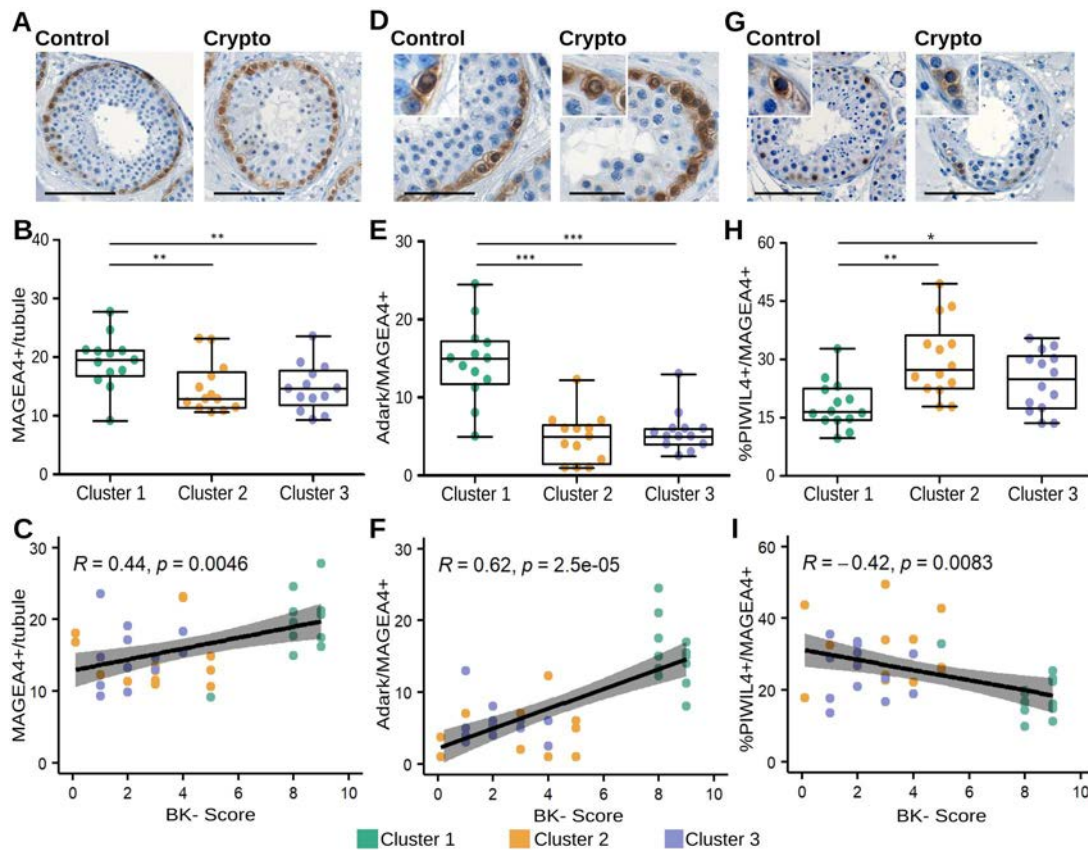


Figure 6. Histological evaluation and comparison of patient samples. From Cluster 1 (teal, $n = 13$ (12 Control and 1 cryptozoospermic patients)), Cluster 2 (orange, $n = 13$ cryptozoospermic patients), and Cluster 3 (purple, $n = 13$ cryptozoospermic patients). Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (A) Representative melanoma-associated antigen 4 (MAGEA4) immunohistochemical stainings in tubules from a Control and a cryptozoospermic (Crypto) sample. Scale bars: $100\mu\text{m}$. (B) Boxplot comparing the quantification of MAGEA4⁺ spermatogonia per tubule. (C) Scatterplot and linear regression showing the correlation between the Bergmann Kliesch (BK)-Score (0–10, indicating 0–100% of tubules with elongated spermatids) and the number of MAGEA4⁺ spermatogonia per tubule. (D) Representative MAGEA4 stainings in tubules from a Control and Crypto sample. Inlay showing a representative A_{dark} spermatogonium, discernible by the rarefaction zone in the nucleus. Scale bar: $100\mu\text{m}$. (E) Boxplot comparing the quantification of A_{dark} spermatogonia per MAGEA4⁺ spermatogonia. (F) Scatterplot and linear regression showing the correlation between the BK Score (0–10, indicating 0–100% of tubules with elongated spermatids) and the proportion of A_{dark} spermatogonia per MAGEA4⁺ spermatogonia. (G) Representative Piwi Like RNA-Mediated Gene Silencing 4 (PIWIL4) stainings in tubules from a Control and Crypto sample. Scale bar: $100\mu\text{m}$. (H) Boxplot showing the percentage of PIWIL4⁺ spermatogonia per MAGEA4⁺ spermatogonia after quantification. (I) Scatterplot and linear regression showing the correlation between the BK Score (0–10, indicating 0–100% of tubules with elongated spermatids) and the percentage of PIWIL4⁺ spermatogonia per MAGEA4⁺ spermatogonia. Boxplot elements are defined as follows: center line: median; box limits: upper and lower quartiles; whiskers: $1.5 \times$ interquartile range; points: outliers. Details about the absolute numbers and statistical analyses are provided in [Supplementary Table S3](#).

hypothesis can be confirmed, these patients may benefit from cryopreservation of sperm early after the diagnosis, avoiding the surgical approach of TESE in order to fulfill their wish for fatherhood.

The existence of two subgroups, largely presenting a meiotic-arrest or SCO phenotype, could also support the second scenario of distinct underlying genetic causes in cryptozoospermic patients. To confirm this hypothesis, genetic analyses of candidate genes previously associated with these histological phenotypes ([Oud et al., 2019](#)) are required. One example demonstrating that this approach can be successful involves the recently revealed mutations in the M1AP gene, which result in phenotypes ranging from azoospermia owing to complete meiotic arrest to cryptozoospermia ([Wyrwoll et al., 2020](#)). For counseling of respective patients, it will be relevant to assess whether those sperm that are formed are also carrying respective loss of function variants and could be passed on to the offspring.

We tested for an association with the FSHB-211 G>T polymorphism between Clusters 2 and 3, as patients with unexplained azoospermia presenting the TT-genotype in the FSHB-211 locus show decreased testicular volume and increased FSH levels

compared to GG patients ([Busch et al., 2019](#)). However, we could not find a specific enrichment for this TT-genotype in Cluster 3, despite the similarities regarding testicular volume and FSH levels. In line with our findings, cluster analysis of patients with obstructive and non-obstructive azoospermia also revealed that cluster formation did not depend on FSHB genotypes ([Schubert et al., 2020](#)), although among men with unexplained infertility this parameter was the strongest segregation marker ([Krenz et al., 2021](#)).

Intriguingly, despite these highly distinct histological phenotypes between the two cryptozoospermic clusters, the analysis of the spermatogonial compartment has revealed shared alterations. In both clusters, we found a reduction of A_{dark} spermatogonia and an increase in PIWIL4⁺ spermatogonia, in line with previous data ([Di Persio et al., 2021](#)). This suggests that the alterations of the spermatogonial compartment are a general response mechanism to the increased loss of more advanced germ cells—that is shared in both cryptozoospermic clusters irrespective of the potentially distinct underlying causes and the elevated FSH levels in Cluster 3. Moreover, we found a correlation between the reduction of the A_{dark} spermatogonia and the percentages of

tubules containing elongated spermatids, which ranged from 1% to 46% in both clusters. Those patients with a lower percentage of tubules with elongated spermatids are also those with the lowest numbers of reserve stem cells (A_{dark}); the latter are required to safeguard genetic integrity of the germline and to ensure sustainability of spermatogenesis. Consequently, in cryptozoospermic men the genetic integrity of sperm may worsen over time. Provided that this can be demonstrated in follow-up studies, cryopreservation of sperm should be offered to this patient cohort early after diagnosis. Importantly, the comparative evaluation of MAR results between the two cryptozoospermic groups does not point to an obvious impairment of sperm integrity based on the within range percentages of clinical pregnancies and live births (Nordhoff et al., 2015; Corona et al., 2019; ESHRE Clinic PI Working Group et al., 2021).

While we have uncovered the existence of subgroups within the cohort of cryptozoospermic patients, a limitation of this study is that follow-up analyses are required to uncover the underlying etiologies. As this study was limited to clinical data from humans, we do not have the opportunity to assess testicular biopsies obtained at consecutive time points. However, this would be the prerequisite to draw immediate conclusions on the progressiveness of cryptozoospermia. Another limitation of the current study is that comprehensive genetic analyses remain to be performed to uncover further loss of function variants underlying cryptozoospermia.

In conclusion, employing cluster analysis, we have identified that cryptozoospermic patients can be divided into two distinct subgroups. The leading parameters to distinguish these two clusters are the histological composition of testicular tissues, testicular volumes, and FSH levels, which may reflect distinct stages of a progressive disease phenotype. In line with this, we have shown that the alterations in the spermatogonial compartment are shared in all cryptozoospermic patients and that they become more severe with a lower number of tubules containing full spermatogenesis. This distinction of two clinical phenotypes of cryptozoospermia is, therefore, the first step to evaluate if cryptozoospermia constitutes a progressive disease and to assess potentially distinct underlying genetic causes in follow-up studies.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Acknowledgements

We thank Heidi Kerseboom and Elke Kößer for histological evaluation of testicular tissues. We also thank Sabine Forsthoof and Nicole Terwort for endocrinological measurements and the andrological laboratory for semen analyses. Finally, we thank Dr Sandra Laurentino and Dr Maria Schubert for their expert opinion on this manuscript.

Authors' roles

Study conception and design: SDP, NN; Supervision: SDP, NN; Acquisition and evaluation of clinical data: J-FC, SK, VN, HMB; Lab work: LCS, JW; Data and bioinformatic analyses: LCS, SDP;

Writing-Original Draft; LCS, SDP, and NN; All authors were involved in editing, read and approved the final version of the manuscript.

Funding

This work was supported by the German Research Foundation (grant number CRU326 to SDP and NN). We thank the Faculty of Medicine of the University of Münster for the financial support through the MedK-program. We acknowledge support from the Open Access Publication Fund of the University of Münster.

Conflict of interest

The authors declare no conflict of interest.

References

- Bergmann M, Kliesch S. Testicular biopsy and histology. In: Nieschlag E, Behre H, Nieschlag S (eds). *Andrology*. Berlin, Heidelberg: Springer, 2010;155–167.
- Burgel P-R, Paillasseur J-L, Roche N. Identification of clinical phenotypes using cluster analyses in COPD patients with multiple comorbidities. *Biomed Res Int* 2014;**2014**:420134.
- Busch AS, Tüttelmann F, Cremers J-F, Schubert M, Nordhoff V, Schüring AN, Zitzmann M, Gromoll J, Kliesch S. FSHB-211 G>T polymorphism as predictor for TESE success in patients with unexplained azoospermia. *J Clin Endocrinol Metab* 2019;**104**:2315–2324.
- Busch AS, Tüttelmann F, Zitzmann M, Kliesch S, Gromoll J. The FSHB –211G>T variant attenuates serum FSH levels in the supra-physiological gonadotropin setting of Klinefelter syndrome. *Eur J Hum Genet* 2015;**23**:700–703.
- Caldeira-Brant AL, Martinelli LM, Marques MM, Reis AB, Martello R, Almeida FRCL, Chiarini-Garcia H. A subpopulation of human Adark spermatogonia behaves as the reserve stem cell. *Reproduction* 2020;**159**:437–451.
- Clermont Y. Renewal of spermatogonia in man. *Am J Anat* 1966;**118**:509–524.
- Clermont Y. The cycle of the seminiferous epithelium in man. *Am J Anat* 1963;**112**:35–51.
- Corona G, Minhas S, Giwercman A, Bettocchi C, Dinkelmann-Smit M, Dohle G, Fusco F, Kadioglou A, Kliesch S, Kopa Z et al. Sperm recovery and ICSI outcomes in men with non-obstructive azoospermia: a systematic review and meta-analysis. *Hum Reprod Update* 2019;**25**:733–757.
- Di Persio S, Neuhaus N. Human spermatogonial stem cells and their niche in male (in)fertility: novel concepts from single-cell RNA-sequencing. *Hum Reprod* 2023;**38**:1–13.
- Di Persio S, Tekath T, Siebert-Kuss LM, Cremers J-F, Wistuba J, Li X, Meyer W Hörste G, Drexler HCA, Wyrwoll MJ, Tüttelmann F et al. Single-cell RNA-seq unravels alterations of the human spermatogonial stem cell compartment in patients with impaired spermatogenesis. *Cell Rep Med* 2021;**2**:100395.
- ESHRE Clinic PI Working Group, Vlasisavljevic V, Apter S, Capalbo A, D'Angelo A, Gianaroli L, Griesinger G, Kolibianakis EM, Lainas G, Mardesic T, Motrenko T et al. The Maribor consensus: report of an expert meeting on the development of performance indicators for clinical practice in ART. *Hum Reprod Open* 2021;**2021**:hoab022.
- Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, Guo Y, Takei Y, Yun J, Cai L et al. The adult human testis transcriptional cell atlas. *Cell Res* 2018;**28**:1141–1157.

- Krenz H, Sansone A, Kliesch S, Gromoll J, Schubert M. FSHB genotype identified as a relevant diagnostic parameter revealed by cluster analysis of men with idiopathic infertility. *Front Endocrinol (Lausanne)* 2021;**12**:780403.
- Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Soft* 2008;**25**:1–18.
- Nordhoff V, Fricke RK, Schüring AN, Zitzmann M, Kliesch S. Treatment strategies for severe oligoasthenoteratozoospermia (OAT) (<0.1 million/mL) patients. *Andrology* 2015;**3**:856–863.
- Oud MS, Volozonoka L, Smits RM, Vissers LELM, Ramos L, Veltman JA. A systematic review and standardized clinical validity assessment of male infertility genes. *Hum Reprod* 2019;**34**:932–941.
- R Development Core Team. *R: A Language and Environment for Statistical Computing: Reference Index*. Vienna: R Foundation for Statistical Computing, 2010.
- Schubert M, Kaldewey S, Pérez Lanuza L, Krenz H, Dugas M, Berres S, Kliesch S, Wistuba J, Gromoll J. Does the FSHB c.-211G>T polymorphism impact Sertoli cell number and the spermatogenic potential in infertile patients? *Andrology* 2020;**8**:1030–1037.
- Shami AN, Zheng X, Munyoki SK, Ma Q, Manske GL, Green CD, Sukhwani M, Orwig KE, Li JZ, Hammoud SS. Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis. *Dev Cell* 2020;**54**:529–547.e12.
- Sohni A, Tan K, Song H-W, Burow D, Rooij DGd, Laurent L, Hsieh T-C, Rabah R, Hammoud SS, Vicini E et al. The neonatal and adult human testis defined at the single-cell level. *Cell Rep* 2019;**26**:1501–1517.e4.
- Tüttelmann F, Luetjens CM, Nieschlag E. Optimising workflow in andrology: a new electronic patient record and database. *Asian J Androl* 2006;**8**:235–241.
- van Alphen M, van de Kant HJG, de Rooij DG. Depletion of the spermatogonia from the seminiferous epithelium of the rhesus monkey after X irradiation. *Radiat Res* 1988;**113**:473–486.
- Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab* 1999;**84**:3666–3672.
- Vernaev V, Verheyen G, Goossens A, Van Steirteghem A, Devroey P, Tournaye H. How successful is repeat testicular sperm extraction in patients with azoospermia? *Hum Reprod* 2006;**21**:1551–1554.
- World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*. Geneva: World Health Organization, 2010.
- Wyrwoll MJ, Temel ŞG, Nagirnaja L, Oud MS, Lopes AM, Heijden GWvd, Heald JS, Rotte N, Wistuba J, Wöste M et al.; GEMINI Consortium. Bi-allelic mutations in M1AP are a frequent cause of meiotic arrest and severely impaired spermatogenesis leading to male infertility. *Am J Hum Genet* 2020;**107**:342–351.
- Zhu H, Wu C, Jiang N, Wang Y, Zhao J, Xu D, Wang Q, Li M, Zeng X. Identification of 6 dermatomyositis subgroups using principal component analysis-based cluster analysis. *Int J Rheum Dis* 2019;**22**:1383–1392.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Human Reproduction, 2024, 39, 892–901
<https://doi.org/10.1093/humrep/deae013>
Original Article

Andrology

Microfluidic sperm sorting selects a subpopulation of high-quality sperm with a higher potential for fertilization

Nadia Sheibak ¹, Fatemehsadat Amjadi^{1,2}, Amir Shamloo^{3,4}, Fatemeh Zarei^{3,4}, and Zahra Zandieh ^{1,2,*}

¹Department of Anatomical Sciences, Reproductive Sciences and Technology Research Center, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

²Shahid Akbarabadi Clinical Research Development Unit (ShACRDU), Iran University of Medical Sciences, Tehran, Iran

³School of Mechanical Engineering, Sharif University of Technology, Tehran, Iran

⁴Stem Cell and Regenerative Medicine Institute, Sharif University of Technology, Tehran, Iran

*Correspondence address. Department of Anatomical Sciences, Reproductive Sciences and Technology Research Center, School of Medicine, Iran University of Medical Sciences, Tehran, Iran. Tel: +98-21-88622689; E-mail: zandieh.dvm@gmail.com  <https://orcid.org/0000-0002-7720-9465>

ABSTRACT

STUDY QUESTION: Is a microfluidic sperm sorter (MSS) able to select higher quality sperm compared to conventional methods?

SUMMARY ANSWER: The MSS selects sperm with improved parameters, lower DNA fragmentation, and higher fertilizing potential.

WHAT IS KNOWN ALREADY: To date, the few studies that have compared microfluidics sperm selection with conventional methods have used heterogeneous study population and have lacked molecular investigations.

STUDY DESIGN, SIZE, DURATION: The efficiency of a newly designed MSS in isolating high-quality sperm was compared to the density-gradient centrifugation (DGC) and swim-up (SU) methods, using 100 semen samples in two groups, during 2023–2024.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Semen specimens from 50 normozoospermic and 50 non-normozoospermic men were sorted using MSS, DGC, and SU methods to compare parameters related to the quality and fertilizing potential of sperm. The fertilizing potential of sperm was determined by measurement of phospholipase C zeta (PLC ζ) and post-acrosomal sheath WW domain-binding protein (PAWP) expression using flow cytometry, and the chromatin dispersion test was used to assess sperm DNA damage.

MAIN RESULTS AND THE ROLE OF CHANCE: In both normozoospermic and non-normozoospermic groups, the MSS-selected sperm with the highest progressive motility, PLC ζ positive expression and PLC ζ and PAWP fluorescence intensity the lowest non-progressive motility, and minimal DNA fragmentation, compared to sperm selected by DGC and SU methods ($P < 0.05$).

LIMITATION, REASONS FOR CAUTION: The major limitations of our study were the low yield of sperm in the MSS chips and intentional exclusion of severe male factor infertility to yield a sufficient sperm count for molecular experiments; thus testing with severe oligozoospermic semen and samples with low count and motility is still required. In addition, due to ethical considerations, at present, it was impossible to use the sperm achieved from MSS in the clinic to assess the fertilization rate and further outcomes.

WIDER IMPLICATIONS OF THE FINDINGS: Our research presents new evidence that microfluidic sperm sorting may result in the selection of high-quality sperm from raw semen. This novel technology might be a key to improving clinical outcomes of assisted reproduction in infertile patients.

STUDY FUNDING/COMPETING INTEREST(S): The study is funded by the Iran University of Medical Sciences and no competing interest exists.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: spermatozoa / microfluidics / DNA fragmentation / sperm-specific proteins / sperm motility

Introduction

Today, ICSI is a common choice in the treatment of infertile couples, particularly in cases with male factor infertility. However, the outcome of this technique remains around 30% (Fréour *et al.*, 2017; Wyns *et al.*, 2020; Jones *et al.*, 2022; Tharakan *et al.*, 2022). Bypassing physiologic barriers in sperm–oocyte interaction during the ICSI procedure may be associated with the unintentional injection of sperm with fragmented DNA or other hidden

abnormalities (Henkel, 2012; Fréour *et al.*, 2017; Lara-Cerrillo *et al.*, 2021; Tavalaei *et al.*, 2017c; Zhang *et al.*, 2021) and can therefore result in failed fertilization, defective embryo development, implantation failure, or even pregnancy loss (Khakpour *et al.*, 2019; Setti *et al.*, 2021; Wang *et al.*, 2022).

The *in vivo* selection of sperm in the female reproductive tract (FRT) generally acts in favor of normal sperm with intact DNA to provide the best choices for the released oocytes;

Received: September 29, 2023. Revised: January 12, 2024. Editorial decision: February 20, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

therefore, sperm preparation for ART should simulate the physiological conditions as much as possible (Tavalaee et al., 2017c; Lara-Cerrillo et al., 2021; Pinto et al., 2021). The conventional methods for sperm preparation, including swim-up (SU) and density-gradient centrifugation (DGC) techniques, isolate motile sperm with normal morphology (Baldini et al., 2021; Dai et al., 2021; Pinto et al., 2021). Both of these routine methods target motile sperm and have low efficiency regarding other sperm parameters that are essential for successful fertilization, including sperm DNA fragmentation (SDF), acrosome integrity, and genetic or molecular characteristics (Khakpour et al., 2019; Dai et al., 2021; Lara-Cerrillo et al., 2021).

One of the new advanced methods of sperm selection, microfluidics, controls the behavior of sub-microliter volumes of fluid and mimics the geometrical properties of the FRT, providing a way to isolate high-quality sperm in a process similar to natural sperm selection (Lara-Cerrillo et al., 2021; Pinto et al., 2021; Lara-Cerrillo et al., 2023). Clinical practices on the use of microfluidic sperm sorter (MSS) chips for sperm selection have produced sperm with higher motility (Gode et al., 2020; Ataei et al., 2021), intact acrosomes (de Wagenaar et al., 2015), lower SDF (Anbari et al., 2021; Mirsanei et al., 2022), lower reactive oxygen species generation (Erimşah et al., 2020; Gode et al., 2020), and lower incidence of apoptosis (Vasilescu et al., 2023). The use of MSS-selected sperm in ART has resulted in a higher rate of euploid top-quality embryos (Parrella et al., 2019; Yetkinel et al., 2019; Guler et al., 2021; Budak et al., 2023), and increased fertilization (Yildiz and Yukse, 2019; Mirsanei et al., 2022), implantation (Anbari et al., 2021), pregnancy (Parrella et al., 2019; Patel et al., 2020), and live birth rates (Lara-Cerrillo et al., 2023).

Phospholipase C zeta (PLC ζ) and post-acrosomal sheath WW domain-binding protein (PAWP) are two sperm-born oocyte activating factors (SOAF) located in the perinuclear theca of the sperm head and are recognized as main biomarkers of sperm fertilizing potential and mediators of oocyte fertilization (Anifandis et al., 2016; Tavalaee et al., 2017b,c). It has been reported that lower levels of these factors or any change in their structure or location are associated with a reduced fertilization rate and affect the development of embryos, even in cases with normal sperm parameters (Aarabi et al., 2014; Anifandis et al., 2016; Mirsanei et al., 2022).

Routine semen analysis and conventional sperm preparation methods are inefficient with regard to these preferential biomarkers of sperm quality and predictions of outcomes of ART procedures (Pinto et al., 2021; Vasilescu et al., 2023). Since the use of microfluidics in ART is still in the phase of development, the quality of sperm sorted by this method, especially molecular profiles, has been less addressed. Therefore, we aimed to introduce a novel microfluidic chip and to compare the quality of the isolated sperm with sperm isolated by conventional sperm preparation methods including SU and DGC.

Materials and methods

Study population

Semen samples were obtained from males referred to the IVF clinic, at the Shahid Akbarabadi Hospital, for infertility screening ($n = 100$), as two age-matched groups: normozoospermic ($n = 50$) and non-normozoospermic ($n = 50$) men, based on primary semen analysis. Participants were classified as having normal or abnormal sperm parameters according to the sixth edition of the World Health Organization (WHO) criteria (World Health Organization, 2021). To yield enough sperm for further

experiments, men with severe asthenozoospermia and severe oligozoospermia (total motile sperm count < 5 million) were excluded from the study. The study was conducted in accordance with the Declaration of Helsinki for conducting research involving humans and it was approved by the Research Ethics Committees of the School of Medicine, Iran University of Medical Sciences (IR.IUMS. FMD.REC.1401.069). To use surplus semen samples, written informed consent was obtained from all participants.

Semen analysis

After infertility screening, surplus samples were enrolled in this study. Following the liquefaction, each raw semen sample was divided into four aliquots: the first portion was analyzed to serve as the control, and the other fractions were processed by DGC, SU, and MSS methods. Routine semen analysis including sperm concentration and motility (total, progressive, and non-progressive) was assessed according to WHO criteria (World Health Organization, 2021). Sperm morphology was assessed using the DIFF QUIK (rapid sperm staining) kit (Ideh Varzan Farda, IVF Co., Tehran, Iran), and sperm with normal oval heads, containing 40–70% acrosome, without cytoplasmic residues and without neck, midpiece, or tail defects were considered normal (Supplementary Fig. S1) based on Kruger strict criteria (Kruger et al., 2019).

Sperm preparation

Density-gradient centrifugation

One fraction of semen samples was processed by the DGC method (Xiao et al., 2021) to separate motile sperm using a double-density gradient of 40/80 (ORIGIO, Cooper-surgical).

Swim-up

The SU method was performed using washed samples (quickly centrifuged and resuspended) as described previously (Xiao et al., 2021). The final sperm pellet was placed under medium (ORIGIO Sperm Wash, Cooper-surgical) and incubated for 1 h at 37°C and then the uppermost portion containing motile sperm was collected.

Microfluidics sperm sorting

The MSS chip, designed using the CAD software (Autodesk AutoCAD 2023.0.1), consists of three parts including two identical sinusoidal units at the beginning and end of the route, and a curved path between these two parts. Sinusoidal channels were considered with the idea of sperm sinus movement, and due to the impact of curved paths on improving sorting efficiency, the middle-curved channel was added to the platform. The details of the design dimensions are shown in Fig. 1A.

This sorter was designed based on the passively driven technology that requires no external power source, and sorting occurs only by the driving force of sperm motility. As there is no difference in heights between two columns of fluid (inlet and outlet), then there will be no gradient of pressure and, consequentially, no flow rate (Guzzi et al., 2020) and the sperm movement occurred only due to its intrinsic motility in the specific geometry of the microfluidic platform. It mimics the sperm competition in the *in vivo* environment, and allowing for the best and most competent sperm to be selected. The concept that sperm tend to swim near walls (Heydari et al., 2023) was used to allow sperm that are 100% motile and capable of exploiting boundary-following behavior to reach the target reservoir (outlet), while undesirable sperm remained in the inlet or along the path of the microchannels.

The final designed structure was output to a micro-milling machine (Roland MDX-40A, Japan) for direct engraving on the polymethyl methacrylate (PMMA) substrate and then attached to the second layer of a PMMA sheet using a pressure-sensitive

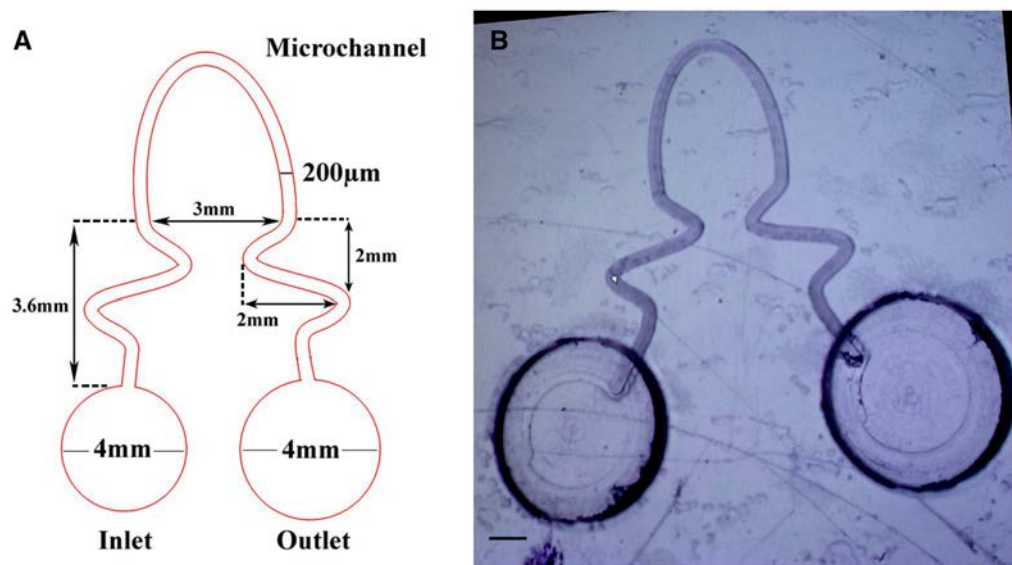


Figure 1. The microfluidics sperm-sorting device. Schematic diagram of the microfluidics device (A) and the fabricated microfluidics sperm sorter made of poly methyl methacrylate (B), Scale bar = 500 µm.

adhesive of medical grade (DFM 200 clear 150 POLY H-9V-95, FLEXcon, USA), as described by Pishbin et al. (2019, 2020) (Fig. 1B). Despite the previous approval of PMMA safety for human sperm (Asghar et al., 2014; Vasilescu et al., 2021, 2023), the impact of the MSS chip on sperm vitality was confirmed again by the human sperm survival assay (Meintjes, 2017) and eosin–nigrosin staining (Supplementary Fig. S2) (World Health Organization, 2021).

To sort sperm by the microfluidic method, the MSS was initially filled with pre-warmed medium (ORIGIO Sperm Wash, Cooper-surgical). Then 5 µl of liquefied semen was gently added to the inlet, and the MSS was incubated at 37°C for 30 min. Because of the low concentration yield of the MSS, each sample was sorted using five chips to achieve a sufficient a number of sperm for the following experiments.

Sperm chromatin dispersion

The sperm chromatin dispersion test was used to evaluate the integrity of sperm DNA, based on halo formation after treating sperm with an acid solution (World Health Organization, 2021), using the SDFA[®] kit (Ideh Varzan Farda, IVF Co., Iran). The sperm with large and medium halos was considered to have no DNA fragmentation. However, the sperm with small halos (smaller than one-third of head minor diameter) or without halos were recorded as sperm with fragmented DNA (Supplementary Fig. S1). At least 500 sperm per sample were evaluated to estimate SDF.

Assessment of PLCζ and PAWP by flow cytometry

To evaluate the fertilizing potential of sperm, the percentage of sperm with positive staining for PLCζ and PAWP, as two SOAF candidates, were measured using flow cytometry, as previously described by Tavalae et al. (2017c). Briefly, after permeabilization of sperm membrane with 0.5% Triton X-100, and blocking with 3% bovine serum albumin (BSA)/phosphate-buffered saline (PBS), samples were separately incubated with anti-PLCζ and anti-PAWP antibodies (respectively Covalab, France, 1:500 and Abcam, UK, 1:200) overnight at 4°C and fluorescence labeling was performed next day by incubation with a FITC-conjugated secondary antibody (goat anti-rabbit IgG; Sigma-Aldrich, 1:100). Propidium iodide (Sigma-Aldrich, 5 µg/ml) was used as a DNA-

binding dye for verifying the specificity of the sperm gate. The mean fluorescence intensity and percentage of positively labeled sperm were determined by a FACS Calibur flow cytometer (Becton-Dickinson, USA), and the resulting data were analyzed by FlowJo software, version 10.0 (TreeStar-Inc., USA). In addition, a Nikon fluorescence microscope was used to localize PLCζ and PAWP in positively stained sperm.

Statistical analysis

At a confidence level of 95% and a test power of 80%, the minimum sample size was estimated as 31 per group (using Stata software version 16). Results of the present study are expressed as mean ± SD. All statistical analyzes were carried out using SPSS Statistics software, version 27.0 (IBM, USA), with a significance level of $P < 0.05$. Statistical comparisons of the mean values between different groups were performed using repeated measures ANOVA and *post-hoc* comparisons were performed with Tukey's HSD test.

Results

Sperm parameters

Sperm parameters in raw semen samples were analyzed and no difference was found between the normozoospermic and non-normozoospermic groups with regard to age, semen volume, and primary SDF ($P > 0.05$) (Supplementary Table S1).

We found a significant difference in all sperm parameters among preparation methods, and all three methods resulted in selecting sperm with better parameters (Fig. 2A, Supplementary Table S2).

Normozoospermic group

In the normozoospermic group (Table 1), there was a significant difference between sperm concentrations after processing by the three different methods, with superiority in the DGC method (respectively $P = 0.001$ and $P < 0.001$ compared to SU and MSS). The total motility in sperm prepared by SU and MSS was higher than the DGC method (respectively $P = 0.023$ and $P < 0.001$). The highest progressive motility was achieved using MSS (respectively $P = 0.001$ and $P < 0.001$ compared to SU and DGC), and

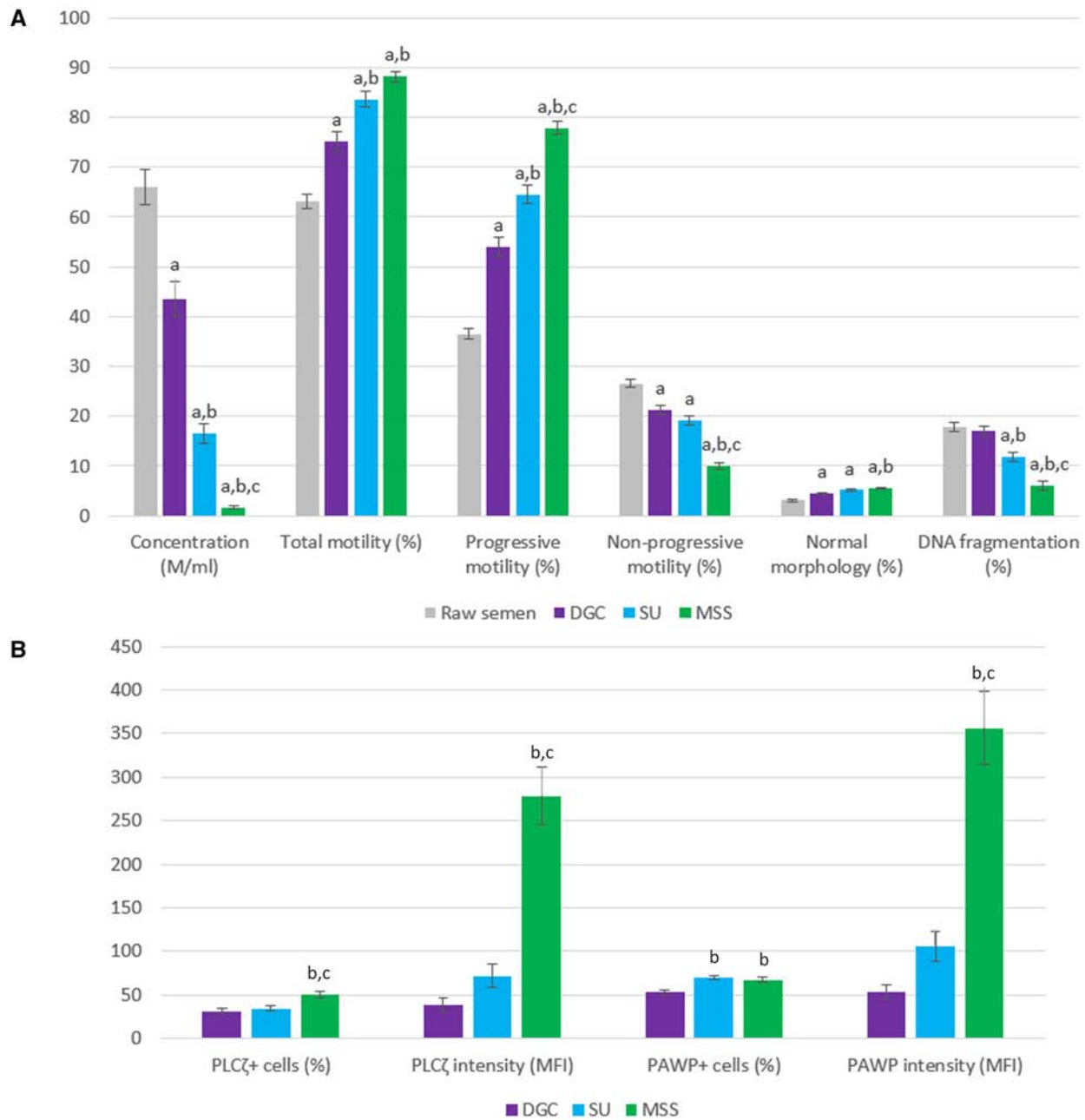


Figure 2. Evaluation of sperm quality in the whole study population. Sperm parameters (A) and sperm fertilizing potential (B) in the whole study population ($n = 100$). Data are presented as mean (SEM). DGC: density-gradient centrifugation; MFI: mean fluorescence intensity; MSS: microfluidic sperm sorter; PAWP: post-acrosomal WW domain-binding protein; PLCζ: phospholipase C zeta; SU: swim-up; $P < 0.05$ was set as significance; a: significant difference compared to raw semen (post-hoc Tukey HSD); b: compared to DGC (post-hoc Tukey HSD); c: compared to SU (post-hoc Tukey HSD).

moreover, the lowest non-progressive motility was also observed using MSS ($P < 0.001$ compared to both SU and DGC). Regarding progressive motility, there was a significant difference between the SU and DGC methods ($P = 0.043$).

Non-normozoospermic group

Evaluation of the non-normozoospermic group also showed a significant difference in all sperm parameters between the three methods (Table 2). The highest sperm concentration resulted from the DGC method ($P < 0.001$ compared to both SU and MSS methods). Sperm total motility showed no significant difference between the SU and MSS methods ($P > 0.05$), despite a significant increase with MSS compared to the DGC method ($P < 0.001$). The

highest progressive motility was observed with MSS with a significant difference compared to DGC ($P < 0.001$) and SU ($P < 0.001$) methods. The sperm sorted by the MSS device had the lowest non-progressive motility (respectively $P < 0.001$ compared to both SU and DGC) and the percentage of sperm with normal morphology was higher in the MSS group compared to the DGC group ($P = 0.003$).

Sperm DNA fragmentation

Regardless of the study grouping, the lowest DNA damage was achieved by MSS, with a significant difference compared to SU and DGC methods (for both $P < 0.001$) (Fig. 2A, Supplementary Table S1).

Table 1. Evaluation of sperm parameters and fertilizing potential in normozoospermic participants (n = 50).

Groups Parameters	Raw semen	DGC	SU	MSS	P-value
Concentration (M/ml)	79.13 ± 22.61 (69.34–88.91)	47.86 ± 24.97 ^a (37.06–58.66)	21.98 ± 14.98 ^{a,b} (15.50–28.46)	2.07 ± 1.92 ^{a,b,c} (1.23–2.90)	<0.001*
Total motility (%)	69.17 ± 6.35 (66.42–71.92)	77.58 ± 13.44 ^a (71.76–83.39)	85.85 ± 9.44 ^{a,b} (81.77–89.93)	90.07 ± 7.71 ^{a,b} (86.73–93.40)	<0.001*
Progressive motility (%)	41.91 ± 4.89 (39.79–44.03)	56.11 ± 13.75 ^a (50.16–62.06)	66.68 ± 12.47 ^{a,b} (61.28–72.07)	79.36 ± 7.18 ^{a,b,c} (76.25–82.46)	<0.001*
Non-progressive motility (%)	27.26 ± 4.81 (25.17–29.34)	21.46 ± 7.32 ^a (18.29–24.63)	19.17 ± 7.67 ^a (15.85–22.49)	10.27 ± 4.43 ^{a,b,c} (8.35–12.19)	<0.001*
Normal morphology (%)	4.68 ± 0.63 (4.40–4.95)	4.75 ± 0.70 (4.44–5.05)	5.38 ± 1.43 (4.76–6.01)	5.60 ± 1.72 ^a (4.86–6.35)	0.035*
DNA fragmentation (%)	17.47 ± 6.70 (14.57–20.36)	17.02 ± 6.99 (14.00–20.05)	13.06 ± 7.55 (9.79–16.32)	6.08 ± 4.96 ^{a,b,c} (3.93–8.32)	<0.001*
PLCζ ⁺ cells (%)	–	37.53 ± 22.61 (27.75–47.30)	37.91 ± 23.50 (27.74–48.07)	56.45 ± 22.59 ^{b,c} (46.68–66.22)	0.009*
PLCζ intensity (MFI)	–	29.27 ± 37.80 (12.92–45.61)	53.67 ± 42.17 (35.44–71.91)	315.23 ± 242.57 ^{b,c} (210.33–420.12)	<0.001*
PAWP ⁺ cells (%)	–	50.16 ± 22.41 (40.47–59.85)	71.35 ± 16.69 ^b (64.13–78.57)	64.57 ± 22.45 ^b (54.86–74.27)	0.012*
PAWP intensity (MFI)	–	38.12 ± 38.75 (21.36–54.88)	67.50 ± 44.53 (48.24–86.76)	374.99 ± 300.36 ^{b,c} (245.10–504.87)	<0.001*

Data are presented as mean ± SD (95% confidence intervals of the means).

DGC: density-gradient centrifugation; MFI: mean fluorescence intensity; MSS: microfluidic sperm sorter; PAWP: post-acrosomal WW domain-binding protein; PLCζ: phospholipase C zeta; SU: swim-up.

* Significant difference between groups at $P < 0.05$ using repeated measures ANOVA test (followed by post-hoc Tukey HSD test).

^a Compared to raw.

^b Compared to DGC.

^c Compared to SU.

Table 2. Evaluation of sperm parameters and fertilizing potential in non-normozoospermic participants (n = 50).

Groups Parameters	Raw semen	DGC	SU	MSS	P-value
Concentration (M/ml)	52.95 ± 17.97 (45.18–60.73)	39.17 ± 22.46 (29.45–48.88)	10.90 ± 9.07 ^{a,b} (6.97–14.82)	1.33 ± 1.51 ^{a,b} (0.67–1.98)	<0.001*
Total motility (%)	57.08 ± 8.52 (53.40–60.77)	72.82 ± 13.18 ^a (67.11–78.52)	81.41 ± 10.59 ^{a,b} (76.83–85.99)	86.28 ± 6.96 ^{a,b} (83.27–89.29)	<0.001*
Progressive motility (%)	31.17 ± 5.70 (28.70–33.63)	51.81 ± 14.30 ^a (45.63–58.00)	62.41 ± 10.96 ^{a,b} (57.67–67.16)	76.44 ± 9.16 ^{a,b,c} (72.47–80.40)	<0.001*
Non-progressive motility (%)	25.91 ± 6.63 (23.04–28.78)	21.00 ± 3.93 ^a (19.30–22.70)	18.99 ± 4.81 ^a (16.90–21.07)	9.84 ± 4.85 ^{a,b,c} (7.74–11.94)	<0.001*
Normal morphology (%)	1.58 ± 0.92 (1.18–1.97)	4.29 ± 0.66 ^a (4.00–4.58)	4.94 ± 1.05 ^a (4.49–5.40)	5.40 ± 1.42 ^{a,b} (4.79–6.02)	<0.001*
DNA fragmentation (%)	18.22 ± 5.71 (15.75–20.69)	16.98 ± 6.41 (14.21–19.75)	10.33 ± 3.86 ^{a,b} (8.66–12.00)	5.40 ± 6.10 ^{a,b,c} (2.76–8.04)	<0.001*
PLCζ ⁺ cells (%)	–	25.49 ± 19.43 (17.09–33.90)	30.29 ± 16.55 (23.13–37.45)	44.00 ± 23.31 ^{b,c} (33.92–54.08)	0.001*
PLCζ intensity (MFI)	–	48.05 ± 63.80 (20.46–75.65)	90.39 ± 115.10 (40.61–140.16)	241.56 ± 197.83 ^{b,c} (156.01–327.12)	<0.001*
PAWP ⁺ cells (%)	–	55.79 ± 20.61 (46.88–64.71)	69.43 ± 16.97 ^b (62.10–76.77)	69.77 ± 16.30 ^b (62.72–76.83)	0.018*
PAWP intensity (MFI)	–	68.47 ± 65.93 (39.96–96.99)	145.19 ± 151.29 (79.77–210.62)	337.66 ± 269.95 ^{b,c} (220.93–454.40)	<0.001*

Data are presented as mean ± SD (95% confidence intervals of the means).

DGC: density-gradient centrifugation; MFI: mean fluorescence intensity; MSS: microfluidic sperm sorter; PAWP: post-acrosomal WW domain-binding protein; PLCζ: phospholipase C zeta; SU: swim-up.

* Significant difference between groups at $P < 0.05$ using repeated measures ANOVA test (followed by post-hoc Tukey HSD test).

^a Compared to raw.

^b Compared to DGC.

^c Compared to SU.

Normozoospermic group

In the normozoospermic samples, the SDF by the MSS device was significantly lower than raw semen ($P < 0.001$), SU ($P = 0.003$), and DGC ($P < 0.001$) methods (Table 1).

Non-normozoospermic group

In non-normozoospermic group also, the MSS-selected sperm with minimal DNA fragmentation compared to the SU ($P = 0.019$) and DGC ($P < 0.001$) methods. On the other hand, SDF

after the SU method was significantly lower than after DGC ($P < 0.001$) (Table 2).

Sperm fertilizing potential

To compare sperm quality between these three methods, the fertilizing potential of sperm was evaluated using flow cytometric measurement of two candidates as sperm-specific factors, PLCζ, and PAWP (Fig. 3). In addition, a fluorescence microscope was

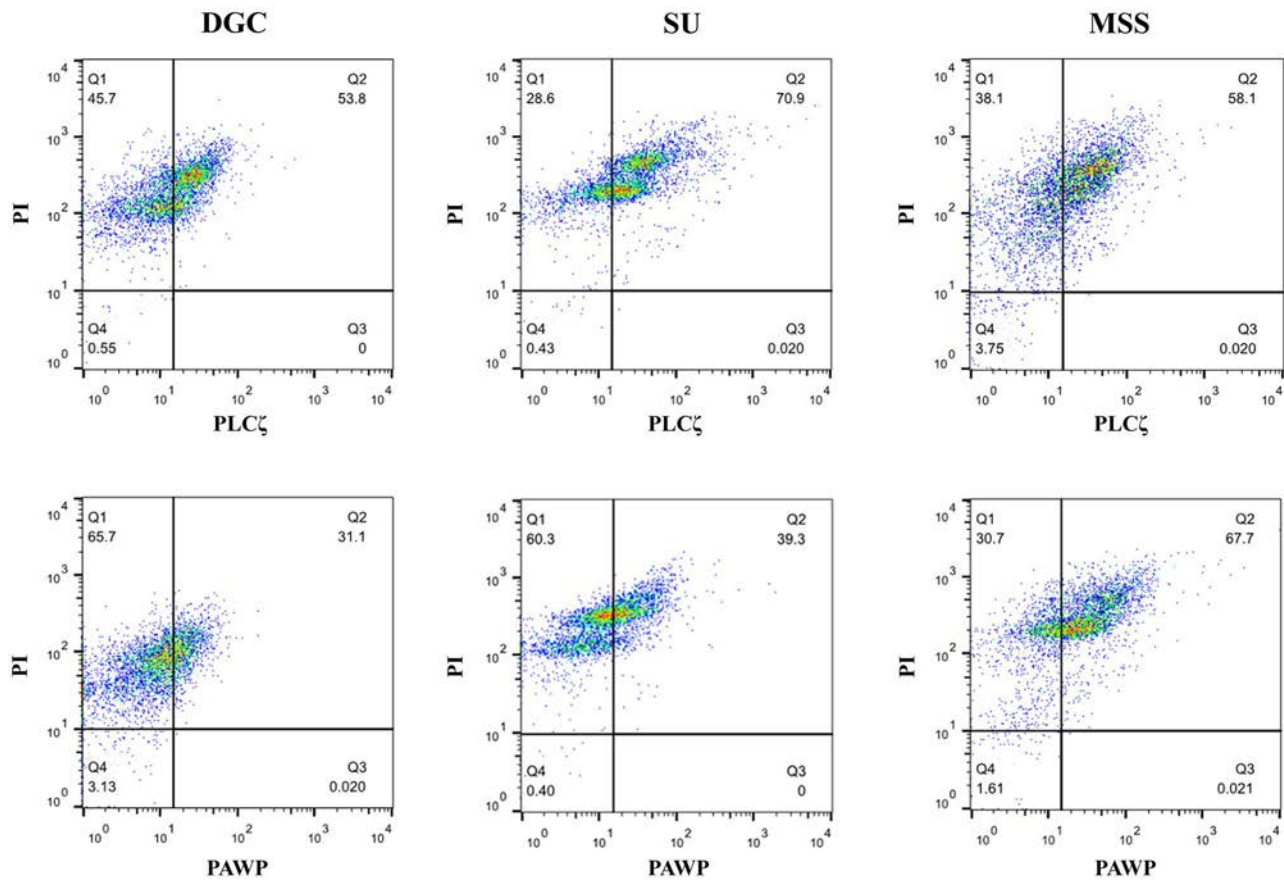


Figure 3. Comparison of sperm fertilizing potential. Sperm prepared by density-gradient centrifugation (DGC), swim-up (SU), and microfluidic sperm sorting (MSS) methods were compared using assessment of phospholipase C zeta (PLC ζ) and post-acrosomal WW domain-binding protein (PAWP) by flow cytometry.

used to localize PLC ζ and PAWP in positively stained sperm (Fig. 4).

The highest percentage of PLC ζ ⁺ sperm was obtained by MSS, with a significant difference compared to DGC and SU methods (respectively $P < 0.001$ and $P = 0.002$, Fig. 2B, Supplementary Table S3). In addition, the highest fluorescence intensity of both PLC ζ and PAWP was observed for MSS, which comes from the higher content of these factors ($P < 0.001$ compared to both SU and DGC) (Fig. 2B, Supplementary Table S3).

Normozoospermic group

In normozoospermic group, the highest percentage of PLC ζ ⁺ sperm was obtained using MSS ($P = 0.018$ and $P = 0.021$, respectively, compared to DGC and SU), with the highest PLC ζ intensity compared to DGC and SU methods (for both $P < 0.001$) (Table 1). Both MSS and SU methods resulted in a higher percentage of PAWP⁺ sperm compared to the DGC method (respectively $P = 0.045$ and $P = 0.003$). There was also a significant difference in the PAWP intensity; in the MSS-obtained sperm, it was higher than after the DGC and SU methods (for both $P < 0.001$) (Table 1).

Non-normozoospermic group

The percentage of PLC ζ ⁺ sperm was higher using MSS compared to DGC ($P = 0.007$) and SU ($P = 0.049$) methods (Table 2). In addition, the highest PLC ζ intensity was found in the MSS group (respectively $P = 0.001$ and $P < 0.001$ compared to SU and DGC). Semen preparation by the DGC method showed a lower percentage of PAWP⁺ sperm compared to the SU ($P = 0.034$) and MSS

($P = 0.029$) methods. Finally, the fluorescence intensity of PAWP was higher in sperm selected by MSS compared to the SU (0.002) and DGC ($P < 0.001$) methods.

Correlations between sperm quality-related factors

To find any correlation between sperm parameters, data related to all participants (regardless of grouping) were analyzed using the Spearman test. There was a strong inverse correlation between sperm total motility and its DNA fragmentation ($r: -0.532$, $P < 0.001$); therefore, any increase in DNA fragmentation could reduce sperm motility. We also found an inverse correlation between DNA fragmentation and the percentage of both PLC ζ ⁺ and PAWP⁺ sperm separately (respectively $r: -0.338$, $P < 0.001$ and $r: -0.287$, $P = 0.001$). SDF was also strongly and inversely correlated with fluorescence intensity of both PLC ζ and PAWP (respectively $r: -0.567$, $P < 0.001$ and $r: -0.533$, $P < 0.001$).

A direct correlation existed between the percentage of PLC ζ ⁺ sperm and sperm total motility and there was a similar correlation with progressive motility (respectively $r: 0.342$, $P < 0.001$, and $r: 0.397$, $P < 0.001$). This correlation also applied for the percentage of PAWP⁺ sperm, respectively with sperm total motility ($r: 0.317$, $P < 0.001$) and progressive motility ($r: 0.374$, $P < 0.001$). In addition, there was an inverse correlation between the percentage of PLC ζ ⁺ sperm and non-progressive motility ($r: -0.276$, $P = 0.001$) and between the percentage of PAWP⁺ sperm and non-progressive motility ($r: -0.239$, $P = 0.005$). A direct correlation was found between the percentage of PLC ζ ⁺ sperm and the

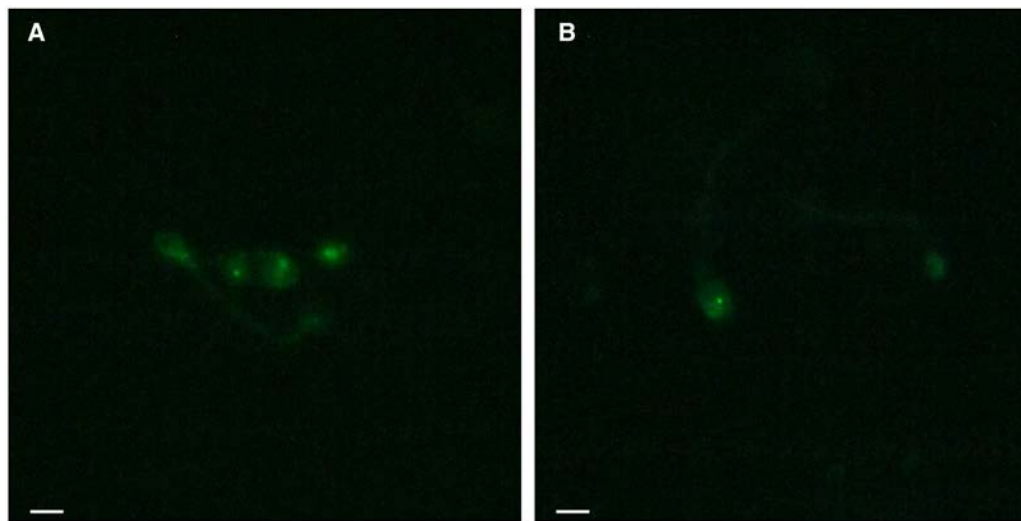


Figure 4. Localization of sperm-specific factors. Phospholipase C zeta (A) and post-acrosomal WW domain-binding protein (B) are located in the perinuclear theca of the sperm heads. Fluorescence microscopy, $\times 1000$ magnification, Scale bar = 5 μm .

percentage of sperm with normal morphology ($r: 0.184$, $P = 0.031$). Statistical analysis also indicated a direct correlation between the percentage of PLC ζ^+ and PAWP $^+$ sperm ($r: 0.484$, $P < 0.001$).

The Eta-coefficient test was used to examine the relationships between sperm preparation methods and sperm parameters and showed that the method used for sperm sorting is associated with all sperm parameters ($P < 0.05$). The result of these tests showed a strong association between preparation method and SDF ($\eta: 0.57$), progressive motility ($\eta: 0.64$), PLC ζ intensity ($\eta: 0.60$), and PAWP intensity ($\eta: 0.59$). Therefore, the preparation method affects sperm progressive motility more than other parameters.

Discussion

The present MSS device provided a simple way to separate clinically relevant concentration of high-quality sperm for ART procedures in a biomimetic manner that selected sperm with high progressive motility and low DNA damage containing higher levels of critical SOAFs (PLC ζ and PAWP) without the need for prior washing, centrifugation, and additional manual work.

To date, limited studies have compared microfluidics sperm selection with conventional methods, and have used mostly heterogeneous populations with a lack of molecular investigations (Erimşah et al., 2020; Jeřeta et al., 2020; Patel et al., 2020; Ozcan et al., 2021; Zaha et al., 2023). We compared the quality of sperm selected by the MSS device with the conventional DGC and SU methods and found that sperm concentration decreased significantly after processing by DGC, SU, and MSS methods, in descending order. This is in agreement with all previous studies that found a lower sperm concentration with a MSS device compared to the unprocessed sample, SU and DGC methods (Quinn et al., 2018; Rappa et al., 2018; Parrella et al., 2019; Vasilescu et al., 2023). It should be considered that the required volume of semen for sperm sorting by microfluidics is significantly lower than for other methods (Parrella et al., 2019; Erimşah et al., 2020; Vasilescu et al., 2023). So, the main reason for the low sperm yield by the MSS is the limited loading volume of semen. On the other hand, only 50 000 highly motile sperm (per oocyte) are suggested to be sufficient for IVF and at least 1000 sperm are required for ICSI procedure (Xiao et al., 2021; Leung et al., 2022; Vasilescu et al.,

2023), as microfluidics devices have been used effectively in ICSI trials (Parrella et al., 2019; Mirsanei et al., 2022; Zaha et al., 2023).

The higher progressive motility and lower non-progressive motility by the present MSS device is not far from expectation in a passive motility-based platform (Shamloo et al., 2019). Previously, a higher percentage of sperm with progressive motility using MSS devices has been achieved in several studies, when compared to raw semen (Mirsanei et al., 2022; Vasilescu et al., 2023), DGC (Parrella et al., 2019; Mirsanei et al., 2022; Zaha et al., 2023), SU (Erimşah et al., 2020; Vasilescu et al., 2023), direct-SU (Anbari et al., 2021), and DGC + SU methods (Shirota et al., 2016; Quinn et al., 2018). It is suggested that the microfluidics method selects motile sperm with more energy and potentially more fecundity, similar to sperm subpopulations that in the *in vivo* condition have a higher chance of migrating along the female genital tract, reaching the oocyte, and contributing to the fertilization process (Vasilescu et al., 2023).

Both the MSS and SU methods showed improved sperm morphology compared to raw semen and the DGC technique. Our previous study also resulted in the selection of a higher percentage of normal-shaped sperm by a commercial MSS device compared to the DGC method or raw semen (Mirsanei et al., 2022). In agreement, Erimşah et al. (2020) reported a significant improvement in sperm morphology by MSS compared to DGC and SU methods. Anbari et al. (2021) using motile sperm organelle morphology examination (MSOME) found a higher portion of good-quality sperm by a MSS compared to direct-SU (Anbari et al., 2021). Zhang et al. (2015) by studying sperm ultrastructure using transmission electron microscopy found the highest percentage of sperm with normal structure and lower head deformity rate in the MSS group compared to raw semen and the DGC method.

The MSS device used here resulted in sperm with a minimum level of SDF compared to both SU and DGC methods. Contrary to our results, the failure of DGC and SU methods in preparing samples with high DFI, and even the induction of DNA damage, has been reported in previous studies (Mirsanei et al., 2022; Pujol et al., 2022; Vasilescu et al., 2023) and has been attributed to induction of oxidative stress by gravitational force during centrifugation and long incubation times due to more manual interventions (Mirsanei et al., 2022; Vasilescu et al., 2023). Similarly, SDF in sperm sorted by microfluidics devices was

evaluated previously using different assays and confirmed that this method as the best way to reduce the percentage of sperm with fragmented DNA (even to 0% without any fragmented sperm) (Asghar *et al.*, 2014; Quinn *et al.*, 2018; Anbari *et al.*, 2021; Mirsanei *et al.*, 2022; Vasilescu *et al.*, 2023). Due to the important role of sperm DNA integrity in embryo development, implantation, and also miscarriage, the selection of sperm with low SDF is essential for a successful live birth following IVF/ICSI (Vasilescu *et al.*, 2023). Lara-Cerrillo *et al.* (2023) achieved higher pregnancy and live births rates using the MSS in couples with high seminal SDF, after they had failed ICSI trials using the DGC and SU methods. Our results showed a strong association between the sperm preparation method and the induction of DNA fragmentation. It was confirmed that DNA damage induced by oxidative stress would negatively affect sperm motility (Quinn *et al.*, 2018; Casanovas *et al.*, 2019; Pujol *et al.*, 2022). Thereby, any sperm-sorting method that selects sperm with the highest motility would result in the lowest DNA fragmentation (Mirsanei *et al.*, 2022; Pujol *et al.*, 2022), as was shown with the MSS. Highly motile sperm with intact DNA (unlike poor-quality sperm) showed an increased tendency to swim close to walls and follow solid boundaries (Nosrati *et al.*, 2014; Nosrati, 2016; Khodamoradi *et al.*, 2021). This could be the main reason for sorting sperm with the highest motility and the lowest SDF in the microfluidic method.

Aarabi *et al.* (2014) have indicated that sperm PAWP content could be a valid biomarker for the prediction of the fertilization rate following ICSI and successful embryo development independent of sperm parameters and oocyte quality. A correlation between PLC ζ expression and the fertilization rate has also been reported (Yelumalai *et al.*, 2015), and impaired sperm function has been associated with low-levels or abnormal forms of this factor, that resulted in failure in fertilization and embryo development (Yan *et al.*, 2020; Mirsanei *et al.*, 2022). Based on our findings, the MSS device sorted more sperm with the positive expression of PLC ζ and PAWP compared to the conventional methods. A decreased percentage of PLC ζ ⁺ and PAWP⁺ sperm and total levels of these factors have been reported in patients with fertilization failure, compared to fertile men (Azad *et al.*, 2018a). However, MSS resulted in a higher level of PLC ζ expression than the DGC method, even in patients with a history of fertilization failure (with lower baseline PLC ζ levels) (Mirsanei *et al.*, 2022). None of the published data concerning the microfluidics method have evaluated the level of PAWP in the sorted sperm. Similar to our findings, Tavalae *et al.* (2017c) have indicated that the percentage of PLC ζ ⁺ and PAWP⁺ sperm after preparation by the DGC method was significantly higher than in fresh semen. We found that the sperm preparation method is associated with the level of PLC ζ and PAWP, and there was a straight correlation between the expressions of these factors; as previously has been reported (Tavalae *et al.*, 2017b; Azad *et al.*, 2018a,b), in cases with a higher percentage of PLC ζ ⁺ sperm, higher PAWP⁺ sperm would also be found. Similarly, it is demonstrated that SDF is inversely correlated with PLC ζ and PAWP expression (Tavalae *et al.*, 2017a,b; Kashir, 2020). Therefore, in sperm with a lower SDF, a higher content of PLC ζ and PAWP, and probably a higher fertilizing potential are expected.

Based on our findings, sperm subpopulations with higher total and progressive motility and lower non-progressive motility showed an increased percentage of expression of both PLC ζ and PAWP. A significant correlation between the PLC ζ level and sperm motility was previously reported (Abadi *et al.*, 2016). We observed a higher percentage of PLC ζ positive expression in sperm with normal morphology. Similarly, Abadi *et al.* (2016)

have indicated that sperm morphology significantly correlates with the expression of both PAWP and PLC ζ .

The MSS device introduced in the present study was fabricated based on passive approaches, which were originally developed to imitate the physiological path of sperm transit and mimic the motility-based competition that exists in the *in vivo* condition of the female genital tract, thus sorting sperm with higher quality and motility analogously (Khodamoradi *et al.*, 2021). The major advantages of the MSS used in our study are its non-invasiveness, its lack of complexity in structure, and its simple design, fabrication, and working procedure.

A major limitation of our study was low yield of sperm in the MSS chips. On the other hand, due to ethical considerations, it was impossible to use the sperm achieved from MSS in the clinic to assess the fertilization rate and further outcomes using human oocytes. Another limitation of our study was the intentional exclusion of severe male factors to yield sufficient sperm count for the molecular experiments; testing with severe oligozoospermic semen and samples with low count and motility is still required.

In conclusion, microfluidics sperm sorting could be a good choice to obtain motile sperm with higher quality and fertilizing potential for IVF/ICSI, without any adverse effects or damage to the sperm compared to conventional sperm preparation methods, and may result in improved ART outcomes.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article are available in the article and in its online [supplementary material](#). More details will be shared on reasonable request to the corresponding author.

Acknowledgements

The authors are grateful to the assistance of all personnel of the IVF clinic of Shahid Akbarabadi Hospital.

Authors' roles

Z.Z., F.A., and N.S. were involved in conception and design of the research, and the microfluidic device was designed and fabricated by A.S. and F.Z. Sample collection, platform testing, and data analysis were performed by N.S. Z.Z. and F.A. interpreted and validated the data and N.S. drafted the manuscript. All authors participated in manuscript revision and approved it.

Funding

This work was supported by the Iran University of Medical Sciences (Project no. 22044-4-3-1400).

Conflict of interest

The authors declare no potential conflict of interest.

References

- Aarabi M, Balakier H, Bashar S, Moskovtsev SI, Sutovsky P, Librach CL, Oko R. Sperm content of postacrosomal WW binding protein

- is related to fertilization outcomes in patients undergoing assisted reproductive technology. *Fertil Steril* 2014;**102**:440–447.
- Abadi MK-D, Tavalaei M, Shahverdi A, Nasr-Esfahani MH. Evaluation of PLC ζ and PAWP expression in globozoospermic individuals. *Cell J* 2016;**18**:438.
- Anbari F, Khalili MA, Ahamed AMS, Mangoli E, Nabi A, Dehghanpour F, Sabour M. Microfluidic sperm selection yields higher sperm quality compared to conventional method in ICSI program: a pilot study. *Syst Biol Reprod Med* 2021;**67**:137–143.
- Anifandis G, Messini CI, Dafopoulos K, Daponte A, Messinis IE. Sperm contributions to oocyte activation: more that meets the eye. *J Assist Reprod Genet* 2016;**33**:313–316.
- Asghar W, Velasco V, Kingsley JL, Shoukat MS, Shafiee H, Anchan RM, Mutter GL, Tüzel E, Demirci U. Selection of functional human sperm with higher DNA integrity and fewer reactive oxygen species. *Adv Healthc Mater* 2014;**3**:1671–1679.
- Ataei A, Lau A, Asghar W. A microfluidic sperm-sorting device based on rheotaxis effect. *Microfluid Nanofluid* 2021;**25**:52.
- Azad N, Nazarian H, Nazari L, Novin MG, Piryaee A, Heidari MH, Farahani RM, Sadjadpour SS. Evaluation of PAWP and PLC ζ Expression in infertile men with previous ICSI fertilization failure. *Urol J* 2018;**15**:116.
- Azad N, Nazarian H, Novin MG, Farahani RM, Piryaee A, Heidari MH, Alitappeh MA. Oligoasthenoteratozoospermic (OAT) men display altered phospholipase C ζ (PLC ζ) localization and a lower percentage of sperm cells expressing PLC ζ and post-acrosomal sheath WW domain-binding protein (PAWP). *Bosn J Basic Med Sci* 2018;**18**:178–184.
- Baldini D, Ferri D, Baldini GM, Lot D, Catino A, Vizziello D, Vizziello G. Sperm selection for ICSI: do we have a winner? *Cells* 2021;**10**:3566.
- Budak Ö, Bostanci MS, Osman K, Akdemir N, Gökçe A, Cevrioğlu S. Evaluation of the clinical results of using microfluidic channel system for sperm selection in IVF cycles in patients with low sperm concentration. *Konuralp Med J* 2023;**15**:9–15.
- Casanovas A, Ribas-Maynou J, Lara-Cerrillo S, Jimenez-Macedo AR, Hortal O, Benet J, Carrera J, García-Peiró A. Double-stranded sperm DNA damage is a cause of delay in embryo development and can impair implantation rates. *Fertil Steril* 2019;**111**:699–707.e691.
- Dai C, Zhang Z, Shan G, Chu L-T, Huang Z, Moskovtsev S, Librach C, Jarvi K, Sun Y. Advances in sperm analysis: techniques, discoveries and applications. *Nat Rev Urol* 2021;**18**:447–467.
- de Wagenaar B, Berendsen JTW, Bomer J, Olthuis W, van den Berg A, Segerink L. Microfluidic single sperm entrapment and analysis. *Lab Chip* 2015;**15**:1294–1301.
- Erimşah S, Gonca S, Kükner A, Duran B, Yazir Y. An ideal sperm selection method for the intrauterine insemination treatment of normozoospermic infertile patients. *Anatol Clin J Med Sci* 2020;**25**:19–26.
- Fréour T, Barragan M, Ferrer-Vaquer A, Rodríguez A, Vassena R. WBP2NL/PAWP mRNA and protein expression in sperm cells are not related to semen parameters, fertilization rate, or reproductive outcome. *J Assist Reprod Genet* 2017;**34**:803–810.
- Gode F, Gürbüz AS, Tamer B, Pala I, Isik AZ. The effects of microfluidic sperm sorting, density gradient and swim-up methods on semen oxidation reduction potential. *Urol J* 2020;**17**:397–401.
- Guler C, Melil S, Ozekici U, Donmez Cakil Y, Selam B, Cincik M. Sperm selection and embryo development: a comparison of the density gradient centrifugation and microfluidic chip sperm preparation methods in patients with astheno-teratozoospermia. *Life* 2021;**11**:933.
- Guzzi F, Candeloro P, Coluccio ML, Cristiani CM, Parrotta EI, Scaramuzzino L, Scalise S, Dattola E, D'Attimo MA, Cuda G et al. A disposable passive microfluidic device for cell culturing. *Biosensors (Basel)* 2020;**10**:18.
- Henkel R. Sperm preparation: state-of-the-art—physiological aspects and application of advanced sperm preparation methods. *Asian J Androl* 2012;**14**:260–269.
- Heydari A, Zabetian Targhi M, Halvaei I, Nosrati R. A novel microfluidic device with parallel channels for sperm separation using spermatozoa intrinsic behaviors. *Sci Rep* 2023;**13**:1185.
- Jeřeta M, Franzová K, Žáková J, Ventruba P, Crha I. Comparison of microfluidic and swim-up sperm separation methods for IVF. *Med J Cell Biol* 2020;**8**:170–175.
- Jones C, Meng X, Coward K. Sperm factors and egg activation: phospholipase C zeta (PLCZ1) and the clinical diagnosis of oocyte activation deficiency. *Reproduction* 2022;**164**:F53–F66.
- Kashir J. Increasing associations between defects in phospholipase C zeta and conditions of male infertility: not just ICSI failure? *J Assist Reprod Genet* 2020;**37**:1273–1293.
- Khakpour S, Sadeghi E, Tavalaei M, Bahadorani M, Nasr-Esfahani MH. Zeta method: a noninvasive method based on membrane charge for selecting spermatozoa expressing high level of phospholipaseC ζ . *Andrologia* 2019;**51**:e13249.
- Khodamoradi M, Rafizadeh Tafti S, Mousavi Shaeigh SA, Aflatoonian B, Azimzadeh M, Khashayar P. Recent microfluidic innovations for sperm sorting. *Chemosensors* 2021;**9**:126.
- Kruger TF, Acosta AA, Simmons KF, Swanson RJ, Matta JF, Oehninger S. Reprint of: predictive value of abnormal sperm morphology in in vitro fertilization. *Fertil Steril* 2019;**112**:e61–e66.
- Lara-Cerrillo S, Muñoz CU, de la Casa Heras M, Fernández-Pacheco SC, de la Santa JG, Lacruz-Ruiz T, Rosado-Iglesias C, Gonçalves-Aponte V, Liébana VB, García-Peiró A. Microfluidic sperm sorting improves ICSI outcomes in patients with increased values of double-strand breaks in sperm DNA. *Rev Int Androl* 2023;**21**:100338.
- Lara-Cerrillo S, Ribas-Maynou J, Rosado-Iglesias C, Lacruz-Ruiz T, Benet J, García-Peiró A. Sperm selection during ICSI treatments reduces single-but not double-strand DNA break values compared to the semen sample. *J Assist Reprod Genet* 2021;**38**:1187–1196.
- Leung ET, Lee C-L, Tian X, Lam KK, Li RH, Ng EH, Yeung WS, Chiu PC. Simulating nature in sperm selection for assisted reproduction. *Nat Rev Urol* 2022;**19**:16–36.
- Meintjes M. *Sperm Survival Assay for Quality Control in the IVF Laboratory*, Vol. 2017. Cambridge, UK: Cambridge University Press, 2017.
- Mirsanei JS, Sheibak N, Zandieh Z, Mehdizadeh M, Aflatoonian R, Tabatabaei M, Mousavi AS, Amjadi F. Microfluidic chips as a method for sperm selection improve fertilization rate in couples with fertilization failure. *Arch Gynecol Obstet* 2022;**306**:901–910.
- Nosrati R. *Microfluidics for Sperm Selection and Male Infertility Diagnostics*, Vol. 2016. Toronto, ON, Canada: University of Toronto, 2016.
- Nosrati R, Vollmer M, Eamer L, San Gabriel MC, Zeidan K, Zini A, Sinton D. Rapid selection of sperm with high DNA integrity. *Lab Chip* 2014;**14**:1142–1150.
- Ozcan P, Takmaz T, Yazici MGK, Alagoz OA, Yesiladali M, Sevketo, Ficioglu C. Does the use of microfluidic sperm sorting for the sperm selection improve in vitro fertilization success rates in male factor infertility? *J Obstet Gynaecol Res* 2021;**47**:382–388.
- Parrella A, Keating D, Cheung S, Xie P, Stewart JD, Rosenwaks Z, Palermo GD. A treatment approach for couples with disrupted sperm DNA integrity and recurrent ART failure. *J Assist Reprod Genet* 2019;**36**:2057–2066.
- Patel K, Patel V, Patel V. Case Study: to evaluate the clinical pregnancy rate of iOAT cases with microfluidic sperm sorter. *Int J Res Eng Sci Manag* 2020;**3**:547–550.

- Pinto S, Carrageta DF, Alves MG, Rocha A, Agarwal A, Barros A, Oliveira PF. Sperm selection strategies and their impact on assisted reproductive technology outcomes. *Andrologia* 2021; **53**:e13725.
- Pishbin E, Eghbal M, Navidbakhsh M, Zandi M. Localized air-mediated heating method for isothermal and rapid thermal processing on lab-on-a-disk platforms. *Sens Actuators B Chem* 2019; **294**:270–282.
- Pishbin E, Kazemzadeh A, Chimera M, Asiaei S, Navidbakhsh M, Russom A. Frequency dependent multiphase flows on centrifugal microfluidics. *Lab Chip* 2020; **20**:514–524.
- Pujol A, García-Peiró A, Ribas-Maynou J, Lafuente R, Mataró D, Vassena R. A microfluidic sperm-sorting device reduces the proportion of sperm with double-stranded DNA fragmentation. *Zygote* 2022; **30**:200–205.
- Quinn MM, Jalalian L, Ribeiro S, Ona K, Demirci U, Cedars MI, Rosen MP. Microfluidic sorting selects sperm for clinical use with reduced DNA damage compared to density gradient centrifugation with swim-up in split semen samples. *Hum Reprod* 2018; **33**:1388–1393.
- Rappa K, Samargia J, Sher M, Pino JS, Rodriguez HF, Asghar W. Quantitative analysis of sperm rheotaxis using a microfluidic device. *Microfluid Nanofluid* 2018; **22**:100.
- Setti AS, Braga DPdAF, Provenza RR, Iaconelli A, Borges E. Oocyte ability to repair sperm DNA fragmentation: the impact of maternal age on intracytoplasmic sperm injection outcomes. *Fertil Steril* 2021; **116**:123–129.
- Shamloo A, Abdorahimzadeh S, Nasiri R. Exploring contraction–expansion inertial microfluidic-based particle separation devices integrated with curved channels. *AIChE J* 2019; **65**:e16741.
- Shirota K, Yotsumoto F, Itoh H, Obama H, Hidaka N, Nakajima K, Miyamoto S. Separation efficiency of a microfluidic sperm sorter to minimize sperm DNA damage. *Fertil Steril* 2016; **105**:315–321.e311.
- Tavalaee M, Kiani-Esfahani A, Nasr-Esfahani MH. Relationship between phospholipase C-zeta, semen parameters, and chromatin status. *Syst Biol Reprod Med* 2017a; **63**:259–268.
- Tavalaee M, Kiani-Esfahani A, Nasr-Esfahani MH. Relationship between potential sperm factors involved in oocyte activation and sperm DNA fragmentation with intra-cytoplasmic sperm injection clinical outcomes. *Cell J* 2017b; **18**:588.
- Tavalaee M, Parivar K, Shahverdi A-H, Ghaedi K, Nasr-Esfahani MH. Status of sperm-born oocyte activating factors (PAWP, PLC ζ) and sperm chromatin in uncapacitated, capacitated and acrosome-reacted conditions. *Hum Fertil (Camb)* 2017c; **20**:96–103.
- Tharakan T, Bettocchi C, Carvalho J, Corona G, Jones TH, Kadioglu A, Salamanca JIM, Serefoglu EC, Verze P, Salonia A et al.; EAU Working Panel on Male Sexual Reproductive Health. European Association of Urology Guidelines Panel on male sexual and reproductive health: a clinical consultation guide on the indications for performing sperm DNA fragmentation testing in men with infertility and testicular sperm extraction in nonazoospermic men. *Eur Urol Focus* 2022; **8**:339–350.
- Vasilescu SA, Ding L, Parast FY, Nosrati R, Warkiani ME. Sperm quality metrics were improved by a biomimetic microfluidic selection platform compared to swim-up methods. *Microsyst Nanoeng* 2023; **9**:37.
- Vasilescu SA, Khorsandi S, Ding L, Bazaz SR, Nosrati R, Gook D, Warkiani ME. A microfluidic approach to rapid sperm recovery from heterogeneous cell suspensions. *Sci Rep* 2021; **11**:7917.
- Wang S, Tan W, Huang Y, Mao X, Li Z, Zhang X, Wei P, Xue L. Sperm DNA fragmentation measured by sperm chromatin dispersion impacts morphokinetic parameters, fertilization rate and blastocyst quality in ICSI treatments. *Zygote* 2022; **30**:72–79.
- World Health Organization. *Laboratory Manual for the Examination and Processing of Human Semen*, Vol. **2021**. Geneva, Switzerland: World Health Organization, 2021.
- Wyns C, Bergh C, Calhaz-Jorge C, De Geyter C, Kupka MS, Motrenko T, Rugescu I, Smeenk J, Tandler-Schneider A, Vidakovic S et al.; European IVF-monitoring Consortium (EIM)† for the European Society of Human Reproduction and Embryology (ESHRE). ART in Europe, 2016: results generated from European registries by ESHRE. *Human Reprod Open* 2020; **2020**:hoaa032.
- Xiao S, Riordon J, Simchi M, Lagunov A, Hannam T, Jarvi K, Nosrati R, Sinton D. FertDish: microfluidic sperm selection-in-a-dish for intracytoplasmic sperm injection. *Lab Chip* 2021; **21**:775–783.
- Yan Z, Fan Y, Wang F, Yan Z, Li M, Ouyang J, Wu L, Yin M, Zhao J, Kuang Y et al. Novel mutations in PLCZ1 cause male infertility due to fertilization failure or poor fertilization. *Hum Reprod* 2020; **35**:472–481.
- Yelumalai S, Yeste M, Jones C, Amdani SN, Kashir J, Mounce G, Da Silva SJM, Barratt CL, McVeigh E, Coward K. Total levels, localization patterns, and proportions of sperm exhibiting phospholipase C zeta are significantly correlated with fertilization rates after intracytoplasmic sperm injection. *Fertil Steril* 2015; **104**:561–568.e564.
- Yetkinel S, Kilicdag EB, Aytac PC, Haydardedeoglu B, Simsek E, Cok T. Effects of the microfluidic chip technique in sperm selection for intracytoplasmic sperm injection for unexplained infertility: a prospective, randomized controlled trial. *J Assist Reprod Genet* 2019; **36**:403–409.
- Yildiz K, Yuksel S. Use of microfluidic sperm extraction chips as an alternative method in patients with recurrent in vitro fertilisation failure. *J Assist Reprod Genet* 2019; **36**:1423–1429.
- Zaha I, Naghi P, Stefan L, Bunescu C, Radu M, Muresan ME, Sandor M, Sachelarie L, Huniadi A. Comparative study of sperm selection techniques for pregnancy rates in an unselected IVF-ICSI population. *J Pers Med* 2023; **13**:619.
- Zhang B, Yin T, Yang J. A novel microfluidic device for selecting human sperm to increase the proportion of morphologically normal, motile sperm with uncompromised DNA integrity. *Anal Methods* 2015; **7**:5981–5988.
- Zhang F, Li J, Liang Z, Wu J, Li L, Chen C, Jin F, Tian Y. Sperm DNA fragmentation and male fertility: a retrospective study of 5114 men attending a reproductive center. *J Assist Reprod Genet* 2021; **38**:1133–1141.

Early pregnancy

First trimester maternal tryptophan metabolism and embryonic and fetal growth: the Rotterdam Periconceptional Cohort (Predict Study)

Sofie K.M. van Zundert^{1,2}, Nina C.M. van Egmond^{1,2}, Lenie van Rossem¹, Sten P. Willemsen^{1,3}, Pieter H. Griffioen², Ron H.N. van Schaik², Mina Mirzaian², and Régine P.M. Steegers-Theunissen^{1,*} 

¹Department of Obstetrics and Gynecology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

²Department of Clinical Chemistry, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

³Department of Biostatistics, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

*Correspondence address. Department of Obstetrics and Gynecology, Erasmus MC, University Medical Center, P O Box 2040, Rotterdam, 3000 CA, The Netherlands. Tel: +31-6-12472643; E-mail: r.steegers@erasmusmc.nl  <https://orcid.org/0000-0002-4353-5756>

ABSTRACT

STUDY QUESTION: What is the association between first trimester maternal tryptophan (TRP) metabolites and embryonic and fetal growth?

SUMMARY ANSWER: Higher 5-hydroxytryptophan (5-HTP) concentrations are associated with reduced embryonic growth and fetal growth and with an increased risk of small-for-gestational age (SGA), while higher kynurenine (KYN) concentrations are associated with a reduced risk of SGA.

WHAT IS KNOWN ALREADY: The maternal TRP metabolism is involved in many critical processes for embryonic and fetal growth, including immune modulation and regulation of vascular tone. Disturbances in TRP metabolism are associated with adverse maternal and fetal outcomes.

STUDY DESIGN, SIZE, DURATION: This study was embedded within the Rotterdam Periconceptional Cohort (Predict Study), an on-going prospective observational cohort conducted at a tertiary hospital from November 2010 onwards.

PARTICIPANTS/MATERIALS, SETTING, METHODS: A total of 1115 women were included before 11 weeks of gestation between November 2010 and December 2020. Maternal serum samples were collected between 7 and 11 weeks of gestation, and TRP metabolites (TRP, KYN, 5-HTP, 5-hydroxytryptamine, and 5-hydroxyindoleacetic acid) were determined using a validated liquid chromatography (tandem) mass spectrometry method. Serial 3D ultrasound scans were performed at 7, 9, and 11 weeks of gestation to accurately assess features of embryonic growth, including crown-rump length (CRL) and embryonic volume (EV) offline using virtual reality systems. Fetal growth parameters were retrieved from medical records and standardized according to Dutch reference curves. Mixed models were used to assess associations between maternal TRP metabolites and CRL and EV trajectories. Linear and logistic regression models were utilized to investigate associations with estimated fetal weight (EFW) and birthweight, and with SGA, respectively. All analyses were adjusted for potential confounders.

MAIN RESULTS AND THE ROLE OF CHANCE: Maternal 5-HTP concentrations and the maternal 5-HTP/TRP ratio were inversely associated with embryonic growth (5-HTP, β CRL = -0.015 , 95% CI = -0.028 to -0.001 ; 5-HTP β EV = -0.009 , 95% CI = -0.016 to -0.003). An increased maternal 5-HTP/TRP ratio was also associated with lower EFW and birthweight, and with an increased risk of SGA (odds ratio (OR) = 1.006 , 95% CI = 1.00 – 1.013). In contrast, higher maternal KYN concentrations were associated with a reduced risk of SGA in the unadjusted models (OR = 0.548 , 95% CI = 0.320 – 0.921).

LIMITATIONS, REASONS FOR CAUTION: Residual confounding cannot be ruled out because of the observational design of this study. Moreover, this study was conducted in a single tertiary hospital, which assures high internal validity but may limit external validity.

WIDER IMPLICATIONS OF THE FINDINGS: The novel finding that maternal 5-HTP concentrations are associated with a smaller embryo and fetus implies that disturbances of the maternal serotonin pathway in the first trimester of pregnancy are potentially involved in the pathophysiology of fetal growth restriction. The association between higher maternal KYN concentrations and a reduced risk of SGA substantiate the evidence that the KYN pathway has an important role in fetal growth. More research is needed to delve deeper into the potential role of the maternal TRP metabolism during the periconception period and pregnancy outcome for mother and offspring.

STUDY FUNDING/COMPETING INTEREST(S): This study was funded by the Department of Obstetrics and Gynecology and the Department of Clinical Chemistry of the Erasmus MC, University Medical Center, Rotterdam, the Netherlands. The authors have no competing interests to disclose.

TRIAL REGISTRATION NUMBER: N/A.

Received: August 25, 2023. Revised: February 17, 2024. Editorial decision: February 22, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: tryptophan metabolism / 5-hydroxytryptophan / kynurenine / periconception period / embryonic growth

Introduction

Embryonic and fetal growth are important determinants for birth outcomes and health later in life. Especially during the periconception period, when critical processes occur and growth rates are the highest, maternal exposures can have profound effects on growth of the embryo (Dickey and Gasser, 1993; Gluckman *et al.*, 2010; Steegers-Theunissen *et al.*, 2013). Indeed, maternal exposures, such as a poor diet quality and smoking, during the periconception period have been found to reduce embryonic growth (van Uitert *et al.*, 2013b; Abdollahi *et al.*, 2021; Pietersma *et al.*, 2022). Previous studies have revealed that smaller embryos are associated with reduced fetal growth, lower birthweight, and cardiovascular risk factors in the future child (Mook-Kanamori *et al.*, 2010; Jaddoe *et al.*, 2014; Ustunyurt *et al.*, 2015; Xu *et al.*, 2022).

The exact biological mechanisms underlying impaired embryonic growth and the subsequent adverse outcomes remain, however, largely unclear. It has been demonstrated that one-carbon metabolism is of key importance in the periconception period because of its role in the biosynthesis of DNA and proteins, and epigenetic modifications (Steegers-Theunissen *et al.*, 2013; Kalhan, 2016). Derangements in one-carbon metabolism as a result of poor maternal exposures have been associated with impaired embryonic and fetal growth, with implications for future health (Steegers-Theunissen *et al.*, 2013). Another potential biological mechanism that contributes to fetal growth is the tryptophan (TRP) metabolism. TRP metabolism is linked to one-carbon metabolism as it provides methyl donors and shares substrates and cofactors (Steegers-Theunissen *et al.*, 2013). TRP is an essential amino acid, and its free concentration is dependent on protein

and energy intake, and other exposures (Wurtman *et al.*, 2003; Badawy, 2017b). Physiologically, 95% of free TRP is metabolized through the kynurenine (KYN) pathway, with 1–5% through the serotonin pathway, and the rest of TRP is reserved for the indole pathway and protein synthesis (Fig. 1) (Bender, 1983; Oxenkrug, 2010). TRP metabolites are involved in processes essential to maintain pregnancy, including immune regulation, anti-oxidative processes, and regulation of vascular tone (Badawy, 2015; Broekhuizen *et al.*, 2021).

Throughout pregnancy, the TRP metabolite concentrations are closely regulated, and alterations of KYN pathway or serotonin pathway metabolites have been associated with pregnancy complications, such as pre-eclampsia and fetal growth restriction (FGR) (Gumusoglu *et al.*, 2021; van Zundert *et al.*, 2022a). In late pregnancy, maternal TRP and KYN concentrations were inversely associated with FGR (van Zundert *et al.*, 2022a; Yaman *et al.*, 2023). Studies investigating TRP metabolism in the placenta showed that in placentas of FGR pregnancies the KYN pathway was downregulated, while the serotonin pathway was upregulated (Ranzil *et al.*, 2019). Animal studies also suggest that increased maternal serotonin pathway metabolite concentrations have a negative effect on fetal growth, illustrated by reduced fetal weight after administration of 5-hydroxytryptophan (5-HTP) and 5-hydroxytryptamine (5-HT) (Salas *et al.*, 2007).

To date, however, human studies investigating the role of the TRP metabolism in fetal growth are scarce and focus on late pregnancy or on placental TRP metabolism (van Zundert *et al.*, 2022a). To expand our understanding of the biological mechanisms involved in (patho)physiological embryonic and fetal growth, we studied the association between first trimester

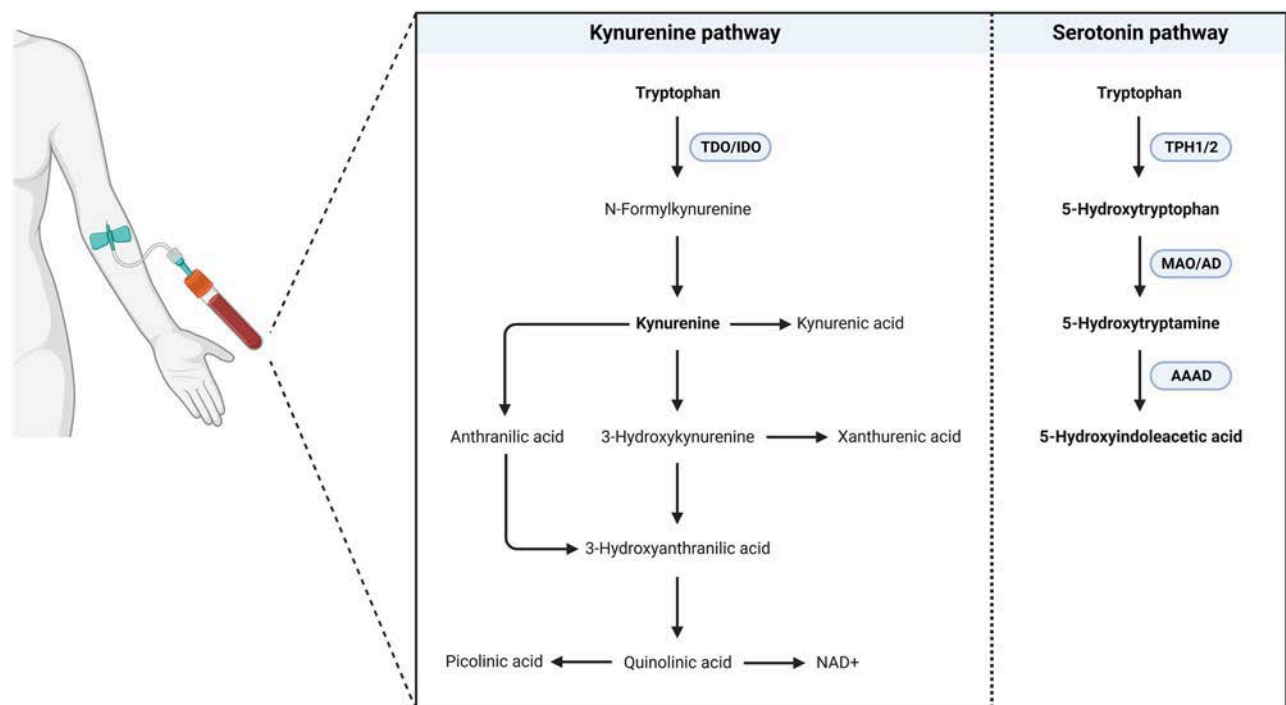


Figure 1. Tryptophan metabolism along the kynurenine and serotonin pathway. The analytes in bold were determined in maternal serum in the present study. AAAD: aromatic amino acid decarboxylase; AD: aldehyde dehydrogenase; IDO: indoleamine-2,3-dioxygenase; MAO: monoamine oxidase; TDO: tryptophan-2,3-dioxygenase; TPH: tryptophan hydroxylase.

maternal TRP metabolites and embryonic growth. The secondary aim of this study was to investigate the associations between first trimester maternal TRP metabolites and fetal growth and birth-weight, and the risk of small-for-gestational age (SGA) at birth.

Materials and methods

Study design and setting

This study was embedded within the Rotterdam Periconceptional Cohort (Predict Study), an ongoing prospective tertiary hospital-based cohort study with the aim to explore determinants of periconception parental health in relation to reproductive and pregnancy course and outcomes, and potential underlying mechanisms. The cohort profile has been described in detail previously (Stegers-Theunissen et al., 2016; Rousian et al., 2021). In short, women were eligible if they were older than 18 years, familiar with the Dutch language, and <11 weeks pregnant.

Study population

In total, 2051 pregnancies were included in the Predict Study between November 2010 and December 2020 (Fig. 2). We excluded pregnancies conceived after oocyte donation ($n = 23$), since no information was collected of the donors. In addition, pregnancies with an increased *a priori* risk of reduced embryonic growth were excluded: multiple pregnancies ($n = 63$), miscarriages ($n = 145$), congenital anomalies ($n = 83$), and fetal and neonatal deaths ($n = 24$). Then, pregnancies with missing first trimester 3D ultrasound imaging ($n = 245$) or missing first trimester maternal serum samples ($n = 35$) were excluded.

Since longitudinal crown-rump length (CRL) was one of the features of embryonic growth assessed in this study, and gestational age is the most important determinant of embryonic growth, gestational age was not based on the CRL. The gestational age was calculated using the first day of the last menstrual period (LMP) when women conceived naturally and had a regular menstrual cycle (25–31 days). The gestational age was adjusted for the duration of the menstrual cycle if the menstrual cycle deviated by 3–7 days from 28 days, with a fixed number of days each cycle.

The gestational age was determined using the CRL at 9 weeks of gestation; however, when women conceived naturally but had an irregular menstrual cycle (<21 days or >35 days), the LMP was missing or when the difference between the gestational age based on the LMP and the gestational age based on the CRL at 9 weeks of gestation was ≥ 7 days. For pregnancies conceived after IVF and ICSI, the gestational age was calculated from the conception date: the oocyte retrieval day minus 14 days for the fresh embryo transfers or minus 19 days for the cryopreserved embryo transfers (Stegers-Theunissen, et al., 2016).

Pregnancies of which the gestational age was based on the CRL were excluded from the analysis with embryonic growth features as outcome ($n = 318$). As a result, the primary analysis was performed on a study sample consisting of 1115 pregnancies, and the secondary analysis was performed on a study sample of 1433 pregnancies.

Data collection

TRP metabolites

Non-fasting venous blood was drawn during the hospital intake at 8.4 (interquartile range (IQR) = 7.3–9.4) weeks of gestation, and TRP metabolites were determined in the serum using a validated liquid chromatography (tandem) mass spectrometry (LC-MS/MS) method (van Zundert et al., 2022b). The following TRP metabolites were determined: TRP, KYN, 5-HTP, 5-HT, and 5-hydroxyindole acetic acid (5-HIAA). The KYN/TRP and 5-HTP/TRP ratios were determined as

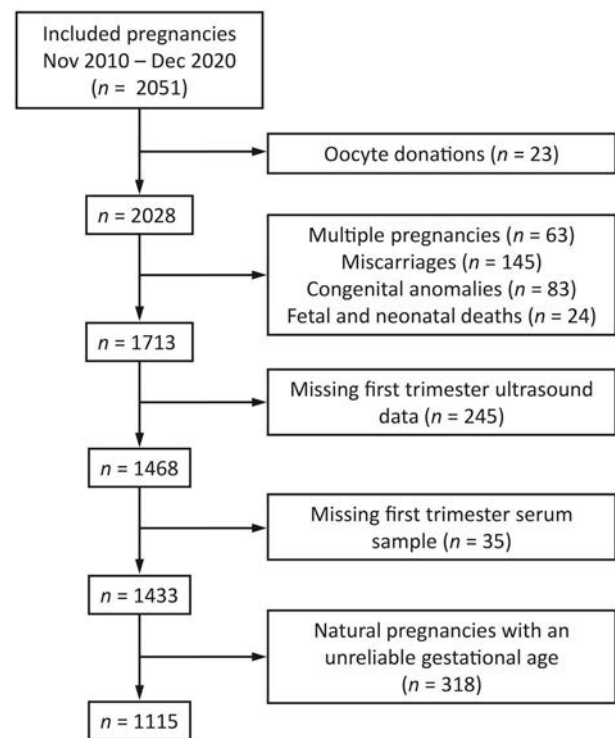


Figure 2. Flowchart of the study population in an investigation of first trimester maternal tryptophan metabolism and embryonic and fetal growth.

they may provide an indication of the KYN pathway and serotonin pathway, respectively (Badawy and Guillemin, 2019).

Embryonic growth features

Embryonic growth features included CRL and embryonic volume (EV), which were measured using 3D ultrasound imaging and virtual reality techniques (Stegers-Theunissen et al., 2016; Rousian et al., 2021). These measurements have been proven reliable and reproducible (Rousian et al., 2010). Transvaginal 3D ultrasounds were performed at ~7, 9, and 11 weeks of gestation using a Voluson E8 or E10 (GE Healthcare, Austria) ultrasound machine. Accurate and precise 3D CRL and EV measurements were performed offline using 4D view software (GE Healthcare, Austria) and the V-Scope volume rendering application. A hologram of the embryo was created, which allowed for interaction and real depth perception. Manually, the signal of the uterus and umbilical cord was removed based on differences in echogenicity between the structures. The CRL was measured manually three times using a tracing tool, and the EV was measured using a semi-automated method (Stegers-Theunissen et al., 2016; Rousian et al., 2021).

Fetal growth parameters and birth outcomes

Fetal growth parameters were retrieved from the routine trans-abdominal mid-pregnancy anomaly scan at 20.1 (IQR = 19.9–20.6) weeks of gestation. Estimated fetal weight (EFW) was calculated using the Hadlock formula (Hadlock et al., 1985):

$$EFW = 10^{(1.3596 - 0.00386 \times AC \times FL + 0.0064 \times HC + 0.00061 \times BPD \times AC + 0.0424 \times AC + 0.174 \times FL)}$$

EFW: estimated fetal weight; AC: abdominal circumference; FL: femur length; HC: head circumference; BPD: biparietal diameter.

Birthweight was retrieved from delivery reports. For both EFW and birthweight, standard scores (percentiles and z-scores) were

calculated based on Dutch reference curves, adjusted for gestational age, and for gestational age and fetal sex, respectively (Gaillard *et al.*, 2011; Hoftiezer *et al.*, 2016). Percentiles were used to identify SGA cases, defined as a birthweight below the 10th percentile (Hoftiezer *et al.*, 2016).

Covariates

Data on maternal characteristics were collected via a general questionnaire and a validated food frequency questionnaire (FFQ), which were both filled out before the hospital intake during the first trimester of pregnancy (Feunekes *et al.*, 1993; Fayet *et al.*, 2011). The general questionnaire collected data on date of birth, geographical background, educational level, smoking, and folic acid supplement use. Based on the date of birth and the conception date, the age at conception was calculated. Geographical background was classified as either Western or non-Western, following the categorization proposed by Alders (2001). Educational levels were categorized into three groups based on the International Standard Classification of Education (ISCED): low (ISCED 0–2), intermediate (ISCED 3–4), and high (ISCED 5–8) (United Nations Educational Scientific and Cultural Organization, 2012). Smoking included any smoking during the periconception period. Folic acid supplement use was considered adequate in this study when initiated before conception. Information on maternal protein and energy intake were obtained from the FFQ taking into account the reliability using the Goldberg cut-off, which compared an individual's reported energy intake to their basal metabolic rate and physical activity (Black, 2000). A comprehensive description of this method can be found in the recent publication of Smit *et al.* (2022). The completeness of the questionnaire was checked by a trained research nurse upon hospital intake. Additionally, anthropometric measurements were performed using a standard protocol and blood was drawn. BMI was calculated by dividing weight (kg) by the square of height (m). Data on conception mode, parity, and fetal sex were retrieved from medical records. Conception mode referred to the method through which pregnancy was achieved, encompassing naturally conceived pregnancies and IVF/ICSI pregnancies.

Statistical analysis

Baseline characteristics, including periconception maternal characteristics, fetal and birth outcomes, and TRP metabolite concentrations, were presented as means with SD or as numbers of individuals. However, for 5-HIAA, owing to its right-skewed distribution, the median and IQR were used instead.

5-HIAA was (natural) log-transformed to obtain an approximate normal distribution for the analyses. In addition, a square root transformation for CRL and a cube root transformation for EV were adopted to obtain a constant variance and an approximate normal distribution of residuals provided the included covariates.

Multivariable mixed models were constructed to investigate the associations between maternal TRP metabolites and embryonic growth trajectories. Mixed models are flexible and can handle unbalanced data and meanwhile account for the correlated (repeated) measurements within each pregnancy by including a random intercept (Laird and Ware, 1982). Model 1 (crude model) was only adjusted for a cubic spline function of gestational age at the moment of the ultrasound. To identify potential confounders, a directed acyclic graph based on existing literature and a correlation matrix were constructed. Covariates additionally included in Model 2 (adjusted model) were gestational age at the blood draw, maternal age, geographical background, educational level, smoking, folic acid supplement use, protein intake/energy intake, BMI, conception mode, parity, and fetal sex. Including the

covariates as nonlinear terms and including all possible two-way interactions did not improve the fit based on the Akaike information criterion and Bayesian information criterion.

Multivariable linear regression models were used to assess the associations between TRP metabolites and EFW and birthweight. Since the EFW z-scores were already adjusted for gestational age and SGA for gestational age and fetal sex, these covariates were not included in the multivariable model. Model 2 included the covariates gestational age at the blood draw, maternal age, geographical background, educational level, smoking, folic acid supplement use, protein intake/energy intake, BMI, conception mode, parity, and fetal sex (only for EFW). Finally, multivariable logistic regression models, using the same covariates as in the linear regression model for birthweight, were built to investigate the associations between TRP metabolites and SGA.

Results were displayed as effect estimates (β or odds ratio (OR)) with 95% CI. All statistical analyses were performed using R version 4.2.1 and IBM SPSS Statistics for Windows version 28 (IBM Corp, 2021; R Core Team, 2022).

Ethical approval

Ethical clearance was sought and obtained from the local Medical Ethical Committee of the Erasmus MC, University Medical Center, Rotterdam and the Central Committee on Research The Hague (15 October 2004, MEC-2004-277; Steegers-Theunissen, *et al.*, 2016). All participants gave written informed consent at enrolment.

Results

Baseline characteristics

Table 1 shows the baseline characteristics of the total study population ($n = 1115$), including information on the periconception maternal characteristics, fetal and birth outcomes, and absolute maternal TRP concentrations. In Supplementary Table S1, the baseline characteristics of the study sample used for the secondary analysis ($n = 1433$) are presented, which were comparable with those described below. Supplementary Table S2 shows the baseline characteristics of the total excluded population and of those excluded based on missing first trimester 3D ultrasound data or missing first trimester serum samples. Those who were excluded from the analysis tended to have higher BMI and lower educational level, and were more likely to smoke, conceive naturally, be multiparous, and have an inadequate folic acid supplement intake.

Periconception maternal characteristics

The mean maternal age at conception was 32.5 (SD = 4.5) years, and the mean BMI was 25.5 (SD = 4.8) kg/m² (Table 1). Women mostly had a Western geographical background ($n = 918$ [86.3%]), were highly educated ($n = 626$ [58.8%]), and often nulliparous ($n = 597$ [54.3%]). Slightly less than half of the women conceived after IVF/ICSI treatment (49.1%) and the others conceived naturally (50.9%). During the periconception period, one-third of the women consumed alcohol ($n = 308$ [29.0%]), 147 (13.8%) women smoked, and a few women ($n = 15$ [1.4%]) used drugs. Most women initiated folic acid supplement use before conception ($n = 896$ [84.4%]), which was considered adequate. Women consumed daily 72.6 g protein (SD = 19.6) and 8215.0 kJ (SD = 2223.2) energy, both of which are in accordance with current recommendations in obstetric practice (Most *et al.*, 2019; Murphy *et al.*, 2021).

Table 1. Periconception maternal characteristics, fetal and birth outcomes, and tryptophan metabolite concentrations of the total study population.

Baseline characteristics	Total study population (n = 1115)
Periconception maternal characteristics	
Age at conception (years)	32.5 (4.5)
Missing	0
BMI (kg/m ²)	25.5 (4.8)
Missing	30
Geographical background	
Western	918 (86.3)
Non-Western	146 (13.7)
Missing	51
Educational level	
Low	73 (6.9)
Middle	366 (34.4)
High	626 (58.8)
Missing	50
Parity	
Nulliparous	597 (54.3)
Multiparous	503 (45.7)
Missing	15
Conception mode	
Natural	567 (50.9)
IVF/ICSI	548 (49.1)
Missing	0
Any smoking	
Yes	147 (13.8)
No	917 (86.2)
Missing	51
Any alcohol use	
Yes	308 (29.0)
No	755 (71.0)
Missing	52
Any drug use	
Yes	15 (1.4)
No	1049 (98.6)
Missing	51
Folic acid supplement use ^a	
Adequate	896 (84.4)
Inadequate	166 (15.6)
Missing	53
Energy intake (kJ/day)	8215.0 (2223.2)
Missing	108
Unreliable	216
Protein intake/energy intake (grams/day)	72.6 (19.6)
Missing	108
Unreliable	216
Fetal and birth outcomes	
Fetal sex	
Girl	530 (50.6)
Boy	518 (49.4)
Missing	67
EFW mid-pregnancy (g)	352.6 (46.9)
Missing	102
EFW mid-pregnancy (z-score)	0.5 (1.0)
Missing	105
Gestational age at delivery (weeks)	38.8 (2.0)
Missing	69
Birthweight (grams)	3280.5 (580.4)
Missing	75
Birthweight (z-score)	-0.10 (1.1)
Missing	80
SGA (birthweight <p10)	
Yes	129 (11.6)
No	906 (81.3)
Missing	80
Tryptophan metabolites	
TRP (μmol/l)	56.3 (9.5)
	28.8–101.5

(continued)

Table 1. Continued

Baseline characteristics	Total study population (n = 1115)
KYN (μmol/l)	1.5 (0.3)
	0.7–3.5
5-HTP (nmol/l)	4.5 (1.7)
	1.1–28.9
5-HT (nmol/l)	676.0 (284.9)
	16.1–2180.2
5-HIAA (nmol/l)	38.8 (31.8–51.3)
	15.9–508.2
Natural log of 5-HIAA (nmol/l)	3.8 (0.5)
	2.8–6.2

^a Folic acid supplement use was considered adequate when initiated before conception. Continuous data are presented as mean with SD or median with interquartile range (5-HIAA), and categorical data as numbers with percentages. For the tryptophan metabolites also the range (minimum–maximum) is given.

5-HIAA: 5-hydroxyindoleacetic acid; 5-HT: 5-hydroxytryptamine; 5-HTP: 5-hydroxytryptophan; AC: abdominal circumference; EFW: estimated fetal weight; KYN: kynurenine; TRP: tryptophan.

Fetal and birth outcomes

Mid-pregnancy, at 20.1 (IQR = 19.9–20.6) weeks of gestation, the EFW in our study population was slightly higher than in the general Dutch population (z-score = 0.5 (SD = 1.0)), while the birthweight was slightly lower than in the general Dutch population (z-score = -0.10 (SD = 1.1)). Women delivered at a mean gestational age of 38.8 (SD = 2.0) weeks, and in 11.6% (n = 129), this concerned a SGA-baby.

Embryonic growth

The associations between maternal TRP metabolite concentrations and CRL and EV trajectories are presented in Table 2. Results described below are based on the adjusted model, unless stated otherwise. An inverse association was found between maternal 5-HTP concentrations and both CRL and EV trajectories ($\sqrt{\text{CRL}}$: $\beta = -0.015$, 95% CI = -0.028 to -0.001; $\sqrt{\text{EV}}$: $\beta = -0.009$, 95% CI = -0.016 to -0.003). The median decrease in CRL and EV during the first trimester of pregnancy for +2SD 5-HTP concentrations compared to mean 5-HTP concentrations is illustrated in Fig. 3. In addition, the maternal 5-HTP/TRP ratio was inversely associated with both CRL and EV trajectories ($\sqrt{\text{CRL}}$: $\beta = -0.843$, 95% CI = -1.597 to -0.088; $\sqrt{\text{EV}}$: $\beta = -0.529$, 95% CI = -0.902 to -0.157). No associations were found between other TRP metabolites and CRL or EV trajectories.

Fetal growth and birth outcomes

Tables 3 and 4 show the associations between maternal TRP metabolite concentrations and EFW and SGA, respectively. In Supplementary Table S3, the associations between maternal TRP metabolite concentrations and birthweight are presented. Inverse associations were found between maternal 5-HTP concentrations and both EFW and birthweight, though not statistically significant. Comparable results were found for the maternal 5-HTP/TRP ratio. In addition, a higher maternal 5-HTP/TRP ratio was associated with an increased risk of SGA (OR = 1.006, 95% CI = 1.000–1.013). In contrast, increased maternal KYN concentrations were associated with higher birthweight and a reduced risk of SGA in the unadjusted models (birthweight z-score: $\beta = 0.188$, 95% CI = 0.010–0.365, SGA: OR = 1.006, 95% CI = 1.000–1.013). The analysis revealed no other associations between TRP metabolites and fetal growth or birth outcomes.

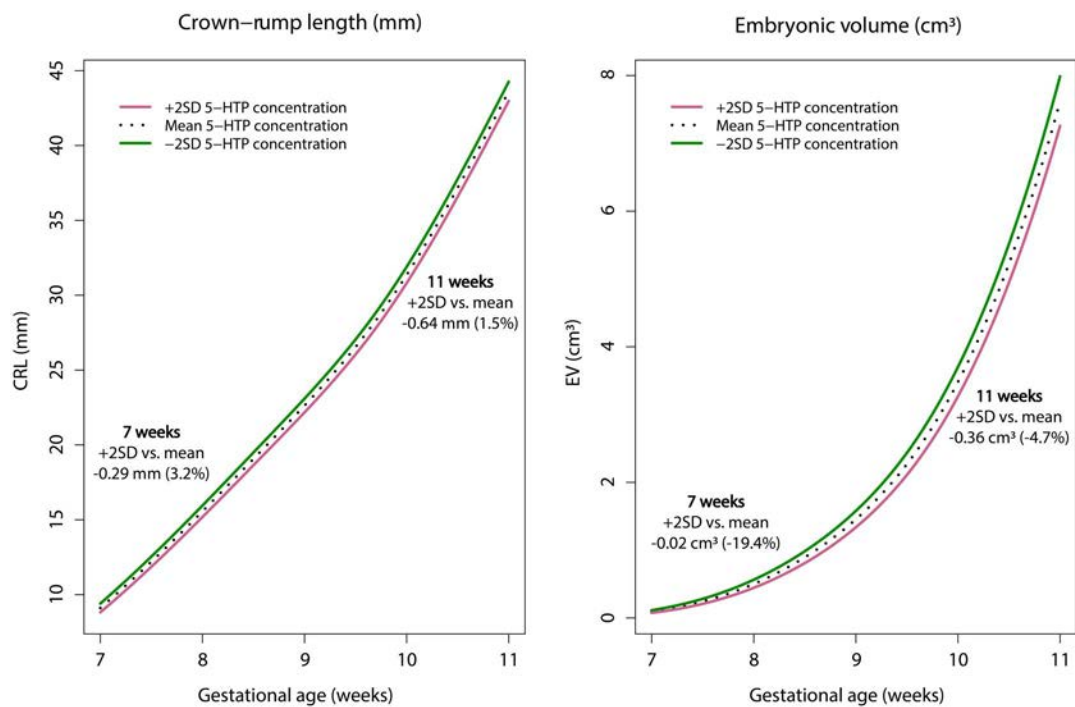
Table 2. Associations between first trimester maternal tryptophan metabolites and embryonic growth (crown–rump length and embryonic volume trajectories) during the first trimester of pregnancy in the total study population.

Study sample primary analysis (n = 1115)		CRL (mm)			³ EV (cm ³)		
		β	95% CI		β	95% CI	
TRP ($\mu\text{mol/l}$)	Model 1	0.000	−0.002	0.002	0.000	−0.001	0.001
	Model 2	0.000	−0.002	0.002	−0.000	−0.001	0.001
KYN ($\mu\text{mol/l}$)	Model 1	−0.008	−0.062	0.046	−0.004	−0.031	0.023
	Model 2	−0.007	−0.078	0.063	0.002	−0.033	0.038
5-HTP (nmol/l)	Model 1	0.001	−0.009	0.011	−0.002	−0.007	0.003
	Model 2	−0.015*	−0.028	−0.001	−0.009*	−0.016	−0.003
5-HT (nmol/l)	Model 1	0.000	0.000	0.000	0.000	0.000	0.000
	Model 2	−0.000	−0.000	0.000	−0.000	−0.000	0.000
Ln 5-HIAA (nmol/l)	Model 1	0.019	−0.016	0.054	0.017	−0.001	0.035
	Model 2	0.001	−0.043	0.045	0.007	−0.015	0.029
KYN/TRP ratio	Model 1	−0.257	−2.955	2.440	−0.030	−1.396	1.337
	Model 2	−0.264	−3.783	3.255	0.325	−1.425	2.075
5-HTP/TRP ratio · 10 ³	Model 1	−0.035	−0.606	0.537	−0.148	−0.436	0.140
	Model 2	−0.843*	−1.597	−0.088	−0.529*	−0.902	−0.157

Mixed models are used. Model 1 is adjusted for gestational age at the ultrasound. Model 2 is additionally adjusted for gestational age at the blood draw, maternal age, geographical origin, educational level, smoking, folic acid supplement use, protein intake/energy intake, BMI, conception mode, parity, and fetal sex. A square root transformation for CRL and a cube root transformation for EV were adopted to obtain a constant variance and an approximate normal distribution of residuals provided the included covariates.

* P-value ≤ 0.05 (statistically significant).

5-HIAA: 5-hydroxyindoleacetic acid; 5-HT: 5-hydroxytryptamine; 5-HTP/TRP ratio: 5-hydroxytryptophan/tryptophan ratio; 5-HTP: 5-hydroxytryptophan; CRL: crown–rump length; EV: embryonic volume; KYN/TRP ratio: kynurenine/tryptophan ratio; KYN: kynurenine; TRP: tryptophan.

**Figure 3.** Median crown–rump length and embryonic volume trajectories during the first trimester of pregnancy stratified for 5-hydroxytryptophan concentration. The trajectories were adjusted for gestational age, maternal age, geographical background, educational level, smoking, folic acid supplement use, protein intake/energy intake, BMI, conception mode, parity, and fetal sex. 5-HTP: 5-hydroxytryptophan; CRL: crown–rump length; EV: embryonic volume.

Discussion

Summary of findings

Table 5 summarizes the associations found in this study between first trimester maternal TRP metabolites and embryonic growth, fetal growth, and birth outcomes. It shows that increased first trimester maternal 5-HTP concentrations are associated with a smaller embryo, expressed by reduced CRL and EV trajectories. In addition, our findings suggest that higher first trimester maternal 5-HTP concentrations are also associated with lower EFV

and birthweight, and with an increased risk of SGA, while higher maternal KYN concentrations are associated with a higher birthweight, and a reduced risk of SGA.

Comparison with existing literature and interpretation

Studies investigating maternal TRP metabolites in relation to fetal growth or birth outcomes are scarce (van Zundert *et al.*, 2022a). In fact, this is the first human study that found an inverse

Table 3. Associations between first trimester maternal tryptophan metabolites and standardized estimated fetal weight (z-scores) in the total study population.

Study sample secondary analysis (n = 1433)		EFW (z-score)		
		β	95% CI	
TRP ($\mu\text{mol/l}$)	Model 1	0.004	-0.002	0.010
	Model 2	0.001	-0.006	0.008
KYN ($\mu\text{mol/l}$)	Model 1	0.045	-0.131	0.221
	Model 2	-0.057	-0.275	0.160
5-HTP (nmol/l)	Model 1	0.023	-0.011	0.057
	Model 2	-0.030	-0.073	0.012
5-HT (nmol/l)	Model 1	0.000	-0.001	0.000
	Model 2	0.000	-0.000	0.000
Ln 5-HIAA (nmol/l)	Model 1	0.063	-0.056	0.182
	Model 2	-0.067	-0.206	0.072
KYN/TRP ratio $\cdot 10^3$	Model 1	-0.002	-0.011	0.006
	Model 2	-0.004	-0.014	0.007
5-HTP/TRP ratio $\cdot 10^6$	Model 1	0.000	-0.002	0.002
	Model 2	-0.002	-0.004	0.000

Linear regression models are used. Model 1 is the crude model. Model 2 is adjusted for gestational age at the blood draw, maternal age, geographical origin, educational level, smoking, folic acid supplement use, protein intake/energy intake, BMI, conception mode, parity, and fetal sex.

* P-value ≤ 0.05 (statistically significant).

5-HIAA: 5-hydroxyindoleacetic acid; 5-HT: 5-hydroxytryptamine; 5-HTTP: 5-hydroxytryptophan; EFW: standardized estimated fetal weight; KYN: kynurenine; TRP: tryptophan.

Table 4. Associations between first trimester maternal tryptophan metabolites and the risk of small-for-gestational age at birth in the total study population.

Study sample secondary analysis (n = 1433)		Small-for-gestational age		
		Odds ratio	95% CI	
TRP ($\mu\text{mol/l}$)	Model 1	0.994	0.978	1.011
	Model 2	0.995	0.973	1.016
KYN ($\mu\text{mol/l}$)	Model 1	0.548*	0.320	0.921
	Model 2	0.679	0.330	1.338
5-HTP (nmol/l)	Model 1	1.035	0.943	1.125
	Model 2	1.101	0.978	1.240
5-HT (nmol/l)	Model 1	1.000	1.000	1.001
	Model 2	1.000	0.999	1.001
Ln 5-HIAA (nmol/l)	Model 1	0.912	0.645	1.259
	Model 2	1.000	0.646	1.498
KYN/TRP ratio $\cdot 10^3$	Model 1	0.987	0.962	1.012
	Model 2	1.001	0.967	1.035
5-HTP/TRP ratio $\cdot 10^6$	Model 1	1.003	0.998	1.008
	Model 2	1.006*	1.000	1.013

Small-for-gestational age at birth was defined as a birthweight below the 10th percentile. Logistic regression models are used. Model 1 is the crude model. Model 2 is adjusted for gestational age at the blood draw, maternal age, geographical origin, educational level, smoking, folic acid supplement use, protein intake/energy intake, BMI, conception mode, and parity.

* P-value ≤ 0.05 (statistically significant).

5-HIAA, 5-hydroxyindoleacetic acid; 5-HT: 5-hydroxytryptamine; 5-HTTP: 5-hydroxytryptophan; KYN: kynurenine; TRP: tryptophan.

association between first trimester maternal 5-HTP concentrations and embryonic and fetal growth. The effect estimates of the association between first trimester maternal 5-HTP and embryonic growth ($\sqrt{\text{CRL}}$: $\beta = -0.015$, 95% CI = -0.028 to -0.001; $\sqrt{\text{EV}}$: $\beta = -0.009$, 95% CI = -0.016 to -0.003) were smaller than those of the association between periconception smoking and embryonic growth ($\sqrt{\text{CRL}}$: $\beta = -0.055$, 95% CI = -0.159 to -0.006; $\sqrt{\text{EV}}$: $\beta = -0.042$, 95% CI = -0.089 to 0.006), but larger than those of the association between periconception alcohol use and embryonic growth ($\sqrt{\text{CRL}}$: $\beta = -0.003$, 95% CI = -0.005 to -0.003; $\sqrt{\text{EV}}$: $\beta =$

Table 5. Summary of associations between first trimester maternal tryptophan metabolites and embryonic growth, fetal growth, and birth outcomes in the total study population.

	CRL	EV	EFW	BW	SGA-risk
TRP	●	●	●	●	●
KYN	●	●	●	↑	↓
5-HTP	↓*	↓*	↓	↓	↑
5-HT	●	●	●	●	●
Ln 5-HIAA	●	●	●	●	●
KYN/TRP	●	●	●	●	●
5-HTP/TRP	↓*	↓*	↓	↓	↑*

↑: positive association; ↓: negative association; ●: no association.

* P-value ≤ 0.05 (statistically significant in the adjusted model).

5-HIAA: 5-hydroxyindoleacetic acid; 5-HT: 5-hydroxytryptamine; 5-HTP: 5-hydroxytryptophan; BW: birthweight; CRL: crown-rump length; EFW: estimated fetal weight; EV: embryonic volume; KYN: kynurenine; SGA: small-for-gestational age; TRP: tryptophan.

-0.005, 95% CI = -0.040 to 0.030) (van Uiter et al., 2013a; van Dijk et al., 2018).

There are several possible explanations for this finding. Insights from animal studies suggest that maternal 5-HTP can directly inhibit fetal growth by passing into the fetal circulation. These animal studies showed that 5-HTP can induce developmental arrest, inhibit growth, and increase cell death (Gordeeva and Gordeev, 2021; Gordeeva and Safandeev, 2021). In humans, there is increasing evidence that initially nutrition of the embryo is histiotrophic, with secretion of nutrients by uterine glands into the intervillous space, which are taken up by the trophoblast (Burton et al., 2001). The nutrients diffuse into the coelomic fluid from which they may be absorbed by the (secondary) yolk sac, and then reach the embryo. There is a transition toward hemotrophic nutrition when the maternal placental circulation is established (Burton et al., 2001). Additional research is needed to determine whether maternal 5-HTP indeed reaches the embryo through histiotrophic nutrition.

Another explanation for the inverse association between maternal 5-HTP concentrations and fetal growth may be the involvement of 5-HTP in placental development. Support for this biological mechanism can be found in an animal study that reported reduced fetal and placental weights after treating pregnant rats with 5-HTP (Salas et al., 2007). This study postulated that treatment with 5-HTP increases placental vascular resistance. Consequently, reduced utero-placental blood flow can potentially affect fetal growth negatively.

Disruption of the one-carbon metabolism could also underlie the inverse association between first trimester maternal 5-HTP and embryonic and fetal growth. TRP metabolism provides one-carbon units for the folate cycle and shares cofactors with one-carbon metabolism, which is crucial for epigenetic programming of cells and tissues (Stegers-Theunissen et al., 2013; Shibata et al., 2015). A deficiency of a shared cofactor, for example vitamin B6, can lead to increased 5-HTP and homocysteine concentrations (Shabbir et al., 2013; Steegers-Theunissen et al., 2013). Increased homocysteine is a marker of derangements in one-carbon metabolism, which has been associated with impaired embryonic and placental development (Parisi et al., 2017; Hoek et al., 2021; Rubini et al., 2021, 2022).

No associations between 5-HT and embryonic and fetal growth were found in the present study. This can be explained by the methodological difficulty to accurately measure serum 5-HT owing to leakage of 5-HT from thrombocytes during the clotting process. Nevertheless, the concentrations found in our study are comparable with those reported in earlier studies (Ernberg et al.,

2000; Comai *et al.*, 2010; Lindström *et al.*, 2018; Zhao *et al.*, 2022). Future research should consider assessing platelet count and 5-HT in other blood fractions as well, such as in platelet-rich plasma (Korse *et al.*, 2017).

Another interesting finding was the association between higher maternal KYN concentrations during the first trimester of pregnancy and a reduced risk of SGA. These associations were only statistically significant in the unadjusted model, and therefore need to be interpreted with caution. Although reference values for maternal KYN concentrations during the first trimester of pregnancy are currently lacking, the concentrations observed in our study align with those recently reported (Sha *et al.*, 2022). Even though maternal KYN concentrations remain relatively stable throughout pregnancy, we accounted for timing of the blood draw in the first trimester of pregnancy by adjusting for gestational age at the moment of the blood draw, which occurred at hospital intake between 7 and 9 weeks of gestation (Schröcksnadel *et al.*, 1996; van Zundert, *et al.*, 2022a). However, as expected, the timing of the blood draw did not affect the results substantially.

In line with our findings, a recent study reported lower maternal KYN concentrations in the third trimester in pregnancies complicated by FGR (Yaman *et al.*, 2023). Explanations for the positive association between first trimester maternal KYN concentrations and SGA involve the several essential functions of the KYN pathway during pregnancy, including its role in neovascularization, vasodilatation of placental arteries, and immune regulation (Wang *et al.*, 2010; Mondal *et al.*, 2016; Badawy, 2017a; Broekhuizen, *et al.*, 2021; Worton *et al.*, 2021). These processes are all important for embryonic and fetal growth, and placental development and functioning. No clear positive associations were found with fetal growth. Since the assessment of fetal growth was based on EFW calculated from the mid-pregnancy anomaly scan, measurements errors cannot be ruled out, which may have weakened the observed effects (Milner and Arezina, 2018).

First trimester maternal KYN concentrations were also not positively associated with embryonic growth in this study. Excess KYN concentrations can induce oxidative stress and dysregulate apoptosis, which may affect embryonic growth negatively (Forrest *et al.*, 2004; Wang *et al.*, 2014). Possibly, the potential harmful effects of increased KYN concentrations may outweigh the relaxing effects on placental arteries, since the embryo is dependent on histiotrophic nutrition rather than on hemotrophic nutrition via the placenta during early pregnancy (Burton, *et al.*, 2001).

Strengths and limitations

To the best of our knowledge, this is the first human study that investigated associations between maternal TRP metabolites in the first trimester of pregnancy and embryonic growth. The primary strengths of our study lie in its periconceptional prospective design, enabling comprehensive data collection on various periconception maternal characteristics through validated general and food questionnaires (Feunekes *et al.*, 1993; Fayet *et al.*, 2011). This allowed building extensive multivariable models to adjust for potential confounders including dietary factors, such as protein and energy intake, although the presence of residual confounding from unobserved and unknown factors, such as comorbidities and medication use (e.g. selective serotonin reuptake inhibitors and interferon alpha), cannot be ruled out (Correia and Vale, 2022). In addition, standardized longitudinal embryonic growth measurements were conducted using 3D ultrasound scans and virtual reality systems, which have been proven highly accurate and reliable (Rousian, *et al.*, 2010, 2021).

Furthermore, the TRP concentrations were determined using a recently published LC-MS/MS method, which has been validated in women in early pregnancy, allowing for sensitive and accurate measurements (van Zundert *et al.*, 2022b). Possible selection bias exists if associations differ between excluded and included participants; however, there is no reason to assume, for example, that lower first trimester maternal 5-HTP concentrations result in reduced embryonic growth in the excluded population compared to the study population, especially since participants were unaware of the study's associations, and the outcome did not impact their pregnancy course or health.

Total TRP was measured instead of free TRP, as free TRP can be influenced by various factors that were not optimal in this study, such as the absence of fasting blood samples (Badawy, 2010). However, this limits the interpretation of maternal TRP concentrations, as only free TRP is available for protein synthesis and metabolic processes. Another potential limitation is the recruitment of women from a tertiary hospital, which may limit external validity. We expect that the direction of the associations will remain consistent in the general population, and there is a possibility that they might even be stronger. Considering the explorative character of this study, we accepted a higher type I error rate, and have not corrected for multiple testing using statistical procedures.

Implications and future research

Our findings provide valuable insights into the role of the maternal TRP metabolism during the first trimester of pregnancy in the physiology of embryonic and fetal growth. Given the pivotal role of embryonic and fetal growth in determining birth outcomes, and long-term health, including potential transgenerational effects, understanding the associated metabolic processes is crucial to achieve the ultimate goal of optimizing embryonic and fetal growth and reducing the occurrence of adverse birth outcomes (Stegers-Theunissen *et al.*, 2013). Our findings suggest that alterations in maternal TRP metabolism, particularly leading to increased 5-HTP concentrations, may contribute to impaired embryonic and fetal growth. This novel finding contributes to our comprehension of the pathophysiology of impaired embryonic and fetal growth, and provides potential preventative and therapeutic targets that warrant further investigation to assess their value in periconception care. A well-balanced diet containing an appropriate intake of macro- and micronutrients is recommended, and vigilance is required when making future dietary recommendations for TRP intake. Future research is needed to identify modifiable factors that influence maternal TRP metabolism, which can be addressed in periconception care to optimize maternal TRP metabolism, and subsequently enhance embryonic and fetal growth. Despite the promising results, additional research within large prospective cohorts and intervention studies is required to validate our findings and to further elucidate the mechanisms through which alterations of maternal TRP metabolites affect embryonic and fetal growth. To develop a full picture of the maternal TRP metabolism, free TRP also needs to be determined in future studies. To measure free TRP accurately, fasting blood samples collected and stored under strictly controlled conditions are required (Badawy, 2010). This remains a challenging task for future studies.

Conclusion

In conclusion, alterations of the maternal TRP metabolism during the first trimester of pregnancy, specifically increased 5-HTP

concentrations, are associated with a smaller embryo, and potentially with a smaller fetus and a higher risk of SGA. In contrast, increased maternal KYN concentrations during the first trimester of pregnancy may be associated with a larger fetus and a lower risk of SGA. The findings of this study indicate that maternal TRP metabolism during the first trimester of pregnancy potentially plays a role in the (patho)physiology of embryonic and fetal growth.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article will be shared upon reasonable request to the corresponding author.

Acknowledgements

We would like to thank all women of the Rotterdam Periconceptional Cohort (Predict Study) for their participation, without whom the study could not have been performed. In addition, we gratefully acknowledge the help provided by the researchers of the Predict Study team for recruiting participants and collecting data, and by Ann Vanrolleghem for coordination of the Predict Study.

Authors' roles

Under supervision of M.M., S.K.M.v.Z. and N.C.M.v.E. drafted the manuscript, and L.v.R., S.P.W., P.H.G., R.H.N.v.S., and R.P.M.S.-T. revised it critically. With help of S.P.W. and L.v.R., S.K.M.v.Z. and N.v.E. performed the data analysis. L.v.R., M.M., R.H.N.v.S., and R.P.M.S.-T. helped with interpretation of the data. R.P.M.S.-T. is principal investigator of the Rotterdam Periconceptional Cohort (Predict Study) and initiated the study aims. M.M. developed and designed the LC-MS/MS method used to determine the TRP metabolites in this study, and P.H.G. prepared all serum samples and analyzed the LC-MS/MS data together with M.M. All authors have read and approved the final version to be published.

Funding

This study was funded by the Department of Obstetrics and Gynecology and the Department of Clinical Chemistry of the Erasmus MC, University Medical Center, Rotterdam, the Netherlands.

Conflict of interest

None disclosed.

References

Abdollahi S, Soltani S, de Souza RJ, Forbes SC, Toupchian O, Salehi-Abargouei A. Associations between maternal dietary patterns and perinatal outcomes: a systematic review and meta-analysis of cohort studies. *Adv Nutr* 2021;**12**:1332–1352.

Alders M. Classification of the population with a foreign background in the Netherlands. In: *Paper for the conference "The measure and mismeasure of populations. The statistical use of ethnic and racial*

categories in multicultural societies". 2001. Paris: Citeseer, p. 18. Statistics Netherlands.

Badawy AA. Plasma free tryptophan revisited: what you need to know and do before measuring it. *J Psychopharmacol* 2010;**24**:809–815.

Badawy AA. Tryptophan metabolism, disposition and utilization in pregnancy. *Biosci Rep* 2015;**35**:1–16.

Badawy AA. Kynurenine pathway of tryptophan metabolism: regulatory and functional aspects. *Int J Tryptophan Res* 2017a;**10**:1178646917691938.

Badawy AA. Tryptophan availability for kynurenine pathway metabolism across the life span: Control mechanisms and focus on aging, exercise, diet and nutritional supplements. *Neuropharmacology* 2017b;**112**:248–263.

Badawy AA, Guillemin G. The plasma [kynurenine]/[tryptophan] ratio and indoleamine 2,3-dioxygenase: time for appraisal. *Int J Tryptophan Res* 2019;**12**:1178646919868978.

Bender DA. Biochemistry of tryptophan in health and disease. *Mol Aspects Med* 1983;**6**:101–197.

Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations. *Int J Obes Relat Metab Disord* 2000;**24**:1119–1130.

Broekhuizen M, Danser AHJ, Reiss IKM, Merkus D. The function of the kynurenine pathway in the placenta: a novel pharmacotherapeutic target? *Int J Environ Res Public Health* 2021;**18**:11545.

Burton GJ, Hempstock J, Jauniaux E. Nutrition of the human fetus during the first trimester—a review. *Placenta* 2001;**22 Suppl A**:S70–S77.

Comai S, Bertazzo A, Carretti N, Podfigurna-Stopa A, Luisi S, Costa CV. Serum levels of tryptophan, 5-hydroxytryptophan and serotonin in patients affected with different forms of amenorrhea. *Int J Tryptophan Res* 2010;**3**:69–75.

Correia AS, Vale N. Tryptophan metabolism in depression: a narrative review with a focus on serotonin and kynurenine pathways. *Int J Mol Sci* 2022;**23**:8493.

Dickey RP, Gasser RF. Ultrasound evidence for variability in the size and development of normal human embryos before the tenth post-insemination week after assisted reproductive technologies. *Hum Reprod* 1993;**8**:331–337.

Ernberg M, Voog U, Alstergren P, Lundberg T, Kopp S. Plasma and serum serotonin levels and their relationship to orofacial pain and anxiety in fibromyalgia. *J Orofac Pain* 2000;**14**:37–46.

Fayet F, Flood V, Petocz P, Samman S. Relative and biomarker-based validity of a food frequency questionnaire that measures the intakes of vitamin B(12), folate, iron, and zinc in young women. *Nutr Res* 2011;**31**:14–20.

Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. *Am J Clin Nutr* 1993;**58**:489–496.

Forrest CM, Mackay GM, Stoy N, Egerton M, Christofides J, Stone TW, Darlington LG. Tryptophan loading induces oxidative stress. *Free Radic Res* 2004;**38**:1167–1171.

Gaillard R, de Ridder MA, Verburg BO, Witteman JC, Mackenbach JP, Moll HA, Hofman A, Steegers EA, Jaddoe VW. Individually customised fetal weight charts derived from ultrasound measurements: the Generation R Study. *Eur J Epidemiol* 2011;**26**:919–926.

Gluckman PD, Hanson MA, Buklijas T. A conceptual framework for the developmental origins of health and disease. *J Dev Orig Health Dis* 2010;**1**:6–18.

Gordeeva O, Gordeev A. Comparative assessment of toxic responses in 3D embryoid body differentiation model and mouse early

- embryos treated with 5-hydroxytryptophan. *Arch Toxicol* 2021; **95**:253–269.
- Gordeeva O, Safandeev V. 5-Hydroxytryptophan (5-HTP)-induced intracellular syndrome in mouse non-neural embryonic cells is associated with inhibited proliferation and cell death. *Neuropharmacology* 2021; **195**:107862.
- Gumusoglu S, Scroggins S, Vignato J, Santillan D, Santillan M. The serotonin-immune axis in preeclampsia. *Curr Hypertens Rep* 2021; **23**:37.
- Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study. *Am J Obstet Gynecol* 1985; **151**:333–337.
- Hoek J, Schoenmakers S, Ringelberg B, Reijnders IF, Willemsen SP, De Rijke YB, Mulders A, Steegers-Theunissen RPM. Periconceptional maternal and paternal homocysteine levels and early utero-placental (vascular) growth trajectories: The Rotterdam periconception cohort. *Placenta* 2021; **115**:45–52.
- Hoftiezer L, Hukkelhoven CW, Hogeveen M, Straatman HM, van Lingen RA. Defining small-for-gestational-age: prescriptive versus descriptive birthweight standards. *Eur J Pediatr* 2016; **175**:1047–1057.
- IBM Corp. IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp, 2021.
- Jaddoe VW, de Jonge LL, Hofman A, Franco OH, Steegers EA, Gaillard R. First trimester fetal growth restriction and cardiovascular risk factors in school age children: population based cohort study. *BMJ* 2014; **348**:g14.
- Kalhan SC. One carbon metabolism in pregnancy: impact on maternal, fetal and neonatal health. *Mol Cell Endocrinol* 2016; **435**:48–60.
- Korse CM, Buning-Kager J, Linders TC, Heijboer AC, van den Broek D, Tesselaar MET, van Tellingen O, van Rossum HH. A serum and platelet-rich plasma serotonin assay using liquid chromatography tandem mass spectrometry for monitoring of neuroendocrine tumor patients. *Clin Chim Acta* 2017; **469**:130–135.
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
- Lindström M, Tohmola N, Renkonen R, Hämäläinen E, Schalin-Jäntti C, Itkonen O. Comparison of serum serotonin and serum 5-HIAA LC-MS/MS assays in the diagnosis of serotonin producing neuroendocrine neoplasms: a pilot study. *Clin Chim Acta* 2018; **482**:78–83.
- Milner J, Arezina J. The accuracy of ultrasound estimation of fetal weight in comparison to birth weight: a systematic review. *Ultrasound* 2018; **26**:32–41.
- Mondal A, Smith C, DuHadaway JB, Sutanto-Ward E, Prendergast GC, Bravo-Nuevo A, Muller AJ. IDO1 is an integral mediator of inflammatory neovascularization. *EBioMedicine* 2016; **14**:74–82.
- Mook-Kanamori DO, Steegers EA, Eilers PH, Raat H, Hofman A, Jaddoe VW. Risk factors and outcomes associated with first-trimester fetal growth restriction. *JAMA* 2010; **303**:527–534.
- Most J, Dervis S, Haman F, Adamo KB, Redman LM. Energy intake requirements in pregnancy. *Nutrients* 2019; **11**:11.
- Murphy MM, Higgins KA, Bi X, Barraj LM. Adequacy and sources of protein intake among pregnant women in the United States, NHANES 2003–2012. *Nutrients* 2021; **13**:795.
- Oxenkrug GF. Metabolic syndrome, age-associated neuroendocrine disorders, and dysregulation of tryptophan-kynurenine metabolism. *Ann N Y Acad Sci* 2010; **1199**:1–14.
- Parisi F, Rousian M, Koning AH, Willemsen SP, Cetin I, Steegers-Theunissen RP. Periconceptional maternal one-carbon biomarkers are associated with embryonic development according to the Carnegie stages. *Hum Reprod* 2017; **32**:523–530.
- Pietersma CS, Mulders A, Sabanovic A, Willemsen SP, Jansen MS, Steegers EAP, Steegers-Theunissen RPM, Rousian M. The impact of maternal smoking on embryonic morphological development: the Rotterdam Periconception Cohort. *Hum Reprod* 2022; **37**:696–707.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2022.
- Ranzil S, Ellery S, Walker DW, Vaillancourt C, Alfaidy N, Bonnin A, Borg A, Wallace EM, Ebeling PR, Erwich JJ et al. Disrupted placental serotonin synthetic pathway and increased placental serotonin: potential implications in the pathogenesis of human fetal growth restriction. *Placenta* 2019; **84**:74–83.
- Rousian M, Koning AH, van Oppenraaij RH, Hop WC, Verwoerd-Dikkeboom CM, van der Spek PJ, Exalto N, Steegers EA. An innovative virtual reality technique for automated human embryonic volume measurements. *Hum Reprod* 2010; **25**:2210–2216.
- Rousian M, Schoenmakers S, Eggink AJ, Gootjes DV, Koning AH, Koster MP, Mulders AG, Baart EB, Reiss IK, Laven JS et al. Cohort profile update: the Rotterdam periconceptional cohort and embryonic and fetal measurements using 3D ultrasound and virtual reality techniques. *Int J Epidemiol* 2021; **50**:1426–1427.
- Rubini E, Baijens IMM, Horánszky A, Schoenmakers S, Sinclair KD, Zana M, Dinnyés A, Steegers-Theunissen RPM, Rousian M. Maternal one-carbon metabolism during the periconceptional period and human foetal brain growth: a systematic review. *Genes (Basel)* 2021; **12**:1634.
- Rubini E, Snoek KM, Schoenmakers S, Willemsen SP, Sinclair KD, Rousian M, Steegers-Theunissen RPM. First trimester maternal homocysteine and embryonic and fetal growth: the Rotterdam Periconception Cohort. *Nutrients* 2022; **14**:112.
- Salas SP, Giacaman A, Romero W, Downey P, Aranda E, Mezzano D, Vío CP. Pregnant rats treated with a serotonin precursor have reduced fetal weight and lower plasma volume and kallikrein levels. *Hypertension* 2007; **50**:773–779.
- Schröcksnadel H, Baier-Bitterlich G, Dapunt O, Wachter H, Fuchs D. Decreased plasma tryptophan in pregnancy. *Obstet Gynecol* 1996; **88**:47–50.
- Sha Q, Madaj Z, Keaton S, Escobar Galvis ML, Smart L, Krzyzanowski S, Fazleabas AT, Leach R, Postolache TT, Achtyes ED et al. Cytokines and tryptophan metabolites can predict depressive symptoms in pregnancy. *Transl Psychiatry* 2022; **12**:35–38.
- Shabbir F, Patel A, Mattison C, Bose S, Krishnamohan R, Sweeney E, Sandhu S, Nel W, Rais A, Sandhu R et al. Effect of diet on serotonergic neurotransmission in depression. *Neurochem Int* 2013; **62**:324–329.
- Shibata K, Hirose J, Fukuwatari T. Method for evaluation of the requirements of B-group vitamins using tryptophan metabolites in human urine. *Int J Tryptophan Res* 2015; **8**:31–39.
- Smit AJP, Hojeij B, Rousian M, Schoenmakers S, Willemsen SP, Steegers-Theunissen RPM, van Rossem L. A high periconceptional maternal ultra-processed food consumption impairs embryonic growth: the Rotterdam periconceptional cohort. *Clin Nutr* 2022; **41**:1667–1675.
- Steegers-Theunissen RP, Twigt J, Pestinger V, Sinclair KD. The periconceptional period, reproduction and long-term health of offspring: the importance of one-carbon metabolism. *Hum Reprod Update* 2013; **19**:640–655.
- Steegers-Theunissen RP, Verheijden-Paulissen JJ, van Uitert EM, Wildhagen MF, Exalto N, Koning AH, Eggink AJ, Duvekot JJ, Laven JS, Tibboel D et al. Cohort profile: the Rotterdam Periconceptional Cohort (Predict Study). *Int J Epidemiol* 2016; **45**:374–381.
- United Nations Educational Scientific and Cultural Organization. *International Standard Classification of Education: ISCED* 2011. Montreal: UNESCO Institute for Statistics, 2012, 85.

- Ustunyurt E, Simsek H, Korkmaz B, Iskender C. First-trimester crown-rump length affects birth size symmetrically. *J Matern Fetal Neonatal Med* 2015;**28**:2070–2073.
- van Dijk MR, Borggreven NV, Willemsen SP, Koning AHJ, Steegers-Theunissen RPM, Koster MPH. Maternal lifestyle impairs embryonic growth: the Rotterdam Periconception Cohort. *Reprod Sci* 2018;**25**:916–922.
- van Uitert EM, Exalto N, Burton GJ, Willemsen SP, Koning AH, Eilers PH, Laven JS, Steegers EA, Steegers-Theunissen RP. Human embryonic growth trajectories and associations with fetal growth and birthweight. *Hum Reprod* 2013a;**28**:1753–1761.
- van Uitert EM, van der Elst-Otte N, Wilbers JJ, Exalto N, Willemsen SP, Eilers PH, Koning AH, Steegers EA, Steegers-Theunissen RP. Periconception maternal characteristics and embryonic growth trajectories: the Rotterdam Predict study. *Hum Reprod* 2013b;**28**:3188–3196.
- van Zundert SK, Broekhuizen M, Smit AJ, van Rossem L, Mirzaian M, Willemsen SP, Danser AJ, De Rijke YB, Reiss IK, Merkus D et al. The role of the kynurenine pathway in the (patho)physiology of maternal pregnancy and fetal outcomes: a systematic review. *Int J Tryptophan Res* 2022a;**15**:11786469221135545.
- van Zundert SKM, Griffioen PH, van Rossem L, Willemsen SP, de Rijke YB, van Schaik RHN, Steegers-Theunissen RPM, Mirzaian M. Simultaneous quantification of tryptophan metabolites by liquid chromatography tandem mass spectrometry during early human pregnancy. *Clin Chem Lab Med* 2022b;**61**:442–451.
- Wang Q, Zhang M, Ding Y, Wang Q, Zhang W, Song P, Zou MH. Activation of NAD(P)H oxidase by tryptophan-derived 3-hydroxy-kynurenine accelerates endothelial apoptosis and dysfunction in vivo. *Circ Res* 2014;**114**:480–492.
- Wang Y, Liu H, McKenzie G, Witting PK, Stasch JP, Hahn M, Changsirivathanathamrong D, Wu BJ, Ball HJ, Thomas SR et al. Kynurenine is an endothelium-derived relaxing factor produced during inflammation. *Nat Med* 2010;**16**:279–285.
- Worton SA, Pritchard HAT, Greenwood SL, Alakrawi M, Heazell AEP, Wareing M, Greenstein A, Myers JE. Kynurenine relaxes arteries of normotensive women and those with preeclampsia. *Circ Res* 2021;**128**:1679–1693.
- Wurtman RJ, Wurtman JJ, Regan MM, McDermott JM, Tsay RH, Breu JJ. Effects of normal meals rich in carbohydrates or proteins on plasma tryptophan and tyrosine ratios. *Am J Clin Nutr* 2003;**77**:128–132.
- Xu Y, Ni M, Zhang Q, Zhao J, Tang Z, Liu Z. Correlation between crown-rump length in the first trimester of pregnancy and neonatal outcomes. *BMC Pediatr* 2022;**22**:386.
- Yaman S, Ceyhan M, Hancerliogullari N, Koc EM, Candar T, Tokmak A. Serum levels of kynurenine in pregnancies with fetal growth restriction and oligohydramnios. *J Perinat Med* 2023;**51**:641–645.
- Zhao YJ, Zhou C, Wei YY, Li HH, Lei W, Boeldt DS, Wang K, Zheng J. Differential distribution of tryptophan-metabolites in fetal and maternal circulations during normotensive and preeclamptic pregnancies. *Reprod Sci* 2022;**29**:1278–1286.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Human Reproduction, 2024, 39, 912–922
<https://doi.org/10.1093/humrep/deae046>
Original Article

Early pregnancy

Morphologic development of the first-trimester utero-placental vasculature is positively associated with embryonic and fetal growth: the Rotterdam Periconception Cohort

E.S. De Vos ¹, A.G.M.G.J. Mulders ^{1,*}, A.H.J. Koning ², S.P. Willemsen ^{1,3}, M. Rousian ¹, B.B. Van Rijn ¹, E.A.P. Steegers ¹, and R.P.M. Steegers-Theunissen ¹

¹Department of Obstetrics and Gynecology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

²Department of Pathology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

³Department of Biostatistics, Erasmus MC University Medical Center, Rotterdam, The Netherlands

*Correspondence address. Department of Obstetrics and Gynecology/Periconception Epidemiology, Erasmus MC University Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands. Tel: +31-10-7033492/+31-6-28914348; Fax: +31-10 7036815; E-mail: a.mulders@erasmusmc.nl  <https://orcid.org/0000-0002-4407-8323>

ABSTRACT

STUDY QUESTION: Is morphologic development of the first-trimester utero-placental vasculature associated with embryonic growth and development, fetal growth, and birth weight percentiles?

SUMMARY ANSWER: Using the utero-placental vascular skeleton (uPVS) as a new imaging marker, this study reveals morphologic development of the first-trimester utero-placental vasculature is positively associated with embryonic growth and development, fetal growth, and birth weight percentiles.

WHAT IS KNOWN ALREADY: First-trimester development of the utero-placental vasculature is associated with placental function, which subsequently impacts embryonic and fetal ability to reach their full growth potential. The attribution of morphologic variations in the utero-placental vascular development, including the vascular structure and branching density, on prenatal growth remains unknown.

STUDY DESIGN, SIZE, DURATION: This study was conducted in the VIRTUAL Placental study, a subcohort of 214 ongoing pregnancies, embedded in the prospective observational Rotterdam Periconception Cohort (Predict study). Women were included before 10 weeks gestational age (GA) at a tertiary referral hospital in The Netherlands between January 2017 and March 2018.

PARTICIPANTS/MATERIALS, SETTING, METHODS: We obtained three-dimensional power Doppler volumes of the gestational sac including the embryo and the placenta at 7, 9, and 11 weeks of gestation. Virtual Reality-based segmentation and a recently developed skeletonization algorithm were applied to the power Doppler volumes to generate the uPVS and to measure utero-placental vascular volume (uPVV). Absolute vascular morphology was quantified by assigning a morphologic characteristic to each voxel in the uPVS (i.e. end-, bifurcation-crossing-, or vessel point). Additionally, total vascular length (mm) was calculated. The ratios of the uPVS characteristics to the uPVV were calculated to determine the density of vascular branching. Embryonic growth was estimated by crown-rump length and embryonic volume. Embryonic development was estimated by Carnegie stages. Fetal growth was measured by estimated fetal weight in the second and third trimester and birth weight percentiles. Linear mixed models were used to estimate trajectories of longitudinal measurements. Linear regression analysis with adjustments for confounders was used to evaluate associations between trajectories of the uPVS and prenatal growth. Groups were stratified for conception method (natural/IVF-ICSI conceptions), fetal sex (male/female), and the occurrence of placenta-related complications (yes/no).

MAIN RESULTS AND THE ROLE OF CHANCE: Increased absolute vascular morphologic development, estimated by positive random intercepts of the uPVS characteristics, is associated with increased embryonic growth, reflected by crown-rump length (endpoints $\beta = 0.017$, 95% CI [0.009; 0.025], bifurcation points $\beta = 0.012$, 95% CI [0.006; 0.018], crossing points $\beta = 0.017$, 95% CI [0.008; 0.025], vessel points $\beta = 0.01$, 95% CI [0.002; 0.008], and total vascular length $\beta = 0.007$, 95% CI [0.003; 0.010], and similarly with embryonic volume and Carnegie stage, all P-values ≤ 0.01 . Density of vascular branching was negatively associated with estimated fetal weight in the third trimester (endpoints: uPVV $\beta = -94.972$, 95% CI [-185.245; -3.698], bifurcation points: uPVV $\beta = -192.601$ 95% CI [-360.532; -24.670]) and birth weight percentiles (endpoints: uPVV $\beta = -20.727$, 95% CI [-32.771; -8.683], bifurcation points: uPVV $\beta = -51.097$ 95% CI [-72.257; -29.937], and crossing points: uPVV $\beta = -48.604$ 95% CI [-74.246; -22.961]), all P-values < 0.05 . After stratification, the associations were observed in natural conceptions specifically.

LIMITATION, REASONS FOR CAUTION: Although the results of this prospective observational study clearly demonstrate associations between first-trimester utero-placental vascular morphologic development and prenatal growth, further research is required before we can draw firm conclusions about a causal relationship.

Received: June 26, 2023. Revised: January 31, 2024. Editorial decision: February 29, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

WIDER IMPLICATIONS OF THE FINDINGS: Our findings support the hypothesis that morphologic variations in utero-placental vascular development play a role in the vascular mechanisms involved in embryonic and fetal growth and development. Application of the uPVS could benefit our understanding of the pathophysiology underlying placenta-related complications. Future research should focus on the clinical applicability of the uPVS as an imaging marker for the early detection of fetal growth restriction.

STUDY FUNDING/COMPETING INTEREST(S): This research was funded by the Department of Obstetrics and Gynecology of the Erasmus MC, University Medical Centre, Rotterdam, The Netherlands. There are no conflicts of interest.

TRIAL REGISTRATION NUMBER: Registered at the Dutch Trial Register (NTR6854).

Keywords: prenatal growth / placenta / 3D / power Doppler / ultrasound / virtual reality / spiral artery / early pregnancy / vascular morphology / fetal growth restriction

Introduction

Worldwide, impaired prenatal growth, defined as fetal growth restriction (FGR) and infants born small-for-gestational-age (SGA), is an important risk factor for perinatal morbidity and mortality, and has life-course health implications (Lawn et al., 2005; Almond and Currie, 2011; Hanson and Gluckman, 2014; Jaddoe et al., 2014; Malhotra et al., 2019). It is generally recognized that utero-placental vascular development contributes to prenatal growth as it dictates the maternal blood flow to the placenta (Junaid et al., 2014; Burton and Jauniaux, 2018). To provide more insight into vascular mechanisms involved in prenatal growth, it is essential to investigate the first-trimester development of the utero-placental vasculature in vivo. The ability to monitor first-trimester utero-placental vascular development in humans will be valuable for the early detection, treatment, and prevention of placenta-related complications.

Ultrasound provides low-cost opportunities for longitudinal non-invasive evaluation of prenatal growth and is widely used in antenatal care. Associations between utero-placental vascular development and prenatal growth have been investigated by a variety of ultrasound markers, including placental volume (PV), vascularization and flow indices, and uterine artery Dopplers (Collins et al., 2013; Reus et al., 2014; Velauthar et al., 2014; Alfievic et al., 2017; Chen et al., 2022). Evidence suggests utero-placental vascular development and prenatal growth are associated with conception method and fetal sex (Clifton, 2010; Broere-Brown et al., 2016; Rizzo et al., 2016; Cavoretto et al., 2020; Reijnders et al., 2021). Combining three-dimensional (3D) power Doppler (PD) ultrasound and virtual reality (VR), a recent study showed positive associations between utero-placental vascular volume (uPVV) development and embryonic growth, with fetal sex specific modifications (Reijnders et al., 2021). Unfortunately, these studies have not resulted in an imaging marker with potential for early detection of pregnancies at risk for FGR or SGA.

Histopathological research has shown that morphologic variations of the placental vasculature, including vessel caliber and number of remodeled spiral arteries in the placental bed, are associated with prenatal growth (Brosens et al., 2011). However, it is unclear to what extent morphologic variations in the first trimester affect prenatal growth. Recently, our study group has combined 3D PD ultrasound, VR and a skeletonization algorithm to generate the utero-placental vascular skeleton (uPVS) as a structural overview of the utero-placental vascular morphology in the first trimester (de Vos et al., 2022). The uPVS provides information on branching patterns, vascular network length and thickness and can be used to calculate density of vascular branching when combined with uPVV measurements. Using the uPVS, we found decreased utero-placental vascular morphologic development in pregnancies with placenta-related complications (de Vos et al., 2022). Therefore, we hypothesize that uPVS development is

associated with prenatal growth and thus may have potential as an imaging marker for the early detection of FGR and SGA.

In this study, we aim to investigate associations between morphologic development of the utero-placental vasculature and prenatal growth by analyzing associations between uPVS characteristics, density of vascular branching and measurements of embryonic and fetal growth and birth weight. Further, our aim is to investigate whether these associations are affected by conception method, fetal sex and the occurrence of placenta-related complications.

Materials and methods

Ethical approval and study design

The Rotterdam Periconception Cohort (Predict study) was approved by the Institutional Review Board of the Erasmus Medical Centre on 15 October 2004 (MEC-2004-227) (Stegers-Theunissen et al., 2016; Rousian et al., 2021). The VIRTUAL Placenta study, embedded in the Rotterdam Periconception Cohort, was approved by the Institutional Review Board of the Erasmus Medical Centre on 2 June 2015 (MEC 2015-494) (Reijnders et al., 2018).

Women who met the following criteria—a minimum age of 18 years old, a singleton pregnancy with a gestational age (GA) of less than 10 weeks and written informed consent—were recruited from the Erasmus Medical Centre, a tertiary referral hospital in the Netherlands. Both natural conceptions and pregnancies achieved via IVF or ICSI using autologous oocytes were eligible for inclusion.

Study parameters

Between January 2017 and March 2018, participants were enrolled and answered a self-administered questionnaire on general characteristics and medical and obstetric history. At least two study visits were scheduled in the first trimester at 7, 9, and 11 weeks GA, and two study visits at 22 and 32 weeks GA. At the first study visit, height and weight were measured methodically to calculate body-mass index (BMI).

For natural conceptions in regular cycles (25–35 days), GA was calculated from the first day of last menstrual period (LMP). In case of unknown LMP or irregular cycle, GA was calculated from Crown-Rump-Length (CRL). If the two methods varied >6 days, the CRL-based GA was used in pregnancy dating. For fresh IVF-ICSI pregnancies, GA was calculated from oocyte pick-up day +14 days. In case of cryopreserved embryo transfer, GA was calculated from transfer date +19 days.

Pregnancy outcomes were collected through a questionnaire filled out by participants and complemented with medical delivery records. Birth weight was standardized according to Hoftiezer percentiles (Hoftiezer et al., 2016).

Ultrasonography

Trained sonographers performed all ultrasounds using the GE Voluson E8 (GE, Zipf, Austria) with a transvaginal 6–12 MHz transducer. Standardized ultrasound settings were previously described (quality: max; pulse repetition frequency (PRF): 0.6; wall motion filter (WMF): low1; compound resolution imaging (CRI): off; power Doppler (PD) gain: -8.0) (Reijnders et al., 2018).

All first-trimester measurements included transvaginal 3D sweeps of the embryo and gestational sac including the placenta. The utero-placental vasculature was visualized with a 3D PD sweep (Reijnders et al., 2018). All ultrasound examinations were performed according to international guidelines on safe use of Doppler ultrasound in the first trimester of pregnancy (ALARA-principle) (Drukker et al., 2020).

At 22 and 32 weeks GA, transabdominal ultrasonography was performed to measure fetal growth in the second and third trimester, defined as estimated fetal weight (EFW) calculated using the Hadlock formula (Hadlock et al., 1985).

Offline 3D measurements

Image quality was scored as optimal, good, average, or poor based on the presence of artefacts such as blurring of the volume or acoustic shadowing, the ability to distinguish between myometrium and trophoblast, and completeness of the placenta. Images with quality score 'poor' were deemed unusable and excluded from the analyses.

Offline, the V-Scope volume rendering application was used to determine CRL, embryonic volume (EV), and Carnegie stage as measures for embryonic growth and to measure uPVV from the 3D PD sweeps using VR-based segmentation (Supplementary Fig. S1A–C) (Verwoerd-Dikkeboom et al., 2008; O'Rahilly and Muller, 2010; Reijnders et al., 2018; Rousian et al., 2018).

Advanced image processing for retrieving morphologic characteristics

A skeletonization algorithm was applied to the uPVV segmentations to automatically generate the uPVS. This algorithm repeatedly peels off the outermost layer of voxels from the uPVV, reducing the diameter of the PD signal at each point in the vascular network until one central voxel remains, thereby creating a network-like structure representing the vascular morphology (Supplementary Fig. S1D) (de Vos et al., 2022).

Following the construction of the network, the skeletonization algorithm classifies each 26-connected voxel based on the number of neighboring voxels. The algorithm identifies endpoints (one neighbor), bifurcation points (three neighbors), crossing points (four neighbors), and normal vessel points (two neighbors). Voxels with >4 neighbors are considered an artefact and excluded from analyses. Further, the algorithm measures total network length and average vascular thickness (Supplementary Fig. S1E). These six uPVS characteristics represent absolute vascular morphologic development of the first-trimester utero-placental vasculature. Finally, we calculated the number uPVS end-, bifurcation-, and crossing points per cm^3 uPVV to identify three imaging markers to represent the density of vascular branching in the uPVV. A graphic representation of the virtual imaging markers of first-trimester utero-placental vascular development (uPVV, uPVS, and density of vascular branching) is depicted in Fig. 1.

Given the small contribution of embryonic capillaries to the VR-based segmentation and the restricted presence of embryonic flow in the first trimester, we assume the contribution of the embryonic blood space to the uPVV and the uPVS is limited. Therefore, we conclude that the uPVV and uPVS mainly include

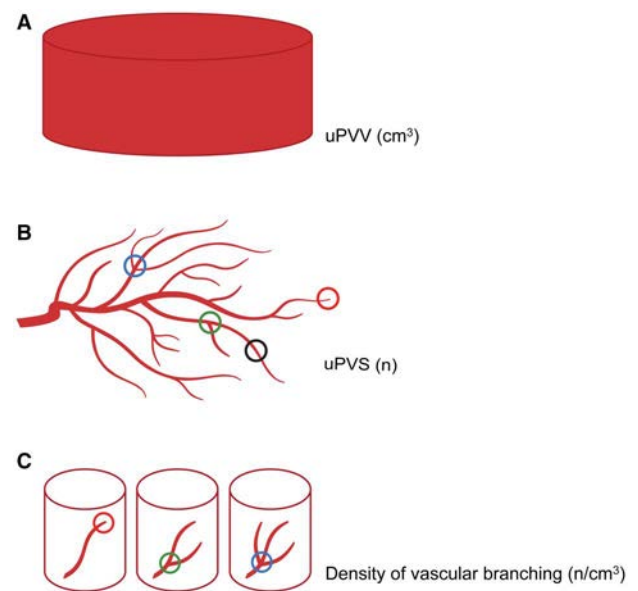


Figure 1. Graphic representation of virtual imaging markers of first-trimester utero-placental vascular development. (A) The utero-placental vascular volume (uPVV) measures the total volume of all power Doppler signal within the first-trimester placental tissue up to the placental-myometrial interface in cm^3 . (B) The utero-placental vascular skeleton (uPVS) measures the number and type of branches within the uPVV. Red circle: voxels that are endpoints. Green circle: voxels that are bifurcation points. Blue circle: voxels that are crossing points. Black circle: voxels that are vessel points (no branch or endpoint). In addition, the total vascular length and average vascular thickness are calculated. (C) Density of vascular branching is calculated by dividing the number of endpoints, bifurcation points and crossing points by the uPVV.

the distal segments of the maternal spiral and basal arteries and communicating anastomoses (arteriovenous shunts), see Fig. 2A and B (Reijnders et al., 2018; de Vos et al., 2022).

Definitions of placenta-related complications

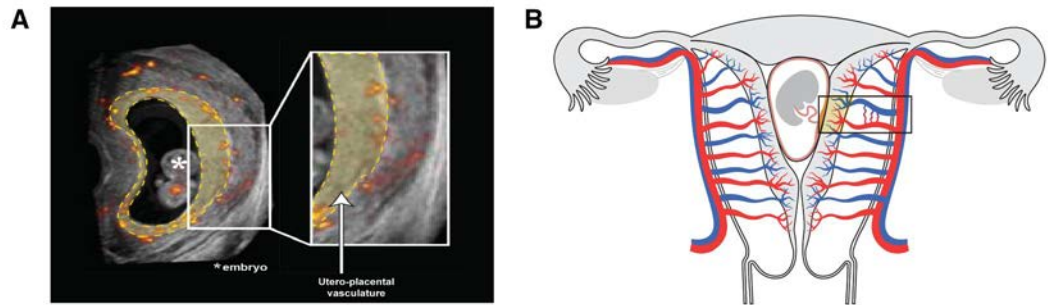
Placenta-related complications were specified as pregnancy-induced hypertension (PIH), preeclampsia (PE), and/or FGR, SGA, or preterm birth (PTB). PIH was defined as new-onset hypertension after 20 weeks GA with systolic blood pressure ≥ 140 mmHg and/or diastolic blood pressure ≥ 90 mmHg (Croke, 2019). PE was defined as PIH with the presence of ≥ 300 mg proteinuria in a 24-h period (Croke, 2019). FGR was defined as a fetal abdominal circumference and/or EFW < 10 th percentile on the Hadlock curve, or > 20 percentile decrease compared to a previous measurement with a minimal timespan of 2 weeks (Gordijn et al., 2016). SGA was defined as birth weight < 10 th percentile on growth curves specific for GA at birth, parity, and fetal sex (Zeve et al., 2016). PTB was defined as a GA at birth $< 37 + 0$ weeks (Vogel et al., 2018).

Statistical analyses

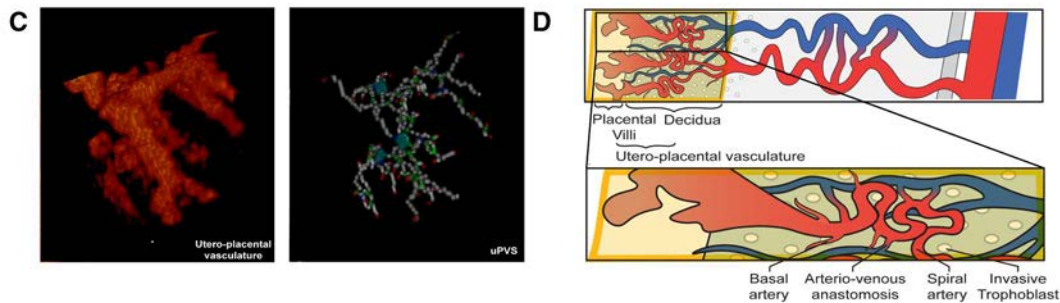
Baseline characteristics were presented as mean with standard deviation or median with first and third quartiles. If needed, non-volumetric parameters, volumetric parameters, and ratios were transformed using a square root, cubic root and natural log transformation, respectively, to approximate a normal distribution.

Linear mixed models were used to estimate individual trajectories of longitudinal measurements using a quadratic relation with GA. In these models, we allowed for subject-specific intercepts and slopes, the random effects. The models with the best fit included random intercepts only. For Carnegie stages a

First-trimester pregnancy



Normal prenatal growth



Restricted prenatal growth

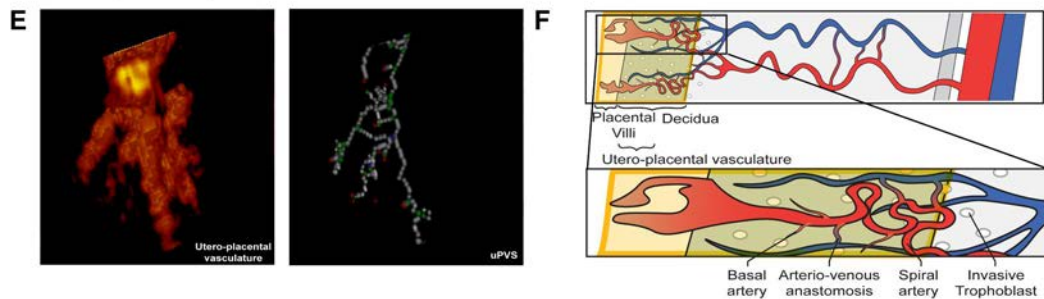


Figure 2. Power Doppler ultrasound and schematic images of the utero-placental vasculature in a first-trimester pregnancy displaying normal and restricted prenatal growth. Overview of the utero-placental vasculature. (A) Two-dimensional (2D) power Doppler (PD) ultrasound image of a first-trimester pregnancy; all recorded vasculature is depicted in red-yellow colour. The yellow-highlighted section forms the virtual reality (VR)-based segmentation and corresponds with the yellow box in the schematic images. All vasculature in the VR-based segmentation is included in the utero-placental vascular skeleton (uPVS). (B) Schematic image of a first-trimester pregnant uterus including the uterine and placental vasculature (red, arterial vasculature; blue, venous vasculature). The yellow box contains decidua, invasive extravillous trophoblast and placental tissue which form the interface of the utero-placental vasculature. The vascular structures within the utero-placental vasculature are confined to the maternal basal and spiral arteries and their communicating arterio-venous anastomoses, and possibly vessels of the embryonic/foetal-placental blood space. (C) Enlarged section of a 3D PD ultrasound image of the utero-placental vasculature and uPVS in pregnancies with normal prenatal growth. The uPVS is characterized by an increased number of end- (red), bifurcation- (green), crossing- (blue), and vessel- (white) points when compared to pregnancies with restricted prenatal growth (as shown in panel E). (D) Schematic image of the utero-placental vasculature in pregnancies with normal prenatal growth, characterized by an increased absolute number of branches and arterio-venous anastomosis when compared to pregnancies with restricted prenatal growth. (E) Enlarged section of a 3D PD ultrasound image of the utero-placental vasculature and uPVS in pregnancies with restricted prenatal growth. The uPVS is characterized by a decreased number of end- (red), bifurcation- (green), crossing- (blue), and vessel- (white) points when compared to pregnancies with normal prenatal growth. (F) Schematic image of the utero-placental vasculature in pregnancies with restricted prenatal growth, characterized by a decreased absolute number of branches and arterio-venous anastomosis when compared to pregnancies with normal prenatal growth.

continuation ratio model with random intercepts was estimated. The random intercepts were used as summaries representing first-trimester utero-placental vascular morphologic development and first-trimester embryonic development.

Hereafter, the summaries of first-trimester utero-placental vascular morphologic development (uPVS characteristics and density of vascular branching) were used as covariates in a

linear regression analysis for measures of embryonic growth (summaries of CRL, EV and Carnegie stage development) and fetal growth (EFW at 22 and 32 weeks GA, and birth weight percentiles).

Associations between utero-placental vascular morphologic development and prenatal growth were first assessed in an unadjusted model. A second model was adjusted for maternal age,

BMI, parity, conception mode, smoking, folic acid supplement use and, if appropriate, GA and fetal sex. Potential confounders were selected from literature and discussion amongst authors. In addition, adjusted models were analyzed in subgroups after stratification for conception mode (natural/IVF-ICSI), fetal sex (male/female), and occurrence of placenta-related complications (yes/no).

At last, we applied the Bonferroni correction for multiple testing to reevaluate all our P -values < 0.05 and assessed whether the new level of significance was met.

All analyses were performed using SPSS (version 25.0; SPSS Inc., Chicago, IL, USA) and R (version 4.0.2, R Core Team, Vienna, Austria, 2020). P -values < 0.05 were considered statistically significant.

Results

Study population and first-trimester utero-placental vascular morphology

The flowchart of participant selection is depicted in [Supplementary Fig. S2](#). In total, 214 women and 394 respective placental datasets were included in the analysis. Participant characteristics and pregnancy outcomes are depicted in [Supplementary Table S1](#). [Supplementary Tables S2 and S3](#) summarize absolute vascular morphology and density of vascular branching at 7, 9, and 11 weeks GA, respectively.

Utero-placental vascular morphologic development and embryonic growth

Increased absolute vascular morphologic development, estimated by trajectories of the uPVS characteristics, was positively associated with trajectories of CRL (adjusted model: endpoints: $\beta = 0.017$, bifurcation points: $\beta = 0.012$, crossing points: $\beta = 0.017$, vessel points: $\beta = 0.01$, total length: $\beta = 0.007$, all $P \leq 0.01$) and EV (adjusted model: endpoints: $\beta = 0.008$, bifurcation points: $\beta = 0.01$, crossing points: $\beta = 0.008$, vessel points: $\beta = 0.002$, total length: $\beta = 0.003$, all $P \leq 0.01$), see [Table 1](#). Similarly, absolute vascular morphologic development was positively associated with Carnegie stages (endpoints: OR = 1.23, bifurcation points: OR = 1.14, crossing points OR = 1.20, vessel points: OR = 1.06, total length OR = 1.07, all $P < 0.05$), see [Table 1](#).

Next, we stratified for conception mode, fetal sex and the occurrence of placenta-related complications, see [Table 2](#). Absolute vascular morphologic development was positively associated with trajectories of CRL and EV in both natural and IVF-ICSI conceptions, but associations were statistically significant in natural conceptions only. The subgroup analyses for fetal sex showed positive associations for both sexes, but association estimates were higher in females. Further, subgroup analyses showed positive associations with CRL and EV trajectories in both pregnancies with and without placenta-related complications, with higher association estimates in pregnancies with placenta-related complications. All P -values < 0.05 . For Carnegie stages, no associations were observed in the subgroup analyses.

Trajectories of density of vascular branching were negatively associated with CRL and EV trajectories, but the associations were not statistically significant and we observed no associations with Carnegie stages, see [Supplementary Table S4](#). After stratification for conception mode, fetal sex, and placenta-related complications, there were no associations between trajectories of the density of vascular branching and trajectories of CRL, EV, or Carnegie stages, see [Supplementary Table S5](#).

After correction for multiple testing using Bonferroni correction, almost all previously significant associations ($P < 0.05$) between uPVS characteristics and CRL and EV remained statistically significant.

Utero-placental vascular morphologic development and fetal growth

Absolute vascular morphologic development, estimated by uPVS trajectories, was positively associated with EFW at 22 weeks GA (adjusted model: uPVS end points: $\beta = 3.379$, bifurcation points: $\beta = 2.134$, crossing points: $\beta = 3.207$, vessel points: $\beta = 0.856$, total length: $\beta = 1.205$, all $P < 0.05$) and EFW at 32 weeks GA (adjusted model: uPVS end points: $\beta = 8.802$, vessel points: $\beta = 2.654$, and total length: $\beta = 3.143$, all $P < 0.05$), see [Table 3](#). We observed no associations between absolute vascular morphologic development and birth weight percentiles.

Again, we stratified for conception mode, fetal sex, and the occurrence of placenta-related complications. Subgroup analyses showed absolute vascular morphologic development was positively associated with EFW at 22 weeks GA and 32 weeks GA in

Table 1. Associations between first-trimester trajectories of absolute utero-placental vascular morphologic development and first-trimester trajectories of embryonic growth and development ($n = 214$).

Morphologic characteristic	Crown rump length (mm)		Embryonic volume (cm ³)		Carnegie stage	
	N = 200		N = 192		N = 152	
	β [95% CI]		β [95% CI]		OR [95% CI]	
	Crude	Adjusted	Crude	Adjusted	Crude	Adjusted
√ uPVS endpoints (n)	0.017 [0.009; 0.025]**	0.017 [0.009; 0.025]**	0.008 [0.004; 0.011]**	0.008 [0.004; 0.012]**	1.244 [1.030; 1.504]*	1.233 [1.013; 1.502]*
√ uPVS bifurcation points (n)	0.011 [0.006; 0.017]**	0.012 [0.006; 0.018]**	0.005 [0.003; 0.008]**	0.005 [0.004; 0.008]**	1.155 [1.017; 1.312]*	1.139 [1.002; 1.295]*
√ uPVS crossing points (n)	0.016 [0.009; 0.024]**	0.017 [0.008; 0.025]**	0.007 [0.004; 0.011]**	0.008 [0.004; 0.012]**	1.218 [1.022; 1.452]*	1.196 [1.001; 1.429]*
√ uPVS vessel points (n)	0.005 [0.002; 0.007]**	0.005 [0.002; 0.008]**	0.002 [0.001; 0.003]**	0.002 [0.001; 0.004]**	1.065 [1.008; 1.125]*	1.062 [1.001; 1.126]*
√ Total length (mm)	0.006 [0.004; 0.009]**	0.007 [0.004; 0.010]**	0.003 [0.002; 0.004]**	0.003 [0.002; 0.005]**	1.079 [1.013; 1.149]*	1.077 [1.005; 1.154]*
√ Average thickness (mm)	-0.270 [-0.614; 0.074]	-0.249 [-0.611; 0.114]	-0.084 [-0.260; 0.093]	-0.084 [-0.267; 0.099]	0.453 [0.000; 369.526]	0.274 [0.000; 596.959]

Depicted are the crude and adjusted estimates: adjusted for maternal age, parity, conception mode, BMI, smoking, preconception initiation of folic acid supplement use, and fetal sex. Carnegie stages ($n = 23$). uPVS, utero-placental vascular skeleton. Bold text indicates a statistically significant association. Significance at * $P \leq 0.05$, ** $P \leq 0.01$.

Table 2. Associations between first-trimester trajectories of absolute utero-placental vascular morphologic development and first-trimester trajectories of embryonic growth and development, stratified for conception mode, fetal sex, and occurrence of placenta-related complications.

Conception method	Crown rump length (mm)		Embryonic volume (cm ³)		Carnegie stage	
	Adjusted model		Adjusted model		Adjusted model	
	β [95% CI]		β [95% CI]		OR [95% CI]	
	IVF/ICSI (n = 81)	Natural (n = 117)	IVF/ICSI (n = 81)	Natural (n = 112)	IVF/ICSI (n = 77)	Natural (n = 120)
Morphologic characteristic						
✓ uPVS endpoints (n)	0.007 [−0.004; 0.017]	0.023 [0.010; 0.035]**	0.004 [−0.000; 0.008]	0.011 [0.004; 0.017]**	1.105 [0.929; 1.316]	1.296 [0.969; 1.733]
✓ uPVS bifurcation points (n)	0.004 [−0.003; 0.011]	0.017 [0.008; 0.026]**	0.003 [−0.001; 0.006]	0.007 [0.003; 0.012]**	1.076 [0.953; 1.215]	1.197 [0.973; 1.473]
✓ uPVS crossing points (n)	0.005 [−0.005; 0.015]	0.024 [0.012; 0.036]**	0.004 [−0.001; 0.008]	0.010 [0.004; 0.017]**	1.092 [0.933; 1.278]	1.243 [0.958; 1.613]
✓ uPVS vessel points (n)	0.002 [−0.001; 0.005]	0.007 [0.003; 0.011]**	0.001 [−0.000; 0.003]	0.003 [0.001; 0.005]**	1.035 [0.979; 1.094]	1.084 [0.989; 1.187]
✓ Total length (mm)	0.002 [−0.002; 0.006]	0.010 [0.006; 0.014]**	0.002 [−0.000; 0.003]	0.004 [0.002; 0.006]**	1.037 [0.977; 1.101]	1.105 [0.992; 1.231]
✓ Average thickness (mm)	−0.142 [−0.632; 0.349]	−0.035 [−0.860; 0.167]	−0.046 [−0.259; 0.167]	−0.129 [−0.402; 0.145]	0.269 [0.000; 86.603]	0.269 [0.000; 5038.740]
Fetal sex	Crown rump length (mm)		Embryonic volume (cm ³)		Carnegie stage	
	Adjusted model		Adjusted model		Adjusted model	
	β [95% CI]		β [95% CI]		OR [95% CI]	
	Male (n = 99)	Female (n = 97)	Male (n = 97)	Female (n = 94)	Male (n = 102)	Female (n = 95)
Morphologic characteristic						
✓ uPVS endpoints (n)	0.012 [−0.000; 0.024]	0.021 [0.009; 0.033]**	0.007 [0.000; 0.014]*	0.009 [0.004; 0.014]**	1.108 [0.885; 1.387]	1.452 [0.997; 2.115]
✓ uPVS bifurcation points (n)	0.010 [0.002; 0.018]*	0.013 [0.004; 0.023]*	0.005 [0.000; 0.009]*	0.006 [0.002; 0.010]**	1.071 [0.923; 1.242]	1.342 [0.994; 1.812]
✓ uPVS crossing points (n)	0.014 [0.004; 0.025]**	0.019 [0.006; 0.032]**	0.007 [0.001; 0.013]*	0.008 [0.004; 0.014]**	1.103 [0.896; 1.357]	1.468 [0.965; 2.233]
✓ uPVS vessel points (n)	0.008 [0.001; 0.008]*	0.006 [0.002; 0.010]**	0.002 [0.000; 0.004]*	0.003 [0.001; 0.004]**	1.033 [0.968; 1.103]	1.134 [0.997; 1.290]
✓ Total length (mm)	0.006 [0.002; 0.010]**	0.008 [0.003; 0.012]**	0.003 [0.000; 0.005]*	0.004 [0.002; 0.006]**	1.041 [0.967; 1.121]	1.151 [0.999; 1.327]
✓ Average thickness (mm)	−0.078 [−0.543; 0.386]	−0.490 [−1.084; 0.103]	−0.064 [−0.348; 0.221]	0.124 [−0.373; 0.125]	82.980 [0.000; 759573.720]	0.000 [0.000; 17.565]
Placenta-related complications	Crown rump length (mm)		Embryonic volume (cm ³)		Carnegie stage	
	Adjusted model		Adjusted model		Adjusted model	
	β [95% CI]		β [95% CI]		OR [95% CI]	
	With (n = 50)	Without (n = 148)	With (n = 45)	Without (n = 148)	With (n = 51)	Without (n = 146)
Morphologic characteristic						
✓ uPVS endpoints (n)	0.025 [0.006; 0.045]*	0.014 [0.005; 0.023]**	0.011 [−0.000; 0.023]	0.008 [0.004; 0.012]**	1.399 [0.796; 2.458]	1.152 [0.961; 1.380]
✓ uPVS bifurcation points (n)	0.022 [0.007; 0.036]**	0.009 [0.003; 0.016]**	0.008 [−0.001; 0.017]	0.005 [0.002; 0.008]**	1.334 [0.865; 2.057]	1.102 [0.956; 1.270]
✓ uPVS crossing points (n)	0.036 [0.015; 0.056]**	0.013 [0.004; 0.022]**	0.013 [0.000; 0.027]*	0.007 [0.003; 0.011]**	1.604 [0.839; 3.067]	1.131 [0.938; 1.364]
✓ uPVS vessel points (n)	0.009 [0.003; 0.015]**	0.004 [0.001; 0.007]**	0.003 [−0.000; 0.007]	0.002 [0.001; 0.003]**	1.131 [0.936; 1.366]	1.044 [0.982; 1.110]
✓ Total length (mm)	0.012 [0.005; 0.019]**	0.006 [0.002; 0.009]**	0.005 [0.000; 0.009]*	0.003 [0.002; 0.004]**	1.183 [0.941; 1.487]	1.030 [0.987; 1.076]
✓ Average thickness (mm)	−0.501 [−1.375; 0.373]	−0.115 [−0.538; 0.308]	−0.153 [−0.800; 0.494]	−0.030 [−0.219; 0.157]	0.045 [0.000; 8325774.645]	0.38 [0.000; 2481.524]

Depicted are the adjusted estimates: adjusted for maternal age, parity, conception mode, BMI, smoking, folic acid supplement use, and fetal sex. Groups stratified for conception mode were not adjusted for conception mode. Groups stratified for fetal sex were not adjusted for fetal sex. Placenta-related complications are specified as PIH or PE and/or FGR, PTB, and SGA; diagnoses may be overlapping. Carnegie stages (n = 23). uPVS, utero-placental vascular skeleton; PIH, pregnancy-induced hypertension; PE, preeclampsia; FGR, fetal growth restriction; PTB, pre-term birth (including both spontaneous and iatrogenic PTB); SGA, small-for-gestational-age. Bold text indicates a statistically significant association. Significance at * $P \leq 0.05$, ** $P \leq 0.01$.

Table 3. Associations between first-trimester trajectories of absolute utero-placental vascular morphologic development and second and third trimester of fetal growth, birthweight percentiles.

Morphologic characteristic	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles	
	N = 198		N = 193		N = 197	
	β [95% CI]		β [95% CI]		β [95% CI]	
	Crude	Adjusted	Crude	Adjusted	Crude	Adjusted
✓ uPVS endpoints (n)	2.729 [0.398; 5.060]*	3.379 [0.945; 5.813]**	6.396 [-1.449; 14.241]	8.802 [0.336; 17.267]*	-0.081 [-1.164; 1.001]	0.020 [-1.159; 1.199]
✓ uPVS bifurcation points (n)	1.677 [0.011; 3.343]*	2.134 [0.370; 3.897]*	4.012 [-1.628; 9.653]	5.697 [-0.438; 11.831]	0.223 [-0.573; 1.019]	0.293 [-0.569; 1.155]
✓ uPVS crossing points (n)	2.608 [0.295; 4.921]*	3.207 [0.744; 5.670]*	5.499 [-2.384; 13.381]	7.454 [-1.155; 16.063]	0.421 [-0.696; 1.537]	0.502 [-0.713; 1.717]
✓ uPVS vessel points (n)	0.666 [-0.055; 1.388]	0.856 [0.094; 1.618]*	1.879 [-0.556; 4.314]	2.654 [0.019; 5.290]*	0.091 [-0.251; 0.434]	0.124 [-0.246; 0.494]
✓ Total length (mm)	0.977 [0.144; 1.810]*	1.205 [0.317; 2.092]**	2.398 [-0.434; 5.230]	3.143 [0.044; 6.243]*	0.025 [-0.028; 0.078]	0.026 [-0.024; 0.0767]
✓ Average thickness (mm)	-39.248 [-141.346; 62.850]	-57.724 [-163.541; 48.093]	-35.841 [-377.592; 305.886]	-81.313 [-446.481; 284.329]	16.363 [-30.929; 63.655]	14.874 [-34.744; 64.4934]

Depicted are the crude and adjusted estimates: adjusted for maternal age, parity, conception mode, BMI, smoking, and preconception initiation of folic acid supplement use. Models for estimated fetal weight (EFW) in the second and third trimesters are additionally adjusted for gestational age and fetal sex. Second-trimester EFW was measured at 22 weeks gestational age (GA) and third-trimester EFW was measured at 32 weeks GA. uPVS, utero-placental vascular skeleton; GA, gestational age. Bold text indicates a statistically significant association. Significance at * $P \leq 0.05$, ** $P \leq 0.01$.

natural conceptions and pregnancies with placenta-related complications only, see Table 4.

Finally, we found negative associations between trajectories of density of vascular branching and EFW at 32 weeks GA (adjusted model: density of endpoints: $\beta = -94.972$, density of bifurcation points: $\beta = -192.601$, all $P < 0.05$) and birth weight percentiles (adjusted model: density of endpoints: $\beta = -20.727$, density of bifurcation points: $\beta = -51.097$, density of crossing points: $\beta = -48.604$, all $P < 0.05$), see Table 5.

After stratification for conception mode, fetal sex, and the occurrence of placenta-related complications, trajectories of density of vascular branching are negatively associated with EFW at 32 weeks GA and birth weight percentiles in natural conceptions only. Further, subgroup analyses showed negative associations between trajectories of density of vascular branching and birth weight percentiles for both males and females, with higher association estimates in females. Finally, density of vascular branching was negatively associated with EFW in the second and third trimester and birth weight percentiles in pregnancies with placenta-related complications. In pregnancies without placenta-related complications negative associations were observed with birth weight percentiles only, see Table 6.

After correction for multiple testing using Bonferroni correction, almost all previously significant associations ($P < 0.05$) between density characteristics and birth weight percentiles remained statistically significant.

Discussion

Principal findings

This study demonstrates associations between absolute vascular morphologic development and density of vascular branching and embryonic and fetal growth using the uPVS as a new imaging marker of first-trimester utero-placental vascular morphologic development. Our findings suggest increased development of absolute vascular morphologic development is favorable for prenatal growth, specifically in natural conceptions and pregnancies with placenta-related complications. Furthermore, negative

associations between first-trimester density of vascular branching and embryonic and fetal growth suggest aberrant vascular development negatively affects prenatal growth, with effects notable in birth weight percentiles. All associations are stronger in pregnancies of a female fetus. Fig. 2C–F provides a graphic summary of our interpretation of this study's findings.

Results in the context of what is known

An important foundation for prenatal growth is established in the periconception period, defined as 14 weeks prior up to 10 weeks after conception (Stegers-Theunissen *et al.*, 2013). During this period, the development of the utero-placental vasculature is mediated by trophoblast invasion, which disrupts spiral artery walls and forms intraluminal plugs that prevent maternal blood flow toward the placenta and developing embryo (Kaufmann *et al.*, 2003; Weiss *et al.*, 2016). The resultant physiological low-oxygen environment promotes angiogenesis which is essential to secure adequate early development of the utero-placental vasculature and safeguards the biological growth potential (Sun *et al.*, 2020).

In cases where trophoblast invasion is defective, remodeling of the spiral arteries is inadequate and trophoblast plugs disintegrate prematurely (Jauniaux *et al.*, 2003; Burton *et al.*, 2009; James *et al.*, 2018; Saghian *et al.*, 2019). Untimely influx of maternal blood flow at high velocity induces oxidative stress and tissue damage to the placental bed, which in the worst case results in miscarriage (Jauniaux *et al.*, 2000, 2003; Burton *et al.*, 2009). In less severe cases, the placental damage results in impaired placental function and can ultimately cause PE and FGR (Burton and Jauniaux, 2004, 2018; Huppertz, 2018; Schoots *et al.*, 2018). We believe the absolute vascular morphologic development reflects spiral artery remodeling and the subsequent (mal-)development of the utero-placental vasculature and is therefore associated with prenatal growth. Furthermore, we postulate an increased density of vascular branching reflects a, conceivably compensatory, abnormal vascular development that follows aberrant spiral artery remodeling, and is accordingly associated with

Table 4. Associations between first-trimester trajectories of absolute utero-placental vascular morphologic development and second and third trimester of fetal growth, birthweight percentiles after stratification for conception method, fetal sex, and occurrence of placenta-related complications.

Conception method Morphologic characteristic	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles	
	Adjusted model		Adjusted model		Adjusted model	
	β [95% CI]		β [95% CI]		β [95% CI]	
	IVF/ICSI (n = 81)	Natural (n = 117)	IVF/ICSI (n = 81)	Natural (n = 112)	IVF/ICSI (n = 77)	Natural (n = 120)
✓ uPVS endpoints (n)	−2.490 [−1.652; 6.631]	4.655 [2.157; 7.152]**	11.090 [−3.506; 25.686]	7.139 [−3.319; 17.578]	0.794 [−0.891; 2.478]	−0.733 [−2.433; 0.967]
✓ uPVS bifurcation points (n)	0.736 [−2.265; 3.736]	3.591 [1.840; 5.342]**	5.834 [−4.820; 16.489]	7.711 [0.287; 15.135]*	0.522 [−0.738; 1.783]	0.087 [−1.140; 1.313]
✓ uPVS crossing points (n)	1.230 [−2.992; 5.451]	4.938 [2.491; 7.385]**	6.840 [−8.314; 21.995]	11.031 −0.698; 21.363]*	0.714 [−1.067; 2.496]	0.278 [−1.460; 2.016]
✓ uPVS vessel points (n)	0.280 [−1.029; 1.589]	1.505 [0.751; 2.259]*	2.8501 [−1.764; 7.464]	3.336 [0.156; 6.517]*	0.263 [−0.282; 0.808]	0.006 [−0.518; 0.530]
✓ Total length (mm)	0.395 [−1.098; 1.888]	1.997 [1.112; 2.882]**	2.681 [−2.631; 7.992]	4.536 [0.738; 8.335]*	0.238 [−0.391; 0.866]	0.202 [−0.424; 0.833]
✓ Average thickness (mm)	−26.855 [−232.422; 178.713]	−70.086 [−172.889; 32.716]	−249.171 [−987.203; 488.861]	−124.407 [−534.432; 285.618]	−42.889 [−129.345; 43.566]	41.509 [−22.021; 105.038]
Fetal sex Morphologic characteristic	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles	
	Adjusted model		Adjusted model		Adjusted model	
	β [95% CI]		β [95% CI]		β [95% CI]	
	Male (n = 99)	Female (n = 97)	Male (n = 97)	Female (n = 94)	Male (n = 102)	Female (n = 95)
✓ uPVS endpoints (n)	2.676 [−0.147; 5.499]	4.258 [0.710; 7.806]*	13.400 [0.705; 26.095]*	4.758 [6.492; 16.007]	0.765 [−1.003; 2.533]	−0.769 [−2.457; 0.920]
✓ uPVS bifurcation points (n)	1.794 [−0.184; 3.772]	1.994 [−0.676; 4.663]	6.879 [−2.072; 15.830]	6.155 [−2.112; 14.423]	0.260 [−0.983; 1.503]	0.331 [−0.944; 1.605]
✓ uPVS crossing points (n)	2.349 [−0.368; 5.066]	3.375 [−0.420; 7.171]	6.959 [−5.477; 19.395]	11.257 [−0.451; 22.966]	0.243 [−1.467; 1.951]	0.798 [−1.042; 2.638]
✓ uPVS vessel points (n)	0.763 [−0.092; 1.617]	1.792 [−0.363; 1.947]	3.204 [−0.649; 7.056]	2.713 [−0.854; 6.280]	0.119 [−0.417; 0.655]	0.135 [−0.413; 0.683]
✓ Total length (mm)	1.024 [0.487; 2.000]*	1.111 [−0.260; 2.482]	3.677 [−0.758; 8.111]	3.215 [−1.043; 7.473]	0.258 [−0.358; 0.874]	0.193 [−0.461; 0.847]
✓ Average thickness (mm)	−45.271 [−158.894; 68.352]	−33.734 [−206.222; 138.755]	−235.663 [−747.501; 276.175]	−93.207 [633.538; 447.124]	10.583 [−59.245; 80.411]	18.265 [−59.741; 96.271]
Placenta-related complications Morphologic characteristic	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles	
	Adjusted model		Adjusted model		Adjusted model	
	β [95% CI]		β [95% CI]		β [95% CI]	
	With (n = 50)	Without (n = 148)	With (n = 45)	Without (n = 148)	With (n = 51)	Without (n = 146)
✓ uPVS endpoints (n)	4.259 [−0.900; 9.418]	1.967 [−0.718; 4.652]	16.548 [−5.147; 38.244]	6.035 [−2.178; 14.247]	−0.209 [−2.830; 2.412]	0.020 [−1.183; 1.224]
✓ uPVS bifurcation points (n)	4.465 [0.810; 8.121]*	0.992 [−0.939; 2.923]	21.757 [6.853; 36.661]**	0.946 [−4.999; 6.891]	0.735 [−1.190; 2.660]	−0.071 [−0.944; 0.802]
✓ uPVS crossing points (n)	6.993 [1.491; 12.496]*	1.663 [−0.986; 4.312]	27.395 [3.117; 51.674]*	1.099 [−7.061; 9.258]	0.743 [−2.196; 3.683]	0.047 [−1.162; 1.256]
✓ uPVS vessel points (n)	1.852 [0.271; 0.432]*	0.375 [−0.458; 1.207]	9.414 [2.984; 15.844]**	0.586 [−1.972; 3.144]	0.348 [−0.479; 1.174]	−0.039 [−0.413; 0.336]
✓ Total length (mm)	2.539 [0.748; 4.331]**	0.582 [−0.395; 1.558]	10.824 [3.123; 18.525]**	0.764 [−2.241; 3.769]	0.339 [−0.641; 1.319]	0.051 [−0.390; 0.492]
✓ Average thickness (mm)	−90.068 [−295.575; 115.439]	−8.738 [−131.113; 113.637]	−515.893 [−1382.235; −350.449]	154.427 [−219.395; 528.249]	−20.486 [−120.205; 79.233]	33.857 [−19.316; 87.030]

Depicted are adjusted estimates: adjusted for maternal age, parity, conception mode, BMI, smoking, and folic acid supplement use. Models for Estimated fetal weight (EFW) in the second and third trimesters are additionally adjusted for gestational age and fetal sex. Groups stratified for conception mode were not adjusted for conception mode. Groups stratified for fetal sex were not adjusted for fetal sex. Second trimester EFW was measured at 22 weeks gestational age (GA) and third trimester EFW was measured at 32 weeks GA. Placenta-related complications are specified as PIH or PE and/or FGR, PTB. uPVS, utero-placental vascular skeleton; GA, gestational age; PIH, pregnancy-induced hypertension; PE, preeclampsia; FGR, fetal growth restriction; PTB, pre-term birth (including both spontaneous and iatrogenic PTB); SGA, small-for-gestational-age. Bold text indicates a statistically significant association. Significance at * $P \leq 0.05$, ** $P \leq 0.01$.

Table 5. Associations between first-trimester trajectories of density of vascular branching (number of uPVS points per cm³ uPVV) and second and third trimesters of fetal growth and birthweight percentiles.

Morphologic characteristic	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles	
	N = 198		N = 193		N = 197	
	β [95% CI]		β [95% CI]		β [95% CI]	
	Crude	Adjusted	Crude	Adjusted	Crude	Adjusted
ln Density of endpoints (n/cm ³)	-10.053	-12.906	-91.573	-94.972	-20.516	-20.727
	[-35.679; 15.574]	[-39.164; 13.351]	[-178.096; -5.050]*	[-185.245; -3.698]*	[-31.972 -9.060]**	[-32.771; -8.683]**
ln Density of bifurcation points (n/cm ³)	-8.957	-10.654	-197.255	-192.601	-49.892	-51.097
	[-56.085; 38.172]	[-59.676; 38.368]	[-354.597]*	[-360.532; -24.670]*	[-70.122; -29.663]**	[-72.257; -29.937]**
ln Density of crossing points (n/cm ³)	28.464	32.211	-175.821	-179.422	-47.356	-48.604
	[-25.889; 82.818]	[-24.756; 89.167]	[-351.500; -0.142]*	[-367.676; 8.832]	[-71.725; -22.988]**	[-74.246; -22.961]**

Depicted are adjusted estimates: adjusted for maternal age, parity, conception mode, BMI, smoking, and folic acid supplement use. Models for estimated fetal weight (EFW) in the second and third trimesters are additionally adjusted for gestational age and fetal sex. Second-trimester EFW was measured at 22 weeks gestational age (GA) and third-trimester EFW was measured at 32 weeks GA. uPVS, utero-placental vascular skeleton; uPVV, utero-placental vascular volume; GA, gestational age. Bold text indicates a statistically significant association. Significance at * $P \leq 0.05$, ** $P \leq 0.01$.

impaired prenatal growth (Borowicz and Reynolds, 2010; Coan et al., 2010; Stenhouse et al., 2019; Huang et al., 2021).

Interestingly, significant associations between uPVS imaging markers and prenatal growth are observed more frequently in natural conceptions than in IVF-ICSI pregnancies. Contrary to our observations, previous studies have shown IVF to be associated with aberrant placental development and consequently an increased risk of FGR (Zhao et al., 2020; Kong et al., 2022). The aberrant placental development in IVF seems predominantly mediated by the hormonal environment and the subsequent epigenetic changes that are the result of ovarian stimulation (Senapati et al., 2018). Another explanation for why we observed no associations in the IVF-ICSI subgroup, may be the fact that there were fewer pregnancies with a placenta-related complication and significantly fewer cases of SGA in this subgroup (3.4% in the IVF-ICSI group vs 12.6% in natural group). We performed an additional analysis in which all pregnancies with a placenta-related complication were excluded. This sensitivity analysis showed similar results to the stratified analysis of the total cohort (data not shown).

Sex differences between males and females contribute to altered prenatal growth and fetal outcomes. While males tend to prioritize physical growth, which ensures a higher birth weight and superior length into adulthood, females tend to invest in placental reserve, making them less vulnerable to placenta-related complications (Eriksson et al., 2010; Meakin et al., 2021). These sex-dependent differences are in accordance with our observations that utero-placental vascular morphologic development seems to be more strongly associated with fetal growth in females.

Using the uPVS, we previously observed decreased absolute vascular morphologic development in pregnancies with placenta-related complications compared to pregnancies without these complications (de Vos et al., 2022). Interestingly, the present study shows embryonic and fetal growth is more positively associated with morphologic development of the utero-placental vasculature in pregnancies with placenta-related complications. These results are in line with our hypothesis and imply that morphologic development of the first-trimester utero-placental vasculature seems to play a more important, perhaps

compensatory, role in regulating prenatal growth in pregnancies with placenta-related complications.

Clinical implications

This prospective observational study suggests that morphologic variations in the utero-placental vascular development play a role in vascular mechanisms involved in embryonic and fetal growth. The clinical applicability of the uPVS as an early imaging marker for detection of FGR and SGA needs further investigation.

Research implications

This study was conducted in a tertiary referral hospital population. Correspondingly, the study population mainly consists of high-risk pregnancies, and over 40% are conceived via IVF-ICSI treatment. Extrapolation of this study's findings to the general population needs to be exercised with consideration. We advise to validate the uPVS in a large population-based cohort.

Strengths and limitations

To our knowledge, this is the first study to investigate associations between *in vivo* morphologic development of first-trimester human utero-placental vasculature and prenatal growth. This study's main strengths are the serial collection of 3D PD ultrasound data in the early first trimester and the use of validated uPVV segmentations (ICC above 0.80 and relative differences of less than 20%) for automated generation of the uPVS (Reijnders et al., 2018). Further, we applied advanced statistical models for adjustment of maternal age, parity, conception mode, BMI, smoking, alcohol consumption, and folic acid supplement use.

According to our study protocol, we used standardized PD gain settings for all participants. As a result, we cannot account for the possible variability, specifically noted for transabdominal ultrasound acquisition, in power Doppler display given the inter-person variability in acoustic properties and placental position, which might have affected variations in the uPVS. However, we believe this issue is mitigated by the use of the transvaginal approach which bypasses the abdominal wall and reduces the distance to the area of interest.

As a result of the observational character of this study, we cannot draw conclusions about causality between first-trimester utero-placental vascular morphologic development and prenatal

Table 6. Associations between first-trimester trajectories of density of vascular branching (number of uPVS points per cm³ uPVV) and second and third trimester of fetal growth, birthweight percentiles after stratification for conception method, fetal sex, and occurrence of placenta-related complications.

Conception method	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles		
	Adjusted model		Adjusted model		Adjusted model		
	β [95% CI]		β [95% CI]		β [95% CI]		
	IVF/ICSI (n = 81)	Natural (n = 117)	IVF/ICSI (n = 81)	Natural (n = 112)	IVF/ICSI (n = 77)	Natural (n = 120)	
Morphologic characteristic							
	In Density of endpoints (n/cm ³)	27.174 [-23.516; 77.864]	-27.897 [-53.137; -2.658]*	-48.744 [231.258; 133.769]	-150.942 [-251.109; -50.776]**	-10.793 [-32.495; -10.910]	-25.548 [-40.676; -10.420]**
	In Density of bifurcation points (n/cm ³)	50.927 [-48.784; 150.641]	-20.889 [-67.410; 15.631]	-95.358 [-455.878; 265.162]	-237.017 [-418.023; -56.012]*	-20.691 [-62.420; 21.038]	-62.802 [-88.457; -37.147]**
	In Density of crossing points (n/cm ³)	85.339 [-23.151; 193.830]	9.319 [-46.081; 64.720]	-110.353; [485.160; 264.454]	-183.700 [-397.124; 29.723]	-7.327 [53.671; 39.017]	-68.725 [100.127; -37.202]**
Fetal sex	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles		
	Adjusted model		Adjusted model		Adjusted model		
	β [95% CI]		β [95% CI]		β [95% CI]		
	Male (n = 99)	Female (n = 97)	Male (n = 97)	Female (n = 94)	Male (n = 102)	Female (n = 95)	
Morphologic characteristics							
	In Density of endpoints (n/cm ³)	-29.912 [-57.883; -1.942]*	27.572 [-13.833; 68.978]	-93.075 [-221.651; 35.501]	-114.636 [-241.440; 12.167]	-17.566 [-35.086; -0.046]*	-27.255 [-45.234; -9.277]**
	In Density of bifurcation points (n/cm ³)	-47.632 [-101.794; 7.070]	46.894 [-26.928; 120.716]	-146.456 [-400.012; 107.100]	-202.562 [-429.245; 24.120]	-48.497 [-81.087; -15.908]**	-57.449 [-88.156; -26.742]**
	In Density of crossing points (n/cm ³)	-34.756 [-99.947; 30.434]	98.098 [15.207; 180.989]*	-208.955 [-498.120; 80.209]	-59.381 [-325.765; 207.003]	-52.002 [-90.449; -13.554]**	-52.682 [-90.611; -14.754]**
Placenta-related complications	Estimated fetal weight second trimester		Estimated fetal weight third trimester		Birth weight percentiles		
	Adjusted model		Adjusted model		Adjusted model		
	β [95% CI]		β [95% CI]		β [95% CI]		
	With (n = 50)	Without (n = 148)	With (n = 45)	Without (n = 148)	With (n = 51)	Without (n = 146)	
Morphologic characteristics							
	In Density of endpoints (n/cm ³)	-63.850 [-107.370; -20.330]**	10.410 [-20.276; 41.089]	-379.193 [-557.827; -200.559]**	37.537 [-56.359; 131.433]	-35.546 [-57.913; -13.178]**	-6.918 [-20.330; 6.494]
	In Density of bifurcation points (n/cm ³)	-102.500 [-185.935; -19.110]*	26.531 [-30.824; 83.885]	-455.436 [-820.730; -90.142]*	3.097 [-173.380; 179.575]	-63.431 [-104.819; -22.042]**	-31.574 [-55.689; -7.458]**
	In Density of crossing points (n/cm ³)	-82.951 [-196.213; 30.311]	55.154 [-8.231; 118.538]	-327.953 [-808.706; 152.800]	-21.290 [-206.993; 164.413]	-63.556 [-114.917; -12.195]*	-28.352 [-56.111; -0.592]*

Depicted are adjusted estimates: adjusted for maternal age, parity, conception mode, BMI, smoking, and folic acid supplement use. Models for estimated fetal weight (EFW) in the second and third trimester are additionally adjusted for gestational age and fetal sex. Groups stratified for conception mode were not adjusted for fetal sex. Second trimester EFW was measured at 22 weeks gestational age (GA) and third trimester EFW was measured at 32 weeks GA. Placenta-related complications are specified as PIH or PE and/or FGR, PTB and SGA; diagnoses may be overlapping. uPVS, utero-placental vascular skeleton; GA, gestational age; PIH, pregnancy-induced hypertension; PE, preeclampsia; FGR, fetal growth restriction; PTB, pre-term birth (including both spontaneous and iatrogenic PTB); SGA, small-for-gestational-age. Bold text indicates a statistically significant association. * $P \leq 0.05$, ** $P \leq 0.01$.

growth. Also, as stratification impacts group size, the results of our subgroup analysis need to be interpreted with consideration. The group of placenta-related complications was highly heterogeneous. We were not able to distinguish between spontaneous and iatrogenic PTB. Because of the size of the study population, we could not analyze SGA, which consists of constitutionally small infants, and FGR, which includes infants that fail to reach their biological growth potential, separately (Audette and Kingdom, 2018). However, we argue that exclusive stratification for FGR-fetuses would likely increase the strength of the associations between vascular morphologic development and prenatal growth.

We performed multiple statistical analyses, mainly to substantiate whether all imaging markers of utero-placental vascular development are associated with embryonic and fetal growth and birth weight percentiles. Previously, uPVV and uPVS characteristics were highly correlated and are inversely correlated with the density of vascular branching (de Vos et al., 2022). Therefore, the positive correlations between the effect estimates of uPVV and uPVS characteristics and the inverse correlation with density of vascular branching found in the present study can be interpreted as internal validation and we assume that multiple testing is not a significant issue in our study.

Conclusions

The uPVS, a novel imaging marker of utero-placental vascular morphology, is associated with embryonic and fetal growth. Positive associations between first-trimester uPVS characteristics and prenatal growth suggest increased absolute vascular morphologic development is favorable for prenatal growth. Furthermore, increased density of vascular branching, apparently reflecting aberrant vascular development, is associated with decreased prenatal growth. Our findings suggest morphologic variations in the utero-placental vascular development play a role in the vascular mechanisms involved in embryonic and fetal growth. Further, these associations seem to be affected by conception method, fetal sex and the occurrence of placenta-related complications. The clinical applicability of the uPVS as an imaging marker for early detection of FGR and SGA requires further investigation.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Acknowledgements

We thank all participants of the Virtual Placenta study, without whom this study could not have been performed. Further, we would like to thank the Rotterdam Periconception Cohort (Predict study) team for their support in the processes of recruitment and data collection, which contributed to the results of this study. We especially thank Igna Reijnders for her contribution in the design and validation of uPVV measurements and for the performance of the VOCAL and VR measurements.

Authors' roles

E.S.D.V. was involved in the study design, data analysis, and wrote the first draft of the manuscript. A.H.J.K., R.P.M.S.-T., S.P.W., M.R., B.B.V.R., E.A.P.S., and A.G.M.G.J.M. were involved in the study design and co-writing of the manuscript. R.P.M.S.-T. is the guarantor of this work. All authors have read and approved the final version of the manuscript.

Funding

This research was funded by the Department of Obstetrics and Gynecology of the Erasmus MC, University Medical Center, Rotterdam, The Netherlands.

Conflict of interest

None declared.

References

- Alfirevic Z, Stampalija T, Dowswell T. Fetal and umbilical Doppler ultrasound in high-risk pregnancies. *Cochrane Database Syst Rev* 2017;**6**:CD007529.
- Almond D, Currie J. Killing me softly: the fetal origins hypothesis. *J Econ Perspect* 2011;**25**:153–172.
- Audette MC, Kingdom JC. Screening for fetal growth restriction and placental insufficiency. *Semin Fetal Neonatal Med* 2018;**23**:119–125.
- Borowicz P, Reynolds LP. 'Placental programming': more may still be less. *J Physiol* 2010;**588**:393–393.
- Broere-Brown ZA, Schalekamp-Timmermans S, Hofman A, Jaddoe VVW, Steegers EAP. Fetal sex dependency of maternal vascular adaptation to pregnancy: a prospective population-based cohort study. *BJOG* 2016;**123**:1087–1095.
- Brosens I, Pijnenborg R, Vercruysse L, Romero R. the "great obstetrical syndromes" are associated with disorders of deep placenta-tion. *Am J Obstet Gynecol* 2011;**204**:193–201.
- Burton GJ, Jauniaux E. Placental oxidative stress: from miscarriage to preeclampsia. *J Soc Gynecol Investig* 2004;**11**:342–352.
- Burton GJ, Jauniaux E. Pathophysiology of placental-derived fetal growth restriction. *Am J Obstet Gynecol* 2018;**218**:S745–S761.
- Burton GJ, Woods AW, Jauniaux E, Kingdom JC. Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy. *Placenta* 2009;**30**:473–482.
- Cavoretto PI, Farina A, Miglio R, Zamagni G, Girardelli S, Vanni VS, Morano D, Spinillo S, Sartor F, Candiani M. Prospective longitudinal cohort study of uterine arteries Doppler in singleton pregnancies obtained by IVF/ICSI with oocyte donation or natural conception. *Hum Reprod* 2020;**35**:2428–2438.
- Chen JY, Chen M, Wu XJ, Sun JM, Zhang Y, Li YF, Zhong LY, Yu BL, Luo J, Liu JH. The value of placental vascularization indices for predicting preeclampsia and fetal growth restriction in different stages of gestation: a prospective and longitudinal study. *Placenta* 2022;**122**:1–8.
- Clifton VL. Review: sex and the human placenta: mediating differential strategies of fetal growth and survival. *Placenta* 2010;**31**:S33–S39.
- Coan PM, Vaughan OR, Sekita Y, Finn SL, Burton GJ, Constancia M, Fowden AL. Adaptations in placental phenotype support fetal growth during undernutrition of pregnant mice. *J Physiol* 2010;**588**:527–538.
- Collins SL, Stevenson GN, Noble JA, Impey L. Rapid calculation of standardized placental volume at 11 to 13 weeks and the

- prediction of small for gestational age babies. *Ultrasound Med Biol* 2013;**39**:253–260.
- Croke L. Gestational hypertension and preeclampsia: a practice bulletin from ACOG. *Am Fam Phys* 2019;**100**:649–650.
- de Vos ES, Koning AHJ, Steegers-Theunissen RPM, Willemsen SP, van Rijn BB, Steegers EAP, Mulders A. Assessment of first-trimester utero-placental vascular morphology by 3D power Doppler ultrasound image analysis using a skeletonization algorithm: the Rotterdam Periconception Cohort. *Hum Reprod* 2022;**37**:2532–2545.
- Drukker L, Droste R, Chatelain P, Noble JA, Papageorgiou AT. (2020). Safety indices of ultrasound: adherence to recommendations and awareness during routine obstetric ultrasound scanning. *Ultraschall Med* **41**:138–145.
- Eriksson JG, Kajantie E, Osmond C, Thornburg K, Barker DJP. Boys live dangerously in the womb. *Am J Hum Biol* 2010;**22**:330–335.
- Gordijn SJ, Beune IM, Thilaganathan B, Papageorgiou A, Baschat AA, Baker PN, Silver RM, Wynia K, Ganzevoort W. Consensus definition of fetal growth restriction: a Delphi procedure. *Ultrasound Obstet Gynecol* 2016;**48**:333–339.
- Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study. *Am J Obstet Gynecol* 1985;**151**:333–337.
- Hanson MA, Gluckman PD. Early developmental conditioning of later health and disease: physiology or pathophysiology? *Physiol Rev* 2014;**94**:1027–1076.
- Hoftiezer L, Hukkelhoven CW, Hogeveen M, Straatman HM, van Lingen RA. Defining small-for-gestational-age: prescriptive versus descriptive birthweight standards. *Eur J Pediatr* 2016;**175**:1047–1057.
- Huang Z, Huang S, Song T, Yin Y, Tan C. Placental angiogenesis in mammals: a review of the regulatory effects of signaling pathways and functional nutrients. *Adv Nutr* 2021;**12**:2415–2434.
- Huppertz B. The critical role of abnormal trophoblast development in the etiology of preeclampsia. *Curr Pharm Biotechnol* 2018;**19**:771–780.
- Jaddoe VW, de Jonge LL, Hofman A, Franco OH, Steegers EA, Gaillard R. First trimester fetal growth restriction and cardiovascular risk factors in school age children: population based cohort study. *BMJ* 2014;**348**:g14.
- James JL, Saghian R, Perwick R, Clark AR. Trophoblast plugs: impact on utero-placental haemodynamics and spiral artery remodeling. *Hum Reprod* 2018;**33**:1430–1441.
- Jauniaux E, Hempstock J, Greenwold N, Burton GJ. Trophoblastic oxidative stress in relation to temporal and regional differences in maternal placental blood flow in normal and abnormal early pregnancies. *Am J Pathol* 2003;**162**:115–125.
- Jauniaux E, Watson AL, Hempstock J, Bao YP, Skepper JN, Burton GJ. Onset of maternal arterial blood flow and placental oxidative stress—a possible factor in human early pregnancy failure. *Am J Pathol* 2000;**157**:2111–2122.
- Junaid TO, Brownbill P, Chalmers N, Johnstone ED, Aplin JD. Fetoplacental vascular alterations associated with fetal growth restriction. *Placenta* 2014;**35**:808–815.
- Kaufmann P, Black S, Huppertz B. Endovascular trophoblast invasion: implications for the pathogenesis of intrauterine growth retardation and preeclampsia. *Biol Reprod* 2003;**69**:1–7.
- Kong F, Fu Y, Shi H, Li R, Zhao Y, Wang Y, Qiao J. Placental abnormalities and placenta-related complications following in-vitro fertilization: based on national hospitalized data in China. *Front Endocrinol (Lausanne)* 2022;**13**:924070.
- Lawn JE, Cousens S, Zupan J. Lancet neonatal survival steering T. 4 million neonatal deaths: when? Where? Why? *Lancet* 2005;**365**:891–900.
- Malhotra A, Allison BJ, Castillo-Melendez M, Jenkin G, Polglase GR, Miller SL. Neonatal morbidities of fetal growth restriction: pathophysiology and impact. *Front Endocrinol (Lausanne)* 2019;**10**:55.
- Meakin AS, Cuffe JSM, Darby JRT, Morrison JL, Clifton VL. Let's talk about placental sex, baby: understanding mechanisms that drive female- and male-specific fetal growth and developmental outcomes. *IJMS* 2021;**22**:6386.
- O'Rahilly R, Muller F. Developmental stages in human embryos: revised and new measurements. *Cells Tissues Organs* 2010;**192**:73–84.
- Reijnders IF, Mulders A, Koster MPH, Koning AHJ, Frudiger A, Willemsen SP, Jauniaux E, Burton GJ, Steegers-Theunissen RPM, Steegers EAP. New imaging markers for preconceptual and first-trimester utero-placental vascularization. *Placenta* 2018;**61**:96–102.
- Reijnders IF, Mulders A, Koster MPH, Kropman ATM, de Vos ES, Koning AHJ, Willemsen SP, Rousian M, Steegers EAP, Steegers-Theunissen RPM. First-trimester utero-placental (vascular) development and embryonic and fetal growth: the Rotterdam Periconception Cohort. *Placenta* 2021;**108**:81–90.
- Reus AD, Klop-van der Aa J, Rifouna MS, Koning AH, Exalto N, van der Spek PJ, Steegers EA. Early pregnancy placental bed and fetal vascular volume measurements using 3-D virtual reality. *Ultrasound Med Biol* 2014;**40**:1796–1803.
- Rizzo G, Aiello E, Pietrolucci ME, Arduini D. Are there differences in placental volume and uterine artery doppler in pregnancies resulting from the transfer of fresh versus frozen-thawed embryos through in vitro fertilization. *Reprod Sci* 2016;**23**:1381–1386.
- Rousian M, Koster MPH, Mulders A, Koning AHJ, Steegers-Theunissen RPM, Steegers EAP. Virtual reality imaging techniques in the study of embryonic and early placental health. *Placenta* 2018;**64**:S29–S35.
- Rousian M, Schoenmakers S, Eggink AJ, Gootjes DV, Koning AHJ, Koster MPH, Mulders A, Baart EB, Reiss IKM, Laven JSE et al. Cohort profile update: the Rotterdam Periconceptional Cohort and embryonic and fetal measurements using 3D ultrasound and virtual reality techniques. *Int J Epidemiol* 2021;**50**:1426–1427.
- Saghian R, Bogle G, James JL, Clark AR. Establishment of maternal blood supply to the placenta: insights into plugging, unplugging and trophoblast behaviour from an agent-based model. *Interface Focus* 2019;**9**:20190019.
- Schoots MH, Gordijn SJ, Scherjon SA, van Goor H, Hillebrands JL. Oxidative stress in placental pathology. *Placenta* 2018;**69**:153–161.
- Senapati S, Wang F, Ord T, Coutifaris C, Feng R, Mainigi M. Superovulation alters the expression of endometrial genes critical to tissue remodeling and placentation. *J Assist Reprod Genet* 2018;**35**:1799–1808.
- Steegers-Theunissen RP, Twigt J, Pestinger V, Sinclair KD. The periconceptional period, reproduction and long-term health of offspring: the importance of one-carbon metabolism. *Hum Reprod Update* 2013;**19**:640–655.
- Steegers-Theunissen RP, Verheijden-Paulissen JJ, van Uitert EM, Wildhagen MF, Exalto N, Koning AH, Eggink AJ, Duvekot JJ, Laven JS, Tibboel D et al. Cohort profile: the Rotterdam Periconceptional Cohort (Predict Study). *Int J Epidemiol* 2016;**45**:374–381.
- Stenhouse C, Hogg CO, Ashworth CJ. Associations between fetal size, sex and placental angiogenesis in the pig. *Biol Reprod* 2019;**100**:239–252.
- Sun C, Groom KM, Oyston C, Chamley LW, Clark AR, James JL. The placenta in fetal growth restriction: what is going wrong? *Placenta* 2020;**96**:10–18.

- Velauthar L, Plana MN, Kalidindi M, Zamora J, Thilaganathan B, Illanes SE, Khan KS, Aquilina J, Thangaratinam S. First-trimester uterine artery Doppler and adverse pregnancy outcome: a meta-analysis involving 55,974 women. *Ultrasound Obstet Gynecol* 2014;**43**:500–507.
- Verwoerd-Dikkeboom CM, Koning AH, van der Spek PJ, Exalto N, Steegers EA. Embryonic staging using a 3D virtual reality system. *Hum Reprod* 2008;**23**:1479–1484.
- Vogel JP, Chawanpaiboon S, Moller AB, Watananirun K, Bonet M, Lumbiganon P. The global epidemiology of preterm birth. *Best Pract Res Clin Obstet Gynaecol* 2018;**52**:3–12.
- Weiss G, Sundl M, Glasner A, Huppertz B, Moser G. The trophoblast plug during early pregnancy: a deeper insight. *Histochem Cell Biol* 2016;**146**:749–756.
- Zeve D, Regelman MO, Holzman IR, Rapaport R. Small at birth, but how small? the definition of SGA revisited. *Horm Res Paediatr* 2016;**86**:357–360.
- Zhao L, Sun L, Zheng X, Liu J, Zheng R, Yang R, Wang Y. In vitro fertilization and embryo transfer alter human placental function through trophoblasts in early pregnancy. *Mol Med Rep* 2020; **21**:1897–1909.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Human Reproduction, 2024, 39, 923–935
<https://doi.org/10.1093/humrep/deae056>
Original Article

Embryology

Disposables used cumulatively in routine IVF procedures could display toxicity

Lucie Delaroche ^{1,2,*}, Lucile Besnard², Valérie Ouary³, Fabienne Bazin⁴, and Guy Cassuto⁵

¹Ramsay Santé, Hôpital Privé de Parly 2, Institut Fertilité Maternité Parly 2, Le Chesnay-Rocquencourt, France

²Centre de Biologie Médicale BIOGROUPE, Hôpital Privé de Parly 2, Le Chesnay-Rocquencourt, France

³C2R SAS, Paris, France

⁴HORIANA, Bordeaux, France

⁵Laboratoire Drouot, Paris, France

*Correspondence address. Ramsay Santé, Hôpital Privé de Parly 2, Institut Fertilité Maternité Parly 2, 21 rue Moxouris, 78150 Le Chesnay-Rocquencourt, France.

Tel: +33-1-39-54-44-51; Fax: +33-1-39-54-04-49; E-mail: lucie.delaroche@biogroup.fr  <https://orcid.org/0000-0002-4768-3051>

ABSTRACT

STUDY QUESTION: Is there a cumulative toxicity of disposables used in IVF procedures?

SUMMARY ANSWER: A toxicity may be detected when consumables are used cumulatively, while no toxicity is detected when the same consumables are used and tested individually.

WHAT IS KNOWN ALREADY: Many components of items used in IVF laboratories may impair human embryonic development. Consequently, it is necessary to screen all reagents and materials which could be in contact with gametes and embryos. Toxicity tests, such as the mouse embryo assay and the human sperm motility assay (HsMA), are used by manufacturers as quality control tools to demonstrate the safety of their products. This evaluation is currently individually performed for each single consumable. However, during an IVF cycle, several devices are used sequentially, potentially creating a cumulative exposure to chemical contaminants, which could not be detected for individually tested consumables.

STUDY DESIGN, SIZE, DURATION: The objective of this observational study conducted from March 2021 to October 2022 was to evaluate with the HsMA methodology if there was a cumulative toxicity when several disposables are sequentially used. Fourteen categories of consumables currently used in routine IVF procedures were studied, which included devices used for sperm and oocyte collection (cups, condoms, and oocyte aspiration needles), manipulation (flasks, tubes, tips, pipettes, embryo transfer catheters, syringes, and gloves), culture (dishes), and storage (straws).

PARTICIPANTS/MATERIALS, SETTING, METHODS: After obtaining patient consent, the surplus semen assessed as having normal parameters according to the [World Health Organization 2010](#) criteria were used to perform the HsMAs. First, each consumable was tested individually. Then, associations of three, four, and five consumables, previously validated as non-toxic when tested individually, were analyzed. HsMAs were conducted three times to ensure reproducibility, with a defined toxicity threshold of a sperm motility index (SMI) below 0.85 in at least two of three tests.

MAIN RESULTS AND THE ROLE OF CHANCE: Thirty-six references of disposables were first individually tested across 53 lots. Forty-nine (92%) demonstrated compliance. However, four (8%) devices revealed toxicity: one lot of 1 ml syringes, two lots of sperm cups, and one lot of 25 cm² flasks. These four references were excluded from the IVF routine procedures. A total of 48 combinations of consumables were assessed, involving 41 lots from 32 references that were previously individually tested. Among the evaluated combinations, 17 out of 48 (35%) associations exhibited toxicity with a SMI below 0.85 for two of the three tests ($n = 8$) or all the three tests ($n = 9$). Notably, three out of 17 (18%) of the three-consumable associations, five out of 16 (31%) of the four-consumable associations, and nine out of 15 (60%) of the five-consumable associations were found not compliant. The toxicity did not originate from a single consumable, because only consumables that were individually pre-validated as non-toxic were included in the combinations, but the toxicity had a cumulative origin. The risk of cumulative toxicity increased with the number of consumables included in the association (Cochran–Mantel–Haenszel statistic, $P = 0.013$).

LIMITATIONS, REASONS FOR CAUTION: The high proportion of non-compliant combinations of disposables can be attributed directly to the extreme rigorous extraction conditions employed during the tests, which could deviate from the conditions encountered in routine clinical use. Also, the methodology employed in the HsMAs (e.g. toxicity extraction duration, sperm concentrations, and protein supplementation of the medium) can influence the sensitivity of the tests.

WIDER IMPLICATIONS OF THE FINDINGS: This study highlights the significance of performing toxicity testing on devices before introducing them into clinical practice. Disposables should be tested individually to detect immediate toxicities and also in combination. Our results advocate rationalizing the number of consumables used in each IVF procedure and re-evaluating the use of glass consumables.

STUDY FUNDING/COMPETING INTEREST(S): This study received fundings from GCS Ramsay Santé pour l'Enseignement et la Recherche (Paris, France) and the Centre de Biologie Médicale BIOGROUPE (Le Chesnay-Rocquencourt, France). The authors declare

Received: October 20, 2023. Revised: January 24, 2024. Editorial decision: February 7, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

that they have no conflict of interest that could be perceived as prejudicing the impartiality of the reported research.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: human sperm motility assay / IVF / disposable devices / plastics / cumulative / toxicity

Introduction

Plastics have become an integral part of modern society, revolutionizing various industries, and fundamentally changing our way of life. Plastics are raw polymers (base resins) typically enhanced with additives, colorants, plasticizers, and fillers (either soluble or insoluble within the polymer matrix). According to the most common definition, plastics are materials made from synthetic polymers. This definition excludes natural plastics that are derived from biological sources, such as natural rubber initially derived from latex, a milky colloid produced by some plants in tropical countries, primarily *Hevea brasiliensis*. However, a broader interpretation of plastics encompasses both natural and synthetic plastics derived from fossil or renewable resources, better reflecting the diversity of plastics in contemporary usage. Plastics possess remarkable properties such as durability, flexibility, and resistance to degradation. These characteristics have made them incredibly versatile and cost-effective, leading to their widespread use in packaging, construction, electronics, healthcare, and numerous other sectors. The production and consumption of plastics have witnessed an exponential growth over the years, with global production surpassing 460 megatons annually in 2019 and this is on track to triple by 2060 (Landrigan *et al.*, 2023).

However, the dark side of plastics begins to emerge as their environmental impact becomes increasingly apparent. Plastics release toxic chemicals, including additives and residual monomers, both during their use and disposal, leading to critical concerns for human health (Landrigan *et al.*, 2023; Meng *et al.*, 2023). Studies in both animals and humans have revealed associations between plastic exposure and adverse reproductive hormone levels (Yang *et al.*, 2011; Sedha *et al.*, 2021; Panagopoulos *et al.*, 2023). Furthermore, evidence suggests that prenatal exposure to plastic-associated chemicals may have long-term consequences for offspring, including developmental abnormalities and increased susceptibility to certain diseases later in life (Seymore *et al.*, 2022).

The efficacy of ART relies heavily on maintaining stringent control over environmental conditions to ensure the optimal development of *in vitro* embryos (Wale and Gardner, 2016; Cairo Consensus Group, 2020). However, the combined toxicity resulting from various environmental pollutants, chemicals, and physical factors presents a significant threat to the viability and health of developing embryos (Scarica *et al.*, 2022). The use of devices during IVF procedures has raised concerns regarding their potential embryotoxic effects. They have the potential to introduce a complex mixture of associated substances into the culture media through leaching processes (McDonald *et al.*, 2008; Olivieri *et al.*, 2012; Gatimel *et al.*, 2016; Togola *et al.*, 2021). Additionally, IVF plastic devices can release volatile organic compounds which accumulate in the incubators and can alter the culture conditions (Cohen *et al.*, 1997; Mortimer *et al.*, 2018; Vardhan and Shukla, 2018; Chu *et al.*, 2020). All these chemicals contaminants can adversely affect gametes and embryos, both in immediate and long-term ways (Fernández-Gonzalez *et al.*, 2004; Alonso-Magdalena *et al.*, 2010; Chianese *et al.*, 2018; Kouakou *et al.*, 2023) even if human embryos have plasticity and self-correction abilities (Ramos-Ibeas *et al.*, 2019; Coticchio *et al.*, 2021).

Consequently, ESHRE and the American Society for Reproductive Medicine (ASRM) suggest that all medical equipment that may come in contact with gametes and embryos should undergo appropriate quality control tests to prevent the release of embryotoxins (ESHRE Guideline Group on Good Practice in IVF Labs *et al.*, 2016; Practice Committees of the American Society for Reproductive Medicine (ASRM) and the Society for Reproductive Biologists and Technologists (SRBT). Electronic address: asrm@asrm.org, 2022). Ensuring the safety of ART requires the development of risk assessment methods that provide functionally relevant assays for quality control testing. The mouse embryo assay (MEA) and the human sperm motility assay (HSMA) are the two primary tests used to evaluate the suitability of reagents and materials for ART, independent of clinical factors (Ackerman *et al.*, 1985; Critchlow *et al.*, 1989; Claassens *et al.*, 2000; Nijs *et al.*, 2009; Long *et al.*, 2018). Currently, these tests are individually performed for each consumable, ensuring that each consumable has a certificate of conformity. However, during an IVF procedure, multiple devices are used consecutively, potentially leading to cumulative exposure to chemical contaminants that may go undetected when assessing for single consumables. To date, the potential cumulative toxic effects of combinations of consumables have not been studied. The aim of this study was to evaluate the cumulative toxicity resulting from the use of consumables used during an IVF procedure.

Materials and methods

Study design

This study was conducted from March 2021 to October 2022 at the Fertility and Maternity Institute, Private Hospital of Parly 2, Le Chesnay-Rocquencourt, France, and was approved by the French South-West and Overseas Personal Protection Committee 4 on 11 December 2020 (Protocol Number ID-RCB: 2020-A01972-37).

Semen samples

Informed consent was obtained from patients aged between 18 and 65 years who underwent diagnostic sperm tests, granting permission to use any surplus semen samples. Semen samples analyzed in compliance with the Björndahl guidelines and assessed as having normal parameters according to the World Health Organization (WHO) 2010 were included in the study (World Health Organization, 2010; Björndahl *et al.*, 2016). These criteria encompassed a minimum concentration of 15 million/ml or a total number of spermatozoa in the ejaculate of at least 39 million, a progressive motility of at least 32% or a total motility of at least 40%. Within 1 h of collection, the samples underwent processing using a discontinuous density gradient technique with two layers of 40/80% Isolate[®] (Irvine Scientific, Irvine, CA, USA) for 20 min at 350 g, followed by washing in MHM-C[®] (Irvine Scientific, Irvine, CA, USA) for 10 min at 200 g. After selection, samples with a total number of progressive motile spermatozoa recovered from ejaculate exceeding 10 million and a progressive motility of at least 70% were used to perform the HSMAs. The final sperm preparation had an adjusted

concentration of 4 million/ml motile progressive spermatozoa. The same sperm preparation could be used several times to test individual consumables and/or combinations of consumables.

Disposables

The study included 14 categories of disposables used in routine IVF procedures, provided by major companies in the field. They were devices used for sperm and oocyte collection (cups, condoms, and oocyte aspiration needles), manipulation (flasks, tubes, tips, pipettes, embryo transfer catheters, syringes, and gloves), culture (dishes), and storage (straws). Some of the included consumables lacked a certificate of conformity or CE (conformity with European health, safety, and environmental protection standards) marking from manufacturers. Mainly used in andrology procedures, these consumables could, however, have been used as a temporary alternative in IVF procedures in the event of stock shortages, subject to prior validation by HSMAs, particularly in the context of the supply difficulties caused by the coronavirus disease 2019 crisis.

Table 1 provides a comprehensive list of the consumable categories, anonymized references (a reference is a model of a consumable that has a specific trade name known under a specific code), and lot numbers, published HSMA/MEA certificates, and CE marking.

HSMA

All experiments were conducted at a room temperature of 20–24°C (Meintjes, 2017). Initially, single consumables were screened individually for toxicity. Following this, combinations of three, four, or five consumables were tested. A consumable could be included in combinations only if it demonstrated no toxicity during individual testing. Each individual item or combination of disposables was tested three times to ensure the reproducibility of the results. As the volume of one semen preparation was not sufficient to test three items or three associations of items at the same time, HSMAs were conducted using one distinct semen sample for each item or association of items separately. Also, the number of patients who had given their consent or the number of semen samples that met the inclusion criteria was insufficient to conduct tests on all three items or associations on the same day. Consequently, the experiments were performed on different days.

Extraction of toxicity

All materials and supplies used to perform the HSMAs were common to the test and the control groups or were known to be non-toxic.

The extraction protocols for assessing the toxicity of the consumables were developed using established methodologies as described in previous studies (Nijs et al., 2009; Meintjes, 2017) (Table 2). For sperm cups and gloves, HSMAs were also performed with alternative extraction conditions that are indicated in italics.

To test an individual consumable, the test item was either loaded with the sperm preparation or placed into the sperm suspension at room temperature for a specified duration (Fig. 1A). The exposure duration was determined to ensure the consumables were tested under rigorous/stringent extraction conditions, with a minimum exposure time of 30 min, except for gloves. For tubes and dishes, the lids and caps were also tested for a minimum of 30 min. The volume of the sperm preparation used to test consumables depended on the category of consumables; it had to be sufficient to avoid evaporation during incubation, but just enough to avoid diluting toxicity. After the extraction period,

the sperm preparation was returned to the test tube of the same reference (same trade name, lot number, and expiry date) as the control tube. The control tube contained the other half of the sperm preparation which remained unchanged. Both tubes were incubated at room temperature for 24 h.

To evaluate combinations of consumables, the sperm preparation was successively brought into contact with each of the consumables included. The duration and extraction method of each consumable included in the combinations were identical to those used for individual testing, with the exception that the 24-h exposure period was reduced to 30 min owing to time constraints and to standardize all exposure durations. Likewise, the volume of the sperm preparation used to test a combination of consumables was adjusted (1 ml minimum and 3 ml maximum) to ensure a sufficient volume in the final test tube.

For example, to test an association of three consumables, such as a pasteurette, a tube, and a dish, 3 ml of the sperm preparation was first aspirated for 30 min into the pasteurette. It was then flushed into the tube, which was initially placed upright to test the bottom of the tube for 30 min, then turned upside down to test the cap for 30 min. Subsequently, the sperm preparation was transferred to the lid of a dish and incubated for 30 min before being placed in the bottom of the same dish for an additional 30 min. Following the successive extraction of the three consumables, the sperm preparation was returned to the test tube and incubated for 24 h (Fig. 1B).

The majority of consumables combinations were evaluated independently, with the exception of seven combinations that underwent interconnected testing (Fig. 1C). To elaborate, the same sperm preparation was initially used to assess the toxicity of the combination of three consumables. Subsequently, the sperm preparation was divided, with one portion allocated to the three-consumable test tube, which was then incubated for 24 h. Meanwhile, the remaining portion of the sperm preparation was used to continue the HSMAs experiments with the addition of a fourth consumable. Following this, the sperm preparation was again divided, with one portion allocated to the four-consumable test tube, which was incubated for 24 h. The remaining portion was then used to continue the experiments with the addition of a fifth consumable. At the end of the experiments and after 24 h of incubation, the progressive motilities of the three-, four-, and five-consumable test tubes were assessed and compared to the progressive motility of the unique control.

Sperm motility assessment

Following 24 h of incubation at room temperature, the progressive motility of both test and control tubes for each replicate was independently assessed by two operators at 5-min intervals. To ensure the validity of the test, the control tube had to exhibit progressive motility in more than 50% of the spermatozoa.

The sperm motility index (SMI) was calculated by dividing the mean progressive motility of the test sample by the mean progressive motility of the control after the 24-h incubation period. An SMI value exceeding 0.85, observed in at least two out of three tests, confirmed the absence of toxicity (Critchlow et al., 1989).

The reproducibility of motility assessments between operators was verified through monthly internal quality controls and bi-annual external quality controls.

Table 1. Details of 53 disposable items individually assessed for toxicity using the human sperm motility assay method.

Consumable category	Consumable name	Reference	Lot number	Composition	Use	MEA certificate	HSMA certificate	CE marking	ID graphic
Condom	Spermicid free condom	A	A1	PU	Andrology and IVF	No	No	No	1
Cryotube	Cryotube vials	A	A1	PP (included cap)	Andrology	No	No	No	2
Catheter	Soft Catheter	A	A1	PE	IVF	Yes	No	Yes	3
		B	B1	PP	IVF	Yes	No	Yes	4
Dish	100 mm dish	A	A1 A2	PS Non-treated surface	IVF	Yes	Yes	Yes	5 6
	35 mm dish	A	A1	PS Non-treated surface	IVF	Yes	Yes	Yes	7
	4 well dish	A	A1 A2 A3	PS Non-treated surface	IVF	Yes	Yes	Yes	8 9 10
	60 mm dish	A	A1	PS Non-treated surface	IVF	Yes	Yes	Yes	11
Flask	ICSI dish	A	A1	PS Non-treated surface	IVF	Yes	Yes	Yes	12
	25 cm ² flask	A	A1 A2	PS Treated surface (body) Phenol-formaldehyde (cap)	Andrology and IVF	No	No	No	14 15
		B	B1	PS Treated surface (body) PE (cap)	Andrology and IVF	No	No	No	13
Follicle aspiration needle	Follicule aspiration set	A	A1	Stainless steel (all metal parts) Polyolefin based TPE (tubing) Silicone rubber (plug) MABS polymer (luer connectors) Epoxy resin (glue)	IVF	Yes	No	Yes	16
		B	B1	Steel (needle) PVC (connection tube) ABS polymer TheroPlastic polymer Crosslinked polyolefin Silicone	IVF	Yes	No	Yes	17
		C	C1	Stainless steel (needle) PU (puncture tubing and vacuum tubing)	IVF	Yes	No	Yes	18

(continued)

Table 1. Continued

Consumable category	Consumable name	Reference	Lot number	Composition	Use	MEA certificate	HSMA certificate	CE marking	ID graphic
Glove	Examination gloves	A	A1	Nitrile	IVF	No	No	Yes	19
	Surgical gloves	B	B1	Natural rubber latex with nitrile polymer	IVF	No	No	Yes	20
Pasteurette	Pasteur pipette 3 ml	A	A1	PE	IVF	Yes	Yes	Yes	22
			A2						23
		B	B1	PE	IVF	Yes	Yes	Yes	21
		C	C1	PE	Andrology	No	No	No	24
Serological pipette	Serological pipette 10 ml	A	A1	PS	Andrology	No	No	No	25
	Serological pipette 5 ml	A	A1	PS	IVF	Yes	Yes	Yes	26
Sperm cup	Sperm collection cup	A	A1	PP (cap and cup) Non-treated surface	Andrology and IVF	Yes	Yes	Yes	27
			A2						28
			A3						29
		B	B1	PS (body) PE (cap)	Andrology and IVF	Yes	Yes	Yes	30
Sperm straw	Sperm straw	A	B2	Ionomer resin. Braided cap (composition not disclosed by the manufacturer)	IVF	Yes	No	Yes	31
			B3						32
			A1						33
			A2						34
Syringe	Syringe 1 ml	A	A1	PP (body and piston) PI (seal)	IVF	Yes	Yes	Yes	37
		B	B1	PP (body and piston) Silicone (seal)	IVF	Yes	No	Yes	35
		C	C1	PP (body) PE (piston) No seal	IVF	No	No	Yes	36
Tips	Cones 1000 µl	A	A1	PP	Andrology and IVF	No	No	No	39
		B	B1	PP	Andrology	No	No	No	38
		C	C1	PP	Andrology and IVF	No	No	Yes	40
	Cones 0.5–200 µl	A	A1	PP	Andrology and IVF	No	No	No	41

(continued)

Table 1. Continued

Consumable category	Consumable name	Reference	Lot number	Composition	Use	MEA certificate	HSMA certificate	CE marking	ID graphic
Tube	Andrology tube 5 ml	A	A1	PS Non-treated surface	IVF	Yes	Yes	Yes	44
		B	B1	PS	Andrology and IVF	No	No	42	
			B2	Non-treated surface				43	
	Centrifuge tube 15 ml	A	A1	PP Non-treated surface	Andrology and IVF	Yes	Yes	45	
			A2					46	
		B	B1	PS Non-treated surface	Andrology and IVF	Yes	Yes	50	
	OPU tube	A	A1	PS Non-treated surface	IVF	Yes	Yes	Yes	47
			A2						48
			A3						49
	Vitrification straw	B	B1	PP	IVF	No	No	No	51
A		A1	Ionomer resin	IVF	Yes	No	Yes	52	
B		B1	Ionomer resin	IVF	Yes	No	Yes	53	

CE: conformity with European health, safety, and environmental protection standards; OPU: Oocyte pickup; HSMA: human sperm motility assay; MEA: mouse embryo assay; PE: polyethylene; PI: polyisoprene; PP: polypropylene; PS: polystyrene; PU: polyurethane.

Treated surface: surface coating involves altering the surface characteristics of a product to enhance its ability to either adhere to or repel liquids. During the treatment process, the surface undergoes a high-voltage electric shock, which induces a change in the surface energy of the material. This modification imparts adhesive properties, allowing a drop of water placed on the treated surface to spread out and form a flatter shape. In contrast, a drop of water on an untreated surface may retain a more spherical shape but can easily detach when the dish is moved.

Table 2. Summary of human sperm motility assay extraction protocols, by consumable category.

Consumable category		Exposure method	Exposure duration	Volume of sperm suspension
Catheter		Filled with sperm suspension	30 min	0.4 ml
Condom				2 ml
Cryotube			24 h including 30 min in the cap	1 ml
Cup		Filled with sperm suspension	24 h including 30 min in the cap (30 min excluding the cap)	1 ml
Dish	35, 60, 100 mm and ICSI	Surface contact with sperm suspension	30 min in the bottom then 30 min in the cap	2 ml
	4 wells	2 wells filled with sperm suspension	24 h	2 × 0.5 ml
Flask		Filled with sperm suspension	24 h including 30 min in the cap	1 ml
Follicle aspiration needle		10 × flushing with sperm suspension then filled with sperm suspension	30 min	
Gloves		0.5 × 0.5 cm sample in sperm suspension	5 min (1, 10, 30 min)	
Pasteurette		10 × flushing with sperm suspension then filled with sperm suspension	30 min	
Serological pipette		Filled with sperm suspension		1.5 ml
Sperm straw				0.4 ml
Syringe				1 ml
Tip	1000 μl	10 × flushing with sperm suspension then filled with sperm suspension		0.2 ml
	0.5–200 μl			
Tube	5 ml	Filled with sperm suspension	24 h including 30 min in the cap	1 ml
	14 or 15 ml		30 min in the bottom then 30 min in the cap	
Vitrification straw		Surface contact with sperm suspension	30 min	

Alternative extraction conditions are indicated in italics.

Controls

In order to ensure the reliability and robustness of our methodologies, we conducted preliminary tests to assess whether the sequential transfer of the sperm preparation from one consumable to another during the association tests could lead to a mechanical reduction in sperm motility, potentially introducing bias to our results. A series of control experiments was conducted, involving successive mechanical aspirations of sperm preparations. The objective of this successive mechanical pipetting was to replicate the mechanical transfer of the sperm preparation between consumables. The initial sperm preparation was divided into separate tubes, and in each tube, the sperm preparation was successively aspirated either 10, 20, 30, or 40 times. Subsequently, the tubes were incubated at room temperature for 24 h, and the progressive motility of the sperm in each tube was compared to the control sample, which had not undergone any aspiration. In total, six controls were performed with three sperm suspensions, consisting of two 10-pipetting controls, two 20-pipetting controls, one 30-pipetting control, and one 40-pipetting control.

Additionally, the potential impact of the final volume (0.2 or 0.5 ml) on sperm motility was assessed. Similarly, a sperm preparation was divided into two tubes with final volumes of 0.2 and 0.5 ml that were then incubated at room temperature for 24 h.

Statistical analysis

For each HSMA, the homogeneity of the three sperm samples used was tested as follows: the SD was calculated for each set of three SMIs. Outliers were determined with the Tukey test, which is a nonparametric outlier detection method calculated by creating a ‘fence’ boundary a distance of 1.5 interquartile range beyond the 1st and 3rd quartiles. Any data beyond these fences were considered to be outliers and were excluded from the analyses.

The distribution of SMI SD was compared between compliant HSMA (tests with a SMI > 0.85) and non-compliant HSMA using the Wilcoxon test, for single consumables and associations of consumables.

The trend between number of consumables used in the associations and non-compliance with the HSMA test was assessed using the Cochran–Mantel–Haenszel statistic.

Results

Semen samples

A total of 826 patients provided their consent to participate in the study. Based on the inclusion criteria, only 239 semen samples were eligible for inclusion and subsequently used for testing individual consumables and/or combinations of consumables. The

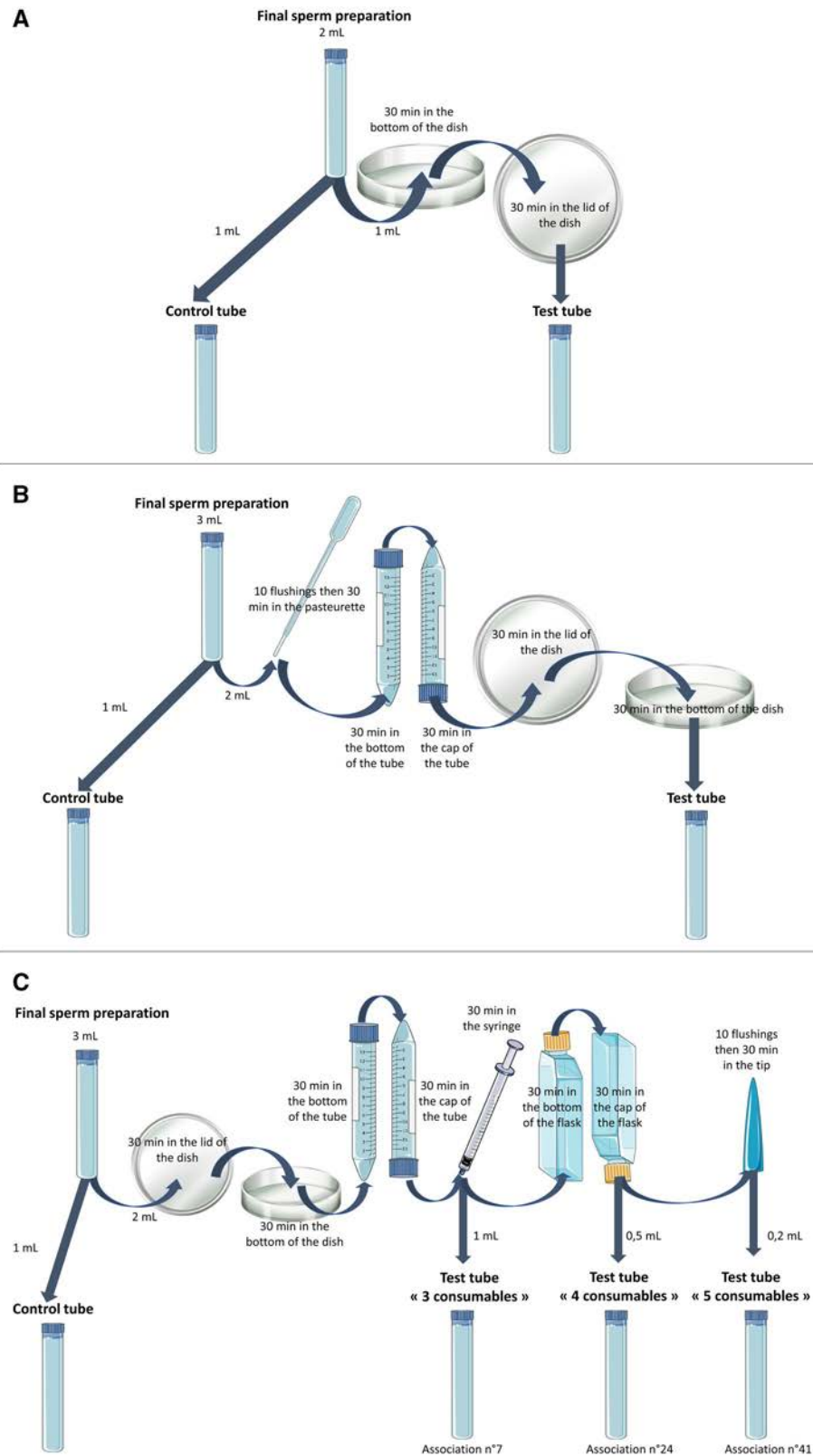


Figure 1. Toxicity extraction methodologies of the human sperm motility assays. (A) Individually tested consumable. Example of a dish. (B) Association of three consumables. Example of an association involving a pasteurette, a tube, and a dish. (C) Association of consumables. Interconnexion of three, four, and five consumables. Example of an associations involving a dish, a tube, a syringe, a flask, and a tip.

other 587 samples did not have sufficient sperm parameters or failed the semen collection. The mean sperm parameters are presented in the [Supplementary Table S1](#).

Controls

No significant differences in progressive motilities were observed among the control groups of four pipetting conditions and of two

volume conditions after 24 h of incubation (Supplementary Fig. S1).

HSMA

The Tukey test considered SMIs to be homogeneous for eight controls, 53 single consumables and 48 associations of consumables. Seven outliers were excluded, two HSMA on single consumables and five on associations of consumables (Supplementary Fig. S2).

HSMA on single consumables

In total, 36 references of consumables were tested across 53 lots.

Forty-nine (92%) items demonstrated compliance with an SMI above 0.85 in at least two out of three tests (Fig. 2). However, four (8%) consumables exhibited toxicity in two out of three tests: the lot A1 of 25 cm² flasks, the lot C1 of 1 ml syringes, and the lots A1 and A2 of sperm collection cups. Consequently, these four references were excluded from our routine andrology and IVF procedures. Of note, two out of these four non-compliant consumables (25cm² flasks and 1 ml syringes) were marketed without MEA or HSMA certificates.

For further evaluation, HSMA experiments were repeated on sperm collection cups using alternative extraction conditions. Specifically, the duration of the toxicity extraction for the sperm collection cups was adjusted to 30 min without testing the cap, as opposed to the previous protocol of 24 h of extraction, including 30 min with the cap. This alternative protocol corresponds to the methodology described on the certificate of conformity used by the manufacturer to market the two lots (A1 and A2) which demonstrated toxicity in the initial 24-h extraction conditions. An additional lot (A3) from the same brand and two lots from another brand (B2 and B3) were also implemented and subjected to both the original and modified extraction conditions (Fig. 3A). Remarkably, none of the lots exhibited toxicity when tested under the 30-min extraction condition and without testing the cap.

Moreover, the surgical gloves from lot B1 underwent testing using four different extraction conditions: 1, 5, 10, and 30 min. It was observed that the gloves exhibited toxicity when the extraction duration exceeded 10 min (Fig. 3B).

HSMA on combinations of consumables

A total of 48 combinations of consumables were assessed, consisting of 17 combinations of three consumables, 16 combinations of four consumables, and 15 combinations of five consumables. These combinations involved 41 lots from 32 references that had previously been found compliant when individually tested (Table 3).

Among the evaluated combinations, 17 out of 48 (35%) associations exhibited toxicity with a SMI below 0.85 for two of the three tests ($n=8$) or all three tests ($n=9$). Notably, three out of 17 (18%) of the three-consumable associations, five out of 16 (31%) of the four-consumable associations, and nine out of 15 (60%) of the five-consumable associations were found not to be compliant (Fig. 4). Thus, the risk of toxicity increased with the number of consumables included in the association (Cochran–Mantel–Haenszel statistic, $P=0.013$).

Among the 74 items included in the 17 non-compliant associations, dishes were highly frequent (35.4%), especially the lot A1 for 35 mm dishes (9.5% frequency), the lot A1 for ICSI dishes (8.1% frequency), and the lot A2 for 100 mm dishes (8.1% frequency). Additionally, tubes were found accounting for 29.8% of 74 items, while pasteurettes were present at a 14.9% frequency.

For the seven associations that were consecutively tested with the addition of a fourth and/or a fifth consumable, we observed that some combinations which did not show toxicity with three or four consumables became toxic when a fourth or a fifth consumable was added. Also, the associations which were toxic with four consumables remained toxic when a fifth consumable was added. On the other hand, some associations did not become toxic with addition of new consumables (Table 4).

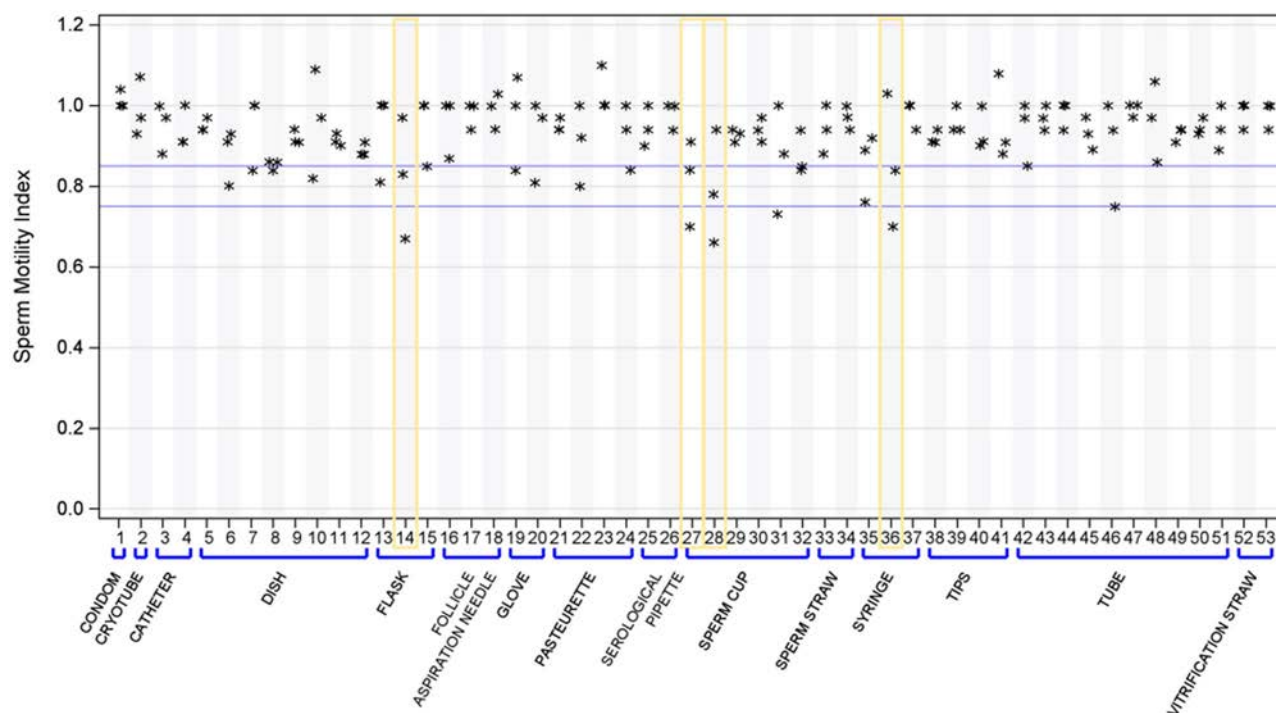


Figure 2. Sperm motility index results individually tested consumables.

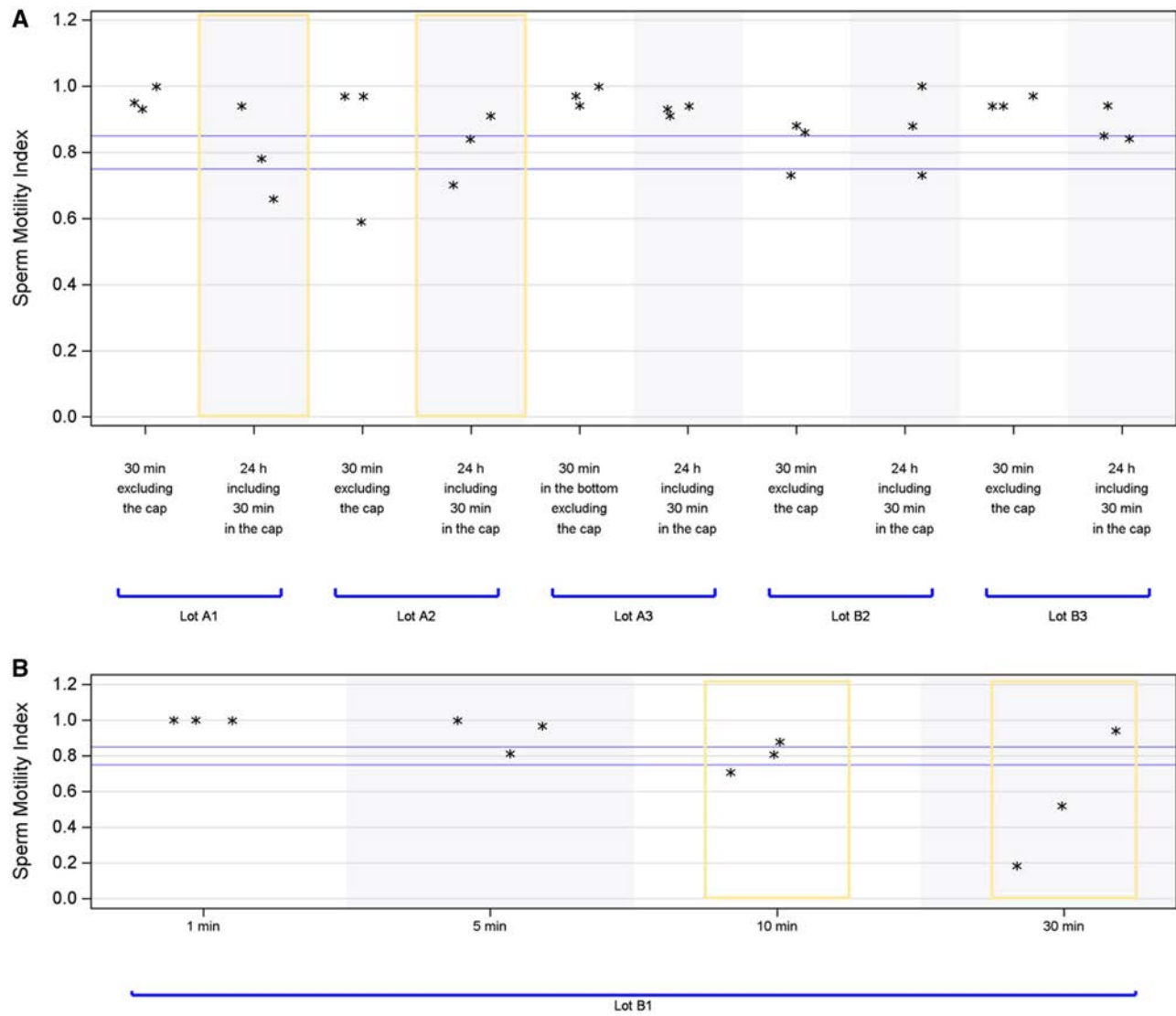


Figure 3. Sperm motility index results according to the toxicity extraction conditions. (A) Sperm cups. (B) Gloves.

In total, the variability between the three SMI of a same test was significantly higher for non-compliant tests than for compliant tests, both for consumables alone (Wilcoxon, $P=0.002$) and for associations of consumables (Wilcoxon, $P=0.002$). These differences suggest that sperm samples may have different degrees of sensitivity to the toxic effects of consumables.

Discussion

This study aimed to assess the potential cumulative toxicity arising from combinations of devices used in IVF procedures. Notably, it was found that 17 out of 48 (35%) associations of three, four, or five disposables, tested using the HSMA methodology, exhibited toxicity toward spermatozoa. This toxicity had a cumulative origin: it did not originate from one single consumable because only consumables that were individually pre-validated as non-toxic were included in the combinations, and increased with the number of consumables included in the association. However, the precise source of the toxicity remained unknown. It could potentially stem from various factors, including the materials' composition, production process, handling and packaging, sterilization, or transport process (Nijs et al., 2009). All these factors could lead to an accumulation of toxic molecules

leaching from disposable plasticware into the sperm preparation as it successively came into contact with plastic consumables (McDonald et al., 2008). These molecules could reduce or stop the cell metabolism responsible for motility in human spermatozoa (Nijs et al., 2009). The release of toxic molecules could have a fairly immediate effect on sperm motility as observed in this study, but could be more insidious in case of embryo culture, as it has already been demonstrated, reducing fertilization and blastulation rates or influencing negatively the post-implantation development and the health of newborns (Takai et al., 2001; Susiarjo et al., 2015; Kouakou et al., 2023). The susceptibility to toxicity may differ among biological samples.

Furthermore, we identified that four out of 53 (8%) devices exhibited toxicity when individually tested. This percentage of toxic devices was lower than that reported in a previous study where 13 out of 36 (36%) products were found to be toxic (Nijs et al., 2009). It is important to note that at the time of the Nijs et al. (2009) study, the consumables did not possess certificates of conformity. This disparity is reassuring, as it suggests an improvement in the safety of consumables over time. Since the recommendations by the ESHRE and ASRM societies, the majority of IVF devices are now subjected to validation for the absence of toxicity before being introduced to the market (ESHRE Guideline

Table 3. Details of 48 associations of disposable items assessed for cumulative toxicity using the human sperm motility assay method.

Number of association	Number of consumables	Consumable category	Consumable name	Reference	Lot number	MEA certificate	HSMA certificate	CE marking
1	3	Dish	100 mm dish	A	A2	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	OPU tube	A	A2	Yes	Yes	Yes
2	3	Dish	100 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	B	B1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	B	B1	No	No	No
3	3	Dish	4 well dish	A	A3	Yes	Yes	Yes
			ICSI dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	A	A1	Yes	Yes	Yes
4	3	Dish	100 mm dish	A	A1	Yes	Yes	Yes
			60 mm dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	B	B1	No	No	No
5	3	Dish	35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
6	3	Dish	35 mm dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
7	3	Dish	60 mm dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	B	B1	No	No	No
			Centrifuge tube 15 ml	A	A1	Yes	Yes	Yes
8	3	Dish	ICSI dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
9	3	Dish	ICSI dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	B	B1	Yes	Yes	Yes
		Tube	Centrifuge tube 15 ml	B	B1	Yes	Yes	Yes
10	3	Dish	100 mm dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	B	B1	No	No	No
		Sperm cup	Sperm collection cup	B	B3	Yes	Yes	Yes
11	3	Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
			Centrifuge tube 15 ml	A	A1	Yes	Yes	Yes
		Sperm cup	Sperm collection cup	B	B1	Yes	Yes	Yes
12	3	Dish	ICSI dish	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	A	A1	Yes	Yes	Yes
13	3	Dish	4 well dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
14	3	Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Andrology tube 5 ml	B	B1	No	No	No
			Centrifuge tube 15 ml	B	B1	Yes	Yes	Yes
15	3	Dish	ICSI dish	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
16	3	Tube	OPU tube	B	B1	No	No	No
			Andrology tube 5 ml	A	A1	Yes	Yes	Yes

(continued)

Table 3. Continued

Number of association	Number of consumables	Consumable category	Consumable name	Reference	Lot number	MEA certificate	HSMA certificate	CE marking
17	3	Tube	OPU tube	B	B1	No	No	No
			Centrifuge tube 15 ml	B	B1	Yes	Yes	Yes
		Flask	25 cm ² flask	B	B1	No	No	No
18	4	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			4 well dish	A	A1	Yes	Yes	Yes
			ICSI dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
19	4	Dish	100 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	B	B1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	B	B1	No	No	No
		Catheter	Soft catheter	B	B1	Yes	No	Yes
20	4	Dish	4 well dish	A	A1	Yes	Yes	Yes
			ICSI dish	A	A1	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
21	4	Dish	100 mm dish	A	A1	Yes	Yes	Yes
			4 well dish	A	A2	Yes	Yes	Yes
			ICSI dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
22	4	Dish	35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
23	4	Dish	35 mm dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
		Flask	25 cm ² flask	A	A2	No	No	No
24	4	Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Centrifuge tube 15 ml	B	B1	Yes	Yes	Yes
		Sperm cup	Sperm collection cup	B	B2	Yes	Yes	Yes
		Serological pipette	Serological pipette 5 ml	A	A1	Yes	Yes	Yes
25	4	Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
			Centrifuge tube 15 ml	A	A1	Yes	Yes	Yes
		Sperm cup	Sperm collection cup	B	B1	Yes	Yes	Yes
		Sperm straw	Sperm straw	A	A2	Yes	No	Yes
26	4	Dish	100 mm dish	A	A1	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	B	B1	Yes	Yes	Yes
		Tube	OPU tube	A	A3	Yes	Yes	Yes
27	4	Dish	4 well dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
		Catheter	Soft catheter	A	A1	Yes	No	Yes
28	4	Dish	ICSI dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	OPU tube	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	A	A1	Yes	Yes	Yes
29	4	Dish	ICSI dish	A	A1	Yes	Yes	Yes

(continued)

Table 3. Continued

Number of association	Number of consumables	Consumable category	Consumable name	Reference	Lot number	MEA certificate	HSMA certificate	CE marking
		Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	OPU tube	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
30	4	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A2	Yes	Yes	Yes
31	4	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A1	Yes	Yes	Yes
32	4	Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Centrifuge tube 15 ml	A	A2	Yes	Yes	Yes
		Sperm cup	Sperm collection cup	B	B2	Yes	Yes	Yes
		Serological pipette	Serological pipette 5 ml	A	A1	Yes	Yes	Yes
33	4	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A2	Yes	Yes	Yes
			Andrology tube 5 ml	A	A1	Yes	Yes	Yes
34	5	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			4 well dish	A	A1	Yes	Yes	Yes
			ICSI dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Vitrification straw	Vitrification straw	B	B1	Yes	No	Yes
35	5	Dish	100 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	B	B1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	B	B1	No	No	No
		Catheter	Soft catheter	B	B1	Yes	No	Yes
		Tips	Cones 0.5–200 µl	A	A1	No	No	No
36	5	Dish	100 mm dish	A	A1	Yes	Yes	Yes
			60 mm dish	A	A1	Yes	Yes	Yes
		Tube	OPU tube	B	B1	No	No	No
			Andrology tube 5 ml	B	B1	No	No	No
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
37	5	Dish	100 mm dish	A	A1	Yes	Yes	Yes
			4 well dish	A	A2	Yes	Yes	Yes
			ICSI dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Vitrification straw	Vitrification straw	A	A1	Yes	No	Yes
38	5	Dish	35 mm dish	A	A1	Yes	Yes	Yes
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
		Flask	25 cm ² flask	A	A2	No	No	No
		Tips	Cones 0.5–200 µl	A	A1	No	No	No
39	5	Dish	ICSI dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	B	B1	Yes	Yes	Yes
		Tube	Centrifuge tube 15 ml	B	B1	Yes	Yes	Yes

(continued)

Table 3. Continued

Number of association	Number of consumables	Consumable category	Consumable name	Reference	Lot number	MEA certificate	HSMA certificate	CE marking
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
		Serological pipette	Serological pipette 5 ml	A	A1	Yes	Yes	Yes
40	5	DISH	ICSI dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
			Centrifuge tube 15 ml	A	A1	Yes	Yes	Yes
		Sperm straw	Sperm straw	A	A2	Yes	No	Yes
41	5	Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Andrology tube 5 ml	A	A1	Yes	Yes	Yes
			Centrifuge tube 15 ml	A	A2	Yes	Yes	Yes
		Tips	Cones 1000 µl	C	C1	No	No	Yes
42	5	Condom	Spermicid free condom	A	A1	No	No	No
		Dish	100 mm dish	A	A1	Yes	Yes	Yes
			OPU tube	B	B1	No	No	No
		Tube	Andrology tube 5 ml	B	B1	No	No	No
		Sperm cup	Sperm collection cup	B	B3	Yes	Yes	Yes
43	5	Tips	Cones 1000 µl	A	A1	No	No	No
		Dish	100 mm dish	A	A1	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	B	B1	Yes	Yes	Yes
		Tube	OPU tube	A	A3	Yes	Yes	Yes
44	5	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A2	Yes	Yes	Yes
45	5	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Pasteurette	Pasteur pipette 3 ml	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A2	Yes	Yes	Yes
46	5	Pasteurette	Pasteur pipette 3 ml	C	C1	No	No	No
		Tube	Andrology tube 5 ml	B	B1	No	No	No
			Centrifuge tube 15 ml	B	B1	Yes	Yes	Yes
		Syringe	Syringe 1 ml	B	B1	Yes	No	Yes
		Sperm straw	Sperm straw	A	A1	Yes	No	Yes
47	5	Dish	100 mm dish	A	A2	Yes	Yes	Yes
			4 well dish	A	A3	Yes	Yes	Yes
			35 mm dish	A	A1	Yes	Yes	Yes
		Tube	OPU tube	A	A2	Yes	Yes	Yes
			Andrology tube 5 ml	A	A1	Yes	Yes	Yes
48	5	Dish	4 well dish	A	A2	Yes	Yes	Yes
		Tube	OPU tube	B	B1	No	No	No
			Centrifuge tube 15 ml	B	B1	Yes	Yes	Yes
		Flask	25 cm ² flask	B	B1	No	No	No
		Serological pipette	Serological pipette 5 ml	A	A1	Yes	Yes	Yes

MEA: mouse embryo assay; HSMA: human sperm motility assay.

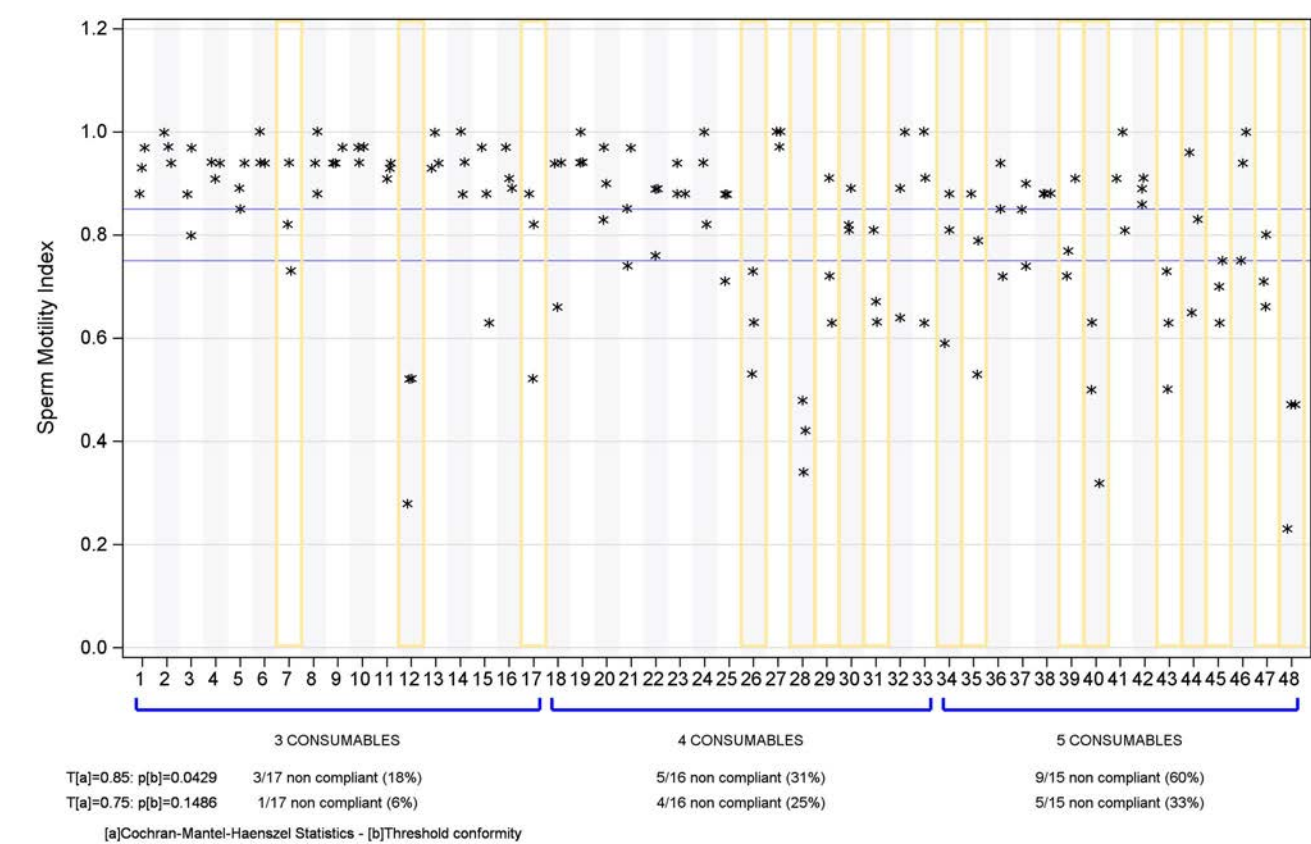


Figure 4. Sperm motility index results—combinations of disposables.

Table 4. Results of conformity assessment for seven associations of three, four, and five consumables tested consecutively using the human sperm motility assay method.

Conformity with three consumables (Association number)	Conformity when a fourth consumable was added to the same previous association	Conformity when a fifth consumable was added to the same previous association
Compliant (7)	Compliant (24)	Compliant (40)
Compliant (3)	Compliant (20)	Non-compliant (36)
NA	Compliant (22)	Compliant (38)
NA	Compliant (19)	Non-compliant (35)
NA	Compliant (34)	Non-compliant (50)
NA	Non-compliant (27)	Non-compliant (45)
NA	Non-compliant (31)	Non-compliant (46)

NA: non-applicable.

Group on Good Practice in IVF Labs et al., 2016; Practice Committees of the American Society for Reproductive Medicine (ASRM) and the Society for Reproductive Biologists and Technologists (SRBT). Electronic address: asrm@asrm.org, 2022). Suppliers are becoming increasingly aware of the importance of ensuring the safety, reliability, and traceability of their products. For some, this involves controlling both their raw materials and finished products. Meanwhile, a growing number of high-quality IVF laboratories are using the HSMA to assess new disposables and evaluate each lots of accepted disposables. In this study, two

out of the four consumables were found to be toxic, specifically the 1 ml syringes and the 25 cm² flasks, which both lacked MEA/HSMA certificates of conformity. This indicates that a consumable that has not undergone prior testing for embryotoxicity is more prone to being identified as toxic. The cap made of phenol and formaldehyde in the 25 cm² flasks could be the source of their toxicity (Thrasher and Kilburn, 2001). Our results further confirmed the previously reported toxicity of gloves (Critchlow et al., 1989; Lierman et al., 2007; Nijs et al., 2009). Even a contact duration of ~10 min between gloves and the sperm preparation was found to have a significant impact on motility. Caution should be exercised, as previous studies have indicated that rapid contact of gloves with a device could be detrimental owing to the potential detachment of fine particles from the gloves (Critchlow et al., 1989).

Additionally, our study has confirmed that the methodologies employed for toxicity extraction did have a significant impact on the sensitivity of the HSMA. We observed that two lots of sperm collection cups, which exhibited toxicity when tested using a 24-h extraction including 30 min of cap testing, no longer demonstrated toxicity when subjected to a 30-min extraction without cap testing. In a previous study (Delaroche et al., 2020), we identified significant variation among manufacturers in the methodologies used for toxicity testing of IVF disposables, including the MEA and HSMA. Importantly, certain parameters, such as sperm concentration, protein supplementation of the media, and incubation duration, were found to impact the sensitivity of the results (Claassens et al., 2000; Lierman et al., 2007). These methodological differences may contribute to the divergent toxicity outcomes observed for the same consumable. To date, there are no established guidelines for the methodology of conducting HSMA, unlike MEAs, which have Food and Drug Administration

(FDA) recommendations (Mouse Embryo Assay for Assisted Reproduction Technology Devices—Guidance for Industry and Food and Drug Administration Staff, n.d.). These FDA guidelines suggest testing only the parts of the article directly in contact with gametes and embryos during clinical use. Applying these guidelines to HSMAs would imply that the lids or caps of plasticware should not undergo routine testing. However, practical scenarios in IVF laboratories may lead to unavoidable contact between samples and lids or caps. For example, sperm collection containers may occasionally reach the laboratory with the sample touching the cap, and tubes used in oocyte retrievals might be overfilled, causing the cap to contact the follicular fluid. There is also the possibility of a tube accidentally tipping over, with its contents coming into contact with the cap. Furthermore, our rationale for choosing a 30-min contact was based on the FDA guidelines for MEA, which recommend a minimum contact time of 30 min for all consumables not used in embryo culture (Mouse Embryo Assay for Assisted Reproduction Technology Devices—Guidance for Industry and Food and Drug Administration Staff, n.d.). Other guidelines advocate that the exposure duration should be at least double the expected exposure time of sperm to the material (for pipette tips, exposure for a few seconds, and for sample collection receptacles, exposure from 30 to 60 min) (EN ISO 23162:2021, n.d.). The comprehensive testing approach employed for our study was grounded in a rigorous evaluation of toxicity, guided by the principle that a robust bioassay should not only be sensitive but also reliable, reproducible, and consistent in detecting compromised contact materials. Thus, quality controls need to be conducted under conditions that exceed the routine use conditions (Meintjes, 2017). Additionally, the choice of toxicity thresholds can lead to different conclusions regarding compliance. Initially, the toxicity threshold was defined as an SMI above 0.85 after 96 h of incubation (Critchlow et al., 1989). However, the currently accepted toxicity threshold is an SMI above 0.75 after 24 h of incubation (Meintjes, 2017; Delaroche et al., 2020). When the toxicity threshold was lowered to 0.75, no more single consumables and only 10 (21%) associations were found to be toxic in this study.

The majority of IVF consumables included in this study are composed of polymers from the thermoplastics family, specifically polystyrene, polypropylene, and polyethylene. The existing literature suggests a potential decline in fertility associated with environmental exposure to microparticles and nanoparticles derived from these polymers. They may accumulate in mammalian reproductive organs, exerting toxic effects on the reproductive systems of both genders, potentially leading to transgenerational impacts (Afreen et al., 2023; He and Yin, 2024). However, there is a noticeable gap in research concerning the direct toxicity of these polymers on gametes and embryos. In contrast, bisphenols (BPs), particularly bisphenol A (BPA), have been extensively studied. BPs, widely used as plasticizers, are primarily employed in polycarbonate synthesis, which is used in the production of medical devices (Gatimel et al., 2016). Concerns about the impact of BPA exposure on reproductive health outcomes have led to numerous publications based on *in vitro*, animal, and human studies. BPs are known to affect fertility as endocrine disruptors, exhibiting weak estrogenic activity. They also exert epigenetic effects in both male and female reproduction, with implications across multiple generations (Welshons et al., 2006; Susiarjo et al., 2015; Chianese et al., 2018; Ryu et al., 2023). Exposure to BPA during the final stages of meiosis can disrupt the meiotic spindle, cause chromosomal misalignment, increase the incidence of meiotic arrest, and lead to aneuploidy in mouse oocytes (Can

et al., 2005). In human studies, a negative dose–response association of BPA exposure with altered oocyte IVM has been observed (Machtinger et al., 2013). Owing to restrictive guidelines, BPA is now being replaced by its structural analogues, bisphenol S (BPS) and bisphenol F (BPF). While their impact on fertility is less well studied, initial reports indicate they also exhibit negative effects on reproductive functions, albeit through distinct intracellular pathways and mechanisms compared to BPA (Rochester and Bolden, 2015; Abouhamzeh et al., 2023). In various animal models, BPS has been shown to disrupt granulosa cell steroidogenesis (Téteau et al., 2020, 2023), impair the *in vitro* early developmental competence of oocytes and embryos (Desmarchais et al., 2020; Žalmanová et al., 2023) and reduce the motility of spermatozoa (Torres-Badía et al., 2023; Téteau et al., 2023). Similarly, BPF has been shown to affect motility, capacitation, and mitochondrial membrane potential in cryopreserved bovine sperm (Nguyen et al., 2022; Davis et al., 2023). In our study, none of the included consumables were made of polycarbonates. Therefore, the toxicity we observed could not be related to the potential leaching of BPs. However, studies have shown that plastic materials, especially strippers, can leach BPA into the surrounding medium (Gatimel et al., 2016). The leaching of BPs, observed at levels below the previously detected upper limit in women's serum and follicular fluid, did not appear to exert a significant impact on embryo development (Moreau et al., 2021). Moreover, contamination of gametes and embryos by BPs could also originate from culture media (Togola et al., 2021). Togola et al. (2021) revealed that the majority of culture media used in IVF procedures contained BPS, with concentrations approaching the nanomolar range. This raises concerns about the potential cocktail effect of culture media and consumables, which has not been adequately investigated to date. Further research is required to fully comprehend the cumulative risks associated with these media and disposables and to develop safer alternatives.

The present investigation has several limitations. First, because of constraints related to the limited availability of semen samples (approximately eight semen diagnoses per day), we could not use a single semen sample for all replicates. Instead, we were limited to using three different semen samples for the three replicates. This limitation may introduce variability in the results owing to natural variations in semen quality. It would have been valuable, when possible, to pool the three semen samples to have a unique final sperm preparation that could be used for the three tests. However, previous research has shown that the influence of semen source on sperm motility assay quality is minimal (Claassens et al., 2000) and to address the issue, we statistically checked the homogeneity between the three semen samples included in each test. The second limitation arises from the possibility that conducting experiments for the same test over multiple days may have influenced the results, even with controlled environmental and technical conditions. Variations in parameters such as temperature or humidity can impact equipment performance and biological reactions, contributing to result variations (Obermeyer and Pope, 2021). Standardized procedures may not eliminate internal variability, as different operators following the same protocols could introduce subtle variations in sample handling (for example accuracy of pipetting) (Crawford et al., 2023). Furthermore, mobility assessments were blindly conducted blind by two operators, and the average of their values was used for analysis. Despite regular internal and external quality checks ensuring a well-controlled process (Siddharth et al., 2023), the variability between operators may explain some random SMI values exceeding 1 in some single

consumable tests. Finally, despite the implementation of controls, it was not possible to fully eliminate the potential mechanical influence of transferring the sperm preparation from one consumable to another. It was apparent that a fraction of the sperm preparation was lost with each contact made with the consumables being tested in the associations. To complete our results, it would have been valuable to quantify the cumulative levels of toxins, such as BPs or phthalates, that could be released from associations of devices. Nevertheless, there is currently no exhaustive compilation of molecules resulting from the degradation of plastic and materials. Taking plastic packaging as an example, Meng and colleagues reported that plastic packaging materials were associated with a minimum of 10 259 chemicals, encompassing substances used in manufacturing, such as solvents, non-intentionally added substances, impurities, oligomers, and degradation products (Meng et al., 2023).

In conclusion, our preliminary observations document some cumulative toxicity on spermatozoa from disposables sequentially used in ART procedures. The quite high proportion of non-compliant combinations of devices can be attributed directly to the extreme rigorous extraction conditions employed during the tests, which deviated from the conditions encountered in clinical routines. Further work is warranted to investigate whether the cumulative toxicity observed under strict extraction conditions is still found in the real-life conditions.

Moreover, our study confirms the significance of conducting toxicity testing on consumables before introducing them into routine clinical practice. Integrating internal testing of consumables without a certificate of conformity into ongoing quality control programs is crucial. While individual screening remains essential, consumables should also be evaluated in combination with culture media, using methodologies consistent with real-life clinical practices. Another question raised by our study is the lack of regulation and standardization in HSMA testing protocols. To address the lack of harmonization, we recommend that IVF laboratories conduct testing of their associations of consumables and media under their own established protocols. Finally, this study advocates rationalizing the number of consumables used in each IVF procedure and re-evaluating the use of glass consumables (Kouakou et al., 2023).

Understanding the mechanisms underlying the toxicity of synthetic materials on human health and reproduction is crucial for mitigating their detrimental effects and formulating effective regulatory measures (Eladak et al., 2015). While there has been some progress in acknowledging and restricting the use of certain hazardous plastic and synthetic compounds, further research is required to elucidate the long-term consequences of exposure to synthetic materials and identify alternatives that pose fewer risks.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

Relevant data are available in the article and in its supplementary material. The complete database, including complementary data not presented in the article, is recorded on a secured server with limited access only to authorized users. If an access to these complementary data is needed, they can be shared on reasonable request to the corresponding author.

Acknowledgements

The authors thank Mrs Sabine Grelet, Mrs Océane Steinmetz, and Mrs Chloé Radoicic for their participation in collecting patient consent and Mrs Sandrine Barthéléry, Mrs Lucile Cavarec, Mrs Valérie Desbrée, Mrs Corinne Fillion, Mrs Clara Di Santo, Mrs Annael Steppe, and Mrs Angélique Kine for their participation in the experiments. The authors thank Mrs Hélène Barba, Mr Jacques Acker, and Mrs Magalie Valteau for their valuable database control. The authors thank Prof. Etienne Ruppé (Université Paris Cité, France) and Mr Johan Sterckx (Brussels IVF, Belgium) for their careful reviewing of the article.

Authors' roles

L.D. designed the study and wrote the article. L.B. did the experiments. V.O. checked the database and F.B. did the statistical analysis. G.C. revised the article and approved the final draft. All the authors contributed to the scientific discussions and reviewed the article.

Funding

This study received fundings from GCS Ramsay Santé pour l'Enseignement et la Recherche (Paris, France) and the Centre de Biologie Médicale BIOGROUPE (Le Chesnay-Rocquencourt, France).

Conflict of interest

The authors declare that they have no conflict of interest that could be perceived as prejudicing the impartiality of the reported research.

References

- Abouhamzeh B, Zare Z, Mohammadi M, Moosazadeh M, Nourian A. Bisphenol-S influence on oxidative stress and endocrine biomarkers of reproductive system: a systematic review and meta-analysis. *Int J Prev Med* 2023;**14**:37.
- Ackerman SB, Stokes GL, Swanson RJ, Taylor SP, Fenwick L. Toxicity testing for human in vitro fertilization programs. *J In Vitro Fert Embryo Transf* 1985;**2**:132–137.
- Afreen V, Hashmi K, Nasir R, Saleem A, Khan MI, Akhtar MF. Adverse health effects and mechanisms of microplastics on female reproductive system: a descriptive review. *Environ Sci Pollut Res Int* 2023;**30**:76283–76296.
- Alonso-Magdalena P, Vieira E, Soriano S, Menes L, Burks D, Quesada I, Nadal A. Bisphenol A exposure during pregnancy disrupts glucose homeostasis in mothers and adult male offspring. *Environ Health Perspect* 2010;**118**:1243–1250.
- Björndahl L, Barratt CLR, Mortimer D, Jouannet P. "How to count sperm properly": checklist for acceptability of studies based on human semen analysis. *Hum Reprod* 2016;**31**:227–232.
- Cairo Consensus Group. "There is only one thing that is truly important in an IVF laboratory: everything" Cairo Consensus Guidelines on IVF Culture Conditions. *Reprod Biomed Online* 2020;**40**:33–60.
- Can A, Semiz O, Cinar O. Bisphenol-A induces cell cycle delay and alters centrosome and spindle microtubular organization in oocytes during meiosis. *Mol Hum Reprod* 2005;**11**:389–396.
- Chianese R, Troisi J, Richards S, Scafuro M, Fasano S, Guida M, Pierantoni R, Meccariello R. Bisphenol A in reproduction: epigenetic effects. *Curr Med Chem* 2018;**25**:748–770.

- Chu Y, Zhou J, Ge D, Lu Y, Zou X, Xia L, Huang C, Shen C, Chu Y. Variable VOCs in plastic culture flasks and their potential impact on cell volatile biomarkers. *Anal Bioanal Chem* 2020;**412**:5397–5408.
- Claassens OE, Wehr JB, Harrison KL. Optimizing sensitivity of the human sperm motility assay for embryo toxicity testing. *Hum Reprod* 2000;**15**:1586–1591.
- Cohen J, Gilligan A, Esposito W, Schimmel T, Dale B. Ambient air and its potential effects on conception in vitro. *Hum Reprod* 1997;**12**:1742–1749.
- Coticchio G, Barrie A, Lagalla C, Borini A, Fishel S, Griffin D, Campbell A. Plasticity of the human preimplantation embryo: developmental dogmas, variations on themes and self-correction. *Hum Reprod Update* 2021;**27**:848–865.
- Crawford ML, Shuford CM, Grant RP. The pipetting olympics: propagating proper pipetting a priori in clinical LC-MS/MS analysis. *J Mass Spectrom Adv Clin Lab* 2023;**29**:16–20.
- Critchlow JD, Matson PL, Newman MC, Horne G, Troup SA, Lieberman BA. Quality control in an in-vitro fertilization laboratory: use of human sperm survival studies. *Hum Reprod* 1989;**4**:545–549.
- Davis OS, Truong VB, Hickey KD, Favetta LA. Quality of fresh and cryopreserved bovine sperm is reduced by BPA and BPF exposure. *Reprod Fertil* 2023;**4**:e230018.
- Delaroché L, Oger P, Genauzeau E, Meicler P, Lamazou F, Dupont C, Humaidan P. Embryotoxicity testing of IVF disposables: how do manufacturers test? *Hum Reprod* 2020;**35**:283–292.
- Desmarchais A, Tétéau O, Papillier P, Jaubert M, Druart X, Binet A, Maillard V, Elis S. Bisphenol S impaired in vitro ovine early developmental oocyte competence. *Int J Mol Sci* 2020;**21**:E1238.
- Eladak S, Grisin T, Moison D, Guerquin M-J, N'Tumba-Byn T, Pozzi-Gaudin S, Benachi A, Livera G, Rouiller-Fabre V, Habert R. A new chapter in the bisphenol A story: bisphenol S and bisphenol F are not safe alternatives to this compound. *Fertil Steril* 2015;**103**:11–21.
- EN ISO 23162:2021. Basic Semen Examination—Specification and Test Methods (ISO 23162:2021). *iTeh Standards*. <https://standards.iteh.ai/catalog/standards/cen/e2cfd734-bcae-41ad-b087-8bd1a-b9ae51e/en-iso-23162-2021> (24 January 2024, date last accessed).
- ESHRE Guideline Group on Good Practice in IVF Labs, De los Santos MJ, Apter S, Coticchio G, Debrock S, Lundin K, Plancha CE, Prados F, Rienzi L, Verheyen G. Revised guidelines for good practice in IVF laboratories (2015). *Hum Reprod* 2016;**31**:685–686.
- Fernández-González R, Moreira P, Bilbao A, Jiménez A, Pérez-Crespo M, Ramírez MA, Rodríguez De Fonseca F, Pintado B, Gutiérrez-Adán A. Long-term effect of in vitro culture of mouse embryos with serum on mRNA expression of imprinting genes, development, and behavior. *Proc Natl Acad Sci USA* 2004;**101**:5880–5885.
- Gatimel N, Lacroix MZ, Chanthavisouk S, Picard-Hagen N, Gayraud V, Parinaud J, Léandri RD. Bisphenol A in culture media and plastic consumables used for ART. *Hum Reprod* 2016;**31**:1436–1444.
- He Y, Yin R. The reproductive and transgenerational toxicity of microplastics and nanoplastics: a threat to mammalian fertility in both sexes. *J Appl Toxicol* 2024;**44**:66–85.
- Kouakou F, Denizot A-L, L'Hostis A, Colet J, Jacques S, Sallem A, Ziyat A, Vaiman D, Wolf J-P. Plastic used in vitro fertilization procedures induces massive placental gene expression alterations. *EBioMedicine* 2023;**91**:104572.
- Landrigan PJ, Raps H, Cropper M, Bald C, Brunner M, Canonizado EM, Charles D, Chiles TC, Donohue MJ, Enck J et al. The Minderoom-Monaco Commission on plastics and human health. *Ann Glob Health* 2023;**89**:71.
- Lierman S, De Sutter P, Dhont M, Van der Elst J. Double-quality control reveals high-level toxicity in gloves used for operator protection in assisted reproductive technology. *Fertil Steril* 2007;**88**:1266–1272.
- Long S, Woodward B, Tomlinson M. Sperm toxicity testing: UK best practice guideline from the Association of Biomedical Andrologists. *Br J Biomed Sci* 2018;**75**:53–60.
- Machtinger R, Combelles CMH, Missmer SA, Correia KF, Williams P, Hauser R, Racowsky C. Bisphenol-A and human oocyte maturation in vitro. *Hum Reprod* 2013;**28**:2735–2745.
- McDonald GR, Hudson AL, Dunn SMJ, You H, Baker GB, Whittall RM, Martin JW, Jha A, Edmondson DE, Holt A. Bioactive contaminants leach from disposable laboratory plasticware. *Science* 2008;**322**:917.
- Meintjes M. Sperm survival assay for quality control in the IVF laboratory. In: Morbeck DE, Montag MHM (eds). *Principles of IVF Laboratory Practice: Optimizing Performance and Outcomes*. Cambridge: Cambridge University Press, 2017, 73–78. <https://www.cambridge.org/core/books/principles-of-ivf-laboratory-practice/sperm-survival-assay-for-quality-control-in-the-ivf-laboratory/1D752528CB43CFF3D7367BD736E7C047>.
- Meng W, Sun H, Su G. Plastic packaging-associated chemicals and their hazards—an overview of reviews. *Chemosphere* 2023;**331**:138795.
- Moreau J, Gatimel N, Lippi Y, Tavenier G, Fauque P, Guilleman M, Naylies C, Huesca AA, Gayraud V, Parinaud J et al. Impact of the polycarbonate strippers used in assisted reproduction techniques on embryonic development. *Hum Reprod* 2021;**36**:331–339.
- Mortimer D, Cohen J, Mortimer ST, Fawzy M, McCulloh DH, Morbeck DE, Pollet-Villard X, Mansour RT, Brison DR, Doshi A et al. Cairo consensus on the IVF laboratory environment and air quality: report of an expert meeting. *Reprod Biomed Online* 2018;**36**:658–674.
- Mouse Embryo Assay for Assisted Reproduction Technology Devices—Guidance for Industry and Food and Drug Administration Staff. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/mouse-embryo-assay-assisted-reproduction-technology-devices> (24 January 2024, date last accessed).
- Nguyen M, Sabry R, Davis OS, Favetta LA. Effects of BPA, BPS, and BPF on oxidative stress and antioxidant enzyme expression in bovine oocytes and spermatozoa. *Genes (Basel)* 2022;**13**:142.
- Nijs M, Franssen K, Cox A, Wissmann D, Ruis H, Ombelet W. Reprotoxicity of intrauterine insemination and in vitro fertilization-embryo transfer disposables and products: a 4-year survey. *Fertil Steril* 2009;**92**:527–535.
- Obermeyer Z, Pope D. Variation in common laboratory test results caused by ambient temperature. *Med* 2021;**2**:1314–1326.e2.
- Olivieri A, Degenhardt OS, McDonald GR, Narang D, Paulsen IM, Kozuska JL, Holt A. On the disruption of biochemical and biological assays by chemicals leaching from disposable laboratory plasticware. *Can J Physiol Pharmacol* 2012;**90**:697–703.
- Panagopoulos P, Mavrogianni D, Christodoulaki C, Drakaki E, Chrelas G, Panagiotopoulos D, Potiris A, Drakakis P, Stavros S. Effects of endocrine disrupting compounds on female fertility. *Best Pract Res Clin Obstet Gynaecol* 2023;**88**:102347.
- Practice Committees of the American Society for Reproductive Medicine (ASRM) and the Society for Reproductive Biologists and Technologists (SRBT). Electronic address: asrm@asrm.org. Comprehensive guidance for human embryology, andrology, and endocrinology laboratories: management and operations: a committee opinion. *Fertil Steril* 2022;**117**:1183–1202.
- Ramos-Ibeas P, Heras S, Gómez-Redondo I, Planells B, Fernández-González R, Pericuesta E, Laguna-Barraza R, Pérez-Cereales S, Gutiérrez-Adán A. Embryo responses to stress induced by assisted reproductive technologies. *Mol Reprod Dev* 2019;**86**:1292–1306.
- Rochester JR, Bolden AL. Bisphenol S and F: a systematic review and comparison of the hormonal activity of Bisphenol A substitutes. *Environ Health Perspect* 2015;**123**:643–650.
- Ryu D-Y, Pang W-K, Adegoke EO, Rahman MS, Park Y-J, Pang M-G. Bisphenol-A disturbs hormonal levels and testis mitochondrial activity, reducing male fertility. *Hum Reprod Open* 2023;**2023**:hoad044.

- Scarica C, Monaco A, Borini A, Pontemezzo E, Bonanni V, De Santis L, Zacà C, Cotichio G; SIERR, Società Italiana di Embriologia Riproduzione e Ricerca. Use of mineral oil in IVF culture systems: physico-chemical aspects, management, and safety. *J Assist Reprod Genet* 2022;**39**:883–892.
- Sedha S, Lee H, Singh S, Kumar S, Jain S, Ahmad A, Bin Jordan YA, Sonwal S, Shukla S, Simal-Gandara J et al. Reproductive toxic potential of phthalate compounds—state of art review. *Pharmacol Res* 2021;**167**:105536.
- Seymore TN, Rivera-Núñez Z, Stapleton PA, Adibi JJ, Barrett ES. Phthalate exposures and placental health in animal models and humans: a systematic review. *Toxicol Sci* 2022; **188**:153–179.
- Siddharth K, Kumar T, Zabihullah M. Interobserver variability in semen analysis: findings from a quality control initiative. *Cureus* 2023;**15**:e46388.
- Susiarjo M, Xin F, Bansal A, Stefaniak M, Li C, Simmons RA, Bartolomei MS. Bisphenol A exposure disrupts metabolic health across multiple generations in the mouse. *Endocrinology* 2015; **156**:2049–2058.
- Takai Y, Tsutsumi O, Ikezuki Y, Kamei Y, Osuga Y, Yano T, Taketan Y. Preimplantation exposure to bisphenol A advances postnatal development. *Reprod Toxicol* 2001;**15**:71–74.
- Téteau O, Jaubert M, Desmarchais A, Papillier P, Binet A, Maillard V, Elis S. Bisphenol A and S impaired ovine granulosa cell steroidogenesis. *Reproduction* 2020;**159**:571–583.
- Téteau O, Vitorino Carvalho A, Papillier P, Mandon-Pépin B, Jouneau L, Jarrier-Gaillard P, Desmarchais A, Lebachelier de la Riviere M-E, Vignault C, Maillard V et al. Bisphenol A and bisphenol S both disrupt ovine granulosa cell steroidogenesis but through different molecular pathways. *J Ovarian Res* 2023;**16**:30.
- Thrasher JD, Kilburn KH. Embryo toxicity and teratogenicity of formaldehyde. *Arch Environ Health* 2001;**56**:300–311.
- Togola A, Desmarchais A, Téteau O, Vignault C, Maillard V, Buron C, Bristeau S, Guerif F, Binet A, Elis S. Bisphenol S is present in culture media used for ART and cell culture. *Human Reproduction* 2021;**36**:1032–1042.
- Torres-Badia M, Martin-Hidalgo D, Serrano R, Garcia-Marin LJ, Bragado MJ. Bisphenol S reduces pig spermatozoa motility through different intracellular pathways and mechanisms than its analog bisphenol A. *Int J Mol Sci* 2023;**24**:9598.
- Vardhan PV, Shukla LI. FT-IR investigations on effect of high doses of gamma radiation-induced damage to polystyrene and mechanism of formation of radiolysis products. *Radiat Environ Biophys* 2018;**57**:301–310.
- Wale PL, Gardner DK. The effects of chemical and physical factors on mammalian embryo culture and their importance for the practice of assisted human reproduction. *Hum Reprod Update* 2016;**22**:2–22.
- Welshons WV, Nagel SC, Vom Saal FS. Large effects from small exposures. III. Endocrine mechanisms mediating effects of bisphenol A at levels of human exposure. *Endocrinology* 2006;**147**:S56–S69.
- World Health Organization. WHO Laboratory Manual for the Examination and Processing of Human Semen. Geneva: World Health Organization, 2010. <https://apps.who.int/iris/handle/10665/44261>.
- Yang CZ, Yaniger SI, Jordan VC, Klein DJ, Bittner GD. Most plastic products release estrogenic chemicals: a potential health problem that can be solved. *Environ Health Perspect* 2011;**119**:989–996.
- Žalmanová T, Hošková K, Prokešová Š, Nevoral J, Jeřeta M, Benc M, Yi Y-J, Moravec J, Močáryová B, Martínková S et al. The bisphenol S contamination level observed in human follicular fluid affects the development of porcine oocytes. *Front Cell Dev Biol* 2023; **11**:1145182.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Human Reproduction, 2024, 39, 936–954
<https://doi.org/10.1093/humrep/deae028>
Original Article

Embryology

Abnormal cleavage up to Day 3 does not compromise live birth and neonatal outcomes of embryos that have achieved full blastulation: a retrospective cohort study

Tammy Lee ^{1,2,*}, Kelli Peirce ², Jay Natalwala², Vincent Chapple ², Peter J. Mark ¹, Katherine Sanders ¹, and Yanhe Liu ^{1,2,3,4}

¹School of Human Sciences, The University of Western Australia, Crawley, WA, Australia

²Fertility North, Joondalup Private Hospital, Joondalup, WA, Australia

³School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA, Australia

⁴School of Health Sciences and Medicine, Bond University, Robina, QLD, Australia

*Correspondence address. Fertility North, Joondalup Private Hospital, Suite 30, Level 2, 60 Shenton Ave, Joondalup, WA 6027, Australia. Tel: +61-8-9301-1075; Fax: +61-8-9300-1000; E-mail: tammy.lee@research.uwa.edu.au  <https://orcid.org/0000-0003-2138-9302>

ABSTRACT

STUDY QUESTION: Do embryos displaying abnormal cleavage (ABNCL) up to Day 3 have compromised live birth rates and neonatal outcomes if full blastulation has been achieved prior to transfer?

SUMMARY ANSWER: ABNCL is associated with reduced full blastulation rates but does not impact live birth rates and neonatal outcomes once full blastulation has been achieved.

WHAT IS KNOWN ALREADY?: It is widely accepted that ABNCL is associated with reduced implantation rates of embryos when transferred at the cleavage stage. However, evidence is scarce in the literature reporting birth outcomes from blastocysts arising from ABNCL embryos, likely because they are ranked low priority for transfer.

STUDY DESIGN, SIZE, DURATION: This retrospective cohort study included 1562 consecutive autologous *in vitro* fertilization cycles (maternal age 35.1 ± 4.7 years) performed at Fertility North, Australia between January 2017 and June 2022. Fresh transfers were performed on Day 3 or 5, with remaining embryos cultured up to Day 6 before vitrification. A total of 6019 embryos were subject to blastocyst culture, and a subset of 664 resulting frozen blastocysts was included for live birth and neonatal outcome analyses following single transfers.

PARTICIPANTS/MATERIALS, SETTING, METHODS: ABNCL events were annotated from the first mitotic division up to Day 3, including direct cleavage (DC), reverse cleavage (RC) and <6 intercellular contact points at the 4-cell stage (<6 ICCP). For DC and RC in combination, the ratios of affected blastomeres over the total number of all blastomeres up to Day 3 were also recorded. All pregnancies were followed up until birth with gestational age, birthweight, and sex of the baby being recorded.

MAIN RESULTS AND THE ROLE OF CHANCE: Full blastulation rates for embryos showing DC (19.5%), RC (41.7%), <6 ICCP (58.8%), and mixed (≥ 2) ABNCL types (26.4%) were lower than the rates for those without ABNCL (67.2%, $P < 0.01$ respectively). Subgroup analysis showed declining full blastulation rates with increasing ratios of combined DC/RC affected blastomeres over all blastomeres up to the 8-cell stage (66.2% when 0 affected, 47.0% when 0.25 affected, 27.4% when 0.5 affected, 14.5% when 0.75 affected, and 7.7% when all affected, $P < 0.01$). However, once full blastulation had been achieved, no difference was detected between DC, RC, <6 ICCP, and no ABNCL blastocysts following single frozen transfers in subsequent live birth rates (25.9%, 33.0%, 36.0% versus 30.8%, $P > 0.05$, respectively), gestational age (38.7 ± 1.6 , 38.5 ± 1.2 , 38.3 ± 3.5 versus 38.5 ± 1.8 weeks, $P > 0.05$, respectively) and birthweight (3343.0 ± 649.1 , 3378.2 ± 538.4 , 3352.6 ± 841.3 versus 3313.9 ± 509.6 g, $P > 0.05$, respectively). Multiple regression (logistic or linear as appropriate) confirmed no differences in all of the above measures after accounting for potential confounders.

LIMITATIONS, REASONS FOR CAUTION: Our study is limited by its retrospective nature, making it impossible to control every known or unknown confounder. Embryos in our dataset, being surplus after selection for fresh transfer, may not represent the general embryo population.

WIDER IMPLICATIONS OF THE FINDINGS: Our findings highlight the incremental impact of ABNCL, depending on the ratio of affected blastomeres up to Day 3, on subsequent full blastulation. The reassuring live birth and neonatal outcomes of ABNCL blastocysts imply a potential self-correction mechanism among those embryos reaching the blastocyst stage, which provides valuable guidance for clinical practice and patient counseling.

Received: December 20, 2023. Revised: February 22, 2024. Editorial decision: March 10, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

STUDY FUNDING/COMPETING INTEREST(S): This research is supported by an Australian Government Research Training Program (RTP) Scholarship. All authors report no conflict of interest.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: direct cleavage / reverse cleavage / <6 intercellular contact points / blastulation / live birth

Introduction

Embryo selection is a critical step in the *in vitro* fertilization (IVF) process to diminish time to pregnancy, as well as for reducing psychological stress and costs to the patient. A shift to single embryo transfer has been gaining a wider global consensus to minimize multiple pregnancies (Practice Committee of the American Society for Reproductive Medicine and the Practice Committee for the Society for Assisted Reproductive Technologies, 2021), consequently decreasing a myriad of adverse complications associated with multiple births, such as mortality and morbidity (D'Antonio et al., 2013), shorter gestational age (Blondel et al., 2006), low birthweight (Giorgione et al., 2021), and long-term health outcomes (Wainstock et al., 2023). However, selection of the embryo with the best implantation potential by non-invasive approaches remains a challenge (Gardner et al., 2015; Glatstein et al., 2023).

Time-lapse incubation (TLI) allows uninterrupted culture and continuous monitoring of embryos. This offers new promise for better selection compared to traditional static morphological assessment (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011) and extended culture to the Day 5 or 6 blastocyst stage (Gardner and Schoolcraft, 1999). Many TLI embryo selection algorithms utilize both qualitative and quantitative measures to select an embryo with the highest implantation potential (Meseguer et al., 2011; Liu et al., 2016; Petersen et al., 2016). These quantitative and qualitative parameters can be annotated in TLI but are otherwise missed in standard benchtop incubation. Quantitative parameters, such as embryo morphokinetics, include times to reach a range of biological milestones such as pronuclear appearance, 2-cell stage, and start of blastulation (Ciray et al., 2014). Qualitative traits include morphometric features as early as the pronuclear stage (Orevich et al., 2022), spontaneous collapsing at the blastocyst stage (Bickendorf et al., 2023), and abnormal cleavage (ABNCL) patterns at the cleavage stage such as direct cleavage (DC) (Rubio et al., 2012; Ciray et al., 2014), reverse cleavage (RC) (Liu et al., 2014) and less than six intracellular contact points at the 4-cell stage (<6ICCP) (Liu et al., 2015). A recent meta-analysis has highlighted the preferable inter-laboratory transferability of qualitative parameters over quantitative parameters (Liu et al., 2020), potentially due to the varying embryo morphokinetics in response to different culture conditions (Kasterstein et al., 2013) and patient populations (Fréour et al., 2013).

Unlike other categories of qualitative parameters used for TLI embryo selection, ABNCL patterns are relatively better defined and clinically applied (ESHRE Working group on Time-lapse technology, 2020). It is well reported that ABNCL leads to a myriad of adverse outcomes such as reduced blastulation (Meseguer et al., 2011; Athayde Wirka et al., 2014; Shavit et al., 2023), lower blastocyst quality (Ho et al., 2018; Jin et al., 2022), and decreased implantation potential when transferring at the cleavage stage (Rubio et al., 2012; Liu et al., 2014, 2015; Fan et al., 2016; Barrie et al., 2017). Previous studies have focused on the incidence of DC at the 1-cell and 2-cell stage (Athayde Wirka et al., 2014; Fan et al., 2016), while few have considered the combined impacts of different ABNCL types up to Day 3 or beyond the 4-cell stage by a

structured classification system. Moreover, the role of blastulation in embryos displaying ABNCL remain underexplored, with a lack of live birth and neonatal outcome data in the literature. Therefore, the current study aims to: (i) quantify the combined impacts of ABNCL on full blastulation, according to the ratio of affected blastomeres over total blastomeres up to Day 3, and (ii) evaluate live birth rates and neonatal outcomes of ABNCL embryos that have successfully blastulated.

Materials and methods

This retrospective cohort study included 1562 consecutive autologous IVF and intra-cytoplasmic sperm injection (ICSI) cycles (maternal age 35.1 ± 4.7 years) performed at Fertility North (Perth, Australia) between January 2017 and June 2022. All included cycles led to either a Day 3 or 5 fresh transfer or freezing of all embryos as decided by the treating clinician, with all remaining embryos being cultured up to Day 6. Live birth outcomes and neonatal analyses were conducted based on a subset of 664 resulting blastocysts which had subsequently undergone a single frozen transfer. Table 1 describes baseline characteristics of cycles included in the study. Retrospective data analysis was approved by the Ramsay Health Care Human Research Ethics Committee (2022/ETH/0073) and the Human Research Ethics Office of The University of Western Australia (2023/ET000715).

Gamete preparation, fertilization, and embryo culture

Ovarian stimulation, transvaginal oocyte aspiration, sperm preparation (either partner or donor), and insemination via either conventional IVF or ICSI were performed according to previous publication (Liu et al., 2014). Oocyte-cumulus complexes (OCCs) were washed and held in G-IVF™ Plus media (Vitrolife, Sweden) at 6% CO₂ and 37°C until completion of oocyte collection. OCCs were then transferred to wells containing G-IVF™ Plus media (Vitrolife) overlaid with Ovoil™ (Vitrolife) at 6% CO₂, 5% O₂, and 89% N₂ at 37°C for culture before conventional IVF insemination or denudation for ICSI.

Oocytes for insemination by ICSI were denuded by brief exposure to SynVibro™ Hyadase (Origio, Denmark) for a maximum of 10s, followed by mechanical removal of cumulus cells in G-MOPS™ (Vitrolife). Oocytes at the metaphase II stage were injected with a single spermatozoon and subsequently loaded into the EmbryoScope+ incubator (Vitrolife) for culture at 6% CO₂, 5% O₂, and 89% N₂ at 37°C. From July 2020, the laboratory shifted from a sequential (G1-Plus™/G2-Plus™ with a changeover on Day 3) to a single-step (G-TL™ Plus up to Day 6 without refresh) culture strategy. For conventional IVF, prepared sperm were mixed with OCCs for overnight co-incubation, before fertilization was assessed in 16–18 h. Oocytes with confirmed two pronuclei were then placed into the EmbryoScope+ for culture up to Day 6.

Time-lapse annotation for ABNCL events

A total of 9451 fertilized oocytes were cultured up to Day 3. Amongst these embryos, 3432 were either transferred on Day 3 or classified as grade F, the lowest grade, according to our

Table 1. Cycle characteristics.

Characteristic	Value
Number of oocyte collection cycles (n)	1562
Female age at oocyte collection (years, mean $\pm$ SD and range)	35.1 $\pm$ 4.7 21.4–48.1
IVF/ICSI ratio	636/926
Number of fertilized oocytes (n)	9451
Number of Day 3 embryos subject to blastocyst culture (n)	6019
Number of blastocysts formed (n)	3239
Number of single frozen blastocyst transfers (n)	664
Female age at oocyte collection (years, mean $\pm$ SD and range)	34.0 $\pm$ 4.4 23.8–45.7
Body mass index (mean $\pm$ SD and range)	25.5 $\pm$ 4.8 17.2–39.1
IVF/ICSI ratio	241/423
Sperm type (partner/donor)	512/152
Number of Day 5 blastocysts (n)	486
Number of Day 6 blastocysts (n)	178
Number of single frozen blastocyst transfer cycles leading to a live birth (n/%)	208 (31.3%)
Live birth babies included for analysis (n)*	204
Baby sex ratio (male/female)	97/107
Gestational age (weeks, mean $\pm$ SD)	38.5 $\pm$ 1.9
Birthweight (grams, mean $\pm$ SD)	3327.4 $\pm$ 550.4
Number of preterm deliveries (<38 weeks, n/%)	24 (11.8%)
Number of low birthweight (<2500 g, n/%)	9 (4.4%)

* Note: Four cycles were excluded from the neonatal analysis, including one with monozygotic twins and three with unknown birthweight outcomes.

previously published algorithm (Liu *et al.*, 2016), and were subsequently excluded from analysis. The remaining 6019 embryos were fully annotated for morphokinetic features and subjected to blastocyst culture.

Three ABNCL events, DC, RC, and <6ICCP, were manually annotated up to Day 3 with the EmbryoViewer® software (Vitrolife). We defined DC as the division of one blastomere into three daughter blastomeres at any stage, previously termed trichotomous mitosis by Ciray *et al.* (2014). RC was defined as either type I: blastomeric fusion following cell division, or type II: failed cytokinesis after karyokinesis (Liu *et al.*, 2014). The ABNCL event <6ICCP was classified if there were fewer than six intra-cellular contact points at the end of the four-cell stage (i.e. if any two of the four cells were not in contact with each other) (Liu *et al.*, 2015).

Moreover, the ratios of DC/RC affected blastomeres over total blastomeres in each embryo were analyzed up to Day 3. The <6ICCP was excluded from this analysis as it does not occur at the single blastomere level and is only applicable at the 4-cell stage. Specifically, time-lapse footage was reviewed for DC, RC, or a combination of both, and each embryo was classified into one of five possible ratio groups (0, 0.25, 0.5, 0.75, or 1 as demonstrated in Table 2). For example, if DC or RC occurred at the 1-cell stage, a ratio of 1 was noted and if it occurred in one blastomere at the 2-cell stage, this accounted for a ratio of 0.5. In a 4-cell stage embryo, a ratio of 0.75 could be either a result of three out of the four blastomeres demonstrating DC/RC or a combined effect of one affected blastomere at the 2-cell stage plus one affected blastomere at the 4-cell stage which arose from the non-affected 2-cell stage blastomere. The division of blastomeres at the 4-cell stage was considered up until Day 3.

Blastocyst vitrification, warming, and frozen transfer

Blastocysts formed on Day 5 or 6 were graded as per the Gardner system (Gardner and Schoolcraft, 1999) and full blastulation was

Table 2. Ratios of affected blastomeres over total blastomeres up to the completion of each blastomere division at the 4-cell stage.

Ratio	Scenarios of number of blastomere(s) affected by direct cleavage, reverse cleavage, or mixed at each developmental stage		
	1-cell stage	2-cell stage	4-cell stage
0	0	0	0
0.25	–	–	1
0.5	–	1	2
0.75	–	–	3
1	1	2	4

defined as at least stage 3 in expansion with an A/B grade inner cell mass. Suitable blastocysts were vitrified on Day 5 or 6 immediately after grading. The Rapid-i™ device, RapidVit™, and RapidWarm™ Blast media (Vitrolife) were used for vitrification and warming as per manufacturer's protocol. A total of 664 resulting blastocysts underwent single frozen transfers (hormone replacement therapy or natural cycles) and were included for live birth and neonatal analysis. All pregnancies were followed up until birth, with the sex of the baby, gestational age and birthweight outcomes being recorded as appropriate.

Statistical analysis

Continuous data were analyzed via t-test (expressed as mean $\pm$ standard deviation) or a multiple linear regression model. Linear regression statistics were expressed as standardized coefficient (β) coupled with a 95% confidence interval (CI). Proportional data were evaluated by chi-squared analysis or logistic regression to determine adjusted odds ratios (aOR) with a 95% CI. Statistical analysis was performed using Statistical Package for the Social Sciences (version 25, IBM, Washington, USA) and $P < 0.05$ was considered as statistically significant.

Results

Incidence of abnormal cleavage

Out of 6019 fully annotated Day 3 embryos, the majority presented with no ABNCL (57.2%, 3443/6019), followed by RC (16.2%, 976/6019), DC (13.8%, 830/6019), and <6ICCP (7.8%, 471/6019), while 5.0% (299/6019) showed a mixture of two or more ABNCL events (Fig. 1A). In reference to the Day 3 embryo dataset of no ABNCL embryos (57.2%, Fig. 1A), there was a significant difference with the proportion of embryos present in the blastocyst dataset (71.4%, $P < 0.01$, Fig. 1B), in the transferred blastocyst dataset (74.7%, $P < 0.01$, Fig. 1C), and in the live birth dataset (73.6%, $P < 0.05$, Fig. 1D). The compositions of different ABNCL types were similar ($P > 0.05$) between the transferred blastocysts (Fig. 1C) and live birth datasets (Fig. 1D),

Full blastulation rate

Overall, full blastulation rates for different groups are shown in Fig. 2. Further analysis was performed to investigate the combined impacts of DC/RC on full blastulation based on the ratio of affected blastomeres over all blastomeres. Overall, there was an inverse relationship between the full blastulation rate and the ratio of affected blastomeres over all of the blastomeres (Fig. 3). In the DC group, full blastulation rates were 61.2% (3053/4989), 38.9% (70/180), 19.6% (97/495), 0%, and 5.4% (19/355) for ratios of 0, 0.25, 0.5, 0.75, and 1, respectively; 57.9% (2756/4761), 44.1% (346/784), 32.0% (122/381), 25.0% (5/20), and 13.7% (10/73) for the

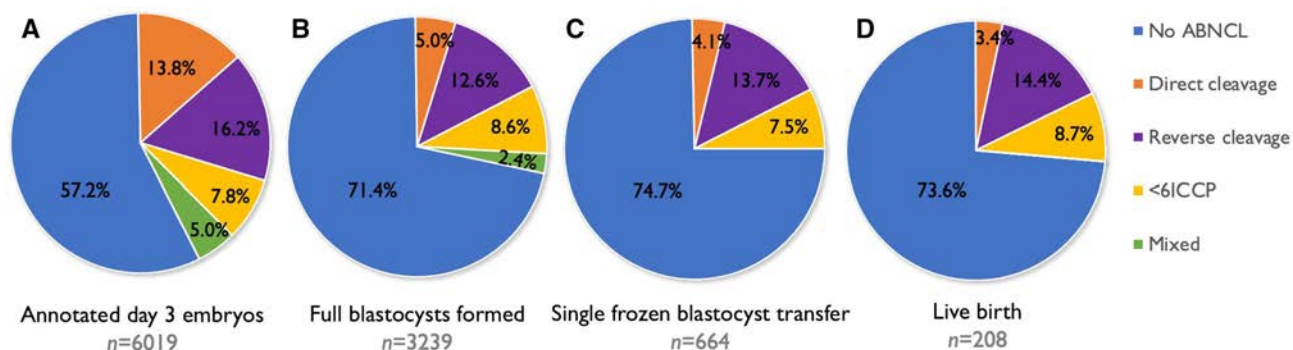


Figure 1. Incidences of different types of abnormal cleavage between different datasets. Note: Proportion of embryos showing no abnormal cleavage or abnormal cleavage events in all: (A) annotated Day 3 embryos; (B) blastocysts formed; (C) blastocysts undergoing single frozen transfer; and (D) blastocysts leading to a live birth. <6ICCP, less than six intracellular contact points; ABNCL, abnormal cleavage; mixed, embryos showing at least two types of abnormal cleavage patterns.

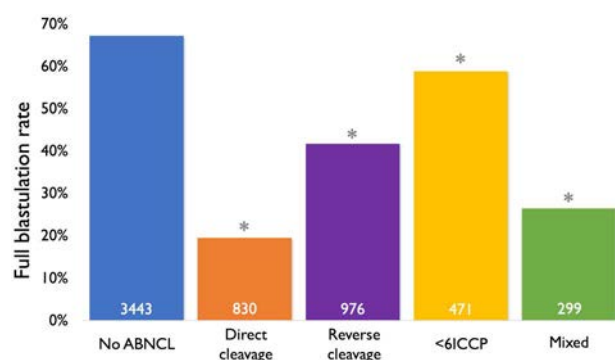


Figure 2. Blastulation rates of embryos with or without abnormal cleavage events. ABNCL, abnormal cleavage; <6ICCP, less than six intracellular contact points at the 4-cell stage; mixed, embryos showing at least two types of abnormal cleavage patterns. Numbers within each bar represent the total number of embryos for each group. Statistics and P values were based on Chi squared comparisons with the No ABNCL group. * $P < 0.01$ compared to no ABNCL.

RC group; and 66.2% (2591/3914), 47.0% (396/842), 27.4% (209/764), 14.5% (10/69), and 7.7% (33/430) in combination of DC and RC groups. All sub-groups of 0.25, 0.5, 0.75, and 1 demonstrated statistical significance compared to their respective 0 ratio group ($P < 0.05$).

Live birth rate

A total of 208 single blastocyst transfers resulted in a live birth outcome. There was no significant difference in live birth rates following single frozen transfer of blastocysts with (32.7%) or without ABNCL (30.8%, $P > 0.05$). Live birth rates were comparable between the DC, RC, <6ICCP, and no ABNCL groups ($P > 0.05$ respectively) (Table 3). No transfers were performed for the mixed ABNCL group. Logistic regression confirmed no significant impact of DC, RC, or <6ICCP on live birth in comparison to the no ABNCL, after accounting for potential confounders including sequential/single-step culture, maternal age at oocyte collection, female body mass index (BMI), vitrification day, blastocyst expansion stage, inner cell mass/trophoblast grades, insemination method, and sperm type (partner or donor) (Supplementary Table S1).

Further analysis comparing live birth rates of single transferred frozen blastocysts between different ratios of blastomeres affected by DC or RC is shown in Table 4. No significant difference was detected between any categories ($P > 0.05$).

Gestational age and birthweight

Four live birth cycles (out of 208 total live births) were excluded from neonatal analysis: one resulted in monozygotic twins and three had unknown birthweight outcomes. Neonatal outcomes were comparable between the blastocysts with DC, RC, <6ICCP, and no ABNCL for gestational age and birthweight (Table 3).

Multiple linear regression showed no significant impact on the gestational age by DC, RC, and <6ICCP in reference to the no ABNCL group. Similarly, there was no significant impact on birthweight by DC, RC, and <6ICCP in reference to the no ABNCL group. All of the above standard coefficients were adjusted for sequential/single-step culture, maternal age at oocyte collection, female BMI, vitrification day, blastocyst expansion stage, inner cell mass/trophoblast grades, insemination method, and sperm type (partner or donor) (Supplementary Table S2). Furthermore, no significant difference ($P > 0.05$, respectively) was detected using either: (i) Z-score combining sex of the baby, birthweight, and gestational age (Supplementary Table S2), (ii) pre-term delivery (<37 weeks), or (iii) low birthweight (<2500 g), as separate endpoints (Supplementary Table S1).

Discussion

This study investigated live birth rates and neonatal outcomes of embryos with or without ABNCL following a single frozen blastocyst transfer. Embryos displaying ABNCL (DC, RC, <6ICCP) by Day 3 had reduced full blastulation rates compared to embryos with no ABNCL. There was a linear relationship between the proportion of blastomeres impacted by ABNCL and full blastulation rates for DC, RC, and a combination of both events. However, once full blastulation was achieved, the live birth rate, gestational age, and birthweight were comparable to embryos with no ABNCL following single frozen blastocyst transfer.

It is widely acknowledged that ABNCL in cleavage stage human embryos leads to poorer clinical outcomes (Rubio et al., 2012; Liu et al., 2014, 2015, 2020). In these studies, the embryos were transferred at the cleavage stage and reduced live birth rates were reported. The observed decrease in live birth rate following Day 3 transfers in other studies could well be explained by a reduced blastulation capability, while embryos failing to blastulate would have been excluded from transfer in a blastocyst transfer program. De-prioritizing ABNCL embryos at transfer seems to be widely adopted in clinical practice, even if they have successfully blastulated (Gorodeckaja et al., 2020). This may have partly contributed

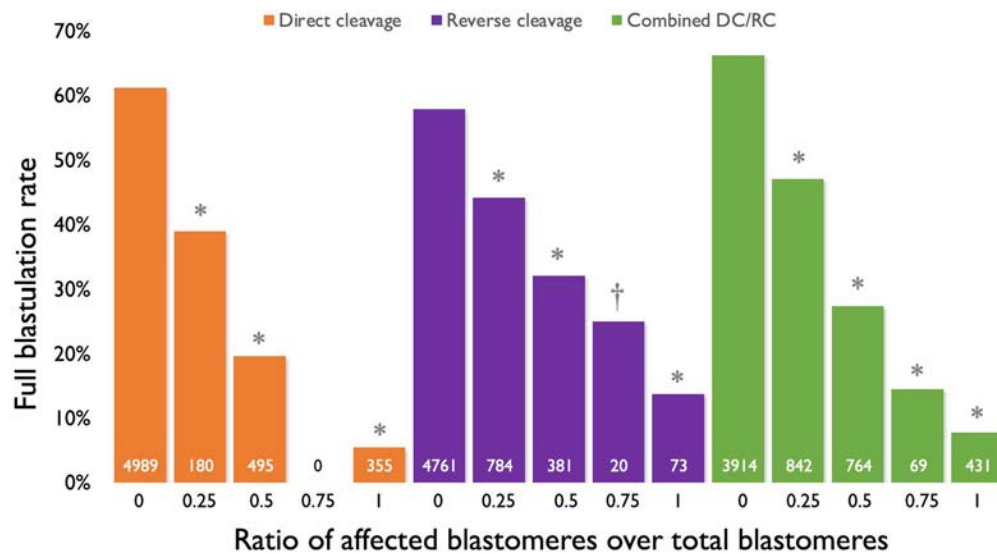


Figure 3. Blastulation rates of embryos according to ratios of blastomeres affected by direct cleavage, reverse cleavage, or both, over the total blastomeres in the 1- to 8-cell embryos. DC, direct cleavage; RC, reverse cleavage. Numbers within each bar represent the total number of embryos for each group. Statistics and P values were based on Chi squared comparisons with corresponding 0 ratio. *P < 0.001, †P < 0.05 compared to no affected blastomeres (ratio of 0) within each abnormal cleavage category.

Table 3. Live birth rate, gestational age, and birthweight following single transfer of blastocysts displaying no abnormal cleavage events (ABNCL), direct cleavage, reverse cleavage, and <6 intercellular contact points at the four-cell stage (ICCP).

	No ABNCL (reference)	Direct cleavage	Reverse cleavage	<6ICCP
Live birth rate (n = 664)	30.8% (153/496)	25.9% (7/27)	33.0% (30/91)	36.0% (18/50)
Gestational age (weeks, n = 204)	38.5 ± 1.8	38.7 ± 1.6	38.5 ± 1.2	38.3 ± 3.5
Birthweight (grams, n = 204)	3313.9 ± 509.6	3343.0 ± 649.1	3378.2 ± 538.4	3352.6 ± 841.3

Note: Chi-square analysis was performed for live birth rate and t-test was performed for gestational age and birthweight. No statistical significance was detected in each abnormal cleavage group compared to the No ABNCL group (P > 0.05).

Table 4. Live birth rates according to ratios of blastomeres affected by direct cleavage or reverse cleavage over total blastomeres up to Day 3.

Live birth rates of single frozen transferred blastocysts according to ratio of blastomeres affected by direct cleavage or reverse cleavage over total blastomeres up to Day 3				
	0.25	0.5	1	Total
Direct cleavage	44.4% (4/9)	13.3% (2/15)	33.3% (1/3)	25.9% (7/27)
Reverse cleavage	34.3% (23/67)	29.2% (7/24)	—	33.0% (30/91)
Total	35.5% (27/76)	23.1% (9/39)	33.3% (1/3)	31.4% (37/118)

Note: No embryos with 75% blastomeres affected were transferred during the study period. No significant difference in live birth rate was detected between each categories (P > 0.05).

to the limited data in the literature, especially for measures such as live birth rate and neonatal outcomes of blastocysts exhibiting ABNCL. In our study, no significant differences were found in live birth rates, gestational age and birthweight when comparing ABNCL blastocysts to those without ABNCL.

In a recent study by Shavit *et al.* (2023), live birth rates and neonatal outcomes of DC embryos were reported using a mixture of single, double and triple embryo transfers. In the study, all fresh transfers were performed on Day 3 with untransferred embryos cultured to the blastocyst stage before cryopreservation. A subset of single frozen blastocyst transfers led to similar live birth rates between the DC and no DC groups following single frozen blastocyst transfer (Shavit *et al.*, 2023). This study is in agreement with our results, though the neonatal results were, however, omitted from their frozen blastocyst transfer analysis.

Interestingly, based on a small euploid blastocyst dataset, Ozbek *et al.* (2021) found reduced live birth rates in DC (33.3%, 8/24) and RC blastocysts (23.1%, 6/26) in comparison to the control group (55.9%, 133/238). However, the same study also showed an unexpectedly higher aneuploidy rate in the control blastocysts (57.6%) compared to DC (47.7%) and RC blastocysts (46.2%), implying a potentially biased dataset considering its small sample size. Regardless, our study has shown that there are similar live birth rates and neonatal outcomes despite ABNCL events once full blastulation is achieved. The results of this study are reassuring for counseling patients who are undertaking blastocyst transfers.

Accurate and comprehensive annotations of early developmental events are critical for embryo selection (Lee *et al.*, 2024). Although TLI offers more reliable annotation in comparison to traditional assessment based on static observation (Scott *et al.*, 2021),

the increasing number of layers of cells and decreasing blastomere size create challenges for accurate annotation as the embryo develops into later stages. Our study demonstrated that DC/RC can be reliably tracked from the 1-cell up to 8-cell stage (i.e. completion of cell division by each one of the 4-cell blastomeres), enabling annotation of a more complete early cleavage history of the embryo. Reports including annotation at such a level of detail have been scarce in the literature. [Barrie et al. \(2017\)](#) reported a comprehensive analysis of a range of ABNCL patterns, including DC, RC Type I and Type II (termed absent cleavage in their study). However, neither blastulation data nor the ratios of affected blastomeres over all blastomeres were analyzed. [Shavit et al. \(2023\)](#) reported reduced blastulation when a single early DC event occurred at the first mitotic division or multiple DC events occurred during subsequent mitotic divisions, though a single late DC event (up to the third cleavage) did not affect blastulation. This is in general agreement with the findings of our study and others have also reported similar trends ([Hlinka et al., 2012](#); [Zhan et al., 2016](#)). By considering the combined impacts of DC/RC up to Day 3, we reported a structured classification system based on the ratios of blastomeres affected over all blastomeres and presented incremental impacts on subsequent blastulation. A full blastulation rate as low as 7.7% was observed when all (ratio of 1) blastomeres were affected by DC/RC, in comparison to 66.2% full blastulation when free from any DC/RC. Similar results were reported by [Yang et al. \(2015\)](#) based on a much smaller sample size ($n = 345$), where a decreasing blastulation rate was identified in embryos with a mixture of several different ABNCL patterns (including DC but without a detailed breakdown analysis) as the cleavage cycles of ABNCL occurrence advanced. However, the authors did not quantify the degree of impact by ABNCL. Such quantitative data presented in our study would potentially facilitate additional selection amongst DC/RC embryos at Day 3 transfer, where extended culture is not possible or offered.

Recently, 3D embryo analysis was utilized to report greater blastocyst development of embryos that had a higher number of cell-contact points at both the 4-cell and 8-cell stages ([He et al., 2024](#)). A greater number of contact points at the 8-cell stage was also associated with live birth, but not at the 4-cell stage ([He et al., 2024](#)). There is evidence showing planar embryos at the 4-cell stage, a sub-type of <6ICCP under our criteria, is associated with a range of irregular cleavages and decreased blastulation ([Ebner et al., 2017](#)). Moreover, [Desai and Gill \(2019\)](#) compared 4-cell embryos with non-tetrahedral blastomere spatial arrangements, considered equivalent to <6ICCP under our criteria, to their tetrahedral counterparts. The authors reported suboptimal blastulation outcomes in the non-tetrahedral group, but similar live birth rates based on 79 single transferred blastocysts with a non-tetrahedral spatial arrangement at the 4-cell stage, which is also in line with our findings. The comparable neonatal outcomes reported in our study offers additional reassurance to patients transferring blastocysts showing a history of <6ICCP at the 4-cell stage.

The association between increasing full blastulation capability and decreasing ratios of DC/RC impact supports a potential self-correction mechanism where embryos autonomously exclude or extrude unwanted blastomere(s) ([Coticchio et al., 2021](#); [Parriego et al., 2023](#)). Studies have noted the ejection of abnormal cells resulting from ABNCL blastocysts ([Zhan et al., 2016](#); [Lagalla et al., 2017](#); [Ottoolini et al., 2017](#)) and there is evidence that a euploid embryo can expel aneuploid blastomeres ([Orvieto et al., 2020](#)). It is also hypothesized that cell exclusion or extrusion is a mechanism for aneuploidy rescue and a method of self-correction to keep

euploidy status. This notion is supported by the findings of [Yang et al. \(2015\)](#), where 71.1% of DC daughter cells did not form part of the blastocyst. Our results warrant further follow-up research tracking participation of DC/RC blastomeres in the compaction and blastulation processes to facilitate better understanding of their corresponding clinical outcomes following transfer. Secondly, alternative self-correction mechanisms are suggested by the findings of our study, where embryos 100% affected by DC/RC are still able to reach full blastulation, albeit at a very low rate (5.4% in DC, 13.7% in RC, and 7.7% combined), and further give rise to a live birth. In our dataset, one out of three transferred blastocysts, that had been 100% affected by DC, achieved a live birth. Unfortunately, our dataset does not include any transferred blastocyst that was 100% affected by RC. Self-correction mechanisms in embryos with 100% blastomeres affected by DC/RC is currently unclear. The underlying mechanism may be shared with single- or multi-pronuclear embryos, backed by emerging evidence showing a reasonable euploidy rate and occasional healthy live birth ([Hirata et al., 2020](#); [Bredbacka et al., 2023](#)). Accordingly, further clinical data is required to better understand their true prognosis.

Although the present study's strengths include a large sample size and consistent annotation of ABNCL over many years, there are limitations that must be addressed. Our study is limited by its retrospective nature as it is difficult to control every known and unknown confounder. Endometrial preparation protocols for frozen embryo transfers were not available in our dataset, although hormone replacement cycles were predominantly used. Some studies have reported different live birth rates with the use of hormone replacement and natural cycles ([Zheng et al., 2021](#); [Wang et al., 2023](#)). All ABNCL events were annotated manually, which can have a degree of subjectivity ([Lee et al., 2024](#)). There are inter- and intra-operator variations in the annotation of time-lapse measures, including for ABNCL ([Sundvall et al., 2013](#); [Adolfsson and Andershed, 2018](#)). Although qualitative measurements like ABNCL are more reproducible than quantitative measurements ([Liu et al., 2020](#)), future application of automated annotation of ABNCL events by deep learning would reduce the subjective nature of manual annotation. The wider literature has a range of definitions of DC including rapid cleavage (<5 h for $t3 - t2$) ([Meseguer et al., 2011](#); [Rubio et al., 2012](#)) and direct uneven cleavage (≤ 5 h between mother and daughter cell division, or single blastomere directly dividing into three or more blastomeres) ([Zhan et al., 2016](#)). The present study considered trichotomous mitosis (single cell dividing directly to three daughter cells at any stage) ([Ciray et al., 2014](#)) during annotation and did not consider other types of DC. Furthermore, it is time consuming to build up the live birth dataset for blastocysts involving ABNCL, as they are routinely ranked low priority for transfer in most clinics. Additionally, longitudinal studies of babies born from embryos showing ABNCL are warranted to compare developmental outcomes such as congenital anomalies. Finally, this study was not designed to investigate causal factors of ABNCL occurrence. Two types of culture media were utilized during the study period, and they have been recognized to result in different morphokinetic profiles and the number of good-quality embryos ([Hardarson et al., 2015](#)). However, we have controlled for this issue in our multiple logistic regression. Future focused studies are required to enable better understanding of such ABNCL events.

Conclusion

Our study highlights the important role of blastulation in embryos showing ABNCL up to Day 3, potentially by allowing them

to 'self-correct'. ABNCL embryos have a lower blastocyst formation rate overall; only embryos that self-corrected successfully formed blastocysts. However, once full blastulation has been achieved, ABNCL is not associated with any compromise in subsequent live birth rates, gestational age or birthweight.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Acknowledgements

The authors thank the embryology team at Fertility North for their kind support of this work.

Authors' roles

T.L. and Y.L. conceived this study and drafted the initial manuscript. T.L. and Y.L. conducted the literature search, data analysis, and data interpretation. All authors contributed to editing toward the final manuscript and approved the final version for submission.

Funding

This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

Conflict of interest

All authors report no conflict of interest.

References

- Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Hum Reprod* 2011;**26**:1270–1283.
- Adolfsson E, Andershed AN. Morphology vs morphokinetics: a retrospective comparison of inter-observer and intra-observer agreement between embryologists on blastocysts with known implantation outcome. *JBRA Assist Reprod* 2018;**22**:228–237.
- Athayde Wirka K, Chen AA, Conaghan J, Ivani K, Gvakharia M, Behr B, Suraj V, Tan L, Shen S. Atypical embryo phenotypes identified by time-lapse microscopy: high prevalence and association with embryo development. *Fertil Steril* 2014;**101**:1637–1648.e1631–1635.
- Barrie A, Homburg R, McDowell G, Brown J, Kingsland C, Troup S. Preliminary investigation of the prevalence and implantation potential of abnormal embryonic phenotypes assessed using time-lapse imaging. *Reprod Biomed Online* 2017;**34**:455–462.
- Bickendorf K, Qi F, Peirce K, Natalwala J, Chapple V, Liu Y. Spontaneous collapse as a prognostic marker for human blastocysts: a systematic review and meta-analysis. *Hum Reprod* 2023;**38**:1891–1900.
- Blondel B, Macfarlane A, Gissler M, Breart G, Zeitlin J; PERISTAT Study Group. Preterm birth and multiple pregnancy in European countries participating in the PERISTAT project. *BJOG* 2006;**113**:528–535.
- Bredbacka P, Capalbo A, Kananen K, Picchetta L, Tomás C. Healthy live birth following embryo transfer of a blastocyst of tetrapronuclear (4PN) origin: a case report. *Hum Reprod* 2023;**38**:1700–1704.
- Ciray HN, Campbell A, Agerholm IE, Aguilar J, Chamayou S, Esbert M, Sayed S; Time-Lapse User Group. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Hum Reprod* 2014;**29**:2650–2660.
- Coticchio G, Barrie A, Lagalla C, Borini A, Fishel S, Griffin D, Campbell A. Plasticity of the human preimplantation embryo: developmental dogmas, variations on themes and self-correction. *Hum Reprod Update* 2021;**27**:848–865.
- D'Antonio F, Khalil A, Dias T, Thilaganathan B; Southwest Thames Obstetric Research Collaborative (STORK). Weight discordance and perinatal mortality in twins: analysis of the Southwest Thames Obstetric Research Collaborative (STORK) multiple pregnancy cohort. *Ultrasound Obstet Gynecol* 2013;**41**:643–648.
- Desai N, Gill P. Blastomere cleavage plane orientation and the tetrahedral formation are associated with increased probability of a good-quality blastocyst for cryopreservation or transfer: a time-lapse study. *Fertil Steril* 2019;**111**:1159–1168.e1151.
- Ebner T, Höggerl A, Oppelt P, Radler E, Enzelsberger SH, Mayer RB, Petek E, Shebl O. Time-lapse imaging provides further evidence that planar arrangement of blastomeres is highly abnormal. *Arch Gynecol Obstet* 2017;**296**:1199–1205.
- ESHRE Working group on Time-lapse Technology; Apter S, Ebner T, Freour T, Guns Y, Kovacic B, Le Clef N, Marques M, Meseguer M, Montjean D, Sfontouris I et al. Good practice recommendations for the use of time-lapse technology. *Hum Reprod Open* 2020;**2020**:hoaa008.
- Fan YL, Han SB, Wu LH, Wang YP, Huang GN. Abnormally cleaving embryos are able to produce live births: a time-lapse study. *J Assist Reprod Genet* 2016;**33**:379–385.
- Fréour T, Dessolle L, Lammers J, Lattes S, Barrière P. Comparison of embryo morphokinetics after in vitro fertilization-intracytoplasmic sperm injection in smoking and nonsmoking women. *Fertil Steril* 2013;**99**:1944–1950.
- Gardner DK, Meseguer M, Rubio C, Treff NR. Diagnosis of human preimplantation embryo viability. *Hum Reprod Update* 2015;**21**:727–747.
- Gardner DK, Schoolcraft WB. Culture and transfer of human blastocysts. *Curr Opin Obstet Gynecol* 1999;**11**:307–311.
- Giorgione V, Briffa C, Di Fabrizio C, Bhat R, Khalil A. Perinatal outcomes of small for gestational age in twin pregnancies: twin vs. singleton charts. *J Clin Med* 2021;**10**:1–9.
- Glatstein I, Chavez-Badiola A, Curchoe CL. New frontiers in embryo selection. *J Assist Reprod Genet* 2023;**40**:223–234.
- Gorodeckaja J, Neumann S, McCollin A, Ottolini CS, Wang J, Ahuja K, Handyside A, Summers M. High implantation and clinical pregnancy rates with single vitrified-warmed blastocyst transfer and optional aneuploidy testing for all patients. *Hum Fertil (Camb)* 2020;**23**:256–267.
- Hardarson T, Bungum M, Conaghan J, Meintjes M, Chantilis SJ, Molnar L, Gunnarsson K, Wikland M. Noninferiority, randomized, controlled trial comparing embryo development using media developed for sequential or undisturbed culture in a time-lapse setup. *Fertil Steril* 2015;**104**:1452–1459.e1451–1454.
- He C, Karpaviciūtė N, Hariharan R, Lees L, Jacques C, Ferrand T, Chambost J, Wouters K, Malmsten J, Miller R et al. Seeking arrangements: cell contact as a cleavage-stage biomarker. *Reprod Biomed Online* 2024;**48**:103654.
- Hirata K, Goto S, Izumi Y, Taguchi M, Hayashi A, Fujioka M, Ishiko A, Nakanishi K, Kimura F, Murakami T. Chromosome analysis of blastocysts derived from single pronuclear zygotes by array CGH and clinical outcomes by the transfer of single pronuclear zygotes. *J Assist Reprod Genet* 2020;**37**:1645–1652.

- Hlinka D, Kařatová B, Uhrinová I, Dolinská S, Rutarová J, Rezáčová J, Lazarovská S, Dudáš M. Time-lapse cleavage rating predicts human embryo viability. *Physiol Res* 2012;**61**:513–525.
- Ho JR, Arrach N, Rhodes-Long K, Salem W, McGinnis LK, Chung K, Bendikson KA, Paulson RJ, Ahmady A. Blastulation timing is associated with differential mitochondrial content in euploid embryos. *J Assist Reprod Genet* 2018;**35**:711–720.
- Jin L, Dong X, Tan W, Huang B. Incidence, dynamics and recurrences of reverse cleavage in aneuploid, mosaic and euploid blastocysts, and its relationship with embryo quality. *J Ovarian Res* 2022;**15**:91.
- Kasterstein E, Strassburger D, Komarovskiy D, Bern O, Komsky A, Razieli A, Friedler S, Ron-El R. The effect of two distinct levels of oxygen concentration on embryo development in a sibling oocyte study. *J Assist Reprod Genet* 2013;**30**:1073–1079.
- Lagalla C, Tarozzi N, Sciajano R, Wells D, Di Santo M, Nadalini M, Distratis V, Borini A. Embryos with morphokinetic abnormalities may develop into euploid blastocysts. *Reprod Biomed Online* 2017;**34**:137–146.
- Lee T, Natalwala J, Chapple V, Liu Y. A brief history of artificial intelligence embryo selection: from black-box to glass-box. *Hum Reprod* 2024;**39**:285–292.
- Liu Y, Chapple V, Feenan K, Roberts P, Matson P. Clinical significance of intercellular contact at the four-cell stage of human embryos, and the use of abnormal cleavage patterns to identify embryos with low implantation potential: a time-lapse study. *Fertil Steril* 2015;**103**:1485–1491.e1481.
- Liu Y, Chapple V, Feenan K, Roberts P, Matson P. Time-lapse deselection model for human day 3 in vitro fertilization embryos: the combination of qualitative and quantitative measures of embryo growth. *Fertil Steril* 2016;**105**:656–662.e651.
- Liu Y, Chapple V, Roberts P, Matson P. Prevalence, consequence, and significance of reverse cleavage by human embryos viewed with the use of the Embryoscope time-lapse video system. *Fertil Steril* 2014;**102**:1295–1300.e1292.
- Liu Y, Qi F, Matson P, Morbeck DE, Mol BW, Zhao S, Afnan M. Between-laboratory reproducibility of time-lapse embryo selection using qualitative and quantitative parameters: a systematic review and meta-analysis. *J Assist Reprod Genet* 2020;**37**:1295–1302.
- Meseguer M, Herrero J, Tejera A, Hilligsøe KM, Ramsing NB, Remohí J. The use of morphokinetics as a predictor of embryo implantation. *Hum Reprod* 2011;**26**:2658–2671.
- Orevich LS, Watson K, Ong K, Korman I, Turner R, Shaker D, Liu Y. Morphometric and morphokinetic differences in the sperm- and oocyte-originated pronuclei of male and female human zygotes: a time-lapse study. *J Assist Reprod Genet* 2022;**39**:97–106.
- Orvieto R, Shimon C, Rienstein S, Jonish-Grossman A, Shani H, Aizer A. Do human embryos have the ability of self-correction? *Reprod Biol Endocrinol* 2020;**18**:98.
- Ottolini CS, Kitchen J, Xanthopoulou L, Gordon T, Summers MC, Handyside AH. Tripolar mitosis and partitioning of the genome arrests human preimplantation development in vitro. *Sci Rep* 2017;**7**:9744.
- Ozbek IY, Mumusoglu S, Polat M, Bozdag G, Sokmensuer LK, Yarali H. Comparison of single euploid blastocyst transfer cycle outcome derived from embryos with normal or abnormal cleavage patterns. *Reprod Biomed Online* 2021;**42**:892–900.
- Parriego M, Coll L, Carrasco B, Garcia S, Boada M, Polyzos NP, Vidal F, Veiga A. Blastocysts from partial compaction morulae are not defined by their early mistakes. *Reprod Biomed Online* 2023;**48**:103729.
- Petersen BM, Boel M, Montag M, Gardner DK. Development of a generally applicable morphokinetic algorithm capable of predicting the implantation potential of embryos transferred on Day 3. *Hum Reprod* 2016;**31**:2231–2244.
- Practice Committee of the American Society for Reproductive Medicine and the Practice Committee for the Society for Assisted Reproductive Technologies. Guidance on the limits to the number of embryos to transfer: a committee opinion. *Fertil Steril* 2021;**116**:651–654.
- Rubio I, Kuhlmann R, Agerholm I, Kirk J, Herrero J, Escibá MJ, Bellver J, Meseguer M. Limited implantation success of direct-cleaved human zygotes: a time-lapse study. *Fertil Steril* 2012;**98**:1458–1463.
- Scott N, Barrie A, Smith R, Best L, Davis N, Duffy S, Krokos S, Lodge Y, Montgomery S, O'Boyle S et al. P-792 The introduction of an embryo morphokinetics annotation quality assurance scheme across ten fertility clinics including 59 participants. *Hum Reprod* 2021;**36**:Suppl_1.
- Shavit M, Gonen D, Atzmon Y, Aslih N, Bilgory A, Shibli Abu-Raya Y, Sharqawi M, Estrada Garcia D, Michaeli M, Polotov D et al. Cleavage patterns of 9600 embryos: the importance of irregular cleavage. *J Clin Med* 2023;**12**:5656.
- Sundvall L, Ingerslev HJ, Breth Knudsen U, Kirkegaard K. Inter- and intra-observer variability of time-lapse annotations. *Hum Reprod* 2013;**28**:3215–3221.
- Wainstock T, Yoles I, Sergienko R, Sheiner E. Twins vs singletons-long-term health outcomes. *Acta Obstet Gynecol Scand* 2023;**102**:1000–1006.
- Wang Z, Zhang Y, Shang X, Miao R, Yin M, Yang H, Yu Y, Wei D. The likelihood of a healthy live birth after frozen embryo transfer with endometrium prepared by natural ovulation regimen vs programmed regimen: a propensity-score matching study. *AJOG Glob Rep* 2023;**3**:100210.
- Yang ST, Shi JX, Gong F, Zhang SP, Lu CF, Tan K, Leng LZ, Hao M, He H, Gu YF et al. Cleavage pattern predicts developmental potential of day 3 human embryos produced by IVF. *Reprod Biomed Online* 2015;**30**:625–634.
- Zhan Q, Ye Z, Clarke R, Rosenwaks Z, Zaninovic N. Direct unequal cleavages: embryo developmental competence, genetic constitution and clinical outcome. *PLoS One* 2016;**11**:e0166398.
- Zheng Q, Zhang H, Xu S, Xiao S, Wang X, Mo M, Zeng Y. Optimal endometrial preparation protocols for frozen-thawed embryo transfer cycles by maternal age. *Reprod Sci* 2021;**28**:2847–2854.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Human Reproduction, 2024, 39, 955–962
<https://doi.org/10.1093/humrep/deae062>
Original Article

Infertility

Risk factors for poor oocyte yield and oocyte immaturity after GnRH agonist triggering

S. Gambini^{1,†}, C. Sonigo^{2,†}, G. Robin³, I. Cedrin-Durnerin¹, C. Vinolas¹, C. Sifer⁴, Y. Boumerdassi⁴, A. Mayeur⁵, V. Gallot², M. Grynberg^{1,2}, and M. Peigné^{1,*}

¹Assistance Publique–Hôpitaux de Paris (APHP), Université Sorbonne Paris Nord, Service de Médecine de la Reproduction et Préservation de la Fertilité, Hôpital Jean-Verdier, Bondy, France

²Assistance Publique–Hôpitaux de Paris (APHP), Université Paris Saclay, Service de Médecine de la Reproduction et Préservation de la Fertilité, Hôpital Antoine Bécélère, Clamart, France

³CHU de Lille, Université de Lille, Service de Médecine de la Reproduction et Préservation de la Fertilité, Hôpital Jeanne de Flandre, Lille, France

⁴Assistance Publique–Hôpitaux de Paris (APHP), Université Sorbonne Paris Nord, Unité de Biologie de la Reproduction, Hôpital Jean-Verdier, Bondy, France

⁵Assistance Publique–Hôpitaux de Paris (APHP), Université Paris Saclay, Service de Biologie de la Reproduction, Hôpital Antoine Bécélère, Clamart, France

*Correspondence address. Assistance Publique–Hôpitaux de Paris (APHP), Université Sorbonne Paris Nord, Service de Médecine de la Reproduction et Préservation de la Fertilité, Hôpital Jean-Verdier, Avenue du 14 juillet, 93140 Bondy, France. E-mail: maeliss.peigne@aphp.fr <https://orcid.org/0000-0001-9987-0740>

[†]These authors contributed equally to this work.

ABSTRACT

STUDY QUESTION: What are the potential risk factors for poor oocyte recuperation rate (ORR) and oocyte immaturity after GnRH agonist (GnRHa) ovulation triggering?

SUMMARY ANSWER: Lower ovarian reserve and LH levels after GnRHa triggering are risk factors of poor ORR. Higher BMI and anti-Müllerian hormone (AMH) levels are risk factors of poor oocyte maturation rate (OMR).

WHAT IS KNOWN ALREADY: The use of GnRHa to trigger ovulation is increasing. However, some patients may have a suboptimal response after GnRHa triggering. This suboptimal response can refer to any negative endpoint, such as suboptimal oocyte recovery, oocyte immaturity, or empty follicle syndrome. For some authors, a suboptimal response to GnRHa triggering refers to a suboptimal LH and/or progesterone level following triggering. Several studies have investigated a combination of demographic, clinical, and endocrine characteristics at different stages of the treatment process that may affect the efficacy of the GnRHa trigger and thus be involved in a poor endocrine response or efficiency but no consensus exists.

STUDY DESIGN, SIZE, DURATION: Bicentric retrospective cohort study between 2015 and 2021 (N = 1747).

PARTICIPANTS/MATERIALS, SETTING, METHODS: All patients aged 18–43 years who underwent controlled ovarian hyperstimulation and ovulation triggering by GnRHa alone (triptorelin 0.2 mg) for ICSI or oocyte cryopreservation were included. The ORR was defined as the ratio of the total number of retrieved oocytes to the number of follicles >12 mm on the day of triggering. The OMR was defined as the ratio of the number of mature oocytes to the number of retrieved oocytes. A logistic regression model with a backward selection method was used for the analysis of risk factors. Odds ratios (OR) are displayed with their two-sided 95% confidence interval.

MAIN RESULTS AND THE ROLE OF CHANCE: In the multivariate analysis, initial antral follicular count and LH level 12-h post-triggering were negatively associated with poor ORR (i.e. below the 10th percentile) (OR: 0.61 [95% CI: 0.42–0.88]; P = 0.008 and OR: 0.86 [95% CI: 0.76–0.97]; P = 0.02, respectively). A nonlinear relationship was found between LH level 12-h post-triggering and poor ORR, but no LH threshold was found. A total of 25.3% of patients suffered from oocyte immaturity (i.e. OMR < 75%). In the multivariate analysis, BMI and AMH levels were negatively associated with an OMR < 75% (OR: 4.34 [95% CI: 1.96–9.6]; P < 0.001 and OR: 1.22 [95% CI: 1.03–1.12]; P = 0.015, respectively). Antigonadotrophic pretreatment decreased the risk of OMR < 75% compared to no pretreatment (OR: 0.72 [95% CI: 0.57–0.91]; P = 0.02).

LIMITATIONS, REASONS FOR CAUTION: Our study is limited by its retrospective design and by the exclusion of patients who had hCG retriggerers. However, this occurred in only six cycles. We were also not able to collect information on the duration of pretreatment and the duration of wash out period.

WIDER IMPLICATIONS OF THE FINDINGS: In clinical practice, to avoid poor ORR, GnRHa trigger alone should not be considered in patients with higher BMI and/or low ovarian reserve, balanced by the risk of ovarian hyperstimulation syndrome. In the case of a low 12-h post-triggering LH level, practitioners must be aware of the risk of poor ORR, and hCG retriggering could be considered.

STUDY FUNDING/COMPETING INTEREST(S): None.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: GnRH agonist trigger / oocyte recuperation rate / oocyte yield / oocyte maturity / immaturity

Received: June 21, 2023. Revised: December 20, 2023. Editorial decision: February 7, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved. For permissions, please email: journals.permissions@oup.com

Introduction

The aim of controlled ovarian hyperstimulation (COH) for IVF or fertility preservation (FP) cycles is to induce multifollicular ovarian development leading to the yield of an optimal number of mature oocytes to optimize the outcomes while avoiding the risk of ovarian hyperstimulation syndrome (OHSS). Indeed, recent studies have shown that a positive association exists between the number of oocytes retrieved and the cumulative live birth rate (Cobo et al., 2021; Law et al., 2021).

Final oocyte maturation after COH has been induced for several decades by hCG, which acts directly on LH receptors (Fauser et al., 2002). hCG is a surrogate of natural LH but has a longer half-life (>24 h vs 60 min for LH) and consequently a prolonged stimulatory effect on multiple corpora lutea, which can induce the release of vascular endothelial growth factor (VEGF), causing capillary hyperpermeability that may lead to OHSS (Pfeifer et al., 2016).

The use of a single bolus of GnRH agonist (GnRHa) to trigger ovulation instead of hCG was suggested for the first time in 1988 in a conference abstract (Itskovitz et al., 1991). However, with the advent of the long GnRHa downregulation protocol, the use of GnRHa triggering was not further explored because the simultaneous use of GnRHa for downregulation and triggering is not possible.

When the GnRH antagonist protocol was introduced for the prevention of premature LH surge (Itskovitz-Eldor et al., 1998; The European Orgalutran® Study Group et al., 2000), the use of GnRHa instead of hCG as an ovulation trigger became an option again. Indeed, GnRHa can displace the GnRH antagonist from the receptor and induce an initial activation (flare-up) prior to downregulation of the receptor, leading to a concomitant endogenous LH and FSH surge (Humaidan et al., 2005).

In most women, the natural LH surge is characterized by a rapidly ascending phase of 14 h, followed by a 14-h plateau and then, a descending phase of 20 h (i.e. 48 h in total) (Hoff et al., 1983). After GnRHa administration, LH level increases 4 h after the GnRHa injection and FSH levels 12 h after, and stay elevated for 24–36 h. Compared to the natural cycle, the amplitude of the surges is similar, but their duration is shorter (Humaidan et al., 2011a). This results in acute irreversible luteolysis and thus decreased expression of VEGF and luteal phase insufficiency. Therefore, GnRHa, as an ovulation trigger, has been introduced in COH cycles in antagonist protocols, to prevent OHSS (Engmann et al., 2008; Youssef et al., 2014), and is now routinely used. Due to luteal phase insufficiency, the first randomized controlled trials reported poor pregnancy outcomes with a high early pregnancy loss rate when GnRHa was used to trigger ovulation (Humaidan et al., 2005; Kolibianakis et al., 2005). Currently, several options exist for luteal phase support after GnRHa triggering and result in good live birth rates that are not significantly different from those after hCG triggering (Benadiva and Engmann, 2018), once again increasing the use of GnRHa triggering. However, to avoid this abnormal luteal phase, many teams utilize the “freeze all” policy, freezing all embryos and transferring them in subsequent cycles.

Compared to hCG, another possible advantage of ovulation triggering by GnRHa is the simultaneous releasing of FSH similar to the natural gonadotropin ovulation surges of the natural cycle. Physiologically, in a natural cycle, a concomitant LH and FSH rise induces final oocyte and follicular maturation and ovulation. In luteinizing granulosa cells, FSH induces the formation of LH receptors, thus participating to the function of the corpus luteum. FSH also seems to enhance the resumption of meiosis

(Andersen, 1999) and the synthesis of a matrix of hyaluronic acid that contributes to the expansion of cumulus cells, allowing the oocyte–cumulus cell mass to be free-floating in the antrum (Orvieto, 2015). Differences in the follicular fluid of preovulatory follicles are also observed after GnRHa compared to hCG triggering: there are higher intrafollicular levels of gonadotropin, lower levels of progesterone (Andersen et al., 2006), and a lower level of epidermal growth factor-like amphiregulin (Humaidan et al., 2011b), which could explain why GnRHa triggering seemed to increase oocyte quality (by increasing the total number of embryos obtained) in a recent meta-analysis (Bourdon et al., 2021).

However, some patients may have a suboptimal response after GnRHa triggering. This suboptimal response can refer to any negative endpoint, such as suboptimal oocyte recovery (Shapiro et al., 2011; Chen et al., 2012; Kummer et al., 2013; Popovic-Todorovic et al., 2019), oocyte immaturity (Shapiro et al., 2011; Kummer et al., 2013), or empty follicle syndrome (EFS; complete lack of oocytes) (Deepika et al., 2018). Indeed, different ways and measures exist to assess the efficacy of GnRHa triggering: oocyte yield, oocyte recovery rate, and oocyte maturation rate (OMR). For some authors, a suboptimal response to GnRHa triggering refers to a suboptimal LH and/or progesterone level following triggering (Meyer et al., 2015; Chang et al., 2016), usually defined as a serum LH level below 15 IU/l or serum progesterone levels below 3 ng/ml 8–12 h post-triggering (Russo et al., 2020).

Several studies (Popovic-Todorovic et al., 2019; Lainas et al., 2020; Humaidan and Kol, 2021; Wiltshire et al., 2021; Ganer Herman et al., 2022) have investigated or reviewed different clinical and hormonal characteristics at several points of the COH process that could affect the GnRHa trigger efficacy and thus be involved in a poor endocrine response or efficiency, but no consensus exists.

The objective of the present study was to explore the risk factors for poor oocyte yield and oocyte immaturity after GnRHa trigger in a large cohort of patients undergoing COH in the GnRH antagonist protocol.

Materials and methods

Study design

This retrospective cohort study was conducted in two medically assisted reproductive centers (Antoine Bécclère Hospital, Clamart and Jean Verdier Hospital, Bondy).

Patients' inclusion and exclusion criteria

We included all patients aged 18–43 years who underwent a COH with a GnRH antagonist protocol for IVF with ICSI or FP (only for medical reasons, as social freezing was not authorized in France before 2022), who were triggered by GnRHa only and who had a “freeze all” strategy (freezing of all mature oocytes or embryos obtained) between January 2015 and December 2021.

For the best evaluation of oocyte maturity, we selected only patients who had undergone ICSI or FP with cryopreservation of mature oocytes, assessed on the day of oocyte retrieval.

Patients who received hCG (in combination with GnRHa or in case of suboptimal hormonal response to GnRHa), patients who underwent a fresh embryo transfer after GnRHa trigger, patients who underwent IVF–ICSI cycles with testicular sperm, patients who had no oocyte retrieval despite triggering, patients who underwent cycles with the use of letrozole, and patients in whom the pretreatment before COH was unknown were excluded (Fig. 1).

An initial workup including antral follicular count (AFC) and hormone assays at the beginning of the cycle was performed for

most of the patients. AFC was evaluated using a transvaginal ovarian ultrasound and included all ovarian follicles between 2- and 9-mm. Serum anti-Müllerian hormone (AMH) levels were determined using an ultrasensitive ELISA. We also retrospectively extracted data from the medical records, including chronological age, BMI, and smoking status.

Ovarian stimulation protocols and cryopreservation of mature oocytes and embryos

A pretreatment before starting COH could be administered either with antigonadotrophic medication (estrogen–progestin pill or macroprogestative treatment), non-antigonadotrophic (progestin-only pill, dydrogesterone (Duphaston®, Abbott, Netherlands), or estradiol (Provames®, Delpharm Lille SAS, France)).

Ovarian stimulation was initiated in the follicular phase for ICSI cycles and in either the follicular or luteal phase for FP cycles.

COH was achieved with the use of FSH ± LH. The starting gonadotropin dose (between 75 and 600 IU/day) was based on patient chronological age, BMI, AFC, AMH level, and prior response to gonadotropins and could be adapted during stimulation according to the patient's ovarian response.

GnRH antagonist at a daily dose of 0.25 mg was added in a fixed manner on the 6th day of the stimulation or in a flexible

manner depending on serum estradiol (E2) and leading follicle size. COH was monitored every 48–72 h from the 6th or 8th day of stimulation, by blood sampling (determination of E2, LH, and progesterone levels) and transvaginal ultrasonography.

Trigger was decided when three or more 17–20 mm follicles were detected on transvaginal ultrasonography. Final oocyte maturation was triggered with GnRH α : 0.2 mg of subcutaneous triptorelin (Decapeptyl®, Ipsen, France). LH, E2, and progesterone plasma levels were measured on the day of the triggering in most patients. Missing LH levels on the day of trigger were extrapolated from values one or two days before triggering. As LH levels on the day of trigger were evaluated by classes (≤ 0.5 ; 0.51–0.99; 1.00–2.99; ≥ 3 IU/l), we extrapolated that, for women who did not have hormonal dosages on the day of triggering ($n=1125$), serum LH level on the trigger day was in the same range as 1 or 2 days before. E2, LH, and progesterone levels were determined ~12 h after GnRH α trigger.

Oocytes were retrieved transvaginally 36–40 h after triptorelin administration, using a 17- or a 19-gauge needle (Cook, France) guided by transvaginal ultrasonography, under local anesthesia (in most cases) or general anesthesia. Aspiration pressure was between 120 and 150 mmHg. All follicles with diameters >12 mm were aspirated. Oocyte retrievals were performed by two teams

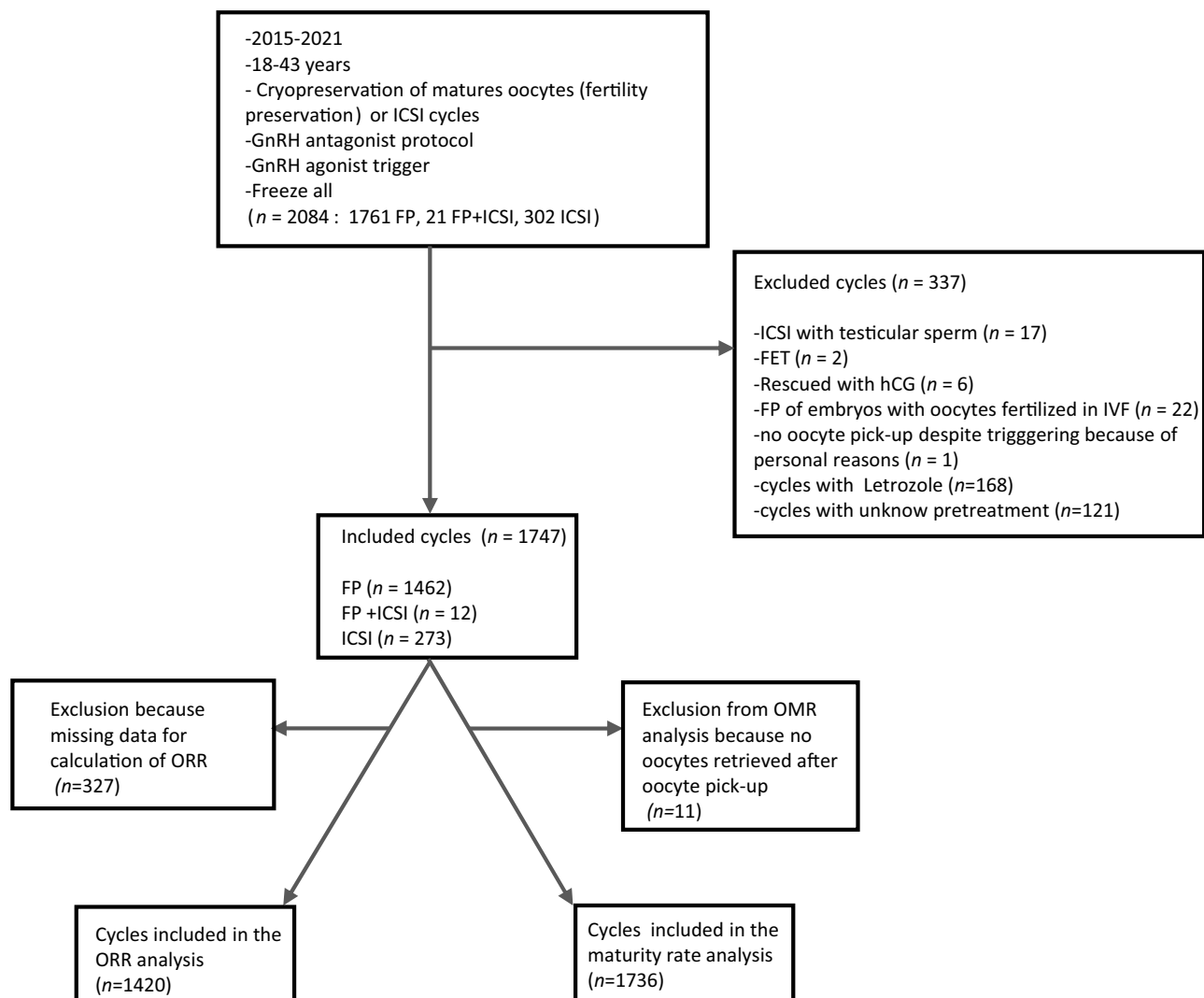


Figure 1. Flow chart summarizing patient inclusion and exclusion criteria. FET, fresh embryo transfer; FP, fertility preservation; GnRH α , GnRH agonist; OMR: oocyte maturation rate; ORR: oocyte recuperation rate.

(from the two different centers) using the same protocol. In this protocol, we had a live report of the number of oocytes recovered. Flushing was performed if no oocyte was recovered in the first two follicles and, if needed, was performed three times flushing for each follicle.

The search for cumulo-oocyte complexes (COCs) was carried out upon reception of the follicular fluid tubes, using a binocular magnifying glass. COCs were then placed in G-Gamet® (Vitrolife) in incubators, at 37°C, under 6% CO₂. Between 1 and 2 h after retrieval, the COCs were denuded by enzymatic treatment (hyaluronidase 80 UI/ml, FertiPro) and mechanical stripping (Stripper®), Origio. Oocyte maturity was then assessed. Mature (MII) oocytes were vitrified or microinjected with one spermatozoa.

Outcome variables

The primary end point was the oocyte recuperation rate (ORR), the ratio of the total number of collected oocytes to the number of follicles measuring ≥ 12 mm on the day of triggering. If no ultrasound had been performed on the day of triggering, the follicle size was extrapolated from the ultrasound results available 1 or 2 days before triggering (adding 2 mm in diameter per day at each follicle).

As previously defined in the literature, a poor ORR was defined as an oocyte yield below the 10th percentile (Popovic-Todorovic et al., 2019; Ganer Herman et al., 2022). The secondary endpoints were the OMR and the number of COCs retrieved.

OMR was defined as the ratio of mature (MII) oocytes to the number of total retrieved oocytes. As previously described, oocyte immaturity was defined as OMR <75% (Bar-Ami et al., 1994). Impaired GnRHa triggering was defined as an ORR <10th percentile or an OMR <75%.

Statistical analysis

Continuous data are summarized by medians (interquartile range, 25–75th) and categorical data are presented as frequency (percentage) of documented categories. Groups were compared by the Mann-Whitney test for continuous data and the chi-square test for categorical data (*P* without *). A logistic regression model was used to identify factors associated with an impaired GnRHa trigger. When a normal distribution could not be assumed, natural logarithmic or square root transformations were used to try to normalize the distribution. When it failed to normalize the distribution, variables were divided into classes (in terciles or according to literature data). After univariate logistic regression, “*P*” is presented with *.

A multivariate logistic regression model was used for the analysis of risk factors of impaired GnRHa trigger. A backwards stepwise selection method was used, starting with all parameters with a *P* value ≤ 0.10 in univariable analysis. The final model included all parameters with a *P* value <0.05. Two-way interactions were tested on the final model. In the multivariate analysis, only COS cycles with completed data of interest were analyzed. Odds ratios (OR) are displayed with their two-sided 95% Wald confidence intervals (95% CIs). Statistical analysis was performed with NCSS 2021 (Statistical Software, NCSS, LLC, Kaysville, UT, USA).

Ethical approval

All patients gave their written consent for the use of their medical data for scientific research purposes. A local ethic committee approved this study (CLEA-2022-257).

Results

Out of the 2084 patients who attended the centers between January 2015 and December 2021, 1747 patients fulfilled the selection criteria and were included in the present analysis. Overall, 1462 cycles for oocyte vitrification for FP (300 (20.5%) for oncologic indications), 273 ICSI cycles (1 oncologic FP indication), and 12 cycles for oocyte and embryo vitrification after ICSI for FP (9 for oncologic indication) were included (Fig. 1). A total of 1420 cycles were included in the ORR analysis and 1736 cycles were included in the OMR analysis.

Patient characteristics

Characteristics of the patients and of the cycles are summarized in Table 1. Half of the patients (881) had non-antigonadotrophic pretreatment (estradiol (817; 46.8%); microprogestin-only pill (43; 2.4%); and dydrogesterone (21; 1.2%)), and 25% (436) had antigonadotrophic pretreatment (estrogen–progestin pill (398; 22.8%) or a macroprogestin pill (38; 2.2%)). No pretreatment was administered to 24.6% (430) of the patients.

The median ORR and median OMR were 75.0% (57.7–100.0) and 85.7% (73.9–100.0), respectively. Oocyte immaturity (i.e., OMR <75%) was shown in 25.3% of the cycles (440/1736). The characteristics of the patients according to COH indication are shown in Supplementary Table S1.

Table 1. Characteristics of the whole population.

Patients characteristics	n = 1747
Age (years)	34.5 (29.9–37.0)
BMI (kg/m ²) (n = 1567)	21.97 (20.03–24.61)
Tobacco use (yes) (n = 1570)	305 (19.4)
AMH (ng/ml) (n = 1534)	2.10 (1.20–3.7)
AFC (n = 1668)	17 (11–25)
Cycle description	
Pretreatment	
Not antigonadotrophic	881 (50.4)
Antigonadotrophic	436 (25.0)
No pretreatment	430 (24.6)
Type of gonadotropin	
FSH	1477 (84.5)
FSH + LH activity	270 (15.5)
Initial dose of gonadotropin (IU/day)	300 (225–450)
Total dose of gonadotropin (IU)	3000 (2200–4300)
Timing of gonadotropin start (n = 1738)	
Follicular phase	1637 (94.2)
Luteal phase	101 (5.8)
Duration of stimulation (days)	11 (10–12)
Duration of GnRH antagonist administration (days)	5 (4–7)
Indication for GnRH agonist trigger	
OHSS	208 (11.9)
Fertility preservation	1486 (85.1)
Elevated progesterone on DT	32 (1.9)
Other	21 (1.2)
Oocyte recuperation rate (ORR) (%) (n = 1420)	75.0 (57.7–100.0)
COC obtained	11 (7–17)
Oocyte maturation rate (OMR) (%) (n = 1736)	85.7 (73.9–100.0)
Oocyte immaturity (OMR <75%) (yes) (n = 1736)	440 (25.3)

Data are presented as median (25th–75th percentile) or n (%). If some data are missing, the frequency of documented data is given in brackets. AFC: antral follicle count; AMH: anti-Müllerian hormone; Antigonadotrophic: estrogen–progestin pill or macroprogestin pill; COC: cumulo-oocyte complexes; DT: day of triggering; Not antigonadotrophic: microprogestin-only pill or dydrogesterone or estradiol; OHSS: ovarian hyperstimulation syndrome; OMR: ratio of mature oocytes to the number of total COC obtained; ORR: ratio of the number of COC obtained to the number of follicles measuring ≥ 12 mm on the day of the triggering.

Oocyte recuperation rate

The value of the 10th percentile of ORR was 36%. Patients with a poor ORR (below the 10th percentile) had fewer oocytes retrieved (4 (2–7) vs 12 (8–18); $P < 0.001^*$) and fewer mature oocytes retrieved (3 (2–6) vs 10 (6–14); $P < 0.001^*$) than the group with an ORR $\geq$ 10th percentile, despite a similar number of follicles $\geq$ 12 mm the day of triggering (15 (10–24) vs 15 (11–21); $P = 0.37$).

Patient and cycle characteristics according to ORR status ($<$ 10th percentile, $\geq$ 10th percentile) are presented in Table 2.

The time from GnRHa injection to oocyte retrieval was also not significantly different between the two groups ($P = 0.55^*$).

In multivariate analysis (Table 3), poor ORR was no longer statistically significantly associated with LH level at the beginning of stimulation ($P = 0.20$), time of gonadotropin start ($P = 0.47$), duration of GnRH antagonist administration ($P = 0.36$), and E2 level 12-h post-triggering ($P = 0.51$). Only, LH level 12-h post-triggering ($P = 0.0001$) and AFC ($P = 0.006$) remained in the final model. Lower ovarian reserve and lower LH level 12-h post-triggering were predictive of lower ORR (OR: 0.80 [95% CI: 0.68–0.94]) and

Table 2. Patient and cycle characteristics according to oocyte recuperation rate.

	Oocyte recuperation rate		P
	$<$ 10th percentile	$\geq$ 10th percentile	
Patient's characteristics at the start of the cycle	n = 139	n = 1281	
Age (years)			0.43
<30	39 (28.1%)	320 (25.0%)	
30–34	44 (31.6%)	372 (29.0%)	
$\geq$ 35	56 (40.3%)	589 (46.0%)	
BMI (kg/m ²)	22.5 (20.1–24.9)	22.1 (20.1–24.8)	0.77
Tobacco use (yes)	28 (21.7%)	215 (18.6%)	0.40
AMH (ng/ml)	1.91 (0.96–3.51)	2.14 (1.24–3.78)	0.01
AFC	14 (8–21)	17 (11–25)	<0.001
Cycle description			
Pretreatment			0.25
Not antagonodotropic	67 (48.2%)	664 (51.8%)	
Antigonadotropic	42 (30.2%)	304 (23.7%)	
No pretreatment	30 (21.6%)	313 (24.5%)	
Time of gonadotropin start			0.07
Follicular phase	126 (90.6%)	1206 (94.7%)	
Luteal phase	13 (9.4%)	68 (5.3%)	
Type of gonadotropin			0.18
FSH	111 (79.9%)	1081 (84.4%)	
FSH + LH	28 (20.1%)	200 (15.6%)	
Initial dose of gonadotropin (IU/day)			0.07
$\leq$ 225	35 (25.2%)	424 (33.1%)	
226–299	47 (33.8%)	442 (34.6%)	
$\geq$ 300	57 (41.6%)	414 (32.3%)	
LH at the beginning of stimulation (IU/l)			0.016
$\leq$ 0.5	9 (8.0%)	37 (3.6%)	
0.51–0.99	7 (6.2%)	28 (2.8%)	
1.00–2.99	36 (31.9%)	274 (27.0%)	
$\geq$ 3	61 (53.9%)	677 (66.6%)	
Duration of stimulation (days)			0.23
$\leq$ 10	58 (41.7%)	605 (47.2%)	
11	22 (15.8%)	227 (17.7%)	
$\geq$ 12	59 (42.5%)	449 (35.1%)	
Total dose of gonadotropin (IU)	3375.0 (2456.2–4875.0)	3000.0 (2100.0–4200.0)	<0.001
Duration of GnRH antagonist administration (days)			0.01
$\leq$ 5	67 (48.2%)	667 (52.21%)	
6	17 (12.2%)	246 (19.2%)	
$>$ 6	55 (36.6%)	368 (28.7%)	
LH the DT (IU/l)			0.14
$\leq$ 0.5	65 (46.7%)	516 (40.3%)	
0.51–0.99	18 (13.0%)	187 (14.6%)	
1.00–2.99	34 (24.5%)	419 (32.7%)	
$\geq$ 3	22 (15.8%)	158 (12.4%)	
Follicles $>$ 12 mm the DT (n)			0.37
$<$ 20	91 (65.5%)	886 (69.2%)	
$\geq$ 20	48 (34.5%)	395 (30.8%)	
E2 12-h post-triggering (pg/ml)	2378 (1449–3660)	2778 (1649–4470)	0.02
LH 12-h post-triggering (IU/l)	34.9 (23.3–51.1)	44.4 (30.4–65.0)	<0.001
Progesterone 12-h post-triggering (ng/ml)			0.37
$<$ 3	8 (6.8%)	51 (4.8%)	
$\geq$ 3	110 (93.2%)	1015 (95.2%)	

Data are presented as median (25th–75th percentile) or n (%).

P after univariate logistic regression (for logistic regression, data with non-normal repartition were either transformed to obtain normal repartition or transformed into classes).

AFC: antral follicle count; AMH: anti-Müllerian hormone; Antigonadotropic: estrogen–progestin pill or macroprogestin pill; COC: cumulo-oocyte complexes; DT: day of triggering; E2: estradiol; not antagonodotropic: microprogestin-only pill or dydrogesterone or estradiol.

Table 3. Risk factors of an oocyte recuperation rate <10th percentile.

	OR univariable analysis*	Multivariable analysis			
		Start of backward		End of backward = final model	
		OR*	P adjusted	OR*	P adjusted
Patients' characteristics at the start of the cycle					
AFC	0.77 [0.66–0.87]	0.74 [0.60–0.90]	0.003	0.80 (0.68–0.94)	0.006
Cycle characteristics					
Time of gonadotropin start			0.47		
Follicular phase	–	–			
Luteal phase	1.35 [0.99–1.84]	1.21 [0.73–1.99]			
LH at the beginning of stimulation (IU/l)			0.20		
≤0.5	–	–			
0.5–1	1.53 [0.80–2.94]	1.78 [0.85–3.76]			
1–2.99	0.80 [0.55–1.18]	0.81 [0.53–1.24]			
≥3	0.55 [0.39–0.78]	0.63 [0.41–0.96]			
Duration of GnRH antagonist administration (days)			0.36		
6 vs ≤5	0.68 [0.48–0.97]	0.75 [0.49–1.14]			
>6 vs ≤5	1.47 [1.13–1.92]	1.12 [0.81–1.57]			
12-h post-triggering E2 (pg/ml)	0.75 [0.59–0.95]	0.90 [0.67–1.22]	0.51		
12-h post-triggering LH (IU/l)	0.80 [0.73–0.89]	0.87 [0.78–0.99]	0.03	0.80 [0.73–0.89]	0.0001

AFC: antral follicular count; E2: estradiol; OR: odds ratio.

* Odds ratio [95% CI].

0.80 [0.73–0.89], respectively) (Table 3). There was no significant interaction between AFC and 12-h post-triggering LH level on poor ORR ($P = 0.23$). A nonlinear relationship was found between LH level 12-h post-triggering and poor ORR, but no LH threshold could be found (data not shown).

Oocyte maturation rate

The number of follicles over 12 mm on the day of triggering (16 (11–22) vs 15 (10–22); $P = 0.84$) and the number of oocytes retrieved were not significantly different between the OMR <75% group and the OMR ≥75% group (11 (7–18) vs 11 (7–17); $P = 0.06^*$). However, the number of mature oocytes retrieved was significantly lower in the OMR <75% group than in the other group (7 (4–11) vs 10 (6–15); $P < 0.001^*$). Patient and cycle characteristics according to OMR are shown in Table 4. There was no significant difference between the two groups regarding the time from GnRHa injection to oocyte retrieval ($P = 0.87^*$).

In the multivariate analysis (Table 5), BMI and AMH level were associated with an OMR <75%: women with a higher BMI and higher AMH levels had an increased risk of oocyte immaturity, respectively (OR: 4.34 [95% CI: 1.96–9.6]; $P < 0.001$) and (OR: 1.22 [95% CI: 1.03–1.12]; $P = 0.015$). Antigonadotrophic pretreatment significantly decreased the risk of poor OMR (<75%) compared to no pretreatment (OR: 0.72 [95% CI: 0.57–0.91]; $P = 0.02$) when there was no statistically significant difference between other than antigonadotrophic pretreatment and no pretreatment (OR: 1.13 [0.95–1.35]) (Table 5). No significant interaction was shown between the parameters ($P = 0.36$ – 0.46). LH level 12-h post-triggering was not predictive of OMR in multivariate analysis (Table 5).

Discussion

In this study, we aimed to search for risk factors for suboptimal outcomes after GnRHa triggering in a GnRH antagonist COH protocol assessed by ORR <10th percentile or OMR <75%. We found that AFC and LH levels 12-h after GnRHa triggering were negatively associated with poor ORR. Higher BMI and AMH levels were positively associated with poor OMR, as well as the use of antigonadotrophic pretreatment vs no pretreatment.

The significant negative association between ovarian reserve and poor ORR after GnRHa is described for the first time in this study. In previous studies that have taken ovarian reserve into account, no difference in AFC was found between patients with an appropriate response to GnRHa (defined as LH level post-triggering ≥15 IU/l) and those without an appropriate response (Chen et al., 2012; Lu et al., 2016). In a study by O'Neill et al. (2015), the post-triggering hormonal response and OMR were non different in patients with polycystic ovary syndrome and controls suggesting that the response to GnRHa triggering was equally effective in those two populations. Some studies have suggested that low ovarian reserve may be involved in EFS due to ovarian dysfunction resulting in impaired folliculogenesis quality and impaired ovulation (Madani and Jahangiri, 2015). Therefore, it can be assumed that this same mechanism could explain the lower ORR that we found in our study.

Because of its close relationship with AFC, the total gonadotrophin dose was not included in our multivariate analysis, but a higher initial dose of gonadotrophin, duration of stimulation, days of antagonist treatment, and logically higher total dose of gonadotrophin were found in the suboptimal ORR group after univariate analysis. Consistent with these results, other studies found that higher doses of gonadotropins for COH and more stimulation days were described in the suboptimal oocyte retrieval group (Popovic-Todorovic et al., 2019) or that the total amount of gonadotrophins was associated with a suboptimal LH surge after GnRHa triggering (Meyer et al., 2015; Chang et al., 2016).

In our multivariate analysis, LH level 12-h post-triggering was significantly associated with ORR. A significant nonlinear association was found between LH levels and ORR, but no satisfying threshold could be found. Two previous studies were not able to identify a strict cutoff post-trigger LH level that would be useful clinically in predicting oocyte yield (Kummer et al., 2013; Chang et al., 2016), even if women with a post-trigger LH value <15 IU/l had significantly fewer total oocytes retrieved. Two other studies reported a significantly lower ORR associated with a LH level 12-h post-trigger <15 IU/l (Chen et al., 2012; Lu et al., 2016). However, Shapiro et al. (2011) found a 12-h post-trigger LH concentration below 52 IU/l to be suboptimal, causing a modest

Table 4. Patient and cycle characteristics according to oocyte maturation rate.

	Oocyte maturation rate		P
	<75%	≥75%	
Patients characteristics at the start of the cycle	n = 440	n = 1296	
Age (years)			0.53*
≤30	118 (26.8)	317 (24.5)	
30–34	130 (29.6)	379 (29.2)	
≥35	192 (43.6)	600 (74.7)	
BMI (kg/m ²)	22.59 (20.46–25.28)	21.77 (19.95–24.46)	<0.001*
Tobacco use (yes)	62/391 (15.9)	240/1169 (20.5)	0.04
AMH (ng/ml)	2.33 (1.24–4.08)	2.02 (1.19–3.45)	0.004*
AFC	17 (11–26)	17 (11–25)	0.27
Cycle description			
Pretreatment			0.01*
Not antigonadotropic	229 (52.1)	644 (49.7)	
Antigonadotropic	89 (20.2)	346 (26.7)	
No pretreatment	122 (27.7)	306 (23.6)	
Type of gonadotropin	n = 440	n = 1296	0.61*
FSH	369 (83.9)	1100 (84.9)	
FSH + LH activity	71 (16.1)	196 (15.1)	
Time of gonadotropin start	n = 437	n = 1290	0.53*
Follicular phase	409 (93.6)	1218 (94.4)	
Luteal phase	28 (6.4)	72 (5.6)	
	n = 352	n = 1024	
LH at the beginning of stimulation (IU/l)			0.39*
≤0.5	13 (3.7)	38 (3.7)	
0.51–0.99	6 (1.7)	34 (3.3)	
1.00–2.99	93 (26.4)	280 (27.3)	
≥3	240 (68.2)	672 (65.6)	
	n = 438	n = 1295	
Initial dose of gonadotropin (IU/day)			0.08*
≤225	154 (35.0)	386 (29.8)	
226–300	153 (34.8)	464 (5.8)	
>300	131 (29.8)	445 (34.3)	
Total dose of gonadotropin (IU)	2900 (2000–4200)	3150 (2250–4400)	0.003*
Duration of GnRH antagonist administration (days)			0.20*
≤5	231 (53.0)	638 (49.2)	
6	101 (23.0)	283 (21.8)	
>6	108 (25.0)	375 (29.0)	
Duration of stimulation (days)			0.12*
≤10	217 (27.5)	572 (44.1)	
11	95 (24.9)	286 (22.1)	
>11	128 (22.6)	438 (33.8)	
	n = 438	n = 1293	
LH the DT (IU/l)			0.87*
≤0.5	16 (36.8)	534 (41.3)	
0.51–0.99	68 (15.5)	173 (13.4)	
1.00–2.99	150 (34.2)	421 (32.6)	
≥3	59 (13.5)	165 (12.8)	
	n = 361	n = 1048	
Follicles >12 mm the DT (n)	16 (11–22)	15 (10–22)	0.84*
	n = 364	n = 1078	
E2 12-h post-triggering (pg/ml)	2803 (1775.75–4420)	2708 (1608–4288.75)	0.06*
LH 12-h post-triggering (IU/l)	46.20 (31.20–68.65)	43 (30.40–61.90)	0.07*
Progesterone 12-h post-triggering (ng/ml)			0.41*
<3	22 (6.0)	53 (4.9)	
≥3	344 (94.0)	1028 (95.1)	

Data are presented as median (25th–75th percentile) or n (%).

* P after univariate logistic regression (for logistic regression, data with non-normal repartition were either transformed to obtain normal repartition or transformed into classes).

AFC: antral follicle count; AMH: anti-Müllerian hormone; Antigonadotropic: estrogen–progestin pill or macroprogestin pill; DT: day of triggering; E2: estradiol. Not antigonadotropic: microprogestin-only pill or dydrogesterone or estradiol.

reduction in ORR and OMR, and a level <12 IU/l resulted in a dramatic reduction in ORR and OMR. Finally, a recent study by Wiltshire *et al.* (2021), who performed analysis using a threshold of 15 IU/l for post-trigger LH level, found no difference between the two groups in terms of oocytes retrieved, maturation rate, or fertilization rate. In a natural cycle in humans, the amplitude of the LH surge required to initiate oocyte maturation is not known. In rats, only 5% of the normal LH surge amplitude is necessary for resumption of meiosis, but 85% of peak levels are needed for follicular

rupture (Peluso, 1990). Studies on the relationship between the LH surge and oocyte maturation in humans have shown that meiosis resumes ~18 h after the LH surge and that metaphase II oocytes are obtainable 28 h after the onset of the LH surge (Seibel *et al.*, 1982). Thus, it is possible that higher levels of LH lasting more than 18 h might be necessary to induce cumulus expansion and detachment from the follicular wall, allowing ovulation in natural cycles or oocyte retrieval in IVF cycles. Therefore, it is believable that a single factor alone cannot explain a suboptimal response.

Table 5. Risk factors of oocyte immaturity <75%.

		Multivariate analysis before backward		Multivariate analysis after backward	
	OR univariate analysis*	OR multivariate analysis*	P adjusted	OR multivariate analysis*	P adjusted
Patients' characteristics at the start of the cycle					
BMI (kg/m ²)	3.34 [1.70–6.55]	4.34 [1.96–9.6]	<0.001	4.34 [1.96–9.6]	<0.001
Tobacco use (yes)	0.85 [0.73–0.99]	0.86 [0.72–1.04]	0.11		
AMH (ng/ml)	1.23 [1.07–1.42]	1.15 [0.91–1.46]	0.24	1.22 [1.03–1.12]	0.015
Cycle description					
Pretreatment			0.006		0.02
Not antagonadotropic	1.07 [0.93–1.24]	1.15 [0.96–1.38]		1.13 [0.95–1.35]	
Antigonadotropic	0.78 [0.65–0.93]	0.69 [0.54–0.87]		0.72 [0.57–0.91]	
No pretreatment	–				
Total dose of gonadotropin (IU)	0.72 [0.57–0.90]	0.86 [0.72–1.04]	0.46		
LH 12-h post-triggering (IU/l)	1.05 [0.99–1.11]	1.06 [0.99–1.14]	0.11		

AMH: anti-Müllerian hormone; Antigonadotropic: estrogen–progesterone pill or macroprogesterone pill; Not antagonodotropic: microprogesterone-only pill or dydrogesterone or estradiol; OMR: ratio of mature oocytes to the number of total cumulo-oocyte complex obtained; OR: odds ratio.

* Odds ratio [95% CI].

In our study, LH level at the beginning of stimulation was not correlated with ORR after multivariate analysis. In contrast, [Popovic-Todorovic et al. \(2019\)](#) found that the baseline LH level was an independent predictor of suboptimal oocyte yield (ORR < 10th percentile) and that baseline LH levels <0.5 IU/l were associated with a 39.1% risk of inadequate oocyte yield. Other authors have also shown that a low level of LH at the beginning of stimulation was a risk factor for poor endocrine response to GnRHa triggering (i.e. LH level below 15 IU/l 12-h post-triggering) ([Chang et al., 2016](#); [Lu et al., 2016](#); [Humaidan and Kol, 2021](#); [Ganer Herman et al., 2022](#)).

LH levels on the day of triggering were not found to be predictive of ORR in our study, which is consistent with other studies ([Chen et al., 2012](#); [Lu et al., 2016](#); [Popovic-Todorovic et al., 2019](#)). Conversely, [Meyer et al. \(2015\)](#) found that an undetectable level of LH (0.1 IU/l) on the day of ovulation trigger carried a 25% chance of a suboptimal LH surge after GnRHa triggering.

After GnRHa triggering, the progesterone concentration reached 3 ng/ml 6 h post-injection ([Itskovitz et al., 1991](#)). Some studies noticed significantly higher progesterone levels 12-h after GnRHa triggering in optimal responders (i.e. in patients without EFS vs patients with EFS ([Deepika et al., 2018](#)) and in patients with LH at 12-h >15 vs ≤15 IU/l ([Chang et al., 2016](#); [Deepika et al., 2018](#))). Another study found the rise in progesterone level following GnRHa triggering to be the strongest hormonal predictor of the number of mature oocytes retrieved ([Abbara et al., 2020](#)). Thus, this threshold of 3 ng/ml is used by some authors to indicate an adequate response to the GnRHa trigger, although no strict cutoff has been shown to successfully or unsuccessfully predict oocyte retrieval ([Kummer et al., 2013](#)). In addition, the progesterone level reflects follicular luteinization and therefore depends on the number of corpora lutea. In our multivariate analysis, the progesterone level 12-h post-triggering was not associated with ORR.

Complete failure of the GnRHa trigger, defined as complete oocyte recuperation failure or EFS, is a rare event; it has been demonstrated to occur in ~1.2% of patients (range 0.5–3.4%) triggered with GnRHa ([Ganer Herman et al., 2022](#)). In our study, it occurred in 11 oocyte retrievals (0.6%). Among those 11 patients, 9 had low ovarian reserve with AMH <1 ng/ml; 3 patients ovulated before the oocyte pick-up; 5 patients had LH levels at 12-h between 15 and 52 IU/l or >52 IU/l and progesterone >3 ng/ml, similar to the cases described by [Chang et al.](#) who defined a “borderline area” based on post-trigger laboratory measurements

in which oocyte retrieval was attempted but abandoned after recovering no oocytes from the first several follicles; 2 patients had progesterone levels <3 ng/ml despite an LH surge >52 IU/l (pretreatment by estradiol and AMH <1 ng/ml); and 1 patient had low LH levels <15 IU/l but progesterone levels >3 ng/ml (pretreatment by estrogen–progesterone pill and AMH 1.35 ng/ml). It was suggested that EFS may be secondary to abnormal folliculogenesis or early oocyte atresia.

In our study, 25.3% of patients suffered from oocyte immaturity. After multivariate analysis, three parameters were related to OMR: BMI, AMH level, and antagonodotropic pretreatment before COH.

BMI was positively associated with poor OMR, with an increased risk of oocyte immaturity in overweight and obese patients. In accordance with our results, [Lainas et al. \(2020\)](#) found that BMI was negatively associated with oocyte maturity after GnRHa triggering. This could be due to different pharmacokinetic characteristics (for example, clearance or bioavailability) of triptorelin in women with different BMI. In our study, all patients were triggered by triptorelin 0.2 mg based on a GnRHa trigger dose-finding study in oocyte donors showing that triptorelin 0.2 mg was probably the more cost effective with comparable numbers of mature oocytes retrieved after the use of triptorelin 0.2, 0.3, or 0.4 mg ([Vuong et al., 2016](#)). Since then, other studies have shown that different GnRHa doses are not related to triggering efficacy. However, we may assume that the standard dose for the GnRHa trigger may not be sufficient to induce appropriate oocyte recovery and maturation in patients with higher BMI, but future studies should test this hypothesis with different sites of injection or different administration modes.

AMH level was also associated with poor OMR. Women with higher AMH levels had a higher proportion of poor OMRs. This can be explained by a possible higher number of intermediate follicles aspirated in this subgroup with a lower proportion of mature oocytes. Indeed, in follicles between 12 and 14 mm, only 30% of oocytes are believed to be mature ([Royère, 2006](#)).

Surprisingly, we found that antagonodotropic pretreatment (including OCP) compared to no pretreatment decreased the risk of oocyte immaturity. This may be due to the fact that the follicular cohort was more synchronous with antagonodotropic treatment, leading to better OMR. Hence, it is not the trigger alone that determines the OMR. LH levels at the beginning of stimulation, on the day of trigger, or post-triggering were not predictive for OMR in our study. In the same way, [Benmachiche et al. \(2019\)](#)

failed to find any significant impact of LH levels on the day of GnRHa triggering on the number of mature oocytes. [Chen et al. \(2012\)](#), [Kummer et al. \(2013\)](#), and [Lu et al. \(2016\)](#) also demonstrated that the OMR and fertilization rate were not linked to serum LH levels at 12-h after GnRHa trigger. [Dunne et al. \(2018\)](#) evaluated 97 IVF/ICSI cycles triggered by 0.5 mg buserelin in 2018 and did not find a relationship between the absolute post-trigger LH level (or the percentage change in LH level) and the proportion of mature oocytes collected.

Our study highlighted some clinical risk factors for poor ORR or OMR after GnRHa trigger such as low ovarian reserve and higher BMI. For those patients, a dual trigger with both GnRHa and a full or partial dose of hCG may be considered instead of GnRHa alone. Indeed, [Lu et al.](#) showed that in a group of suboptimal responders to GnRHa (defined as LH <15 IU/l at 12-h), a dual trigger significantly increased the oocyte retrieval rate. In addition, in a recent meta-analysis, it was found that a dual trigger allowed the retrieval of a significantly higher number of oocytes (+1.2) and mature oocytes (+0.5) with comparable clinical outcomes compared to hCG ([Bourdon et al., 2021](#)). Nevertheless, the risk of OHSS must be taken into account before considering dual- or hCG triggers.

In the case of suboptimal LH response 12-h after GnRHa trigger, retriggering with hCG 24 (or 48) h after failed GnRHa trigger had no obvious negative impact on IVF outcomes ([Asada et al., 2013](#); [Chang et al., 2016](#); [Li et al., 2022](#)).

Our study is limited by its retrospective design and by the exclusion of patients who had hCG retriggerers, which underestimated our results regarding the relationship between post-trigger LH and progesterone levels and ORR and OMR. However, this occurred in only six cycles. We were also not able to collect information on the duration of pretreatment and the duration of wash out period, which can potentially influence the insufficient response to GnRHa trigger due to a prolonged downregulation. Finally, we included two populations (COH for FP and infertility) that are different in many points. However, each parameter that differs between the two populations was individually analyzed in the risk factor evaluation.

The strengths of the present study are that it was a bicentric study, with a large number of included patients and that we performed a multivariate analysis to adjust for potential confounding factors. In addition, compared to most previous studies, we evaluated GnRHa trigger efficacy by selecting clinically relevant endpoints such as ORR and OMR, and not surrogate markers, such as LH and/or progesterone level after GnRHa triggering.

In conclusion, we found that low ovarian reserve and low LH levels 12-h after GnRHa triggering were risk factors for poor ORR and that BMI and AMH level were negatively associated with OMR. None of the other usual hormonal markers of response to the GnRHa trigger were found to be significantly related to OMR or ORR. In clinical practice, in patients with low ovarian reserve or higher BMI, a dual trigger or an hCG trigger may be considered, balanced by the risk of OHSS. Post GnRHa hormonal measurements should be taken with caution as surrogate markers of ORR or OMR, but women with extremely low post-trigger levels may benefit from retriggering with hCG. Unfortunately, no consensus exists as to the cutoff for these "low levels," and the traditional LH level of 15 IU/l post-triggering must be challenged.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

Data are available upon reasonable request.

Acknowledgements

The authors thank Bernadette Darné for her help with the statistical analysis and review of the manuscript.

Authors' roles

S.G., C.S., and M.P. designed the research; S.G., C.S., I.C.-D., C.V., C.Sif., Y.B., A.M., V.G., M.G., and M.P. collected data; C.S., G.R., I. C.-D., M.G., and M.P. contributed with analytic tools and analysis; C.S. performed biostatistical analyses; M.P., C.S., G.R., I.C.-D., M. G., and M.P. analyzed data; and S.G., C.S., and M.P. wrote the paper.

Funding

No external funding was either sought or obtained for this study.

Conflict of interest

None to declare.

References

- Abbara A, Hunjan T, Ho VNA, Clarke SA, Comninou AN, Izzi-Engbeaya C, Ho TM, Trew GH, Hramyka A, Kelsey T et al. Endocrine requirements for oocyte maturation following hCG, GnRH agonist, and kisspeptin during IVF treatment. *Front Endocrinol (Lausanne)* 2020;**11**:537205.
- Andersen CY. FSH-induced resumption of meiosis in mouse oocytes: effect of different isoforms. *Mol Hum Reprod* 1999;**5**:726–731.
- Andersen CY, Humaidan P, Ejdrup HB, Bungum L, Grøndahl ML, Westergaard LG. Hormonal characteristics of follicular fluid from women receiving either GnRH agonist or hCG for ovulation induction. *Hum Reprod* 2006;**21**:2126–2130.
- Asada Y, Itoi F, Honnma H, Takiguchi S, Fukunaga N, Hashiba Y, Baba T, Endo T. Failure of GnRH agonist-triggered oocyte maturation: its cause and management. *J Assist Reprod Genet* 2013;**30**:581–585.
- Bar-Ami S, Zlotkin E, Brandes JM, Itskovitz-Eldor J. Failure of meiotic competence in human oocytes1. *Biol Reprod* 1994;**50**:1100–1107.
- Benadiva C, Engmann L. Luteal phase support after gonadotropin-releasing hormone agonist triggering: does it still matter? *Fertil Steril* 2018;**109**:763–767.
- Benmachiche A, Benbouhedja S, Zoghmar A, Humaidan P. Low LH level on the day of GnRH agonist trigger is associated with reduced ongoing pregnancy and live birth rates and increased early miscarriage rates following IVF/ICSI treatment and fresh embryo transfer. *Front Endocrinol (Lausanne)* 2019;**10**:639.
- Bourdon M, Peigné M, Solignac C, Darné B, Languille S, Pocate-Cheriet K, Santulli P. Gonadotropin-releasing hormone agonist (alone or combined with human chorionic gonadotropin) vs. human chorionic gonadotropin alone for ovulation triggering during controlled ovarian stimulation for in vitro fertilization/intracytoplasmic sperm injection: a systematic review and meta-analysis. *F&S Reviews* 2021;**2**:353–370.
- Chang FE, Beall SA, Cox JM, Richter KS, DeCherney AH, Levy MJ. Assessing the adequacy of gonadotropin-releasing hormone agonist leuprolide to trigger oocyte maturation and management of inadequate response. *Fertil Steril* 2016;**106**:1093–1100.e3.

- Chen S-L, Ye D-S, Chen X, Yang X-H, Zheng H-Y, Tang Y, He Y-X, Guo W. Circulating luteinizing hormone level after triggering oocyte maturation with GnRH agonist may predict oocyte yield in flexible GnRH antagonist protocol. *Hum Reprod* 2012; **27**:1351–1356.
- Cobo A, García-Velasco JA, Remohí J, Pellicer A. Oocyte vitrification for fertility preservation for both medical and nonmedical reasons. *Fertil Steril* 2021; **115**:1091–1101.
- Deepika K, Sindhumma D, Kiran B, Ravishankar N, Gautham P, Kamini R. Empty follicle syndrome following GnRHa trigger in PCOS patients undergoing IVF cycles. 2018; **19**:10.
- Dunne C, Shan A, Nakhuda G. Measurement of luteinizing hormone level after gonadotropin-releasing hormone agonist trigger is not useful for predicting oocyte maturity. *J Obstet Gynaecol Can* 2018; **40**:1618–1622.
- Engmann L, DiLuigi A, Schmidt D, Nulsen J, Maier D, Benadiva C. The use of gonadotropin-releasing hormone (GnRH) agonist to induce oocyte maturation after cotreatment with GnRH antagonist in high-risk patients undergoing in vitro fertilization prevents the risk of ovarian hyperstimulation syndrome: a prospective randomized controlled study. *Fertil Steril* 2008; **89**:84–91.
- Fauser BC, de Jong D, Olivennes F, Wramsby H, Tay C, Itskovitz-Eldor J, van Hooren HG. Endocrine profiles after triggering of final oocyte maturation with GnRH agonist after cotreatment with the GnRH antagonist ganirelix during ovarian hyperstimulation for in vitro fertilization. *J Clin Endocrinol Metab* 2002; **87**:709–715.
- Ganer Herman H, Horowitz E, Mizrahi Y, Farhi J, Razieli A, Weissman A. Prediction, assessment, and management of suboptimal GnRH agonist trigger: a systematic review. *J Assist Reprod Genet* 2022; **39**:291–303.
- Hoff JD, Quigley ME, Yen SSC. Hormonal dynamics at midcycle: a reevaluation. *J Clin Endocrinol Metab* 1983; **57**:792–796.
- Humaidan P, Bredkjaer HE, Bungum L, Bungum M, Grøndahl ML, Westergaard L, Andersen CY. GnRH agonist (buserelin) or hCG for ovulation induction in GnRH antagonist IVF/ICSI cycles: a prospective randomized study. *Hum Reprod* 2005; **20**:1213–1220.
- Humaidan P, Kol S. Suboptimal response to GnRH agonist trigger: causes and practical management. *Curr Opin Obstet Gynecol* 2021; **33**:213–217.
- Humaidan P, Kol S, Papanikolaou E, Copenhagen GnRH Agonist Triggering Workshop Group. GnRH agonist for triggering of final oocyte maturation: time for a change of practice? *Hum Reprod Update* 2011a; **17**:510–524.
- Humaidan P, Westergaard LG, Mikkelsen AL, Fukuda M, Yding Andersen C. Levels of the epidermal growth factor-like peptide amphiregulin in follicular fluid reflect the mode of triggering ovulation: a comparison between gonadotrophin-releasing hormone agonist and urinary human chorionic gonadotrophin. *Fertil Steril* 2011b; **95**:2034–2038.
- Itskovitz J, Boldes R, Levron J, Erlik Y, Kahana L, Brandes JM. Induction of preovulatory luteinizing hormone surge and prevention of ovarian hyperstimulation syndrome by gonadotropin-releasing hormone agonist. *Fertil Steril* 1991; **56**:213–220.
- Itskovitz-Eldor J, Kol S, Mannaerts B, Coelingh Bennink H. First established pregnancy after controlled ovarian hyperstimulation with recombinant follicle stimulating hormone and the gonadotrophin-releasing hormone antagonist ganirelix (Org 37462). *Hum Reprod* 1998; **13**:294–295.
- Kolibianakis EM, Schultze-Mosgau A, Schroer A, van Steirteghem A, Devroey P, Diedrich K, Griesinger G. A lower ongoing pregnancy rate can be expected when GnRH agonist is used for triggering final oocyte maturation instead of HCG in patients undergoing IVF with GnRH antagonists. *Hum Reprod* 2005; **20**:2887–2892.
- Kummer NE, Feinn RS, Griffin DW, Nulsen JC, Benadiva CA, Engmann LL. Predicting successful induction of oocyte maturation after gonadotropin-releasing hormone agonist (GnRHa) trigger. *Hum Reprod* 2013; **28**:152–159.
- Lainas GT, Lainas TG, Sfontouris IA, Venetis CA, Bosdou JK, Chatzimeletiou A, Grimbizis GF, Tarlatzis BC, Kolibianakis EM. Association between body mass index and oocyte maturation in patients triggered with GnRH agonist who are at high risk for severe ovarian hyperstimulation syndrome: an observational cohort study. *Reprod Biomed Online* 2020; **40**:168–175.
- Law YJ, Zhang N, Kolibianakis EM, Costello MF, Keller E, Chambers GM, Venetis CA. Is there an optimal number of oocytes retrieved at which live birth rates or cumulative live birth rates per aspiration are maximized after ART? A systematic review. *Reprod Biomed Online* 2021; **42**:83–104.
- Li X-F, Wang Q-F, He Q-Q, Wang X-J, Lv X-Y, Tang X-J, Zhong Z-H, Ding Y-B, Wan Q. Low LH level does not indicate poor IVF cycle outcomes with GnRh-a single trigger: a retrospective analysis. *BMC Pregnancy Childbirth* 2022; **22**:951.
- Lu X, Hong Q, Sun L, Chen Q, Fu Y, Ai A, Lyu Q, Kuang Y. Dual trigger for final oocyte maturation improves the oocyte retrieval rate of suboptimal responders to gonadotropin-releasing hormone agonist. *Fertil Steril* 2016; **106**:1356–1362.
- Madani T, Jahangiri N. Empty follicle syndrome: the possible cause of occurrence. *Oman Med J* 2015; **30**:417–420.
- Meyer L, Murphy LA, Gumer A, Reichman DE, Rosenwaks Z, Cholst IN. Risk factors for a suboptimal response to gonadotropin-releasing hormone agonist trigger during in vitro fertilization cycles. *Fertil Steril* 2015; **104**:637–642.
- O'Neill KE, Senapati S, Dokras A. Use of gonadotropin-releasing hormone agonist trigger during in vitro fertilization is associated with similar endocrine profiles and oocyte measures in women with and without polycystic ovary syndrome. *Fertil Steril* 2015; **103**:264–269.
- Orvieto R. Triggering final follicular maturation—hCG, GnRH-agonist or both, when and to whom? *J Ovarian Res* 2015; **8**:60.
- Peluso JJ. Role of the amplitude of the gonadotropin surge in the rat. *Fertil Steril* 1990; **53**:150–154.
- Pfeifer S, Butts S, Dumesic D, Fossum G, Gracia C, La Barbera A, Mersereau J, Odem R, Paulson R, Penzias A et al. Prevention and treatment of moderate and severe ovarian hyperstimulation syndrome: a guideline. *Fertil Steril* 2016; **106**:1634–1647.
- Popovic-Todorovic B, Santos-Ribeiro S, Drakopoulos P, De Vos M, Racca A, Mackens S, Thorrez Y, Verheyen G, Tournaye H, Quintero L et al. Predicting suboptimal oocyte yield following GnRH agonist trigger by measuring serum LH at the start of ovarian stimulation. *Hum Reprod* 2019; **34**:2027–2035.
- Royère D. Maturation ovocytaire: peut-on définir la compétence d'un ovocyte? *J Gynécob Obstét Biol Reprod (Paris)* 2006; **35**:8–13.
- Russo M, Liu K, Chan C. Suboptimal response to GnRH-agonist trigger during oocyte cryopreservation: a case series. *Reprod Biol Endocrinol* 2020; **18**:59.
- Seibel MM, Smith DM, Levesque L, Borten M, Taymor ML. The temporal relationship between the luteinizing hormone surge and human oocyte maturation. *Am J Obstet Gynecol* 1982; **142**:568–572.
- Shapiro BS, Daneshmand ST, Restrepo H, Garner FC, Aguirre M, Hudson C. Efficacy of induced luteinizing hormone surge after “trigger” with gonadotropin-releasing hormone agonist. *Fertil Steril* 2011; **95**:826–828.
- The European Orgalutran® Study Group; Borm G, Mannaerts B. Treatment with the gonadotrophin-releasing hormone antagonist ganirelix in women undergoing ovarian stimulation with recombinant follicle stimulating hormone is effective, safe and

- convenient: results of a controlled, randomized, multicentre trial. *Hum Reprod* 2000;**15**:1490–1498.
- Vuong TNL, Ho MT, Ha TD, Phung HT, Huynh GB, Humaidan P. Gonadotropin-releasing hormone agonist trigger in oocyte donors co-treated with a gonadotropin-releasing hormone antagonist: a dose-finding study. *Fertil Steril* 2016;**105**:356–363.
- Wiltshire AM, Tozour JN, Akerman M, Hamer D, Grifo JA, McCulloh DH, Blakemore JK. Serum gonadotropin (Gn) levels predict response to gonadotropin-releasing hormone (GnRH) trigger in antagonist controlled ovarian hyperstimulation (COH) cycles. *Fertil Steril* 2021;**116**:e27–e28.
- Youssef MAFM, Van der Veen F, Al-Inany HG, Mochtar MH, Griesinger G, Nagi Mohesen M, Aboulfoutouh I, van Wely M. Gonadotropin-releasing hormone agonist versus HCG for oocyte triggering in antagonist-assisted reproductive technology. *Cochrane Database Syst Rev* 2014;**2014**:CD008046.

Infertility

Does recurrent implantation failure exist? Prevalence and outcomes of five consecutive euploid blastocyst transfers in 123 987 patients

Pavan Gill ^{1,†}, Baris Ata ^{2,3,†}, Ana Aranz^{4,5}, Danilo Cimadomo ⁶, Alberto Vaiarelli ⁶, Human M. Fatemi ⁷, Filippo Maria Ubaldi⁶, Juan A. Garcia-Velasco⁴, and Emre Seli ^{1,8,*}

¹IVIRMA Global Research Alliance, IVIRMA New Jersey, Basking Ridge, NJ, USA

²ART Fertility Clinics, Dubai, United Arab Emirates

³Department of Obstetrics and Gynecology, Koç University School of Medicine, Istanbul, Turkey

⁴IVIRMA Global Research Alliance, IVIRMA Madrid, Madrid, Spain

⁵Department of Biomedicine and Biotechnology, University of Alcalá, Madrid, Spain

⁶IVIRMA Global Research Alliance, Genera—Clinica Valle Giulia, Rome, Italy

⁷ART Fertility Clinics, Abu Dhabi, United Arab Emirates

⁸Department of Obstetrics, Gynecology and Reproductive Sciences, Yale School of Medicine, New Haven, CT, USA

*Correspondence address. IVIRMA Global Research Alliance, IVIRMA New Jersey, 140 Allen Road, Basking Ridge, NJ 07920, USA. E-mail: emre.seli@ivirma.com

 <https://orcid.org/0000-0001-7464-8203>

[†]These authors contributed equally to this work.

ABSTRACT

STUDY QUESTION: What are the clinical pregnancy and live birth rates in women who underwent up to two more euploid blastocyst transfers after three failures in the absence of another known factor that affects implantation?

SUMMARY ANSWER: The fourth and fifth euploid blastocyst transfers resulted in similar live birth rates of 40% and 53.3%, respectively, culminating in a cumulative live birth rate of 98.1% (95% CI = 96.5–99.6%) after five euploid blastocyst transfers.

WHAT IS KNOWN ALREADY: The first three euploid blastocysts have similar implantation and live birth rates and provide a cumulative live birth rate of 92.6%.

STUDY DESIGN, SIZE, DURATION: An international multi-center retrospective study was conducted at 25 individual clinics. The study period spanned between January 2012 and December 2022. A total of 123 987 patients with a total of 64 572 euploid blastocyst transfers were screened for inclusion.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Patients with a history of any embryo transfer at another clinic, history of any unscreened embryo transfer at participating clinics, parental karyotype abnormalities, the use of donor oocytes or a gestational carrier, untreated intracavitary uterine pathology (e.g. polyp, leiomyoma), congenital uterine anomalies, adenomyosis, communicating hydrosalpinx, endometrial thickness <6 mm prior to initiating of progesterone, use of testicular sperm due to non-obstructive azoospermia in the male partner, transfer of an embryo with a reported intermediate chromosome copy number (i.e. mosaic), preimplantation genetic testing cycles for monogenic disorders, or structural chromosome rearrangements were excluded. Ovarian stimulation protocols and embryology laboratory procedures including trophectoderm biopsy followed the usual practice of each center. The ploidy status of blastocysts was determined with comprehensive chromosome screening. Endometrial preparation protocols followed the usual practice of participating centers and included programmed cycles, natural or modified natural cycles.

MAIN RESULTS AND THE ROLE OF CHANCE: A total of 105 (0.085% of the total population) patients met the criteria and underwent at least one additional euploid blastocyst transfer after failing to achieve a positive pregnancy test with three consecutive euploid blastocyst transfers. Outcomes of the fourth and fifth euploid blastocyst transfers were similar across participating centers. Overall, the live birth rate was similar with the fourth and fifth euploid blastocysts (40% vs 53.3%, relative risk = 1.33, 95% CI = 0.93–1.9, *P* value = 0.14). Sensitivity analyses excluding blastocysts biopsied on Day 7 postfertilization, women with a BMI >30 kg/m², cycles using non-ejaculate or donor sperm, double-embryo transfer cycles, and cycles in which the day of embryo transfer was modified due to endometrial receptivity assay test result yielded similar results. Where data were available, the fourth euploid blastocyst had similar live birth rate with the first one (relative risk = 0.84, 95% CI = 0.58–1.21, *P* = 0.29). The cumulative live birth rate after five euploid blastocyst transfers was 98.1% (95% CI = 96.5–99.6%).

LIMITATIONS, REASONS FOR CAUTION: Retrospective design has its own inherent limitations. Patients continuing with a further euploid embryo transfer and patients dropping out from treatment after three failed euploid transfers can be systematically different, perhaps with regard to ovarian reserve or economic status.

Received: September 27, 2023. Revised: January 5, 2024. Editorial decision: January 18, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

WIDER IMPLICATION OF THE FINDINGS: Implantation failure seems to be mainly due to embryonic factors. Given the stable and high live birth rates up to five euploid blastocysts, unexplained recurrent implantation failure should have a prevalence of <2%. Proceeding with another embryo transfer can be the best next step once a known etiology for implantation failure is ruled out.

STUDY FUNDING/COMPETING INTEREST(s): None.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: in vitro fertilization / assisted reproduction / implantation failure / preimplantation genetic testing / live birth

Introduction

The term 'recurrent implantation failure' (RIF) aims to define a milestone during the course of multiple embryo transfers, all of which have failed to result in implantation. The recurrent nature of the failures presumably suggests a currently unknown etiology that may affect the success of subsequent transfers, perhaps requiring a different diagnostic and therapeutic approach to achieve a live birth (Ata et al., 2021). Until recently, the term has been used loosely with varying definitions based on female age, developmental stage and quality of embryos transferred, fresh or frozen embryo transfers, as well as number of embryos that failed to implant (Shaulov et al., 2020). The heterogeneity of definitions without an underpinning biological rationale has led to several challenges affecting patients and reproductive medicine professionals worldwide. First and foremost, using low thresholds to diagnose RIF has triggered multiple diagnostic and therapeutic interventions to be offered to couples in the absence of sound biological rationale or proper proof of effectiveness (reviewed in Franasiak et al., 2021). Such an approach does not only consume funds available for further treatment, but also puts couples at risk of unintended harm in the form of complications or lower success rates caused by the new intervention. Second, inclusion of women with too few failed transfers in RIF research leads to heterogenous study populations yielding unreliable and conflicting results and conclusions (Ata et al., 2021).

There have been recent efforts to define RIF with an underpinning biological rationale (Ata et al., 2021; Somigliana et al., 2022; ESHRE Working Group on Recurrent Implantation Failure, 2023; (The writing group) for the participants to the 2022 Lugano RIF Workshop, 2023). These new approaches recognize embryo aneuploidy as the most common reason for implantation failure, in the absence of other known factors, e.g. communicating hydrosalpinges or intracavitary lesions. Since the incidence of aneuploidy is strongly dependent on female age, recent definitions consider euploidy status of the embryos transferred, either as directly determined by preimplantation genetic testing for aneuploidy (PGT-A) or by using female age to calculate anticipated probability of euploidy. This information is used to estimate the expected cumulative probability of implantation or live birth to be achieved with the number of embryos transferred. If this figure is above a threshold, that is regarded high enough to rule out implantation failure due to randomly occurring aneuploidy, the patient can be diagnosed with RIF (Ata et al., 2021; ESHRE Working Group on Recurrent Implantation Failure, 2023; (The Writing Group) for the Participants to the 2022 Lugano RIF Workshop, 2023).

An initial landmark study reported the cumulative sustained implantation and live birth rates achieved with three consecutive single euploid blastocyst transfers (SETs) in the absence of other obvious factors (Pirtea et al., 2021). Euploid blastocysts had similar sustained implantation and live birth rates across the first three consecutive single embryo transfers, leading to a 95.2% cumulative probability of implantation and a 92.6% cumulative probability of live birth among the 4429 women included in the

study. These figures were in agreement with the predictions of a mathematical model, which was informed by euploid blastocyst implantation and live birth rates reported in the literature (Ata et al., 2021). Based on these studies, the incidence of 'true' unexplained RIF would be <5% in the IVF population, since a statistically significant progressive decline in either the implantation or live birth rate per euploid blastocyst transferred was not observed between the first three consecutive SETs.

Despite significant progress in our understanding of RIF based on these studies, the fate of the fourth and the fifth euploid blastocyst transfers following the failure of the first three remains unknown. If the fourth and the fifth euploid transfers result in implantation with a similar rate as the first three attempts, the true prevalence of RIF could be much <5%. However, if there is a significant drop in the implantation potential of the embryo at the fourth and the fifth transfer, this finding could be used to identify a subgroup of patients with true RIF and possible biological mechanisms preventing implantation. In the current study, we addressed this question by examining clinical pregnancy and live birth rates in women who underwent at least one additional euploid blastocyst transfer after three failures in the absence of another known factor that affects implantation.

Materials and methods

Study population

This was an international multi-center retrospective study conducted at 26 individual clinics operating as part of Reproductive Medicine Associates (RMA) USA, IVI Spain, Genera, Italy, and ART Fertility Clinics, United Arab Emirates (Supplementary Table S1). Research ethics committee approvals were obtained in each participating institution (RMA Network in the USA: Advarra # 00064197; IVI Spain: PI 229/22; Genera Italy 0747/2023-7283; YALE IRB: 2000032741; and ART Fertility Clinics: REFA095-2304-ABU-005-BA; DOH/ZHCD/2023/1226).

The study period spanned between January 2012 and December 2022. All patients who failed to achieve a positive pregnancy test after the cumulative transfer of three consecutive euploid blastocysts (regardless of the number of transfer procedures) and who subsequently underwent at least one more euploid blastocyst transfers at participating centers were included. Patients were between the age of 18 and 45 years old and had a BMI <40 kg/m². Exclusion criteria included history of any embryo transfer in another clinic, history of any unscreened embryo transfer in participating clinics, parental karyotype abnormalities, the use of donor oocytes or a gestational carrier, intracavity uterine pathology (e.g. polyp, leiomyoma), congenital uterine anomalies (e.g. septum), intramural fibroids distorting uterine cavity, adenomyosis, communicating hydrosalpinx, endometrial thickness <6 mm prior to initiating of progesterone, use of testicular sperm due to non-obstructive azoospermia in the male partner, transfer of an embryo with a reported intermediate chromosome copy number (i.e. mosaic), PGT cycles for monogenic disorders, or structural chromosome

rearrangements. Patients who received a double-euploid blastocyst transfer (DET) were excluded if only one of the embryos implanted.

Participants were stimulated with commonly used ovarian stimulation protocols. Embryology laboratory procedures followed the usual practice of each participating center. Fertilization was achieved with IVF or ICSI depending on sperm characteristics and each center's practice. Embryo culture and trophectoderm biopsy followed each centers' own protocol.

The ploidy status of blastocysts was determined with comprehensive chromosome screening (CCS), using next-generation sequencing (NGS) (ART, RMA, IVI, Genera) or quantitative real-time polymerase chain reaction (qPCR) (RMA, Genera) (Forman et al., 2013; Scott et al., 2013; Capalbo et al., 2015, 2021; de Los Santos et al., 2018; Tiegs et al., 2021; Ata et al., 2023). CCS diagnoses included euploid, whole chromosome aneuploid, segmental aneuploid, or intermediate copy number (i.e. mosaic). Any transfers of blastocysts with a reported intermediate copy number, or segmental or whole chromosome aneuploid were excluded from the analyses.

Embryos were vitrified after biopsy until time of transfer. Endometrial preparation protocols followed the usual practice of participating centers and included programmed cycles with sequential exogenous estrogen and progesterone administration, natural or modified natural cycles (Vaiarelli et al., 2020; Rafael et al., 2022; Ata et al., 2023).

Outcome measures

The primary outcomes were clinical pregnancy rate and live birth rate after the fourth and the fifth consecutive euploid blastocyst transfer. Clinical pregnancy was defined as the presence of an intrauterine gestational sac and yolk sac. Live birth was defined as delivery of a fetus beyond 22 weeks gestational age (Zegers-Hochschild et al., 2017).

Statistical analysis

Continuous variables were defined with mean and standard deviation or median and quartiles and compared with Kruskal-Wallis and ANOVA tests depending on distribution characteristics. Categorical variables were defined with numbers and percentages and compared with chi-squared test and its derivatives as appropriate. Unit of analysis was euploid blastocyst transferred. Clinical pregnancy and live birth rates were calculated for the fourth and fifth euploid blastocysts per participating center and in aggregate. Relative risks (RR) and 95% CIs were calculated for comparisons. Sensitivity analyses were done excluding blastocysts biopsied on Day 7 postfertilization, women with a BMI $>30\text{ kg/m}^2$, cycles using non-ejaculate or donor sperm, double-embryo transfer cycles, and cycles in which the day of embryo transfer was modified due to endometrial receptivity array (ERA) test result. A life table was prepared using the previously published data pertaining to first three consecutive euploid blastocyst transfers and the fourth and fifth euploid blastocyst transfers done in the same center (RMA, New Jersey) with the same CCS platform (Pirtea et al., 2021).

Results

During the study period 123 987 patients underwent an egg retrieval and 94 401 had an embryo transfer, with a total of 64 572 euploid blastocyst transfers performed in the 25 participating clinics. Of the 123 987 patients, 105 (0.085% of the total population) underwent at least one more euploid blastocyst transfer after failing to achieve a pregnancy with three euploid blastocyst

transfers and met the other inclusion criteria to be included in the analyses.

Among the 105 patients, female age was 34.79 ± 3.80 years, BMI was $24.68 \pm 4.74\text{ kg/m}^2$, median serum anti-Müllerian hormone (AMH) level was 2.74 (1.04–4.13) ng/ml at the time of first egg retrieval. Male age was 37.28 ± 6.56 years. The origin of the sperm was ejaculate in 99 cycles, fine-needle aspiration due to obstructive azoospermia in one, testicular sperm extraction due to prior vasectomy in one, and donor sperm in four cycles. Thirty-one women had one, 55 women had two, 11 women had three, and 4 women had four oocyte retrieval procedures.

The median number of oocytes collected was 17 (12–24.5), metaphase two (MII) oocytes was 13 (10–18), fertilization rate per MII oocyte was 82.14% (73.86–90.91%), blastulation rate per MII oocyte was 41.18% (33.33–57.14%), and euploidy rate per biopsied blastocyst was 62.50% (50–83.33%). There were 4 (2–5) euploid blastocysts per stimulation cycle. Table 1 shows baseline demographics and stimulation cycle outcomes per participating center.

Participants received a total of 493 euploid blastocysts in 461 embryo transfer cycles. All 105 women had failed to achieve implantation after receiving three euploid blastocysts, in two or three embryo transfer procedures. There were 105 first (102 SETs, 3 DETs), 105 second (98 SETs, 7 DETs), 105 third (99 SETs, 6 DETs), 100 fourth (90 SETs, 10 DETs), 33 fifth (31 SETs, 2 DETs), 9 sixth (9 SETs), 4 seventh (4 SETs), 2 eighth (2 SETs), and 2 ninth (2 SETs) embryo transfer cycles. One-hundred and ninety-four blastocysts were biopsied on the fifth, 265 on the sixth, and 34 on the seventh day postfertilization; 373 blastocysts were tested with NGS and 120 with qPCR.

Endometrial thickness was $8.6 \pm 1.55\text{ mm}$ on the day of progesterone commencement (in programmed cycles), hCG injection (in modified natural cycles), or the day of ovulation (in natural cycles). ERA was used prior to 57 embryo transfer cycles, and in 34, ET day was changed by 1 day based on ERA result.

The outcomes of the fourth and fifth euploid blastocyst transfers were similar across participating centers and are shown in Table 2. Overall, the live birth rate achieved with the fourth euploid blastocyst was 40% and with the fifth (after four prior failures) was 53.3%. hCG positivity, clinical pregnancy, and live birth rates were also similar with the fourth and fifth euploid blastocysts as shown in Table 3.

Sensitivity analyses excluding blastocysts biopsied on Day 7 postfertilization (Supplementary Tables S2, S3, and S4), women with a BMI $>30\text{ kg/m}^2$ (Supplementary Tables S5, S6, and S7), cycles using non-ejaculate or donor sperm (Supplementary Tables S8, S9, and S10), double-embryo transfer cycles (Supplementary Tables S11, S12, and S13), and cycles in which the day of embryo transfer was modified due to ERA test result (Supplementary Tables S14, S15, and S16) yielded similar results.

The endometrial preparation protocol was changed for 11 patients (24.4%), however, without a systematic trend, i.e., from a natural to programmed cycle in one, from a modified natural to a programmed cycle in six, from a programmed to a modified natural cycle in four. Thirty-four (75.6%) patients underwent the fourth and the fifth euploid blastocyst transfer with same protocol. The distribution of endometrial preparation protocols was similar for fourth and fifth transfers (Supplementary Table S17). The number of patients who underwent six, seven, eight, or nine euploid embryo transfers without achieving a pregnancy was too low for hypothesis testing (Supplementary Table S18).

There were 24 fourth and 7 fifth euploid single blastocyst transfers from RMA New Jersey, which underwent CCS with NGS

Table 1. Demographics and stimulation cycle characteristics per participating center.

IVF center groups	ART fertility clinics (n = 4)	Genera Italy (n = 17)	IVIRMA Spain (n = 18)	IVIRMA USA (n = 66)	P value
Female age in years	31.3 (5.7)	36.8 (2.1)	36.3 (5.3)	34.1 (3.2)	0.03
Male age in years	36.8 (13.8)	41.2 (5.6)	40.4 (10.9)	35.6 (4.1)	0.02
Serum AMH level in ng/ml	3.4 (1.6–3.4)	2.7 (2.1–6.5)	3.0 (2.7–4.4)	3.3 (1.9–4.9)	0.98
Body mass index in kg/m ²	21.4 (18.2–23.1)	22 (17.8–23.6)	23.3 (21.3–25.6)	24.3 (21.9–27.3)	0.04
Cumulus oocyte complexes	15.5 (12.8–24.3)	14 (12–20)	15 (10.8–24.3)	18 (12–26.3)	0.78
Metaphase two oocytes	12 (11.3–17.3)	12 (10–15)	16 (10.5–19.3)	13.5 (10–18)	0.59
Fertilization rate per metaphase two oocyte	78.5% (68.2–89.8)	76.9% (69.6–89.2)	78.4% (72.2–85.8)	83.3% (77.6–91.8)	0.14
Blastulation rate per metaphase two oocyte	45.6% (21.9–64.5)	40% (32.1–54.8)	38.9% (37–51)	45.8% (32.9–59.1)	0.70
Euploidy rate per biopsied blastocyst	50% (40.6–73.9)	75% (43.2–83.3)	53.6% (48.2–62.86)	66.7% (50–86.2)	0.12
Number of euploid blastocysts	2.5 (1.3–7.5)	3 (1.5–5)	3.5 (2–4.3)	4 (3–5.3)	0.51

Numbers are mean (standard deviation) or median (quartiles).
AMH: anti-Müllerian hormone.

Table 2. Outcomes of the fourth and fifth euploid blastocyst transfers subsequent to three prior failed euploid blastocyst transfers.

IVF center groups		ART Fertility Clinics	Genera Italy	IVIRMA Spain	IVIRMA US	Total	P value*
Fourth euploid blastocyst transfer	hCG-positive (%)	2/4 (50)	8/17 (47.1)	12/18 (66.7)	41/66 (62.1)	63/105 (60)	0.61
	Clinical pregnancy (%)	2/4 (50)	8/17 (47.1)	12/18 (66.7)	32/66 (48.5)	54/105 (51.4)	0.57
	Live birth (%)	2/4 (50)	6/17 (35.3)	9/18 (50)	25/66 (37.9)	42/105 (40)	0.75
Fifth euploid blastocyst transfer	hCG-positive (%)	2/3 (66.7)	4/6 (66.7)	5/6 (83.3)	20/30 (66.7)	31/45 (68.9)	0.87
	Clinical pregnancy (%)	2/3 (66.7)	3/6 (50)	5/6 (83.3)	18/30 (60)	28/45 (62.2)	0.66
	Live birth (%)	2/3 (66.7)	3/6 (50)	5/6 (83.3)	14/30 (46.7)	24/45 (53.3)	0.40

* Chi-squared test comparing centers, excluding the column "Total".

Table 3. Comparison of outcomes between the fourth and fifth euploid blastocysts.

	Fourth euploid blastocyst	Fifth euploid blastocyst	RR (95% CI), P value
hCG-positive (%)	63/105 (60)	31/45 (68.9)	1.14 (0.89–1.48), 0.31
Clinical pregnancy (%)	54/105 (51.4)	28/45 (62.2)	1.21 (0.9–1.6), 0.23
Live birth (%)	42/105 (40)	24/45 (53.3)	1.33 (0.93–1.9), 0.14

RR: relative risk.

or qPCR as in the prior study reporting the outcomes of the first three euploid single blastocyst transfers (Pirtea et al., 2021). When these were tabulated together, the fourth euploid blastocyst fared as well as the first euploid blastocyst (RR for livebirth 0.84 with 95% CI between 0.58 and 1.21, $P=0.29$). The number of fifth SETs was too few to make a comparison, yet three out of seven reached live birth with a rate of 42.9%. Overall cumulative live birth rate with five euploid blastocyst transfers reached 98.1% (95% CI = 96.5–99.6%) (Fig. 1).

Discussion

Our results demonstrate that, following the failure with three consecutive euploid blastocyst transfers, the transfer of a fourth and fifth euploid blastocyst provides similar live birth rates at 40% and 53%, respectively. Moreover, when we built up on the prior study, which reported live birth rates with the first three euploid blastocysts conducted at RMA New Jersey (Pirtea et al., 2021) and assessed subsequent euploid blastocyst transfers of the same women at the same institution, the fourth euploid blastocyst achieved similar live birth rate as the first one. While the number of women who underwent a fifth euploid blastocyst transfer at the same institution was too small to make a reliable

comparison, the live birth rate of 42.9% after four prior failures is remarkably high and overall provides a cumulative live birth rate of 98.1% (95% CI = 96.5–99.6%).

The term RIF intends to define implantation failure due to one or more abnormalities that would systematically prevent embryo implantation beyond random events in the absence of a known pathology. Implantation failure can be due to embryonic or uterine factors. In a group of patients undergoing consecutive embryo transfers, if a proportion of them had definitive uterine factor abnormalities, some of the patients without these abnormalities would achieve implantation with each consecutive embryo transfer and would not undergo the next one and drop from its denominator, while patients with these uterine abnormalities would continuously fail and proceed with subsequent transfers and remain in the denominator. This would increase the proportion of women with these abnormalities in each subsequent transfer cycle, and implantation and live birth rates per embryo transfer would progressively fall as the group becomes enriched with them ((The writing group) for the participants to the 2022 Lugano RIF Workshop, 2023). Such an enrichment was not observed in a prior study that excluded the most common random event preventing implantation, i.e. chromosome aneuploidy, by including euploid blastocyst transfers (Pirtea et al., 2021). Their results suggested that the prevalence of true unexplained RIF would be <5% after three euploid blastocyst transfers in women without other known reasons for implantation failure (Pirtea et al., 2021). Our observations provide further evidence against such an enrichment even after the transfer of the fourth and the fifth euploid blastocyst, suggesting that, in the absence of an obvious factor affecting implantation, the embryo is the main determinant of implantation success or failure and a cumulative live birth of 98.1% can be achieved with five euploid blastocysts. In this case, the prevalence of true unexplained RIF would be <2% even if live birth rate is taken into consideration as the outcome measure.

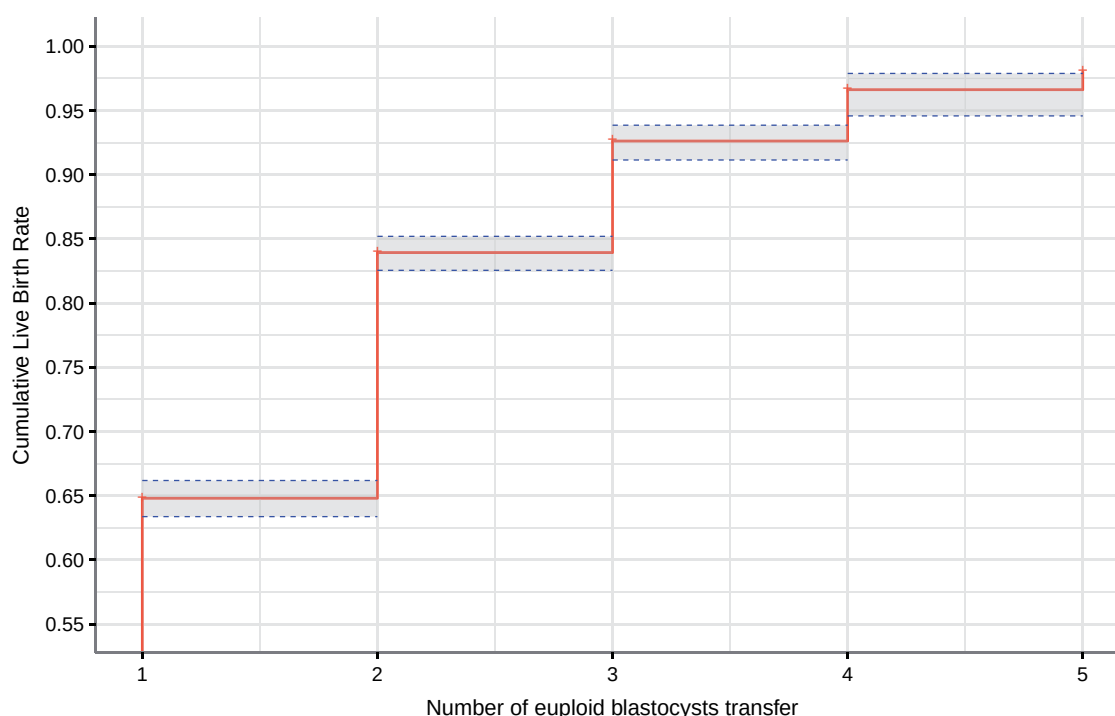


Figure 1. Cumulative live birth rate after subsequent euploid blastocyst transfer. Red line shows the cumulative live birth. Blue dotted lines show upper and lower bounds of the CI.

While the retrospective design is a limitation of the current study, the difficulty of conducting a prospective study should be acknowledged. Based on the prior study by Pirtea et al. (2021), it is reasonable to assume that more than 90% of women will achieve a live birth with the first three euploid blastocysts in the absence of another known reason for failure. Given the inevitable drop-outs from further treatment due to multiple reasons, a minority of eligible patients would continue for the fourth euploid transfer. This is supported by the fact that only <0.1% of the 123 987 women starting treatment in 26 clinics over 10 years were found to be eligible for the current study, when we excluded women who had undergone an embryo transfer in other clinics or with unscreened embryos in the participating clinics or had any of the known factors associated with implantation failure other than randomly occurring embryo aneuploidy in order to limit heterogeneity. Whether patients who continued with a subsequent transfer are systematically different from those who dropped out is unknown, and women with high ovarian reserve and/or financial means might have been more likely to pursue further transfers. There may have been differences between clinical protocols, embryology, and genetic laboratory practices between the participating clinics, but transfer outcomes were not significantly different (Table 2). To ensure the largest possible sample size to provide more reliable estimates, we opted to include Day 7 blastocysts, women with a BMI over 30 mg/m², patients whose transfer date was modified based on the ERA, despite the former two having been associated with lower implantation rates (Cimadomo et al., 2023) and the latter being considered non beneficial (Cozzolino et al., 2020, 2022; Doyle et al., 2022). Such cycles were a minority, and sensitivity analyses excluding these cycles yielded consistent results with the overall analysis. While non-obstructive azoospermia was an exclusion criterion, we included few cases where donor sperm or testicular sperm was used due to other reasons. Sensitivity analysis limited to cases with autologous ejaculate sperm also provided consistent results. Since the

unit of analysis was one euploid blastocyst transferred, not the number of transfer cycles, some patients who received double-embryo transfers were included in the analyses if both embryos implanted or neither implanted to allow tracing of each blastocyst. A sensitivity analysis limited to women who only underwent SETs also yielded similar results with the overall findings. The consistency between all the sensitivity analyses and the overall analysis provide assurance in the reliability of our estimates. The confinement of the study sample with women receiving the fourth euploid blastocyst and not reporting the cumulative results of the first three euploid blastocyst transfers in the participating centers may be regarded another limitation, but the idea was to build up on the former study by Pirtea et al. (2021). While we cannot directly compare the outcomes of the first three euploid blastocysts with the fourth and fifth for the entire study population, similar implantation and live birth rates achieved by the first and the fourth euploid blastocyst among the women who were treated at RMA New Jersey and were also included in the study by Pirtea et al. (2021) provide some assurance for the estimates. The small number of women who underwent a fifth euploid blastocyst transfer in the same center prevents a robust estimate, but 42.9% live birth rate observed in seven women can be considered supportive of our conclusions.

It should be noted that the current study is not about the effectiveness of PGT-A, rather we use euploid blastocyst transfers as a model to study the prevalence and etiology of RIF. In this context, PGT-A serves to exclude blastocysts with full chromosome aneuploidies, which have negligible potential for clinical pregnancy and live birth, to decrease the noise in the data (Tiegs et al., 2021). On the other hand, euploidy alone is not the only determinant of an embryo's potential for live birth (Cimadomo et al., 2023). First and foremost, even though current embryo testing strategies enable detection of segmental aneuploidies, the resolution of the commonly used tests may vary and they cannot detect gene mutations that may prevent implantation even with

a normal chromosomal copy number. There are other embryonic factors that determine an euploid blastocyst's potential for live birth, e.g. morphological quality, developmental rate, as well as those that are yet to be determined (Cimadomo et al., 2023).

Overall, our observation of similar implantation and live birth rates with the fourth and fifth euploid blastocyst transfers lend further credit to the notion that unexplained RIF is most often due to an embryonic factor, while an endometrial factor cannot be completely ruled out. Thus, each new embryo represents another opportunity for live birth and cumulative live birth can reach up to 98.1% after five euploid blastocyst transfers.

Given that obvious maternal reasons for possible implantation failure, such as communicating hydrosalpinges, intracavitary pathology, adenomyosis, and obesity, can be identified even prior to the first transfer cycle or whenever they develop during the course of treatment, and stimulation cycle and embryo-related factors are already observed in the laboratory during each cycle, the term RIF would have limited use in the clinical setting. When proper preconception evaluation, including thyroid function test, is completed and uterine anatomy is reliably assessed, perhaps parental karyotyping is the only remaining assessment to be done in the case of unexplained failures. The recent ESHRE Good Practice Recommendations on RIF advise against investigations for microbiome profiling, uterine or peripheral natural killer cell testing, T lymphocyte assessment, cytokine levels, HLA-C compatibility, mitochondrial DNA testing, sperm DNA fragmentation or FISH analysis as well as against interventions such as endometrial injury, intravenous immunoglobulin, intralipid, peripheral blood mononuclear cells, platelet-rich plasma infusions, human chorionic gonadotropin, or low-molecular-weight heparin injections in routine clinical practice (ESHRE Working Group on Recurrent Implantation Failure, 2023). Honest and compassionate counseling of the patient/couple about possible etiology and future prognosis, including hearing, acknowledging, and respecting their concerns and feelings, avoiding unnecessary tests and interventions of unproven value and offering a new embryo transfer could preserve their motivation and funds, prevent drop-outs, and help them reach their reproductive goals. Preplanning for the possibility of requiring multiple embryo transfers can help patients maintain hope after a single failed transfers and can reduce the chance of treatment discontinuation (Faustini et al., 2023; Harrison et al., 2021, 2023). When the couple already has a frozen embryo available for transfer, the decision to proceed can be easier, but whether starting a new stimulation cycle is worthwhile would be a subjective decision that depends on the probability of achieving a euploid blastocyst, which is determined by female age and ovarian reserve status (Reig et al., 2023). When treatment costs are not covered by third parties, financial constraints also play a role in discontinuing treatment.

In conclusion, our results suggest that once obvious etiology for implantation failure is excluded, in patients who received three prior euploid blastocysts without implantation, the fourth and the fifth euploid blastocysts achieve similar live birth rates. Moreover, among women who were treated at RMA New Jersey, where data from the first three transfers were available, the fourth euploid blastocyst achieved a live birth rate as high as the first one. These observations are against a perceivable enrichment of a true RIF population until five consecutive euploid blastocyst transfers, thereby suggesting the prevalence of such an entity would be <2%. It should also be noted that, whether the sixth euploid blastocyst also achieves a similar implantation/live birth rate remains unknown, and further studies will be needed to determine the outcomes of sixth and subsequent transfers.

Our results in combination with others support encouraging patients with additional attempts with own gametes, whenever clinically reasonable (Cimadomo et al., 2021).

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

Data are not available for public.

Authors' roles

P.G.: data collection and drafting the manuscript. B.A.: study design, data collection and analysis, and drafting the manuscript. A.A.: data collection and critical review of the manuscript. D.C.: data collection and critical review of the manuscript. A.V.: data collection and critical review of the manuscript. H.F.: critical review of the manuscript. F.U.: critical review of the manuscript. J. G.-V.: conception of the study idea, study design, and critical review of the manuscript. E.S.: conception of the study idea, study design, and critical review of the manuscript.

Funding

No funding has been received for this study.

Conflict of interest

None of the authors have any conflict of interest associated with the present work.

References

- Ata B, Kalafat E, Somigliana E. A new definition of recurrent implantation failure on the basis of anticipated blastocyst aneuploidy rates across female age. *Fertil Steril* 2021;**116**:1320–1327.
- Ata B, Liñán A, Kalafat E, Ruíz F, Melado L, Bayram A, Elkhatib I, Lawrenz B, Fatemi HM. Effect of the endometrial thickness on the live birth rate: insights from 959 single euploid frozen embryo transfers without a cutoff for thickness. *Fertil Steril* 2023;**120**:91–98.
- Capalbo A, Poli M, Rienzi L, Girardi L, Patassini C, Fabiani M, Cimadomo D, Benini F, Farcomeni A, Cuzzi J et al. Mosaic human preimplantation embryos and their developmental potential in a prospective, non-selection clinical trial. *Am J Hum Genet* 2021;**108**:2238–2247.
- Capalbo A, Treff NR, Cimadomo D, Tao X, Upham K, Ubaldi FM, Rienzi L, Scott RT Jr. Comparison of array comparative genomic hybridization and quantitative real-time PCR-based aneuploidy screening of blastocyst biopsies. *Eur J Hum Genet* 2015;**23**:901–906.
- Cimadomo D, Capalbo A, Dovere L, Tacconi L, Soscia D, Giancani A, Scepi E, Maggiulli R, Vaiarelli A, Rienzi L et al. Leave the past behind: women's reproductive history shows no association with blastocysts' euploidy and limited association with live birth rates after euploid embryo transfers. *Hum Reprod* 2021;**36**:929–940.
- Cimadomo D, Rienzi L, Conforti A, Forman E, Canosa S, Innocenti F, Poli M, Hynes J, Gemmell L, Vaiarelli A et al. Opening the black box: why do euploid blastocysts fail to implant? A systematic review and meta-analysis. *Hum Reprod Update* 2023;**29**:570–633.
- Cozzolino M, Diaz-Gimeno P, Pellicer A, Garrido N. Evaluation of the endometrial receptivity assay and the preimplantation genetic

- test for aneuploidy in overcoming recurrent implantation failure. *J Assist Reprod Genet* 2020;**37**:2989–2997.
- Cozzolino M, Díaz-Gimeno P, Pellicer A, Garrido N. Use of the endometrial receptivity array to guide personalized embryo transfer after a failed transfer attempt was associated with a lower cumulative and per transfer live birth rate during donor and autologous cycles. *Fertil Steril* 2022;**118**:724–736.
- de Los Santos MJ, Díez Juan A, Mifsud A, Mercader A, Meseguer M, Rubio C, Pellicer A. Variables associated with mitochondrial copy number in human blastocysts: what can we learn from trophectoderm biopsies? *Fertil Steril* 2018;**109**:110–117.
- Doyle N, Jahandideh S, Hill MJ, Widra EA, Levy M, Devine K. Effect of timing by endometrial receptivity testing vs standard timing of frozen embryo transfer on live birth in patients undergoing in vitro fertilization: a randomized clinical trial. *JAMA* 2022;**328**:2117–2125.
- ESHRE Working Group on Recurrent Implantation Failure; Cimadomo D, de Los Santos MJ, Griesinger G, Lainas G, Le Clef N, McLernon DJ, Montjean D, Toth B, Vermeulen N, Macklon N. ESHRE good practice recommendations on recurrent implantation failure. *Hum Reprod Open* 2023;**2023**:hoad023.
- Faustini F, Cimadomo D, Ubaldi FM, Rienzi L. From a cycle-by-cycle to a multicycle treatment planning: the next relevant shift in perspective needed in assisted reproductive technology? *Fertil Steril* 2023;**121**:34–35.
- Forman EJ, Hong KH, Ferry KM, Tao X, Taylor D, Levy B, Treff NR, Scott RT Jr. In vitro fertilization with single euploid blastocyst transfer: a randomized controlled trial. *Fertil Steril* 2013;**100**:100–107.e1.
- Franasiak JM, Alecsandru D, Forman EJ, Gemmell LC, Goldberg JM, Llarena N, Margolis C, Laven J, Schoenmakers S, Seli E. A review of the pathophysiology of recurrent implantation failure. *Fertil Steril* 2021;**116**:1436–1448; Erratum in: *Fertil Steril*. 2022;**117**:653.
- Harrison C, Gameiro S, Boivin J. Patient willingness, preferences and decision-making about planning for three complete cycles of IVF/ICSI treatment. *Hum Reprod* 2021;**36**:1339–1352.
- Harrison C, Gameiro S, Boivin J. Qualitative evaluation of the acceptability and feasibility among healthcare professionals and patients of an ART multi-cycle treatment planning and continuation intervention prototype. *Hum Reprod* 2023;**38**:430–443.
- Pirtea P, De Ziegler D, Tao X, Sun L, Zhan Y, Ayoubi JM, Seli E, Franasiak JM, Scott RT Jr. Rate of true recurrent implantation failure is low: results of three successive frozen euploid single embryo transfers. *Fertil Steril* 2021;**115**:45–53.
- Rafael F, Robles GM, Navarro AT, Garrido N, Garcia-Velasco JA, Bosch E, Nunes SG, Soares SR, Santos-Ribeiro S. Perinatal outcomes in children born after fresh or frozen embryo transfer using donated oocytes. *Hum Reprod* 2022;**37**:1642–1651.
- Reig A, Garcia-Velasco JA, Seli E. Bologna vs. POSEIDON criteria as predictors of the likelihood of obtaining at least one euploid embryo in poor ovarian response: an analysis of 6,889 cycles. *Fertil Steril* 2023;**120**:605–614.
- Scott RT Jr, Upham KM, Forman EJ, Hong KH, Scott KL, Taylor D, Tao X, Treff NR. Blastocyst biopsy with comprehensive chromosome screening and fresh embryo transfer significantly increases in vitro fertilization implantation and delivery rates: a randomized controlled trial. *Fertil Steril* 2013;**100**:697–703.
- Shaulov T, Sierra S, Sylvestre C. Recurrent implantation failure in IVF: a Canadian Fertility and Andrology Society Clinical Practice Guideline. *Reprod Biomed Online* 2020;**41**:819–833.
- Somigliana E, Busnelli A, Kalafat E, Viganò P, Ata B. Recurrent implantation failure: a plea for a widely adopted rational definition. *Reprod Biomed Online* 2022;**45**:183–185.
- (The writing group) for the participants to the 2022 Lugano RIF Workshop; Pirtea P, Cedars MI, Devine K, Ata B, Franasiak J, Racowsky C, Toner J, Scott RT, de Ziegler D, Barnhart KT. Recurrent implantation failure: reality or a statistical mirage?: consensus statement from the July 1, 2022 Lugano Workshop on recurrent implantation failure. *Fertil Steril* 2023;**120**:45–59.
- Tiegs AW, Tao X, Zhan Y, Whitehead C, Kim J, Hanson B, Osman E, Kim TJ, Patounakis G, Gutmann J et al. A multicentre, prospective, blinded, nonselection study evaluating the predictive value of an aneuploid diagnosis using a targeted next-generation sequencing-based preimplantation genetic testing for aneuploidy assay and impact of biopsy. *Fertil Steril* 2021;**115**:627–637.
- Vaiarelli A, Cimadomo D, Alviggi E, Sansone A, Trabucco E, Dusi L, Buffo L, Barnocchi N, Fiorini F, Colamaria S et al. The euploid blastocysts obtained after luteal phase stimulation show the same clinical, obstetric and perinatal outcomes as follicular phase stimulation-derived ones: a multicentre study. *Hum Reprod* 2020;**35**:2598–2608.
- Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, Rienzi L, Sunde A, Schmidt L, Cooke ID et al. The international glossary on infertility and fertility care, 2017. *Hum Reprod* 2017;**32**:1786–1801.

Infertility

Economic evaluations of assisted reproductive technologies in high-income countries: a systematic review

Emily Olive ^{1,*}, Claudia Bull², Adrienne Gordon¹, Miranda Davies-Tuck³, Rui Wang ^{4,†}, and Emily Callander^{5,†}

¹Discipline of Obstetrics, Gynaecology and Neonatology, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia

²Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia

³The Ritchie Centre, Hudson Institute of Medical Research, Melbourne, VIC, Australia

⁴Department of Obstetrics and Gynaecology, Monash University, Melbourne, VIC, Australia

⁵School of Public Health, University of Technology Sydney, Sydney, NSW, Australia

*Correspondence address. Discipline of Obstetrics, Gynaecology and Neonatology, Faculty of Medicine and Health, University of Sydney, Level 2, Charles Perkins Centre D17, Sydney, NSW, Australia. E-mail eoli0633@uni.sydney.edu.au  <https://orcid.org/0000-0002-3366-4204>

[†]These authors contributed equally to this work.

ABSTRACT

STUDY QUESTION: Which assisted reproductive technology (ART) interventions in high-income countries are cost-effective and which are not?

SUMMARY ANSWER: Among all ART interventions assessed in economic evaluations, most high-cost interventions, including preimplantation genetic testing for aneuploidy (PGT-A) for a general population and ICSI for unexplained infertility, are unlikely to be cost-effective owing to minimal or no increase in effectiveness.

WHAT IS KNOWN ALREADY: Approaches to reduce costs in order to increase access have been identified as a research priority for future infertility research. There has been an increasing number of ART interventions implemented in routine clinical practice globally, before robust assessments of evidence on economic evaluations. The extent of clinical effectiveness of some studied comparisons has been evaluated in high-quality research, allowing more informative decision making around cost-effectiveness.

STUDY DESIGN, SIZE, DURATION: We performed a systematic review and searched seven databases (MEDLINE, PUBMED, EMBASE, COCHRANE, ECONLIT, SCOPUS, and CINAHL) for studies examining ART interventions for infertility together with an economic evaluation component (cost-effectiveness, cost-benefit, cost-utility, or cost-minimization assessment), in high-income countries, published since January 2011. The last search was 22 June 2022.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Two independent reviewers assessed publications and included those fulfilling the eligibility criteria. Studies were examined to assess the cost-effectiveness of the studied intervention, as well as the reporting quality of the study. The chosen outcome measure and payer perspective were also noted. Completeness of reporting was assessed against the Consolidated Health Economic Evaluation Reporting Standard. Results are presented and summarized based on the intervention studied.

MAIN RESULTS AND THE ROLE OF CHANCE: The review included 40 studies which were conducted in 11 high-income countries. Most studies ($n = 34$) included a cost-effectiveness analysis. ART interventions included medication or strategies for controlled ovarian stimulation ($n = 15$), IVF ($n = 9$), PGT-A ($n = 7$), single embryo transfer ($n = 5$), ICSI ($n = 3$), and freeze-all embryo transfer ($n = 1$). Live birth was the mostly commonly reported primary outcome ($n = 27$), and quality-adjusted life years was reported in three studies. The health funder perspective was used in 85% ($n = 34$) of studies. None of the included studies measured patient preference for treatment. It remains uncertain whether PGT-A improves pregnancy rates compared to IVF cycles managed without PGT-A, and therefore cost-effectiveness could not be demonstrated for this intervention. Similarly, ICSI in non-male factor infertility appears not to be clinically effective compared to standard fertilization in an IVF cycle and is therefore not cost-effective. Interventions such as use of biosimilars or HMG for ovarian stimulation are cheaper but compromise clinical effectiveness.

LIMITATIONS, REASONS FOR CAUTION: Lack of both preference-based and standardized outcomes limits the comparability of results across studies. The selection of efficacy evidence offered for some interventions for economic evaluations is not always based on high-quality randomized trials and systematic reviews. In addition, there is insufficient knowledge of the willingness to pay thresholds of individuals and state funders for treatment of infertility. There is variable quality of reporting scores, which might increase uncertainty around the cost-effectiveness results.

Received: February 13, 2023. Revised: January 10, 2024. Editorial decision: January 22, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

WIDER IMPLICATIONS OF THE FINDINGS: Investment in strategies to help infertile people who utilize ART is justifiable at both personal and population levels. This systematic review may assist ART funders decide how to best invest to maximize the likelihood of delivery of a healthy child.

STUDY FUNDING/COMPETING INTEREST(S): There was no funding for this study. E.C. and R.W. receive salary support from the National Health and Medical Research Council (NHMRC) through their fellowship scheme (EC GNT1159536, RW 2021/GNT2009767). M.D.-T. reports consulting fees from King Fahad Medical School. All other authors have no competing interests to declare.

REGISTRATION NUMBER: Prospero CRD42021261537.

Keywords: IVF / reproductive techniques / assisted / preimplantation genetic testing / economic analysis / cost-effectiveness / cost-utility / systematic review / ART

Introduction

As education and career opportunities have expanded for women, many women in high-income countries have elected to delay starting a family [Chua et al. \(2020\)](#). With increasing age, there is attrition in oocyte number and increased risk of aneuploidy in the remaining oocytes ([Yatsenko and Rajkovic, 2019](#); [Chua et al., 2020](#)). Concurrently, endometriosis, tubal disease, and increased BMI may coincide as impediments to fertility ([Broughton and Moley, 2017](#); [Zondervan et al., 2020](#)). Infertile women are likely to experience symptoms of anxiety more than others ([Kiani et al., 2020](#)). Combinations of all these factors together lead many women to consider assisted reproduction ([Somigliana et al., 2016](#)).

The term ART encompasses any technique that includes ‘the in vitro (outside of body) handling of human oocytes (eggs) and sperm or embryos for the purposes of establishing a pregnancy’ ([Newman and Chambers, 2021](#)). An audit from The International Federation of Fertility Societies found a total of 6977 clinics in 2022, an increase of 776 clinics worldwide since the previous report in 2019. IVF is the mainstay of work in these clinics: a current estimate of 940 503 IVF cycles were managed in Europe in 2017, an increase of 2.4% since 2014 ([The European IVF Monitoring Consortium for the European Society of Human Reproduction et al., 2021](#)). A linear rise in the number of treatment cycles is also evident in the USA, with more than 160 000 cycles annually in the 2016 update ([Toner et al., 2016](#)). In response to high, and increasing, demand for care, IVF has become a burgeoning industry in all high-income settings ([International Federation of Fertility Societies’ Surveillance \(IFFS\), 2019, 2022](#)).

The economic burden of infertility as well as approaches to reduce costs and increase access have been identified in the top 10 research priorities for future infertility research ([Duffy et al., 2021a](#)). As of 2020, 19 American states require insurers to offer infertility treatment coverage, but IVF is sometimes excluded from the infertility treatments offered, or the number of funded IVF cycles is limited ([Kawwass et al., 2021](#)). The European IVF-Monitoring Consortium notes there is no public funding available in Switzerland, and that of the 39 countries providing some funding, there is a three-cycle limit in 16 countries ([Calhaz-Jorge et al., 2020](#)). The provision or proportion of government funding can also vary with maternal age, existence of older children, and female BMI ([Calhaz-Jorge et al., 2020](#); [Kawwass et al., 2021](#)).

The impact of high cost on individuals or couples in an otherwise publicly funded healthcare system may be one reason why only half of surveyed British adults sought treatment for infertility ([Datta et al., 2016](#)). Furthermore, over one-third of couples surveyed from eight different countries acknowledged cost was a barrier to accessing or proceeding with assisted reproduction ([Domar et al., 2021](#)). There is also sometimes a gulf between preparedness to pay and capacity to pay for help with infertility treatments ([Dyer et al., 2017](#)). IVF access has socioeconomic

associations: access is easier for those with higher incomes and higher educational achievement ([Tierney and Cai, 2019](#)). Ultimately both health care funders and those seeking care for infertility require treatments to be supported by high-quality evidence demonstrating improved outcomes. Better outcomes might justify additional cost.

Over the past decade, there has been an increasing number of ART interventions implemented in routine clinical practice globally, resulting in a shift towards single embryo transfer (SET), blastocyst transfer, freeze-all transfer strategy, the use of preimplantation genetic testing for aneuploidy (PGT-A), and ICSI, as well as other IVF add-ons ([Kushnir et al., 2017](#); [Gleicher et al., 2019, 2021](#); [Bahadur et al., 2020](#); [de Geyter et al., 2020](#); [Lensen et al., 2021](#)). In addition to the concerns about the lack of evidence on the effectiveness and/or safety for some interventions, other concerns have also been raised about the lack of robust assessments of economic evaluations before implementing most interventions in clinical practice ([ESHRE Capri Workshop Group, 2015](#)). Given that the last systematic review on the cost-effectiveness of ART interventions was conducted in 2002 ([Garceau et al., 2002](#)) and ART clinical practice has evolved, it is important to systematically map and assess the latest economic evaluations of ART interventions available in current clinical practice to inform health policy and future research. This systematic review aims to provide an overview of economic evaluations of all ART interventions performed in high-income countries in order to assist both couples and clinicians, and to synthesize the economic evaluations and the reporting quality of these studies.

Materials and methods

Study design

This systematic review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist ([Page et al., 2021](#)). The review protocol was registered with the international prospective register of systematic reviews (PROSPERO) (CRD42021261537).

Study eligibility and search strategy

Economic evaluation studies on any type of ART interventions were included in the review, including cost-effectiveness, cost-benefit, cost-utility, and cost-minimization studies. Studies were included if they reported results from high-income countries as defined by the World Bank ([World Bank, 2021](#)), and were published in English language, peer reviewed journals, and since January 2011. We included studies irrespective of the outcome measure assessed (e.g. live birth). Studies were excluded if they assessed ART for indications other than infertility (e.g. egg freezing).

The following databases were searched for eligible studies ([Supplementary File S1](#)): MEDLINE, PUBMED, CINAHL, ECONLIT,

COCHRANE, EMBASE, and SCOPUS. The reference lists of all included studies were also checked to identify additional eligible studies (Supplementary Table S1). The last search was conducted on 22 June 2022.

Study selection and reporting assessment

Two reviewers (E.O. and C.B.) independently screened titles and abstracts in Covidence systematic review software (Veritas Health Innovation, 2021). The same two reviewers then independently assessed full-text articles for inclusion. Where there were disagreements about article inclusion, reviewers first attempted to resolve this through discussion to reach consensus. Where consensus could not be reached, a third reviewer (E.C. or R.W.) adjudicated the final decision. The professional society for health economics and outcomes research (ISPOR) recommends that health economic studies be evaluated according to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) (Husereau et al., 2022). This checklist is a reporting rather than quality evaluation and contains 24 items to be included for thorough reporting. Documentation of adherence to the CHEERS reporting criteria is included in Supplementary Table S2.

Data extraction and synthesis

Data were extracted to an Excel spreadsheet by one author (E.O.) was reviewed by C.B. and E.C. for accuracy. Extracted information included: author, year, population studied, perspective taken, time horizon, intervention and comparator, costs associated with the intervention and comparator, incremental costs, and discounting of costs. The key effect measure studied was the incremental cost-effectiveness ratio (ICER). We converted all cost outcomes to 2021 costs using <https://www.inflationtool.com>. To enable comparisons across studies, we converted all cost data into US Dollars (USD) using <https://www.xe.com>.

The key characteristics of studies were tabulated descriptively. Types of interventions investigated in economic evaluations were visualized in a sunburst. Findings of different studies were grouped according to the type of interventions evaluated. A synopsis of the findings of economic evaluations for each intervention studied is provided in the Results section below.

Results

Of the 505 papers identified in database searching, 40 papers fulfilled our eligibility criteria of an economic evaluation of an intervention in assisted reproduction in the last decade (Fig. 1).

The key characteristics of included studies are provided in Table 1. Two-thirds of included papers were published after 2016 and half were from European settings. Studies addressed either a general (unselected) population ($n = 26$), poor prognosis patients ($n = 7$), unexplained infertility ($n = 5$), male infertility ($n = 1$), or predicted high responders ($n = 1$). The ART interventions studied included controlled ovarian stimulation ($n = 15$), IVF ($n = 9$), PGT-A ($n = 7$), SET ($n = 5$), ICSI ($n = 3$), and freeze-all embryo transfer ($n = 1$). The details of these interventions are visualized in Fig. 2. Live birth was the most reported primary outcome ($n = 27$), followed by ongoing pregnancy ($n = 6$) and quality-adjusted life years ($n = 3$). In 26/40 studies, the economic evaluation was by way of a model. The clinical efficacy inputs were calculated from at least some trial data in 10/26 studies, while inputs in the other 16 modelling studies were based on observational data. The perspective was solely the health funder in 32 of the 40 papers. A cost-effectiveness analysis (rather than cost-utility or cost minimization) was provided in 34 papers.

Study reporting quality

Comprehensive details of the CHEERS assessment results are provided in Supplementary Table S2. Reporting of both the target population and time horizon of the study was noted in 32 of 40 studies whereas no study reported on patient preferences for outcomes.

Those studies comparing immediate IVF with expectant management, and the group of studies reporting embryo transfer number, had greater adherence to the checklist relative to studies on other interventions.

The perspective of the economic evaluation was noted in 38 of 40 included papers. The source of funding considered in the economic evaluation was explicitly stated in 31 papers. Of the 10 studies comparing agents for ovarian hyperstimulation, nine acknowledged their funding source, and eight of these nine were funded by industry.

The ART interventions evaluated IVF

Among 10 included studies on IVF, nine evaluated IVF versus non-IVF interventions (four on expectant management and five on IUI) and the other study compared different IVF funding policies (Fig. 2).

IVF versus expectant management

Consideration of a delay in IVF start for those with unexplained infertility was explored in four studies (Supplementary Table S3). Pham et al. (2018) assessed the cost-effectiveness of a 6-month delay for women with unexplained infertility aged <40 years, from both the health funder and the patient perspective. Eijkemans assessed the cost-effectiveness of a 12-month delay for couples from the health funder perspective only (Eijkemans et al., 2017). van Eekelen assessed the cost-effectiveness of ovarian hyperstimulation and IUI as an intermediary step between expectant management and IVF (van Eekelen et al., 2020). Finally, Busnelli assessed the cost of IVF rather than expectant management in a group of women pre-defined to be poor responders (Busnelli and Somigliana, 2018).

All papers used the outcome measure of cost per livebirth. An 18-month time frame was used to model the impact of a 6-month period of expectant management in women <40 years of age in Australia. A reduction in livebirth rate of 0.1–0.2% was associated with a delayed start of IVF treatment (Pham et al., 2018). Eijkemans et al. noted that IVF becomes more cost-effective compared to expectant management as women age. van Eekelen noted that expectant management for 1–2 years before beginning IVF lowers the average cost per livebirth compared to an immediate IVF start (van Eekelen et al., 2020).

IVF versus IUI

Four papers used a health funder perspective to evaluate IUI as an alternative to IVF (Supplementary Table S4) (van Rumste et al., 2014; Yu et al., 2014; Tjon-Kon-Fat et al., 2015; Almaslami and Aljunid, 2020). All found IVF to be slightly more effective than IUI to achieve an ongoing pregnancy, livebirth, or healthy child. Three of the four studies found a higher cost associated with IVF, resulting in ICERs of an additional US \$50 492 per additional ongoing pregnancy, an additional \$37 925 per additional live birth or an additional \$50 490 per additional healthy child, respectively.

IVF funding policy

A Canadian study modelled the impact of provincial funding strategies for IVF. Results were stratified for embryo transfer

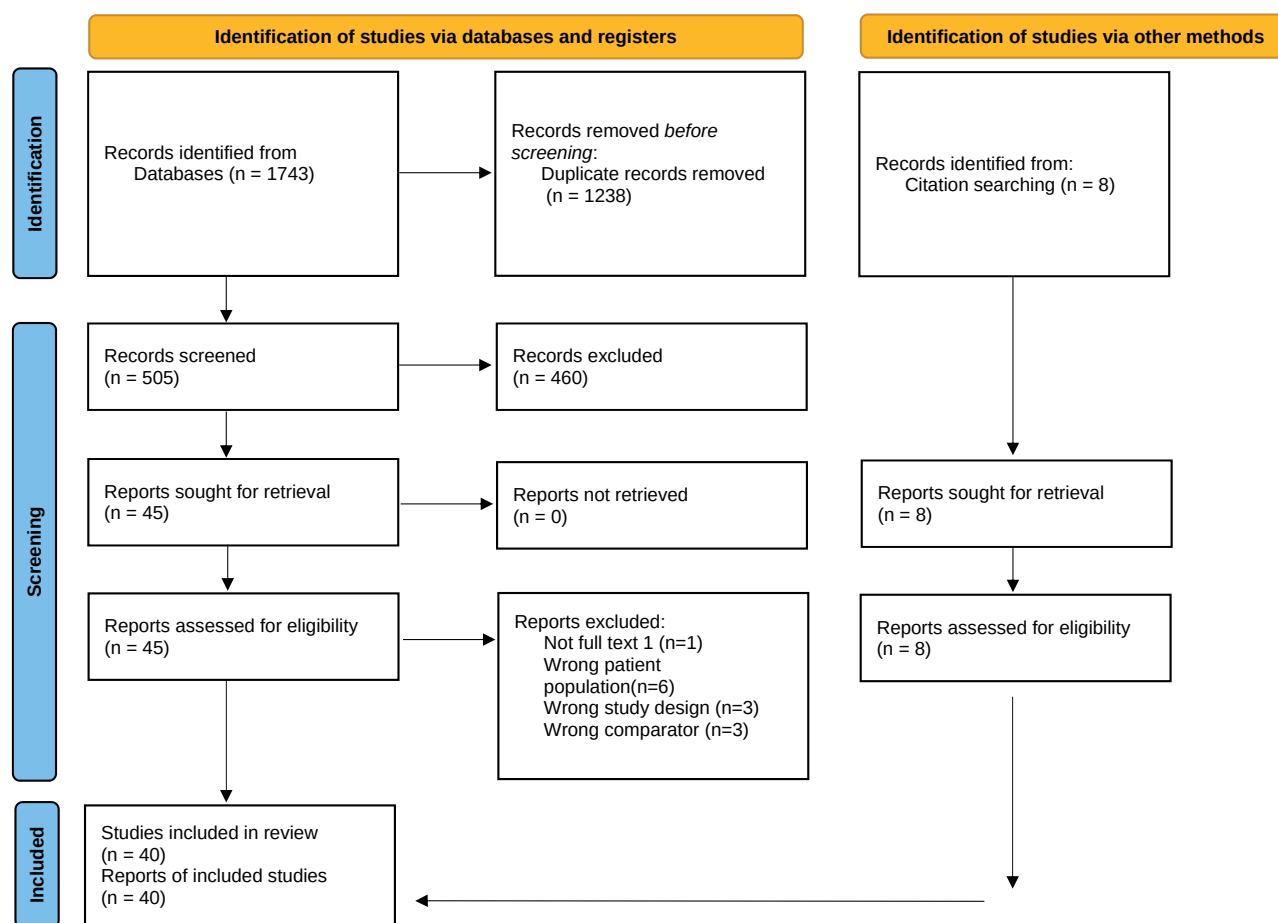


Figure 1. PRISMA flowchart for identifying studies included in a systematic review of economic evaluations of ART in high-income countries.

number and maternal age (Vaidya et al., 2015). Cost-effectiveness was greater when double embryo transfer (DET) was permitted at higher maternal ages.

ICSI

Three papers examined the cost-effectiveness of use of ICSI for fertilization in IVF cycles where male factor infertility was not present (Supplementary Table S5). In models designed from a health funder perspective in the USA and the Netherlands, ICSI had a similar ICER per additional live birth gained for unexplained infertility (\$66 742) and mild male factor infertility (\$72 000) (Vitek et al., 2013; Moolenaar et al., 2015). Rescue ICSI in the setting of total fertilization failure (as an alternative to cycle cancellation) resulted in 11% more live births at lower cost in the third study (Shalom-Paz et al., 2011).

PGT-A

Seven papers included in this review assessed the economics of using PGT-A as an adjunct to IVF, with the outcome measure of live birth (Table 2). In five studies, the funder perspective was chosen, but three of these also calculated costs from the patient perspective (Lee et al., 2019, 2021; Neumann et al., 2020).

In two studies, PGT-A lessened the live birth rate and increased costs (Somigliana et al., 2019; Lee et al., 2021). In the three studies using a health funder perspective and reporting higher live birth rates with PGT-A, the ICER to achieve an additional live birth ranged from an additional \$20 507 to \$607 833 (Lee et al., 2019; Neumann et al., 2020).

Controlled ovarian stimulation

Among 15 studies on controlled ovarian stimulation, three evaluated individualized controlled ovarian stimulation, one evaluated the downregulation protocol, one evaluated natural cycle versus stimulated IVF, and the other 10 studies reported medication for ovarian stimulation (Fig. 2).

Individualized controlled ovarian stimulation

Three papers assessed ovarian reserve testing from a health funder perspective with a view to individualizing the gonadotrophin dose for ovarian hyperstimulation (Supplementary Table S6). Moolenaar modelled both IVF with standard dosing and IVF with individualized dosing compared to no treatment, and concluded that both choices were likely to be considered to be cost-effective (Moolenaar et al., 2011). Yates found individualized dosing resulted in additional ongoing pregnancies with an ICER of an additional \$6067 per ongoing pregnancy (Yates et al., 2011). van Tilborg et al. (2017) subsequently conducted a multicentre prospective cohort study of 1515 women and found that individualized dosing resulted in fewer pregnancies at greater expense.

Downregulation protocol

A cost-minimization study based on a retrospective cohort examined progesterone as a downregulation agent (Supplementary Table S7) (Evans et al., 2019). The cost per livebirth was measured from the perspective of the health funder over a 2-month time horizon. The GnRH antagonist downregulation cost \$6676 less to achieve a live birth compared to progesterone, and the GnRH

Table 1. Overview of 40 included studies of an ART intervention together with a cost analysis.

	Number of studies	Percentage of studies
Publication year		
2011–2015	14	35
2016–2021	26	65
Country of publication		
Netherlands	11	27.5
USA	8	20
Italy	4	10
Germany	3	7.5
Canada	2	5
France	2	5
European collaboration ¹	2	5
UK	2	5
Australia	2	5
Spain	2	5
Saudi Arabia	1	2.5
Greece	1	2.5
Population studied		
Unselected population	26	65
Unexplained infertility	5	12.5
Poor prognosis patients	7	17.5
Male infertility	1	2.5
Predicted high responders	1	2.5
Intervention studied		
Controlled ovarian stimulation	15	37.5
IVF	9	22.5
PGT-A	7	17.5
Embryo transfer number	5	12.5
ICSI	3	7.5
Freeze-all embryo transfer	1	2.5
Study type		
Decision model	26	65
Concurrent with a trial	5	12.5
With a cohort study	9	22.5
Source of effectiveness data		
From randomized trials	14	35
From observational data	26	65
Type of economic analysis		
Cost-effectiveness alone	32	80
Combined cost-effectiveness and cost-utility ¹	2	5
Cost-utility alone	1	2.5
Cost minimization	5	12.5
Payer perspective		
Health funder alone	32	80
Health funder and patient perspective	6	15
Perspective not discussed	2	5

¹ Gizzo *et al.* (2017) examines outcomes in Italy and Spain, and Gizzo *et al.* (2018) examines outcomes in Italy, Spain, and Germany. The papers by Mennini *et al.* (2018) and Scotland *et al.* (2011) measured both cost per additional birth and quality-adjusted life years per additional birth. PGT-A, preimplantation genetic testing for aneuploidy.

agonist downregulation was also \$9753 cheaper than progesterone to achieve a live birth.

Natural cycle IVF versus stimulated IVF

A retrospective cohort study was used to evaluate the cost-effectiveness of six cycles of modified natural cycle for ovulatory women against a single IVF cycle with controlled ovarian hyperstimulation (Supplementary Table S8) (Groen *et al.*, 2013). The cost per live birth was measured from the health funder perspective with a 10-month time horizon. A single IVF cycle with ovarian hyperstimulation was cheaper and as effective as six cycles of modified natural cycle.

Medication for ovarian hyperstimulation

An economic evaluation of agents for ovarian hyperstimulation was the topic of 10 included papers (Supplementary Table S9).

A cost-effectiveness evaluation was described in 9 of 10 papers. The perspective was solely that of the health funder in eight papers, whereas the remaining two reported from both the perspective of the patient and health funder (Fragoulakis *et al.*, 2012, 2016; Gizzo *et al.*, 2017, 2018; Barrenetxea *et al.*, 2018; Barriere *et al.*, 2018; Grynberg *et al.*, 2018; Mennini *et al.*, 2018; Xue *et al.*, 2019; Robins *et al.*, 2020). Four studies were cost-effectiveness analyses of use of a biosimilar compared to recombinant FSH (rFSH). The incremental cost to use rFSH over a comparator varied from an additional \$506 to \$8474 (Gizzo *et al.*, 2018; Xue *et al.*, 2019).

Embryo transfer number

Six papers assessed the cost policies of transferring one or more embryos when more than one is available (Scotland *et al.*, 2011; Hernandez Torres *et al.*, 2015; Vaidya *et al.*, 2015; van Heesch *et al.*, 2016; van Loendersloot *et al.*, 2017; Sitler *et al.*, 2020) (Supplementary Tables S10 and S11). The extreme variation in time horizon chosen for economic evaluation impacts on comparability of these papers.

Transfer of two embryos was found to be increasingly cost-effective with increasing maternal age in three cohort papers (Scotland *et al.*, 2011; Vaidya *et al.*, 2015; van Loendersloot *et al.*, 2017). The incremental cost to attain an additional live birth with transfer of two embryos rather than one fell from an additional \$43 000 for a 32-year-old mother to \$24 000 for a 39-year-old mother (Scotland *et al.*, 2011). Dutch and Canadian studies, also from a funder perspective, showed the same trends (Vaidya *et al.*, 2015; van Loendersloot *et al.*, 2017).

Cost-effectiveness varied with perspective: in the USA, SET is the dominant strategy from the health funder perspective, whereas an incremental cost of \$27 145 is associated with SET for the patient. This is explained by the patient paying for IVF cycles, and the funder paying for obstetric and paediatric care (Sitler *et al.*, 2020).

Freeze-all embryo transfer

A single paper was included addressing this topic for cost-effectiveness from a funder perspective (Supplementary Table S12). Based on observational evidence, a freeze all policy was more effective and less costly than fresh embryo transfer (Papaleo *et al.*, 2017).

Discussion

Summary of key findings

This review identifies and describes studies reporting a contemporary economic evaluation of ART interventions in high-income countries. These evaluations involved a limited number of interventions, including IVF, ICSI, PGT-A, controlled ovarian stimulation, SET, and freeze-all embryo transfer, but most other IVF add-ons were not reported in any economic evaluation study. Dominant interventions are rare in the ART sphere, as incremental gains in clinical efficacy are relatively small while new technologies are relatively expensive for most interventions. Interventions such as use of biosimilars or HMG for ovarian stimulation are cheaper but compromise clinical effectiveness. IVF *per se* is more effective and more expensive than expectant management or IUI. High-cost interventions, including PGT-A or ICSI for non-male factor infertility, seem unlikely to be clinically effective, and are therefore not likely to be cost-effective. Cost-effectiveness around the number of embryos transferred depends on the group studied, the outcomes measured, and the time horizon used. In future work the preferences of those affected should



Figure 2. Sunburst diagram: research landscape of interventions in economic evaluations of ART. SET, single embryo transfer; EM, expectant management; PGT-A, preimplantation genetic testing for aneuploidy; COS, controlled ovarian stimulation. COS n = 15 studies, IVF n = 9, PGT-A n = 7, SET n = 5, ICSI n = 3, freeze-all n = 1.

be reported, as this impacts on the feasibility of practice change implementation.

Interpretation

Economic evaluations are dependent on the quality of evidence on effectiveness. Interpretation of these findings should therefore be cautious. Our review showed that not all included studies used robust evidence from clinical trials and systematic reviews on clinical efficacy to inform their economic evaluations. Some of the claimed efficacy seems to be exaggerated when compared to the latest high-quality evidence from clinical trials and Cochrane reviews. In these cases, the new intervention is either not effective or there is a lack of evidence on effectiveness, and therefore consensus has not been reached prior to conducting economic evaluations. Nevertheless, these economic evaluations were conducted and published, which results in research waste and may misinform practice and policy. An example was the single included study of a freeze-all strategy showing greater efficacy and lower cost than fresh transfer, but with results based on a relatively small observational dataset on effectiveness.

We note that included cohort studies find that DET is increasingly cost-effective in two instances: in populations of advanced

maternal age and also when cost-effectiveness is assessed with a time horizon extending until adulthood of the resultant children (Scotland et al., 2011; Vaidya et al., 2015). In 2014, two embryos were transferred in 16.6% of cycles in Australia but 54.5% of fresh cycles in Europe (Harris et al., 2016; de Geyter et al., 2018). Offering DET to older women aims to create more singleton pregnancies, and to lessen the proportion of couples who would otherwise remain childless. No randomized trial of embryo transfer number in women of advanced maternal age is available to support or refute the findings of the included cohort studies.

Use of PGT-A varies from 2% in the UK to 27% in the USA in 2016. Trials published subsequently have shown no improvement in the live birth rate related to the addition of PGT-A to a standard IVF cycle. (Ozgur et al., 2019; Theobald et al., 2020; Yan et al., 2021). The apparent lack of improvement in clinical efficacy is the driver for the high ICERs noted in this review (Collins et al., 2017; Neal et al., 2018; Lee et al., 2019, 2021; Somigliana et al., 2019; Neumann and Griesinger, 2020; Neumann et al., 2020). ICSI for non-male infertility does not increase live birth rates (Boulet et al., 2015; American Society for Reproductive Medicine, 2020; Dang et al., 2021) and is therefore not a cost-effective intervention. Both PGT-A and ICSI for non-male factor infertility had,

Table 2. Comparison of IVF with and without preimplantation genetic testing for aneuploidy.

Author/year	Country	Economic evaluation	Study Type	Perspective	Intervention	Comparator	Time horizon	Effect measure	ICER (US Dollars 2021)
Collins 2017	USA	Cost-effectiveness	Cohort	Health funder	PGT-A	Standard IVF	1 month	Livebirth	\$115 923
Lee, Evelyn 2019	Australia	Cost-effectiveness	Cohort	Health funder	PGT-A	Standard IVF	Three complete cycles	Livebirth	\$20 507
Lee, Malinda 2021	USA	Cost-effectiveness	Cohort	Patient	PGT-A	Standard IVF	12 months	livebirth	\$45 444
Neal 2018	USA	Cost minimization	Cohort	Patient	PGT-A	Standard IVF	12 months	livebirth	\$80 400
				n/a	PGT-A	Standard IVF	n/a	livebirth	\$121 033
					Single embryo				Cost savings
					Two embryos				\$108 with PGT-A.
					Five embryos				Cost savings
									\$2102 with PGT-A
									Cost savings
									\$2601 with PGT-A
									\$607 833
Neumann 2020	Germany	Cost-effectiveness	RCT	Health funder	PGT-A	Standard IVF	12 months	Livebirth	\$93 069
Neumann and Griesinger 2020	Germany	Cost-effectiveness	RCT	Patient	PGT-A	Standard IVF	12 months	Livebirth	
Somigliana 2019	Italy	Cost-effectiveness	Cohort	Health funder	PGT-A	Standard IVF	n/a	livebirth	
					With a single blastocyst at maternal age (years):				
					35				\$63 899
					38				\$103 830
					42				\$276 893
					PGT-A with six blastocysts at maternal age (years):				
					35				\$220 159
					38				\$357 749
					42				\$954 005

RCT, randomized controlled trial; ICER, incremental cost-effectiveness ratio.

however, become accepted options in some centres prior to availability of high-quality data reporting outcomes (Peipert et al., 2021).

National spending on ART is valuable to society as well as family units (ESHRE Capri Workshop Group, 2015). The 2021 Australian Intergenerational Report notes that fertility rates are directly associated with national productivity. Aged care spending is anticipated to almost double between 2020–2021 and 2060–2061 while the proportion of the working age population falls (The Commonwealth of Australia, 2021). National spending on ART therefore appears supported for the benefit of individuals but also state funders. As there will always be competing demands for funding from a national health budget, careful outcome measurement helps substantiate cost-effectiveness. To this end, adherence to a consistent optimal outcome measure (healthy term infant) and consistent gestation thresholds matter (Duffy et al., 2021b). Whereas infertile women who conceive without ART have an odds ratio (OR) 1.38 (95% CI 1.25–1.54) for preterm birth compared to those who conceive spontaneously, those with an ART conception are even more likely to deliver preterm, with an OR 2.46 (1.46–4.18) compared to women with natural conception (Messerlian et al., 2012; Cavoretto et al., 2022). Strategies to improve these outcomes in turn improve cost-effectiveness of the ART investment.

It is noteworthy that no included study included reference to the preferences of infertile patients. Recent research has noted that women are prepared to accept an increased risk of hyperstimulation if they thought it would increase their likelihood of conception (Stormlund et al., 2019; Abdulrahim et al., 2021). At present there are too few evaluations of the preferences of women for infertility treatments which might result in a more complex pregnancy and higher risks of preterm birth or low birthweight. Decisions around effective use of allocated funds to treat infertility require incorporation of the informed preferences of those affected and the willingness of the payer to pay a premium for small incremental gains. Perceptions of cost-effectiveness will vary, and families should be entitled to spend more if they wish in pursuit of small gains. As care providers, our duty is to provide accurate information on which to base these decisions.

Strengths and limitations

The strength of this paper is the inclusion and comparison of cost-effectiveness studies from countries with different health funding structures, enabling direct financial comparisons. Some included studies use measures of effectiveness that may be thought invalid or exaggerated in the wider literature (Shields and Elvidge, 2020). Variation in the populations studied alter cost-effectiveness figures as clinical efficacy may change considerably depending on the patient cohort studied.

In addition, limiting the review to high-income countries may be seen as a weakness given the enormous growth of ART in emerging economies. Economic analysis to allow selection of clinically effective interventions and to allow selection of the most cost-effective strategy is arguably even more valid in middle- and low-income countries (Chiware et al., 2021).

Conclusion

In economic evaluations, stated gains in effectiveness should be based on high-quality data to avoid a potentially biased economic assessment. In this systematic review, we found that not all economic evaluations are based on such evidence on

effectiveness. In interventions with no proven clinical benefits, economic evaluations could be of limited value.

Outcomes of ART studies should ideally be reported with gestational age at delivery and birthweight data stratified by IVF intervention. There are long-term health and educational costs associated with early delivery and small for gestational age birthweight. Economic evaluation time horizons should be long enough to capture childhood outcomes (Bentley et al., 2018). Use of quality evidence on effectiveness and measurement of obstetric and neonatal outcomes should pave the way for improved cost-effective use of ART.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Authors' roles

E.O. was involved in conception and design of the study, with C. B. and E.C. Data acquisition was by E.O., C.B., and R.W. E.O. was involved in data interpretation with A.G., M.D.-T., and R.W. Draft revisions were written by E.O., with input from all authors. All authors approved the final version.

Funding

No funding was received for this study. E.C. and R.W. receive salary support from the National Health and Medical Research Council (NHMRC) through their fellowship scheme (EC GNT1159536, RW GNT2009767).

Conflict of interest

M.D.-T. reports consulting fees from King Fahad Medical School. All other authors have no competing interests to declare.

References

- Abdulrahim B, Scotland G, Bhattacharya S, Maheshwari A. Assessing couples' preferences for fresh or frozen embryo transfer: a discrete choice experiment. *Hum Reprod* 2021;**36**:2891–2903.
- Almaslami F, Aljunid SM. Cost-effectiveness of assisted reproductive technologies in Saudi Arabia: comparing in vitro fertilization with intrauterine insemination. *SAGE Open Med* 2020;**8**:8.
- American Society for Reproductive Medicine. Intracytoplasmic sperm injection (ICSI) for non-male factor indications: a committee opinion. *Fertil Steril* 2020;**114**:239–245.
- Bahadur G, Homburg R, Bosmans JE, Huirne JAF, Hinstridge P, Jayaprakasan K, Racich P, Alam R, Karapanos I, Illahibuccus A et al. Observational retrospective study of UK national success, risks and costs for 319,105 IVF/ICSI and 30,669 IUI treatment cycles. *BMJ Open* 2020;**10**:e034566.
- Barrenetxea G, Garcia-Velasco JA, Aragon B, Osset J, Brosa M, Lopez-Martinez N, Coroleu B. Comparative economic study of the use of corifollitropin alfa and daily rFSH for controlled ovarian stimulation in older patients: cost-minimization analysis based on the PURSUE study. *Reprod Biomed Soc Online* 2018;**5**:46–59.
- Barriere P, Porcu-Buisson G, Hamamah S. Cost-effectiveness analysis of the gonadotropin treatments HP-hMG and rFSH for assisted

- reproductive technology in France: a Markov model analysis. *Appl Health Econ Health Policy* 2018;**16**:65–77.
- Bentley JP, Schneuer FJ, Lain SJ, Martin AJ, Gordon A, Nassar N. Neonatal morbidity at term, early child development, and school performance: a population study. *Pediatrics* 2018;**141**:e20171726.
- Boulet SL, Mehta A, Kissin DM, Warner L, Kawwass JF, Jamieson DJ. Trends in use of and reproductive outcomes associated with intracytoplasmic sperm injection. *Jama* 2015;**313**:255–263.
- Broughton DE, Moley KH. Obesity and female infertility: potential mediators of obesity's impact. *Fertil Steril* 2017;**107**:840–847.
- Busnelli A, Somigliana E. Prognosis and cost-effectiveness of IVF in poor responders according to the Bologna criteria. *Minerva Ginecol* 2018;**70**:89–98.
- Calhaz-Jorge C, de Geyter CH, Kupka MS, Wyns C, Mocanu E, Motrenko T, Scaravelli G, Smeenk J, Vidakovic S, Goossens V. Survey on ART and IUI: legislation, regulation, funding and registries in European countries: the European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). *Hum Reprod Open* 2020;**2020**:hoz044.
- Cavoretto PI, Giorgione V, Sotiriadis A, Vigano P, Papaleo E, Galdini A, Gaeta G, Candiani M. IVF/ICSI treatment and the risk of iatrogenic preterm birth in singleton pregnancies: systematic review and meta-analysis of cohort studies. *J Matern Fetal Neonatal Med* 2022;**35**:1987–1996.
- Chiware TM, Vermeulen N, Blondeel K, Farquharson R, Kiarie J, Lundin K, Matsaseng TC, Ombet W, Toskin I. IVF and other ART in low- and middle-income countries: a systematic landscape analysis. *Hum Reprod Update* 2021;**27**:213–228.
- Chua SJ, Danhof NA, Mochtar MH, van Wely M, McLernon DJ, Custers I, Lee E, Dreyer K, Cahill DJ, Gillett WR et al. Age-related natural fertility outcomes in women over 35 years: a systematic review and individual participant data meta-analysis. *Hum Reprod* 2020;**35**:1808–1820.
- Collins SC, Xu X, Mak W. Cost-effectiveness of preimplantation genetic screening for women older than 37 undergoing in vitro fertilization. *J Assist Reprod Genet* 2017;**34**:1515–1522.
- Dang VQ, Vuong LN, Luu TM, Pham TD, Ho TM, Ha AN, Truong BT, Phan AK, Nguyen DP, Pham TN et al. Intracytoplasmic sperm injection versus conventional in-vitro fertilisation in couples with infertility in whom the male partner has normal total sperm count and motility: an open-label, randomised controlled trial. *Lancet* 2021;**397**:1554–1563.
- Datta J, Palmer MJ, Tanton C, Gibson LJ, Jones KG, Macdowall W, Glasier A, Sonnenberg P, Field N, Mercer CH et al. Prevalence of infertility and help seeking among 15 000 women and men. *Hum Reprod* 2016;**31**:2108–2118.
- de Geyter C, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, Scaravelli G, Smeenk J, Vidakovic S, Goossens V; European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). ART in Europe, 2014: results generated from European registries by ESHRE: the European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). *Hum Reprod* 2018;**33**:1586–1601.
- de Geyter C, Wyns C, Calhaz-Jorge C, de Mouzon J, Ferraretti AP, Kupka M, Nyboe Andersen A, Nygren KG, Goossens V. 20 years of the European IVF-monitoring Consortium registry: what have we learned? A comparison with registries from two other regions. *Hum Reprod* 2020;**35**:2832–2849.
- Domar A, Vassena R, Dixon M, Costa M, Vegni E, Collura B, Markert M, Samuelsen C, Guiglotto J, Roitmann E et al. Barriers and factors associated with significant delays to initial consultation and treatment for infertile patients and partners of infertile patients. *Reprod Biomed Online* 2021;**43**:1126–1136.
- Duffy JMN, Adamson GD, Benson E, Bhattacharya S, Bhattacharya S, Boffill M, Brian K, Collura B, Curtis C, Evers JLH et al.; Priority Setting Partnership for Infertility. Top 10 priorities for future infertility research: an international consensus development study. *Fertil Steril* 2021a;**115**:180–190.
- Duffy JMN, AlAhwany H, Bhattacharya S, Collura B, Curtis C, Evers JLH, Farquharson RG, Franik S, Giudice LC, Khalaf Y et al.; Core Outcome Measure for Infertility Trials (COMMIT) Initiative. Developing a core outcome set for future infertility research: an international consensus development study. *Fertil Steril* 2021b;**115**:191–200.
- Dyer SJ, Vinoos L, Ataguba JE. Poor recovery of households from out-of-pocket payment for assisted reproductive technology. *Hum Reprod* 2017;**32**:2431–2436.
- Eijkemans MJC, Habbema JDF, Kersten FAM, Lintsen AME, Braat DDM, Hunault CC, Bouwmans CAM, Roijen LH-V. Cost-effectiveness of 'immediate IVF' versus 'delayed IVF': a prospective study. *Hum Reprod* 2017;**32**:999–1008.
- ESHRE Capri Workshop Group. Economic aspects of infertility care: a challenge for researchers and clinicians. *Hum Reprod* 2015;**30**:2243–2248.
- European IVF Monitoring Consortium for the European Society of Human Reproduction, Embryology; Wyns C, de Geyter C, Calhaz-Jorge C, Kupka MS, Motrenko T, Smeenk J, Bergh C, Tandler-Schneider A et al. ART in Europe 2017: results generated from European registries by ESHRE. *Hum Reprod Open* 2021;**2021**:hoab026.
- Evans MB, Parikh T, DeCherney AH, Csokmay JM, Healy MW, Hill MJ. Evaluation of the cost-effectiveness of ovulation suppression with progestins compared with GnRH analogs in assisted reproduction cycles. *Reprod Biomed Online* 2019;**38**:691–698.
- Fragoulakis V, Kourlaba G, Tarlatzis B, Mastrominas M, Maniatakis N. Economic evaluation of alternative assisted reproduction techniques in management of infertility in Greece. *Clinicoecon Outcomes Res* 2012;**4**:185–192.
- Fragoulakis V, Pescott CP, Smeenk MJ, van Santbrink EJP, Oosterhuis GJE, Broekmans FJM, Maniatakis N. Economic evaluation of three frequently used gonadotrophins in assisted reproduction techniques in the management of infertility in the Netherlands. *Appl Health Econ Health Policy* 2016;**14**:719–727.
- Garceau L, Henderson J, Davis LJ, Petrou S, Henderson LR, McVeigh E, Barlow DH, Davidson LL. Economic implications of assisted reproductive techniques: a systematic review. *Hum Reprod* 2002;**17**:3090–3109.
- Gizzo S, Ferrando M, Lispi M, Alvarez Nieto J, Bilger W, Ripellino C, Cataldo N, Eheberg D, Buhler K. A cost-effectiveness modeling evaluation comparing originator follitropin alfa to a biosimilar for the treatment of infertility in Germany, Italy and Spain. *Value Health* 2017;**20**:A519–A520.
- Gizzo S, Ferrando M, Lispi M, Ripellino C, Cataldo N, Buhler K. A cost-effectiveness modeling evaluation comparing a biosimilar follitropin alfa preparation with its reference product for live birth outcome in Germany, Italy and Spain. *J Med Econ* 2018;**21**:1096–1101.
- Gleicher N, Kushnir VA, Barad DH. Worldwide decline of IVF birth rates and its probable causes. *Hum Reprod Open* 2019;**2019**:hoz017.
- Gleicher N, Patrizio P, Brivanlou A. Preimplantation genetic testing for aneuploidy—a castle built on sand. *Trends Mol Med* 2021;**27**:731–742.
- Groen H, Tonch N, Simons AHM, van der Veen F, Hoek A, Land JA. Modified natural cycle versus controlled ovarian

- hyperstimulation IVF: a cost-effectiveness evaluation of three simulated treatment scenarios. *Hum Reprod* 2013;**28**:3236–3246.
- Grynberg M, Murphy C, Doré C, Fresneau L, Paillet S, Petrica N, Frédérique M, Ravonimbola H. A cost-effectiveness analysis comparing the originator follitropin alfa to its biosimilars in patients undergoing a medically assisted reproduction program from a French perspective. *J Med Econ* 2018;**22**:1–15.
- Harris K, Fitzgerald O, Paul R, Macalodow A, Lee E, Chambers G. *Assisted Reproductive Technology in Australia and New Zealand* 2014. Sydney: National Perinatal Epidemiology and Statistics Unit, The University of New South Wales; 2016.
- Hernandez Torres E, Navarro-Espigares JL, Clavero A, López-Regalado M, Camacho-Ballesta JA, Onieva-García M, Martínez L, Castilla JA, Onieva-García M. Economic evaluation of elective single-embryo transfer with subsequent single frozen embryo transfer in an in vitro fertilization/intracytoplasmic sperm injection program. *Fertil Steril* 2015;**103**:699–706.
- Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, Caulley L, Chaiyakunapruk N, Greenberg D, Loder E et al.; CHEERS 2022 ISPOR Good Research Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *Value Health* 2022;**25**:3–9.
- International Federation of Fertility Societies' Surveillance (IFFS). International Federation of Fertility Societies' Surveillance (IFFS) 2019: global trends in reproductive policy and practice, 8th edition. *Glob Reprod Health* 2019;**4**:e29.
- International Federation of Fertility Societies' Surveillance (IFFS). International Federation of Fertility Societies' Surveillance (IFFS) 2022: global trends in reproductive policy and practice, 9th edition. *Glob Reprod Health* 2022;**7**:e58.
- Kawwass JF, Penzias AS, Adashi EY. Fertility-a human right worthy of mandated insurance coverage: the evolution, limitations, and future of access to care. *Fertil Steril* 2021;**115**:29–42.
- Kiani Z, Simbar M, Hajian S, Zayeri F, Shahidi M, Saei Ghare Naz M, Ghasemi V. The prevalence of anxiety symptoms in infertile women: a systematic review and meta-analysis. *Fertil Res Pract* 2020;**6**:7.
- Kushnir VA, Barad DH, Albertini DF, Darmon SK, Gleicher N. Systematic review of worldwide trends in assisted reproductive technology 2004–2013. *Reprod Biol Endocrinol* 2017;**15**:6.
- Lee E, Costello MF, Botha WC, Chambers GM, Illingworth P. A cost-effectiveness analysis of preimplantation genetic testing for aneuploidy (PGT-A) for up to three complete assisted reproductive technology cycles in women of advanced maternal age. *Aust N Z J Obstet Gynaecol* 2019;**59**:573–579.
- Lee M, Lanes A, Ginsburg ES, Hornstein MD, Lofgren KT, Thomas A, Goldman R. The cost-effectiveness of preimplantation genetic testing for aneuploidy in the United States: an analysis of cost and birth outcomes from 158,665 in vitro fertilization cycles. *Am J Obstet Gynecol* 2021;**225**:55.e1.
- Lensen S, Hammarberg K, Polyakov A, Wilkinson J, Whyte S, Peate M, Hickey M. How common is add-on use and how do patients decide whether to use them? A national survey of IVF patients. *Hum Reprod* 2021;**36**:1854–1861.
- Mennini FS, Marcellusi A, Viti R, Bini C, Carosso A, Revelli A, Benedetto C. Probabilistic cost-effectiveness analysis of controlled ovarian stimulation with recombinant FSH plus recombinant LH vs. human menopausal gonadotropin for women undergoing IVF. *Reprod Biol Endocrinol* 2018;**16**:68.
- Messerlian C, MacLagan L, Basso O. Infertility and the risk of adverse pregnancy outcomes: a systematic review and meta-analysis. *Hum Reprod* 2012;**28**:125–137.
- Moolenaar LM, Broekmans FJM, van Disseldorp J, Fauser BCJM, Eijkemans MJC, Hompes PGA, van der Veen F, Mol BWJ. Cost effectiveness of ovarian reserve testing in in vitro fertilization: a Markov decision-analytic model. *Fertil Steril* 2011;**96**:889–894.e2.
- Moolenaar LM, Cissen M, de Bruin JP, Hompes PGA, Repping S, van der Veen F, Mol BWJ. Cost-effectiveness of assisted conception for male subfertility. *Reprod Biomed Online* 2015;**30**:659–666.
- Neal SA, Morin SJ, Franasiak JM, Goodman LR, Juneau CR, Forman EJ, Werner MD, Scott RT Jr. Preimplantation genetic testing for aneuploidy is cost-effective, shortens treatment time, and reduces the risk of failed embryo transfer and clinical miscarriage. *Fertil Steril* 2018;**110**:896–904.
- Neumann K, Griesinger G. An economic analysis of aneuploidy screening of oocytes in assisted reproduction in Germany. *Geburtshilfe Frauenheilkd* 2020;**80**:172–178.
- Neumann K, Griesinger G, Sermon K, Bossuyt P, Goossens V, Geraedts J, Traeger-Synodinos J, Parriego M, Schmutzler A, van der Ven K et al. An economic analysis of preimplantation genetic testing for aneuploidy by polar body biopsy in advanced maternal age. *BJOG* 2020;**127**:710–718.
- Newman JEPR, Chambers GM. *Assisted Reproductive Technology in Australia and New Zealand* 2019. Sydney: National Perinatal Epidemiology and Statistics Unit, The University of New South Wales; 2021.
- Ozgun K, Berkkanoglu M, Bulut H, Yoruk GDA, Canduramaz NN, Coetsee K. Single best euploid versus single best unknown-ploidy blastocyst frozen embryo transfers: a randomized controlled trial. *J Assist Reprod Genet* 2019;**36**:629–636.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 2021;**88**:105906.
- Papaleo E, Pagliardini L, Vanni VS, Delprato D, Rubino P, Candiani M, Viganò P. A direct healthcare cost analysis of the cryopreserved versus fresh transfer policy at the blastocyst stage. *Reprod Biomed Online* 2017;**34**:19–26.
- Peipert BJ, Spinosa D, Howell EP, Weber JM, Truong T, Harris BS. Innovations in infertility: a comprehensive analysis of the ClinicalTrials.gov database. *Fertil Steril* 2021;**116**:1381–1390.
- Pham CT, Karnon JD, Norman RJ, Mol BW. Cost-effectiveness modelling of IVF in couples with unexplained infertility. *Reprod Biomed Online* 2018;**37**:555–563.
- Robins JC, Khair AF, Widra EA, Alper MM, Nelson WW, Foster ED, Sinha A, Ando M, Heiser PW, Daftary GS. Economic evaluation of highly purified human menotropin or recombinant follicle-stimulating hormone for controlled ovarian stimulation in high-responder patients: analysis of the Menopur in Gonadotropin-releasing Hormone Antagonist Single Embryo Transfer-High Responder (MEGASET-HR) trial. *FS Rep* 2020;**1**:257–263.
- Scotland GS, McLernon D, Kurinczuk JJ, McNamee P, Harrild K, Lyall H, Rajkhowa M, Hamilton M, Bhattacharya S. Minimising twins in in vitro fertilisation: a modelling study assessing the costs, consequences and cost-utility of elective single versus double embryo transfer over a 20-year time horizon. *BJOG* 2011;**118**:1073–1083.
- Shalom-Paz E, Alshalati J, Shehata F, Jimenez L, Son WY, Holzer H, Tan SL, Almog B. Clinical and economic analysis of rescue intracytoplasmic sperm injection cycles. *Gynecol Endocrinol* 2011;**27**:993–996.
- Shields GE, Elvidge J. Challenges in synthesising cost-effectiveness estimates. *Syst Rev* 2020;**9**:289.
- Sitler C, Levy G, Pier B, Lustik M. Single embryo transfer versus double embryo transfer: a cost-effectiveness analysis in a non-IVF insurance mandated system. *Mil Med* 2020;**185**:e1700–e1705.

- Somigliana E, Busnelli A, Paffoni A, Vigano P, Riccaboni A, Rubio C, Capalbo A. Cost-effectiveness of preimplantation genetic testing for aneuploidies. *Fertil Steril* 2019;**111**:1169–1176.
- Somigliana E, Paffoni A, Busnelli A, Filippi F, Pagliardini L, Vigano P, Vercellini P. Age-related infertility and unexplained infertility: an intricate clinical dilemma. *Hum Reprod* 2016;**31**:1390–1396.
- Stormlund S, Schmidt L, Bogstad J, Løssl K, Prætorius L, Zedeler A, Pinborg A. Patients' attitudes and preferences towards a freeze-all strategy in ART treatment. *Hum Reprod* 2019;**34**:679–688.
- The Commonwealth of Australia. 2021 *Intergenerational Report*. Canberra: The Commonwealth of Australia; 2021.
- Theobald R, SenGupta S, Harper J. The status of preimplantation genetic testing in the UK and USA. *Hum Reprod* 2020;**35**:986–998.
- Tierney K, Cai Y. Assisted reproductive technology use in the United States: a population assessment. *Fertil Steril* 2019;**112**:1136–1143.e4.
- Tjon-Kon-Fat RI, Bensdorp AJ, Bossuyt PM, Koks C, Oosterhuis GJ, Hoek A, Hompes P, Broekmans FJ, Verhoeve HR, de Bruin JP et al. Is IVF-served two different ways-more cost-effective than IUI with controlled ovarian hyperstimulation? *Hum Reprod* 2015;**30**:2331–2339.
- Toner JP, Coddington CC, Doody K, van Voorhis B, Seifer DB, Ball GD, Luke B, Wantman E. Society for assisted reproductive technology and assisted reproductive technology in the United States: a 2016 update. *Fertil Steril* 2016;**106**:541–546.
- Vaidya A, Stafinski T, Nardelli A, Motan T, Menon D. Assisted reproductive technologies in Alberta: an economic analysis to inform policy decision-making. *J Obstet Gynaecol Can* 2015;**37**:1122–1130.
- van Eekelen R, Eijkemans MJ, Mochtar M, Mol F, Mol BW, Groen H, Wely M. Cost-effectiveness of medically assisted reproduction or expectant management for unexplained subfertility: when to start treatment? *Hum Reprod* 2020;**35**:2037–2046.
- van Heesch MM, van Asselt AD, Evers JL, van der Hoeven MA, Dumoulin JC, van Beijsterveldt CE, Bonsel GJ, Dykgraaf RH, van Goudoever JB, Koopman-Esseboom C et al. Cost-effectiveness of embryo transfer strategies: a decision analytic model using long-term costs and consequences of singletons and multiples born as a consequence of IVF. *Hum Reprod* 2016;**31**:2527–2540.
- van Loendersloot LL, Moolenaar LM, van Wely M, Repping S, Bossuyt PM, Hompes PGA, van der Veen F, Mol BWJ. Cost-effectiveness of single versus double embryo transfer in IVF in relation to female age. *Eur J Obstet Gynecol Reprod Biol* 2017;**214**:25–30.
- van Rumste MME, Custers IM, van Wely M, Koks CA, van Weering HGI, Beckers NGM, Scheffer GJ, Broekmans FJM, Hompes PGA, Mochtar MH et al. IVF with planned single-embryo transfer versus IUI with ovarian stimulation in couples with unexplained subfertility: an economic analysis. *Reprod Biomed Online* 2014;**28**:336–342.
- van Tilborg TC, Oudshoorn SC, Eijkemans MJC, Mochtar MH, van Golde RJT, Hoek A, Kuchenbecker WKH, Fleischer K, de Bruin JP, Groen H et al.; OPTIMIST Study Group. Individualized FSH dosing based on ovarian reserve testing in women starting IVF/ICSI: a multicentre trial and cost-effectiveness analysis. *Hum Reprod* 2017;**32**:2485–2495.
- Veritas Health Innovation. *Covidence Systematic Review Software*. Melbourne, Australia: Veritas Health Innovation, 2021. www.covidence.org
- Vitek WS, Galárraga O, Klatsky PC, Robins JC, Carson SA, Blazar AS. Management of the first in vitro fertilization cycle for unexplained infertility: a cost-effectiveness analysis of split in vitro fertilization-intracytoplasmic sperm injection. *Fertil Steril* 2013;**100**:1381–1388.e1.
- World Bank. *World Bank Country and Lending Groups*. Washington, DC: World Bank Group, 2021.
- Xue W, Lloyd A, Falla E, Roeder C, Papsch R, Bühler K. A cost-effectiveness evaluation of the originator follitropin alpha compared to the biosimilars for assisted reproduction in Germany. *Int J Womens Health* 2019;**11**:319–331.
- Yan J, Qin Y, Zhao H, Sun Y, Gong F, Li R, Sun X, Ling X, Li H, Hao C et al. Live birth with or without preimplantation genetic testing for aneuploidy. *N Engl J Med* 2021;**385**:2047–2058.
- Yates AP, Rustamov O, Roberts SA, Lim HYN, Pemberton PW, Smith A, Nardo LG. Anti-Müllerian hormone-tailored stimulation protocols improve outcomes whilst reducing adverse effects and costs of IVF. *Hum Reprod* 2011;**26**:2353–2362.
- Yatsenko SA, Rajkovic A. Genetics of human female infertility. *Biol Reprod* 2019;**101**:549–566.
- Yu B, Mumford S, Royster GD, Segars J, Armstrong AY. Cost-effectiveness analysis comparing continuation of assisted reproductive technology with conversion to intrauterine insemination in patients with low follicle numbers. *Fertil Steril* 2014;**102**:435–439.
- Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med* 2020;**382**:1244–1256.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Human Reproduction, 2024, 39, 981–991
<https://doi.org/10.1093/humrep/deae039>
Review

Psychology and counselling

An evaluation of sexual function and health-related quality of life following laparoscopic surgery in individuals living with endometriosis

Leesa Van Niekerk ^{1,2,*}, Shaunagh Pugh¹, Antonina Mikocka-Walus ^{2,3}, Cecilia Ng^{2,4}, Rebecca O'Hara⁵, Mike Armour ^{2,6}, Mathew Leonardi^{5,7}, and Subhadra Evans³

¹School of Psychological Sciences, College of Health & Medicine, University of Tasmania, Hobart, TAS, Australia

²National Endometriosis Clinical and Scientific Trials (NECST) Network, UNSW, Sydney, Australia

³School of Psychology, Faculty of Health, Deakin University, Geelong, VIC, Australia

⁴Division of Obstetrics and Gynaecology, School of Clinical Medicine, Medicine and Health, University of New South Wales, Sydney, NSW, Australia

⁵Robinson Research Institute, Adelaide Medical School, University of Adelaide, North Adelaide, SA, Australia

⁶NICM Health Research Institute, Western Sydney University, Penrith, NSW, Australia

⁷Department of Obstetrics and Gynecology, McMaster University, Hamilton, ON, Canada

*Correspondence address. School of Psychological Science, College of Health & Medicine, University of Tasmania, Private Bag 30, Hobart, Tasmania 7001, Australia. Tel: +61-3-6126 6645; E-mail: leesa.vanniekerk@utas.edu.au  <https://orcid.org/0000-0002-6641-5888>

ABSTRACT

STUDY QUESTION: What is the relationship between sexual function, health-related quality of life (HRQoL), and laparoscopic surgery in individuals living with endometriosis?

SUMMARY ANSWER: A higher number of laparoscopic surgeries is significantly associated with poorer HRQoL and greater levels of sexual dysfunction in individuals with endometriosis.

WHAT IS KNOWN ALREADY: Prior research indicates that endometriosis is associated with lowered HRQoL and sexual function and that these outcomes are influenced by endometriosis-related symptom profiles, medical, and surgical management. A limited number of studies have examined changes in sexual function in individuals with endometriosis following laparoscopic surgery or following repeated surgeries.

STUDY DESIGN, SIZE, DURATION: A cross-sectional community-based online survey was used to examine the relationships between sexual function, HRQoL, and laparoscopic surgery (n = 210).

PARTICIPANTS/MATERIALS, SETTING, METHODS: Individuals with a self-reported diagnosis of endometriosis were recruited via online advertising through social media and gynaecology clinics. Endometriosis-specific data (e.g. diagnostic delay, symptom experience) was collected in addition to engagement with laparoscopic surgery, level of HRQoL (EuroQol-5 Dimension: EQ-5D-5L), and sexual function (Female Sexual Function Index: FSFI). Bivariate correlational analyses and hierarchical multiple regression were used to determine the associations between the variables of interest.

MAIN RESULTS AND THE ROLE OF CHANCE: Individuals with endometriosis have substantially poorer HRQoL in comparison to Australian normative samples, with greater levels of endometriosis-related symptom burden, distress, and pain significantly associated with lower levels of HRQoL. The mean FSFI score was suggestive of clinically significant female sexual dysfunction, with the lowest level of function noted in the domain of sexual pain and the highest level of function noted in the sexual satisfaction domain. A greater number of laparoscopic surgeries was significantly associated with poorer overall HRQoL and greater levels of sexual dysfunction.

LIMITATIONS, REASONS FOR CAUTION: The cross-sectional nature of the data precludes direct findings of causality and further longitudinal research is recommended. The information pertaining to engagement in laparoscopic surgery was self-report in nature and was not medically verified.

WIDER IMPLICATIONS OF THE FINDINGS: The study's findings highlight the pervasive impact of endometriosis on all domains of living, emphasizing the need to extend treatment planning beyond that of physical pain management alone. Early referral for assessment and management of sexual wellbeing is recommended prior to, and post-surgical intervention, with a focus on maintaining post-surgical changes, potentially reducing the need for multiple surgeries.

STUDY FUNDING, COMPETING INTEREST(S): The study was not associated with research funding. Author CN reports grant funding from the Australian Government and Medical Research Future Fund (MRFF) and was a previous employee of CSL Vifor (formerly Vifor Pharma Pty Ltd).

TRIAL REGISTRATION NUMBER: N/A.

Keywords: endometriosis / health-related quality of life / sexual function / symptom burden / pain / dyspareunia

Received: October 16, 2023. Revised: February 27, 2024. Editorial decision: March 10, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Endometriosis is a persistent systemic inflammatory condition that impacts approximately one in nine people who were presumed female at birth (Rowlands *et al.*, 2021). Treatment is generally aimed at preventing disease progression and symptom management, with multimodal approaches considered the most effective, which may include hormonal and laparoscopic surgical management (Abbott *et al.*, 2003; Della Corte *et al.*, 2020; Laguerre *et al.*, 2022). Laparoscopic surgery is a minimally invasive surgical technique (ablation or excision) used to remove endometriotic tissue from the internal organs and aims to decrease symptom severity in individuals with endometriosis (Mehedintu *et al.*, 2014). While the difference in outcomes between ablation and excision laparoscopic surgery are still unclear in the literature (Bafort *et al.*, 2020), the goal of surgery is to completely remove endometriotic lesions, for which excision is generally the preferred technique, with some research indicating increased efficacy in managing symptoms such as dyspareunia (Healey *et al.*, 2015; Riley *et al.*, 2019).

The ability to engage with and maintain fulfilling interpersonal relationships is strongly related to psychological wellbeing and is a key aspect of health-related quality of life (HRQoL). Research suggests that sexual satisfaction, orgasmic function, and arousal are all symptoms of endometriosis that are commonly overlooked within clinical settings, especially when considering the heterogeneity of endometriosis symptoms (Fairbanks *et al.*, 2017; Vannuccini *et al.*, 2023). However, the link between poorer sexual function and HRQoL suggests that accounting for sexual satisfaction and its contributors should be a prominent treatment consideration in addressing lowered HRQoL (Dogan *et al.*, 2013; Franck *et al.*, 2018). Overall, the impact of lowered sexual function is strongly associated with multiple domains of HRQoL (e.g. physical, psychological, social, relational), suggesting that sexual function and HRQoL are intrinsically linked for most individuals with endometriosis (van Poll *et al.*, 2020).

Few studies have investigated the improvement in sexual function for individuals with endometriosis following laparoscopic surgery, particularly across time. Similarly, limited research is available regarding the efficacy of repeated surgeries for improving sexual function and HRQoL impacts (D'Alterio *et al.*, 2021). A study by Barbara *et al.* (2017) found that there was an improvement in sexual function following laparoscopic surgery, but that this improvement was likely to be more pronounced if a holistic approach to treatment was adopted, given that physical pain alone is not the sole contributor to decreased sexual satisfaction and function. This is supported by a systematic review conducted by Franck *et al.* (2018), which indicated that sexual function is improved following laparoscopic surgery. However, it was noted within the review that there was substantial heterogeneity within the research in terms of the questionnaires used, the stage of endometriosis, and follow-up duration (Franck *et al.*, 2018). A recent prospective study ($n = 193$) conducted by Martínez-Zamora *et al.* (2021) found that there was a significant improvement in sexual function 6-month post-laparoscopic procedure, and that this improvement was still present but less significant at 36 months (Martínez-Zamora *et al.*, 2021).

Although it is expected that laparoscopic surgery will lead to an initial reduction in endometriosis-related pain and improved HRQoL and sexual functioning in individuals diagnosed with endometriosis, research investigating surgical outcomes has

predominantly focused on changes in sexual pain with limited focus on other aspects of sexual function such as desire and enjoyment (Barbara *et al.*, 2017). As alterations to sexual function are treatment specific, understanding the role that laparoscopic surgery plays in addressing the individual components of sexuality (e.g. desire, arousal, satisfaction) may assist medical practitioners in treatment and referral planning (e.g. engagement in sexual therapy assessment and management pre and/or post-surgery). Furthermore, understanding whether certain endometriosis-related symptom profiles are associated with greater sexual function impairments and lowered HRQoL may also assist in effective treatment planning and management where surgical management may not be readily accessible or is contraindicated.

Therefore, the current study aims to investigate sexual function and HRQoL in individuals with symptomatic endometriosis and the influence of laparoscopic surgery on these outcomes. Based on the existing literature investigating the effects of laparoscopic surgery and symptom improvement (Laguerre *et al.*, 2022), it is hypothesized that sexual function and HRQoL will be significantly negatively associated with the length of time following laparoscopic surgery. To extend the understanding of the relationship between repeated laparoscopic surgeries beyond that of fertility management (Rizk *et al.*, 2015), the current study will also examine the relationship between number of surgeries, HRQoL, and sexual function. Based on the evidence that poorer fertility outcomes may be associated with repeated surgical procedures (Rizk *et al.*, 2015), it is hypothesized that a higher number of laparoscopic procedures will be significantly negatively associated with HRQoL and sexual function. The study will also undertake exploratory analyses to explore whether specific endometriosis-related symptoms, laparoscopic surgery, and HRQoL are significant correlates of sexual function in individuals with endometriosis.

Materials and methods

Study population

The study population consisted of 210 individuals, presumed female at birth, with symptomatic endometriosis and an average age of 31.7 years ($SD = 7.87$, range = 18–50 years). Most of the sample identified as female ($n = 203$, 96.7%), with seven participants (3.3%) self-identifying as non-binary/gender fluid. The study inclusion criteria were: (i) a self-reported diagnosis of endometriosis, (ii) being aged 18 years or older, and (iii) experiencing at least one endometriosis-related symptom in the previous 4-week period. Participants needed to satisfy all three inclusion criteria to be eligible to participate. Participants were recruited via advertisements at gynaecological health services, targeted social media (e.g. Endometriosis Australia, EndoActive, EndoZone), and snowball sampling. Survey data was collected via the RedCap survey platform, hosted by the University of Tasmania (Harris *et al.*, 2019). Participants were directed to a Participant Information Sheet, with submission of the completed survey constituting informed consent.

Measures

Demographic and endometriosis-related information

Demographic information (e.g. country of residence, education, employment) and endometriosis-specific information (e.g. symptom duration, diagnostic delay, disease stage/grading) were collected at the start of the survey. Participants also indicated

the presence/absence of 17 endometriosis-related symptoms in the four weeks prior to survey completion (e.g. dysmenorrhea, dyspareunia, voiding pain, fertility concerns, nausea). The 17 endometriosis-related symptoms were based on a review of the literature of frequently reported/measured symptoms and have been used previously with endometriosis samples (Van Niekerk et al., 2022a,b). Information regarding the number of laparoscopic procedures, date of most recent procedure, and hormonal therapy was also gathered.

Subjective endometriosis pain rating

The Visual Analogue Scale (VAS) pain rating was used as a measure of the subjective experience of pain. The VAS consists of a straight horizontal line, where the endpoints define extreme limits of pain, with 0 = No Pain and 10 = Worst Pain. Participants are asked to use a sliding scale to indicate the severity of endometriosis-related pain experienced: (i) over the past 7 days, and (ii) at the time of survey completion. The VAS is a reliable and valid pain measure in endometriosis populations (Bourdel et al., 2019).

Health-related quality of life

The EuroQol-5 Dimension (EQ-5D-5L) was used to measure mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (Herdman et al., 2011). Each dimension has five levels: no problems (Level 1), slight problems, moderate problems, severe problems, and extreme problems (Level 5). This creates a possible range of 3125 health states when a rating from each dimension is combined, with 11 111 reflecting optimal health and 55 555 reflecting worst health. The EQ-5D-5L health state can be converted to an index 'utility' score based on population-based scoring algorithms. As most of the sample resided in Australia, an Australian scoring algorithm and value set were utilized to provide population-based scores (Norman et al., 2023). Based on this algorithm, values below 0 are considered to represent health states worse than death (Devlin et al., 2018). The EQ-5D-5L also includes a visual rating scale, where the respondent is asked to rate how good or bad their health is at that moment in time, with 0 = Worst Health Imaginable and 100 = Best Health Imaginable. A recently conducted systematic review found the EQ-5D-5L to be a valid and reliable measure of HRQoL across a range of populations, conditions, and settings (Feng et al., 2021).

Sexual function

The Female Sexual Function Index (FSFI) is a 19-item self-report measure that provides an assessment of six sexual function domains: desire, arousal, lubrication, orgasm, global satisfaction, and pain (Rosen et al., 2000). Respondents are asked to rate their responses on a scale from 0 (No Sexual Activity) to 5 (Almost Always/Very High), excluding questions 15 and 16 which are rated on a scale from 1 (Very Dissatisfied) to 5 (Very Satisfied), with higher scores reflecting greater levels of sexual functioning in the individual domain. Participants are given a total FSFI score, ranging from 2 (dissatisfactory sexual function) to 36 (satisfactory sexual function). The FSFI was found to be a reliable and valid measure of sexual function across the lifespan and has been utilized in research relating to medical and psychological disorders and their impacts on sexual function (Bartula & Sherman, 2015).

Design and data analysis

A cross-sectional design was employed to explore the associations between endometriosis-related factors (e.g. symptom duration, average endometriosis-related pain level), sexual function, and HRQoL. Only fully completed and submitted survey

responses were included in the data analyses as per the restrictions of the study's research ethics approval (283 started surveys, 240 completed surveys: 84.81% completion rate). Preliminary data review revealed some statistically significant differences between respondents who had undergone at least one laparoscopic surgery ($n=210$) compared to those without a surgical history ($n=30$) for the sexual health outcomes. The 30 respondents without a surgical history were removed from the analyses to avoid potential confounding influence. GPower priori power analysis advised that 208 responses are required for linear regression models to achieve a power of 0.95, $P \leq 0.05$ (Faul et al., 2007; Mayr et al., 2007), indicating that the remaining sample of 210 responses provides sufficient power. Spearman's Rho bivariate correlations were conducted to assess the associations between the outcome measures and endometriosis-specific variables. Hierarchical multiple regression analysis (MRA) was used to examine the unique contribution of endometriosis-related symptoms (step one), treatment (step two), and HRQoL (step three) have in influencing sexual function in individuals living with endometriosis. Individual regression models were conducted to determine the relative contribution of the examined factors in influencing the different components of sexual function and to extend the findings past that of sexual pain alone. The conceptual model (order of entry) was based on a review of the literature which indicates that endometriosis-related symptoms (block one: e.g. endometriosis-related pain, dysmenorrhea, dyspareunia) and treatment (block two: e.g. number of surgeries, hormonal treatment) have been found to influence HRQoL (block three). Hormonal therapy was included as a treatment of interest as preliminary analyses noted significant differences between participants receiving and not receiving hormonal treatment for level of endometriosis-related pain and HRQoL. HRQoL was entered in the final step (block three) to determine whether HRQoL independently influences sexual function in individuals with endometriosis. Multicollinearity analyses indicated that both the variance inflation factor (VIF) and tolerance scores (TS) were within acceptable ranges: $VIF \leq 3$ and $TS \geq 1$. Statistical significance was assumed at $P \leq 0.05$.

Ethics approval

Ethics approval was provided by the University of Tasmania Human Research Ethics Committee (reference: H0026906) for national and international respondents. Participants were provided with full study details prior to starting the survey and informed consent was indicated by submission of completed surveys.

Results

The average endometriosis-related symptom duration reported by the sample was 14.8 years ($SD=7.79$), with 49.5% of the sample indicating a disease stage of moderate (Stage III) to severe (Stage IV). The average number of laparoscopic procedures was 2 ($SD=1.48$), ranging from one to nine surgeries. The average duration following the most recent laparoscopic surgery was 24.7 months prior to survey completion ($SD=32.94$), ranging from 1 month to 200 months. Fifty-five percent of the respondents ($n=132$) self-reported using hormonal therapy as a treatment for endometriosis (e.g. progestogens, GnRH analogues, combined contraceptives). Participants experienced an average of 8.9 symptoms ($SD=3.10$), ranging from one to 16 symptoms. The average level of endometriosis-related pain was 5.41 ($SD=2.51$), ranging from 1 to 10 where 0 equals no pain and 10 equals worst pain imaginable. Endometriosis-related pain was statistically significantly positively correlated with the length of

time following the most recent laparoscopic surgery, with small effect size, indicating that higher pain levels are linked to longer gaps between surgeries. Level of endometriosis-related pain did not differ significantly for participants receiving hormonal therapy (Mean = 5.41, SD = 2.17) and those not receiving hormonal therapy (Mean = 5.42, SD = 2.83), $t(208)=0.04$, $P=0.970$. Further participant characteristics are outlined in Table 1.

HRQoL outcomes

In the current endometriosis sample, six participants (2.9%) reported a full optimal health state of 11 111 and four participants reported an extremely poor health state of 44 444 (1.9%). As can be seen in Table 2, most participants experienced slight to moderate levels of daily pain and discomfort (72.4%), anxiety and/or depression (62.4%), difficulties engaging in their usual activities (64.8%), but no difficulties completing tasks of self-care such as washing and dressing themselves (69.0%). In terms of overall HRQoL, the EQ-5D-5L VAS score was used to represent the participant's perception of their health at the time of survey completion where 0 equals worst health and 100 equals best health. In general, participants reported an average level of overall HRQoL (Mean = 54.1, SD = 23.27, range = 0.0–89.0). To allow a more meaningful comparison of the HRQoL reported by the

current endometriosis sample compared to the general population, the health state scores were converted to population-based utility scores. The mean utility score was 0.76 (SD = 0.22, range = 0.03–1.0) which is lower than the mean utility score reported in an Australian general population sample (Mean = 0.9, SD = 0.14; Norman et al., 2023), reflecting poorer HRQoL in the current endometriosis sample than that reported in general population samples.

Bivariate correlations were used to determine the relationships between endometriosis-related pain, laparoscopic surgery, and HRQoL (see Table 3). Statistically significant negative correlations were identified between the average level of endometriosis-specific pain experienced in the previous 7 days and the individual health dimensions, Health VAS, and utility score, with small to large effect sizes. This suggests that higher levels of endometriosis-related pain are significantly associated with lower levels of HRQoL as measured by the EQ-5D-5L (see Table 3).

The number of laparoscopic surgeries was found to be statistically significantly negatively correlated with the mobility, usual activities, pain and discomfort, anxiety/depression domains and the utility score, with small effect sizes, with higher numbers of surgical procedures associated with poorer HRQoL in these areas. The length of time since the most recent laparoscopic surgery was found to be statistically significantly positively correlated with all aspects of the EQ-5D-5L, with small effect sizes, indicating that longer duration following surgery was associated with greater HRQoL in the current sample (see Table 3). Independent samples *t*-tests failed to indicate significant differences in HRQoL for those receiving hormonal therapy compared to those not undertaking hormonal therapy.

Sexual function outcomes

As seen in Table 4, the average FSFI overall sexual function score was 16.89 (SD = 9.65, range = 2.0–35.6), which falls within the clinical cut-off range of ≤ 26 and is suggestive of clinically significant female sexual dysfunction. The lowest level of function was noted for the pain subscale (Mean = 2.24, SD = 1.86, range = 0.0–6.0) and the highest level of function for the satisfaction subscale (Mean = 3.25, SD = 1.58, range = 0.80–6.0).

Bivariate correlations were used to determine the relationships between endometriosis-related pain, laparoscopic surgery, and sexual function (see Table 3). The average level of endometriosis-related pain was statistically significantly negatively correlated with orgasmic function, sexual satisfaction, sexual pain, and overall level of sexual function, with small effect sizes. This suggests that higher levels of endometriosis-specific pain are linked to poorer orgasmic function, sexual satisfaction, overall sexual function, and greater sexual pain in the current sample (see Table 3). The number of laparoscopic surgeries and duration following the most recent laparoscopic surgery were not found to be statistically significantly correlated with sexual function in the current endometriosis sample (see Table 3).

Independent *t*-tests revealed several significant differences between participants receiving or not receiving hormonal therapy and all sexual function domains, with moderate effect sizes (excluding orgasmic function). Individuals undertaking hormonal therapy reported significantly poorer overall sexual function and poorer function in the domains of desire, arousal, lubrication, and sexual satisfaction compared to individuals not receiving hormonal therapy.

Table 1. Participants demographic and endometriosis-related characteristics.

	Frequency	Percent (%)
Country of residence		
Australia	184	87.6
Oceania (e.g. New Zealand, Papua New Guinea)	5	2.4
UK	9	4.3
North America	12	5.7
Highest level of academic attainment		
High school or below	30	14.3
Vocational certificate/qualification	50	23.8
Bachelor degree	78	37.1
Postgraduate degree	52	24.8
Employment status		
Student	35	16.7
Employed, part-time	36	17.6
Employed, full-time	111	52.9
Unable to work due to health	27	12.8
Relationship status		
Single	58	27.6
Committed	87	41.4
Married, living together	65	31.0
Sexuality		
Gay, lesbian, or same gender attracted	21	10.0
Straight/heterosexual	161	76.7
Bisexual	17	8.1
Pansexual	7	3.3
Asexual	4	1.9
Perimenopause/menopause status		
I do not experience symptoms of perimenopause	164	78.0
I experience symptoms of perimenopause	21	10.0
I experienced a medical or surgical menopause	25	12.0
Reproductive history		
I have never tried to conceive	121	57.6
I have tried to conceive but have never been pregnant	17	8.1
I am currently trying to conceive	15	7.1
I am currently pregnant	2	1.0
I have been pregnant/no miscarriages	20	9.5
I have been pregnant/1 or more births/1 or more miscarriages	21	10.0
I have been pregnant/Nil live births	14	6.7

n = 210.

Table 2. Dimensions of health-related quality of life (EQ-5D-5L) in an endometriosis sample.

	Mobility, n (%)	Self-care, n (%)	Usual activities, n (%)	Pain/discomfort, n (%)	Anxiety/depression, n (%)
No problems	96 (45.7)	145 (69.0)	47 (22.4)	24 (11.4)	41 (19.5)
Slight problems	66 (31.4)	43 (20.5)	89 (42.4)	79 (37.6)	76 (36.2)
Moderate problems	38 (18.1)	22 (10.5)	47 (22.4)	73 (34.8)	55 (26.2)
Severe problems	10 (4.8)	0 (0.0)	16 (7.6)	25 (11.9)	20 (9.5)
Extreme problems/unable to do	0 (0.0)	0 (0.0)	11 (5.2)	9 (4.3)	18 (8.6)

n = 210. EQ-5D-5L=EuroQol-5 Dimension.

Table 3. Standardized correlation matrix for associations between laparoscopic surgery, health-related quality of life (HRQoL), and sexual function.

Symptom	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1. PainVAS	–	0.059	0.044	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.830	0.203	0.132	0.012	0.025	0.001	0.017
2. SurgeryN	0.13	–	0.753	0.006	0.647	0.013	0.022	0.001	0.058	0.002	0.758	0.556	0.127	0.074	0.085	0.306	0.154
3. SurgeryD	0.14	0.02	–	0.046	0.001	0.001	0.016	0.036	0.013	0.001	0.416	0.823	0.954	0.419	0.937	0.499	0.947
4. Mobility	–0.44	–0.19	0.14	–	0.001	0.001	0.001	0.001	0.001	0.001	0.048	0.001	0.001	0.001	0.001	0.001	0.001
5. Self-Care	–0.37	–0.03	0.23	0.58	–	0.001	0.001	0.001	0.001	0.001	0.191	0.011	0.019	0.001	0.271	0.002	0.004
6. Activity	–0.47	–0.17	0.30	0.59	0.50	–	0.001	0.001	0.001	0.001	0.466	0.026	0.043	0.001	0.006	0.001	0.005
7. Pain/Disc	–0.60	–0.16	0.17	0.62	0.49	0.57	–	0.001	0.001	0.001	0.270	0.001	0.001	0.001	0.001	0.001	0.001
8. Anx/Dep	–0.30	–0.23	0.15	0.33	0.32	0.34	0.41	–	0.001	0.001	0.822	0.221	0.196	0.004	0.009	0.001	0.012
9. HealthVAS	–0.48	–0.13	0.18	0.44	0.34	0.55	0.63	0.51	–	0.001	0.024	0.001	0.001	0.001	0.001	0.001	0.001
10. Utility	–0.59	–0.21	0.25	0.72	0.61	0.73	0.83	0.67	0.69	–	0.155	0.001	0.001	0.001	0.001	0.001	0.001
11. Desire	–0.02	0.01	–0.06	0.14	0.09	0.05	0.08	0.02	0.16	0.10	–	0.001	0.001	0.001	0.001	0.001	0.001
12. Arousal	–0.09	–0.04	0.02	0.23	0.17	0.15	0.26	0.09	0.24	0.27	0.68	–	0.001	0.001	0.001	0.001	0.001
13. Lubrication	–0.10	–0.11	–0.01	0.23	0.16	0.14	0.26	0.09	0.24	0.26	0.54	0.88	–	0.001	0.001	0.001	0.001
14. Orgasm	–0.17	–0.12	0.06	0.33	0.25	0.23	0.35	0.20	0.31	0.39	0.48	0.84	0.77	–	0.001	0.001	0.001
15. Satisfact	–0.16	–0.12	0.01	0.25	0.08	0.19	0.29	0.18	0.36	0.30	0.51	0.72	0.68	0.62	–	0.001	0.001
16. SexPain	–0.24	–0.07	0.05	0.34	0.22	0.26	0.38	0.23	0.38	0.43	0.41	0.66	0.66	0.64	0.60	–	0.001
17. OverallSF	–0.16	–0.10	0.01	0.31	0.20	0.19	0.33	0.17	0.34	0.35	0.71	0.94	0.90	0.87	0.82	0.79	–

PainVAS = Average level of endometriosis-related pain experienced in the last 7 days; SurgeryN = Total number of laparoscopic surgeries; SurgeryD = Approximate length of time since most recent laparoscopic surgery (months); Mobility = EuroQol-5 Dimension Mobility domain; Self-care = EuroQol-5 Dimension Self-care domain; Activity = EuroQol-5 Dimension Usual Activities domain; Pain/Disc = EuroQol-5 Dimension Pain and Discomfort domain; Anx/Dep = EuroQol-5 Dimension Anxiety/Depression domain; HealthVAS = Self-rated HRQoL 0 is worst health imaginable and 100 is best health imaginable; Utility = EuroQol-5 Dimension Utility Norm score; Desire = Female Sexual Function Index desire subscale; Arousal = Female Sexual Function Index sexual arousal subscale; Lubrication = Female Sexual Function Index sexual lubrication subscale; Satisfaction = Female Sexual Function Index sexual satisfaction subscale; SexPain = Female Sexual Function Index sexual pain subscale; OverallSF = Female Sexual Function Index overall level of sexual function; Significance level (2-tailed) reported in upper half of matrix in italics; Spearman's Rho correlation presented in bottom half of matrix. Significant correlations in bold font ($P < 0.05$ to $P < 0.001$).

Table 4. Means and standard deviations for Female Sexual Function Index (FSFI) in women with endometriosis.

FSFI domain	M (SD)	Range
Desire	2.75 (1.35)	1.20–6
Arousal	2.78 (2.06)	0–6
Lubrication	3.18 (2.32)	0–6
Orgasm	2.69 (2.12)	0–6
Satisfaction	3.25 (1.58)	0.80–6
Pain	2.24 (1.86)	0–6
Overall sexual function	16.89 (9.65)	2–35.60

n = 210. M=Mean. Respondents are asked to rate their responses on a scale from 0 (No Sexual Activity) to 5 (Almost Always/Very High) or 1 (Very Dissatisfied) to 5 (Very Satisfied), with higher scores reflecting greater levels of sexual functioning in the individual domain. Overall FSFI score, ranging from 2 (dissatisfactory sexual function) to 36 (satisfactory sexual function).

Associations between HRQoL and sexual function outcomes

Bivariate correlations were used to determine the relationships between HRQoL (Health VAS, utility score) and sexual function (FSFI subscales, overall sexual function) (see Table 3). Statistically positive correlations were found between the Health VAS, utility score, individual FSFI subscales, and overall level of sexual function, with small to medium effect sizes, indicating that higher self-rated HRQoL and utility scores are associated with higher levels of sexual functioning.

Correlates of sexual function for individuals with endometriosis

Hierarchical MRA was used to examine the unique contribution of endometriosis-specific symptoms associated with sexual functioning in existing literature (i.e. endometriosis-related pain, dysmenorrhea, dyspareunia, pain after sexual intercourse, intermenstrual bleeding, fertility concerns, fatigue, bloating, nausea, vulva pain, clitoral pain), treatment factors (i.e. number of laparoscopic surgeries, duration since most recent surgery in months, hormonal therapy), and HRQoL (Health VAS, utility score) on the specific domains of sexual functioning (i.e. sexual desire, sexual arousal, sexual lubrication, orgasmic function, sexual satisfaction, sexual pain) (please see [Supplementary Data File S1](#) for hierarchical model equations).

As seen in Tables 5 and 6, the EQ-5D-5L Health VAS was a significant correlate in the final model for all the domains of sexual function, indicating that higher self-reported HRQoL is linked with greater function in the domains of desire, arousal, lubrication, orgasmic function, sexual satisfaction, and lower levels of sexual pain. As can be seen in Table 5, dysmenorrhea and intermenstrual bleeding were significant negative correlates of desire, arousal, and lubrication, with the presence of intermenstrual bleeding and dysmenorrhea associated with poorer function in these domains. Dysmenorrhea and dyspareunia were also indicated as a significant negative correlate in the final model for

Table 5. Hierarchical multiple regression analyses for correlates of FSFI subscales of sexual desire, sexual arousal, and sexual lubrication in individuals with endometriosis.

Correlate	Sexual desire				Sexual arousal				Sexual lubrication			
	b	β	t	P	b	β	t	P	b	β	t	P
Step 1 ^a												
Pain VAS	-0.01	-0.08	-1.06	0.290	-0.01	-0.11	-1.49	0.137	-0.01	-0.13	-1.74	0.084
Dysmenorrhea	-0.59	-0.22	-2.99	0.003	-0.97	-0.24	-3.41	0.001	-0.97	-0.21	-2.97	0.003
Dyspareunia	-0.53	-0.19	-2.17	0.031	-0.96	-0.24	-2.77	0.006	-0.72	-0.16	-1.79	0.074
Pain after sex	-0.25	-0.09	-1.05	0.297	0.20	0.05	0.56	0.572	0.32	0.07	0.80	0.425
IM bleeding	-0.64	-0.20	-2.76	0.006	-0.87	-0.18	-2.62	0.010	-0.70	-0.13	-1.81	0.072
Fertility concerns	-0.04	-0.01	-0.13	0.896	0.11	0.02	0.25	0.797	0.63	0.09	1.30	0.197
Fatigue	-0.27	-0.07	-0.90	0.372	-0.49	-0.09	-1.15	0.252	-0.44	-0.07	-0.89	0.374
Bloating	0.40	0.11	1.46	0.147	-0.13	-0.02	-0.34	0.738	-0.49	-0.08	-1.07	0.284
Nausea	0.05	0.02	0.23	0.818	-0.46	-0.11	-1.45	0.148	-0.35	-0.07	-0.96	0.339
Vulva pain	0.18	0.06	0.77	0.445	0.11	0.03	0.33	0.739	0.03	0.01	0.09	0.925
Clitoral pain	0.43	0.10	1.21	0.229	0.33	0.05	0.65	0.514	0.37	0.05	0.63	0.529
Step 2 ^b												
Pain VAS	-0.01	-0.10	-1.26	0.210	-0.01	-0.11	-1.53	0.127	-0.01	-0.13	-1.73	0.085
Dysmenorrhea	-0.54	-0.20	-2.69	0.008	-0.85	-0.21	-2.97	0.003	-0.83	-0.18	-2.50	0.013
Dyspareunia	-0.49	-0.18	-1.99	0.048	-0.83	-0.20	-2.39	0.018	-0.56	-0.12	-1.38	0.170
Pain after sex	-0.26	-0.09	-1.02	0.309	0.33	0.08	0.92	0.357	0.50	0.11	1.22	0.224
IM bleeding	-0.60	-0.19	-2.48	0.014	-0.82	-0.17	-2.38	0.019	-0.67	-0.13	-1.67	0.097
Fertility concerns	-0.09	-0.02	-0.31	0.757	-0.04	-0.01	-0.10	0.917	0.43	0.06	0.88	0.379
Fatigue	-0.28	-0.07	-0.92	0.357	-0.52	-0.09	-1.23	0.222	-0.47	-0.07	-0.96	0.337
Bloating	0.39	0.10	1.40	0.163	-0.07	-0.01	-0.18	0.856	-0.41	-0.06	-0.89	0.375
Nausea	-0.15	-0.05	0.66	0.509	-0.36	-0.08	-1.11	0.267	-0.26	-0.05	-0.69	0.491
Vulva pain	0.19	0.07	0.82	0.415	0.12	0.03	0.35	0.728	0.04	0.01	0.11	0.911
Clitoral pain	0.34	0.08	0.94	0.347	0.21	0.03	0.41	0.680	0.26	0.03	0.43	0.666
Number of LES	0.06	0.05	0.66	0.509	-0.05	-0.03	-0.40	0.689	-0.11	-0.06	-0.77	0.444
Time since last LES	-0.01	-0.12	-1.73	0.085	-0.01	-0.06	-0.92	0.359	-0.01	-0.06	-0.92	0.357
Hormonal therapy	-0.34	-0.12	-1.64	0.101	-0.71	-0.17	-2.42	0.017	-0.87	-0.19	-2.56	0.011
Step 3 ^c												
Pain VAS	-0.01	-0.02	-0.20	0.841	0.01	0.02	0.18	0.856	-0.01	-0.01	-0.04	0.964
Dysmenorrhea	-0.55	-0.20	-2.81	0.005	-0.91	-0.22	-3.26	0.001	-0.90	-0.19	-2.78	0.006
Dyspareunia	-0.47	-0.17	-1.93	0.055	-0.74	-0.18	-2.17	0.031	-0.45	-0.10	-1.13	0.261
Pain after sex	-0.32	-0.12	-1.30	0.194	0.23	0.05	0.64	0.520	0.39	0.08	0.97	0.335
IM bleeding	-0.64	-0.20	-2.68	0.008	-0.93	-0.20	-2.74	0.007	-0.78	-0.15	-2.00	0.047
Fertility concerns	-0.05	-0.01	-0.17	0.865	0.01	0.01	0.01	0.990	0.48	0.07	1.01	0.315
Fatigue	-0.24	-0.06	-0.82	0.413	-0.44	-0.07	-1.06	0.289	-0.38	-0.06	-0.80	0.426
Bloating	0.45	0.12	1.63	0.104	-0.04	-0.01	-0.11	0.907	-0.39	-0.06	-0.86	0.389
Nausea	-0.20	-0.07	-0.91	0.362	-0.26	-0.06	-0.81	0.416	-0.15	-0.03	-0.40	0.692
Vulva pain	0.21	0.07	0.92	0.360	0.16	0.04	0.49	0.625	0.09	0.02	0.24	0.810
Clitoral pain	0.42	0.10	1.18	0.239	0.44	0.07	0.86	0.388	0.52	0.07	0.87	0.384
Number of LES	0.08	0.06	0.83	0.405	0.01	0.01	0.06	0.950	-0.05	-0.02	-0.30	0.760
Time since last LES	-0.01	-0.15	-2.09	0.038	-0.01	-0.11	-1.64	0.102	-0.01	-0.11	-1.62	0.106
Hormonal therapy	-0.36	-0.13	-1.77	0.078	-0.73	-0.18	-2.52	0.012	-0.88	-0.19	-2.65	0.009
Health VAS	0.02	0.25	2.65	0.009	0.02	0.21	2.37	0.019	0.02	0.19	2.12	0.035
Utility score	-0.29	-0.05	-0.47	0.637	0.83	0.09	0.96	0.335	1.08	0.11	1.08	0.282

^a=Endometriosis-related symptoms ^b=Endometriosis-related symptoms, Treatment ^c=Endometriosis-related symptoms, Treatment, Health-relation Quality of Life FSFI=Female Sexual Function Index Pain VAS=Pain Visual Analogue Scale (Average level of endometriosis-specific pain experienced in the last 7-day period); IM Bleeding=Intermenstrual Bleeding; Number of LES=Total number of laparoscopic surgeries; Time Since last LES=Approximate duration since last laparoscopic surgery (months); Health VAS=Self-Rated Health-Related Quality of Life (HRQoL) where 0 is worst health imaginable and 100 is best health imaginable; Utility Score=EuroQoL-5 Dimension (EQ-5D-5L) populated normed score; P=0.05 to 0.001; Significant correlates noted in bold font.

sexual satisfaction indicating that the presence of dysmenorrhea is associated with significantly lower levels of orgasmic function and sexual satisfaction. Dyspareunia and hormonal therapy were identified as significant negative correlates of arousal and lubrication, with the presence of dyspareunia and undertaking hormonal therapy associated with poorer function in the domains of sexual arousal and lubrication (see Table 5). Hormonal therapy was also indicated as a significant negative correlate of sexual satisfaction and sexual pain, with the use of hormonal therapy associated with significantly lower levels of satisfaction and greater sexual pain (see Table 6). Intermenstrual bleeding and nausea were found to be significant negative correlates of sexual pain, with the presence of intermenstrual bleeding and nausea associated with significantly higher levels of sexual pain (see Table 6). The length of time since the most recent laparoscopic surgery was identified as a significant negative correlate

of sexual desire, with greater duration post-surgery associated with poorer function in the domain of sexual desire (see Table 5).

Associations between dyspareunia and the FSFI sexual pain domain

Noting that the presence of dyspareunia was not found to be a significant correlate of the FSFI sexual pain domain, bivariate correlations were used to explore whether the distress associated with the experience of dyspareunia may be of greater relevance than its presence alone. A statistically negative correlation was found between the level of distress associated with the presence of dyspareunia and the sexual pain domain, with moderate effect size ($r=-0.38$, $P=0.001$), with higher levels of distress regarding the experience of dyspareunia linked to greater levels of sexual pain (NB: lower scores on FSFI sexual pain subscale equal greater sexual pain).

Table 6. Hierarchical multiple regression analyses for correlates of FSFI subscales of orgasmic function, sexual satisfaction, and sexual pain in individuals with endometriosis.

Correlate	Orgasmic function				Sexual satisfaction				Sexual pain			
	<i>b</i>	β	<i>t</i>	<i>P</i>	<i>b</i>	β	<i>t</i>	<i>P</i>	<i>b</i>	β	<i>t</i>	<i>P</i>
Step 1 ^a												
Pain VAS	−0.02	−0.23	−3.12	0.002	−0.02	−0.27	−3.60	0.001	−0.01	−0.17	−2.27	0.024
Dysmenorrhea	−0.95	−0.23	−3.23	0.001	−0.67	−0.21	−3.02	0.003	−0.30	−0.08	−1.19	0.237
Dyspareunia	−0.98	−0.23	−2.71	0.007	−0.79	−0.25	−2.93	0.004	−0.50	−0.13	−1.57	0.117
Pain after sex	0.11	0.03	0.30	0.765	0.15	0.05	0.55	0.581	0.02	0.01	0.07	0.948
IM bleeding	−0.73	−0.15	−2.10	0.037	−0.22	−0.06	−0.85	0.395	−0.75	−0.17	−2.47	0.014
Fertility concerns	0.46	0.07	1.05	0.295	0.40	0.09	1.23	0.220	0.12	0.02	0.31	0.757
Fatigue	0.70	0.12	1.58	0.116	−0.31	−0.07	−0.92	0.360	−0.29	−0.06	−0.76	0.450
Bloating	0.12	0.02	0.30	0.761	0.25	0.06	0.82	0.411	0.13	0.03	0.36	0.719
Nausea	−0.52	−0.12	−1.58	0.116	0.08	0.02	0.31	0.754	−0.65	−0.17	−2.28	0.024
Vulva pain	−0.31	−0.07	−0.88	0.382	−0.10	−0.03	−0.36	0.716	−0.62	−0.15	−2.02	0.045
Clitoral pain	0.23	0.03	0.42	0.671	0.20	0.04	0.50	0.616	−0.25	−0.04	−0.54	0.590
Step 2 ^b												
Pain VAS	−0.02	−0.22	−2.86	0.005	−0.02	−0.27	−3.57	0.001	−0.01	−0.17	−2.30	0.022
Dysmenorrhea	−0.91	−0.22	−3.03	0.003	−0.56	−0.18	−2.53	0.012	−0.22	−0.06	−0.85	0.395
Dyspareunia	−0.89	−0.21	−2.42	0.017	−0.67	−0.21	−2.47	0.014	−0.41	−0.11	−1.29	0.199
Pain after sex	0.24	0.06	0.62	0.534	0.30	0.09	1.07	0.285	0.10	0.03	0.30	0.763
IM bleeding	−0.87	−0.18	−2.38	0.018	−0.21	−0.06	−0.78	0.437	−0.72	−0.17	−2.29	0.023
Fertility concerns	0.34	0.05	0.76	0.449	0.25	0.05	0.77	0.442	0.01	0.01	0.04	0.971
Fatigue	0.73	0.12	1.62	0.106	−0.33	−0.08	−0.99	0.319	−0.31	−0.06	−0.80	0.427
Bloating	0.19	0.03	0.45	0.651	0.33	0.08	1.06	0.291	0.16	0.03	0.45	0.657
Nausea	−0.61	−0.14	−1.80	0.074	−0.13	0.04	0.52	0.607	−0.58	−0.15	−1.97	0.035
Vulva pain	−0.30	−0.07	−0.85	0.399	−0.09	−0.03	−0.36	0.722	−0.61	−0.15	−1.99	0.047
Clitoral pain	0.34	0.05	0.63	0.533	0.13	0.02	0.33	0.745	−0.33	−0.05	−0.69	0.490
Number of LES	−0.26	−0.14	−1.89	0.061	−0.11	−0.08	−1.09	0.277	−0.03	−0.02	−0.26	0.794
Time since last LES	−0.01	−0.02	−0.22	0.826	−0.01	−0.05	−0.76	0.433	−0.01	−0.06	−0.95	0.343
Hormonal therapy	−0.26	−0.06	−0.82	0.412	−0.63	−0.20	−2.75	0.007	−0.49	−0.12	−1.81	0.071
Step 3 ^c												
Pain VAS	−0.01	−0.08	−0.93	0.354	−0.01	−0.13	−1.63	0.104	−0.01	0.01	0.02	0.987
Dysmenorrhea	−0.98	−0.23	−3.35	0.001	−0.60	−0.19	−2.83	0.005	−0.30	−0.08	−1.20	0.231
Dyspareunia	−0.78	−0.19	−2.17	0.031	−0.62	−0.20	−2.38	0.018	−0.30	−0.08	−1.00	0.321
Pain after sex	0.12	0.03	0.33	0.737	0.19	0.06	0.70	0.488	−0.03	−0.01	−0.11	0.913
IM bleeding	−0.98	−0.20	−2.77	0.006	−0.29	−0.08	−1.13	0.261	−0.85	−0.20	−2.83	0.005
Fertility concerns	0.39	0.06	0.90	0.371	0.32	0.07	1.01	0.311	0.08	0.01	0.21	0.831
Fatigue	0.82	0.14	1.88	0.062	−0.26	−0.06	−0.84	0.403	−0.21	−0.04	−0.57	0.567
Bloating	0.21	0.04	0.51	0.614	0.41	0.10	1.38	0.168	0.20	0.04	0.59	0.558
Nausea	−0.50	−0.11	−1.51	0.134	−0.23	−0.07	0.94	0.348	−0.47	−0.15	−1.89	0.043
Vulva pain	−0.25	−0.06	−0.73	0.465	−0.06	−0.02	−0.22	0.825	−0.56	−0.14	−1.93	0.040
Clitoral pain	0.60	0.09	1.12	0.267	0.30	0.06	0.77	0.441	−0.05	−0.01	−0.12	0.909
Number of LES	−0.19	−0.10	−1.40	0.164	−0.07	−0.05	−0.75	0.457	0.04	0.02	0.36	0.721
Time since last LES	−0.01	−0.07	−0.99	0.319	−0.01	−0.10	−1.51	0.132	−0.01	−0.13	−1.97	0.050
Hormonal therapy	−0.27	−0.06	−0.88	0.379	−0.66	−0.21	−3.01	0.003	−0.51	−0.13	−1.99	0.048
Health VAS	0.10	0.21	2.33	0.021	0.02	0.34	3.93	0.001	0.02	0.30	3.46	0.001
Utility score	1.07	0.11	1.19	0.237	−0.11	−0.02	−0.17	0.87	0.90	0.11	1.18	0.241

^a=Endometriosis-related symptoms ^b=Endometriosis-related symptoms, Treatment ^c=Endometriosis-related Symptoms, Treatment, Health-related Quality of Life FSFI=Female Sexual Function Index Pain VAS=Pain Visual Analogue Scale (Average level of endometriosis-specific pain experienced in the last 7-day period); IM Bleeding=Intermenstrual Bleeding; Number of LES=Total number of laparoscopic surgeries; Time Since last LES=Approximate duration since last laparoscopic surgery (months); Health VAS=Self-Rated Health-Related Quality of Life (HRQoL) where 0 is worst health imaginable and 100 is best health imaginable; Utility Score=EuroQol-5 Dimension (EQ-5D-5L) populated normed score; *P*=0.05 to 0.001; Significant correlates noted in bold font.

Discussion

The current study aimed to investigate HRQoL and sexual function in individuals with symptomatic endometriosis and the influence of laparoscopic surgery on these outcomes. The hypothesis that HRQoL and sexual function would be significantly negatively associated with the length of time following laparoscopic surgery was only partially supported, with a significant negative association only identified for HRQoL. Partial support was found for the hypothesis that a higher number of laparoscopic surgeries would be significantly negatively correlated with HRQoL and sexual function, with a significant negative association found between HRQoL and number of surgeries only.

Of note, level of endometriosis-related pain and HRQoL were not found to be significantly influenced by hormonal therapies. However, sexual function was found to be negatively influenced by hormonal therapies. This finding may be associated with the

surgical focus on the current endometriosis sample, whereby laparoscopic surgery was a key inclusion criterion. Previous research has found that surgery may be necessary for one in three individuals with endometriosis where hormonal therapies are ineffective (Vercellini et al., 2012). Efficacy of hormonal therapies has also been linked to disease stage, with individuals with no history of surgical intervention and early stage endometriosis showing similarities in treatment responsiveness (Jensen et al., 2018). Much of the current sample self-reported a disease stage of moderate to severe, potentially reflecting that both engagement in surgery and disease stage may account for the current findings that use of hormonal therapy was linked to lowered levels of sexual function. The surgical focus and disease stage of the current sample may also explain the nonsignificant differences in current levels of endometriosis-related pain noted between those using or not using hormonal therapies currently.

Reductions in therapeutic effectiveness of hormonal therapies have been associated with limited efficacy in reducing endometriosis-related pain, intolerable side effects, increased intermenstrual bleeding, and lowered sexual desire (Berlanda et al., 2017). The current finding that the use of hormonal therapy was a significant negative correlate of sexual arousal, lubrication, sexual pain, and satisfaction appears consistent with prior findings indicating that, while hormonal therapies are effective for many individuals with endometriosis, they are linked to decreased sexual function in some individuals.

The findings from the current study add further evidence that HRQoL is significantly lower in individuals with endometriosis compared to the Australian public, as measured by the EQ-5D-5L norms (Norman et al., 2023). Poorer HRQoL was especially prevalent in individuals experiencing higher levels of endometriosis-related pain, with poorer HRQoL noted across all individual domains of the EQ-5D-5L, the Health VAS, and utility score. It is also important to note that the current sample had undergone at least one laparoscopic surgery, meaning that the above findings may only be representative of the association between pain and HRQoL of individuals who have received surgical treatment. As such, further research with a greater sample of individuals who have not undergone laparoscopic surgery is required to gain further understanding of the potential influence of surgery on HRQoL outcomes. Future research that measures fluctuations in HRQoL from pre-surgical intervention to regular follow-up intervals is also recommended.

The results indicated that a higher number of laparoscopic surgeries for the treatment of endometriosis was associated with significantly poorer HRQoL in the domains of mobility, usual activity, pain and discomfort, anxiety/depression, and overall HRQoL (utility score). Given the recurrent and pervasive nature of endometriosis, it may be expected that surgical symptom management may be offered more frequently to those with greater physical impairment or disease stage (Pessoa de Farias Rodrigues et al., 2020; Martínez-Zamora et al., 2021; Laguerre et al., 2022). Certainly, within the current sample, over 50% of the sample who reported having at least one laparoscopic surgery self-reported a disease stage of moderate or greater. However, pre-existing research has failed to indicate a clear connection between endometriosis disease stage and HRQoL or surgical outcomes (Leonardi et al., 2020a; Van Niekerk et al., 2022a). Similarly, Warzecha et al. (2020) reported that the presence of common endometriosis-related symptoms (e.g. dysmenorrhea, dyspareunia), type or number of laparoscopic procedures, and psychological wellbeing were not significantly correlated with endometriosis disease stage. They noted that painful defecation and infertility were the only endometriosis-related symptoms positively associated with disease stage, concluding that endometriosis-related symptoms are poorly predictive of disease stage (Warzecha et al., 2020).

Research into the sexual response cycle provides valuable insight into the likelihood that the interplay between psychological and physical factors may act to inhibit the sexual response cycle, negatively impacting the individuals' enjoyment of and willingness to engage in sexual activity (Fairbanks et al., 2017; Della Corte et al., 2020; Vannuccini et al., 2023). A relatively consistent pattern of endometriosis-related symptoms (i.e. dysmenorrhea, dyspareunia, intermenstrual bleeding) were identified as significant negative correlates across multiple domains of sexual function. Endometriosis-related pain and the presence of nausea were also identified as significant negative correlates in specific domains of sexual function.

Dysmenorrhea was a significant negative correlate of sexual function across all subscales of sexual function, except for sexual pain. The significant negative impact of dysmenorrhea on sexual functioning noted in the current study supports the findings of Armour et al., (2020) and Stokes et al. (2023) where individuals with severe or regular dysmenorrhea reported difficulties in their sexually intimate relationships. The results of the current study, and those of Armour et al., (2020) and Stokes et al. (2023), contradict the assumption that the presence of dysmenorrhea is less impactful on sexual function than symptoms such as dyspareunia as it is cyclical in nature and therefore would only be of concern during menstruation (Pluchino et al., 2016).

The presence of dyspareunia was also found to be a significant negative correlate of sexual function in the domains of desire, arousal, orgasmic function, and satisfaction. This finding is consistent with previous research that has highlighted the negative association between dyspareunia and sexual function in individuals with endometriosis (van Poll et al., 2020). The further evidence that dyspareunia is a negative correlate of sexual function, but not specifically in the domain of sexual pain, in individuals with endometriosis highlights the importance of a clinical assessment that provides an opportunity for the individual to discuss their unique symptom presentation and overall sexual response profile with treatment providers. It may be that the presence of dyspareunia inhibits the initial phase of the sexual response cycle (i.e. desire to be intimate), leading to reduced engagement in sexual intimacy. A sexual response screener that focuses solely on the presence of pain during penetrative intercourse may not adequately account for the true impact of dyspareunia on sexual function. Support for extending the assessment beyond the presence or absence of dyspareunia alone is supported by the current finding that greater distress associated with the experience of dyspareunia is significantly linked to greater dysfunction in the sexual pain domain of the FSFI. The experience of dyspareunia has been linked to the experience of negative self-perception, feelings of guilt and inadequacy, and relational distress (Fritzer et al., 2013), further emphasizing the importance of understanding how distressing this symptom is for the person diagnosed. Certainly, Facchin et al., (2021) note that although dyspareunia is a common symptom of endometriosis, the subjective experience of sexual pain remains overlooked in individuals with endometriosis, particularly during patient-doctor interactions. The researchers argue that dyspareunia has a more pervasive impact on sexual function than adequately accounted for and warrants greater clinical attention (Facchin et al., 2021). This is supported by the findings of Schubert et al., (2023) who reported significant improvements in dyspareunia, at a 2-year follow-up time point, in a prospective cohort of 34 women with persistent pelvic pain who had engaged in pelvic physiotherapy.

The unique impact of nausea on the experience of sexual pain is consistent with the finding of reduced sexual function in non-endometriosis samples (Millheiser, 2012) and also supports prior findings regarding the presence of nausea as an endometriosis-related symptom worthy of consideration (Hansen et al., 2014; Evans et al., 2021). Although menstrual irregularities and pelvic pain are often referenced as key symptoms of endometriosis, interestingly over half of the current sample indicated the presence of nausea and its presence was associated with a moderate amount of distress. The experience of nausea may also be related to the process of central sensitization that may occur for individuals with endometriosis, with Evans et al. (2021) finding that abdominal pain, all-over body aches, fatigue, and nausea are predictive of functional pain disability in individuals with

endometriosis. Hansen et al. (2014) reported that individuals with endometriosis frequently reported a constellation of seven symptoms that included painful urination and defecation, constipation or diarrhoea, fatigue, abdominal pain, irregular bleeding, and nausea which they termed 'visceral syndrome', which indicates that the findings in relation to nausea and sexual function warrant further attention.

The findings that specific endometriosis-related symptoms are correlates of sexual function domains highlights the role of multidisciplinary treatment approaches, such as sexual assessment and therapy, rather than relying on medical management alone as well as ensuring that treatment recommendations are person-centred, considering the individual's unique endometriosis symptom profile (Evans et al., 2019; Van Niekerk et al., 2019, 2023). Adopting a holistic treatment approach would involve shifting the onus of care from that of medical professionals to a multidisciplinary team model (Leonardi et al., 2020b; Armour et al., 2022; O'Hara et al., 2022). A multidisciplinary care approach would ensure that surgical intervention is followed by targeted treatment by other medical and allied health practitioners to focus on continued improvement of HRQoL and sexual function outcomes post-surgery. Engagement of multidisciplinary care teams may also be of benefit prior to surgery to assist in managing distress associated with surgical wait times or to provide additional coping and management strategies when further laparoscopic surgery is not deemed to be appropriate. The existing research into endometriosis treatment is largely in support of multidisciplinary management, particularly given the biological, psychological, and social implications of endometriosis, with the consensus being that changes are needed in the diagnosis and management of the condition for improvements in HRQoL to be realized (Della Corte et al., 2020; Evans et al., 2021).

There are some limitations of the current study that are worth noting. Firstly, the correlational research design means that causation cannot be inferred, and the findings from this study are instead useful in that they can help shape future longitudinal research. Sampling bias is also a limitation of the current study's design, as the survey was advertised primarily through endometriosis-related social media platforms and gynaecology services. Attempts were made to mitigate the risk of sampling bias by engaging more than one social media platform, multiple gynaecology clinics, and opening the survey to international respondents. The survey advertisement was also shared by a variety of different health practitioners and networks (e.g. gynaecology, sonography, psychology, physiotherapy, nursing) to broaden participant reach. However, the use of an online survey allowed anonymous responses which aimed to increase the honesty of responses and reduce the risk of social desirability bias impacting the data. The endometriosis diagnosis and surgical history are self-reported only and given the survey design and anonymity of responses was not able to be medically verified. However, Shafrir et al. (2021) found that self-reported endometriosis diagnosis was ~70% accurate and >90% accurate when asked if the diagnosis had been confirmed via laparoscopic surgery, as asked in the current study, concluding that self-reported data can be considered valid in endometriosis samples. The current sample primarily consisted of individuals in heteronormative relationships, meaning that the findings from this study are not generalizable to individuals in sexually diverse and/or non-heteronormative relationships. However, the use of questionnaires with largely gender-neutral and inclusive language facilitates participation from individuals in non-heteronormative relationships. The survey advertisements also used gender

inclusive language to ensure that the experiences of gender diverse people with endometriosis could be accounted for. Future qualitative research focusing on HRQoL and sexual function and any associated changes following laparoscopic surgery is warranted to further understand the current preliminary findings. The current findings relate only to individuals with symptomatic endometriosis and may not be generalizable to individuals with endometriosis who have experienced full symptomatic relief following laparoscopic surgery. It is also noted that most of the sample reported a disease of Stage 3 or greater and therefore further research is required to determine whether the current findings may be influenced by disease stage.

The chosen measure of sexual function in individuals with endometriosis is an important consideration in clinical and research settings. A recent systematic review conducted by Oppenheimer et al. (2024) has found that while the FSFI can be used to discriminate between endometriosis subtypes and has efficacy in evaluating outcomes associated with medical treatment, it may lack specificity for measuring surgical outcomes and low scores may not necessarily be solely reflective of the impacts of endometriosis. In contrast, the use of the Sexual Activity Questionnaire (SAQ; Thirlaway et al., 1996) was noted to have good responsiveness for measuring surgical outcomes in the domains of pleasure and discomfort, has been validated in an endometriosis sample, and allows for a measure of clinically important difference (Oppenheimer et al., 2024). The use of the minimum clinically important difference allows for a closer focus on the domain of concern for the patient and can be used to guide symptom management (Thirlaway et al., 1996). Consistent with the aim of the current study to explore overall sexual function rather than limiting the scope to the experience of dyspareunia, Oppenheimer et al. (2024) argue that the selected sexual function measure must account for the entire sexual response cycle and not be isolated to measurement of pain during penetration alone. While the FSFI does allow for an assessment of the full sexual response cycle, further development of endometriosis-specific scoring norms is recommended. In the absence of FSFI endometriosis-specific norms, use of the FSFI and SAQ together may provide a more comprehensive treatment measure for individuals with endometriosis. The findings from this study support the assertion that care for individuals with endometriosis must extend beyond attempting to alleviate the severity of pain alone and should instead represent a more holistic approach, particularly when considering sexual function and HRQoL. Early referral for assessment and management of sexual concerns is recommended as part of multidisciplinary endometriosis management. Multidisciplinary management of sexual function should address the individual domains of sexual function rather than a limited sexual pain focus. Future studies should look to build upon the small pool of existing research into the unique contribution of specific endometriosis-related symptom profiles on sexual function. Specifically, future research should look longitudinally at the impact of laparoscopic surgery on sexual function from a dyadic perspective, in both heteronormative and non-heteronormative relationships, with the aim of identifying key therapeutic areas for targeted intervention when working with individuals, or couples, with endometriosis.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

Datasets are available on request: The raw data supporting the conclusions of this article will be made available by the authors on written request.

Acknowledgements

The authors wish to acknowledge the support of Endometriosis organizations, including Endometriosis Australia and EndoActive, in assisting with advertising the research. We also wish to acknowledge the medical and allied health professionals who assisted in advertising the research in their patient clinics.

Authors' roles

L.V.N.: Project administration, Conceptualization, Methodology, Data curation, Formal analysis, Validation, Supervision, Writing—original draft; S.P.: Formal analysis, writing—original draft; A.M.-W.: Conceptualization, Methodology, Writing—original draft; C.N.: Conceptualization, Methodology, Writing—original draft; R.O.H.: Conceptualization, Methodology, Writing—original draft; M.A.: Conceptualization, Methodology, Writing—original draft; M.L.: Conceptualization, Methodology, Writing—original draft; S.E.: Conceptualization, Methodology, Writing—original draft.

Funding

The authors declare that no funding sources are associated with the data collection or manuscript submission.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. C.N. reports grant funding from the Australian Government and Medical Research Future Fund (MRFF). This funding is specific to C.N. and was not used to fund the current research.

References

- Abbott JA, Hawe J, Clayton RD, Garry R. The effects and effectiveness of laparoscopic excision of endometriosis: a prospective study with 2–5 year follow-up. *Hum Reprod* 2003;**18**:1922–1927.
- Armour M, Avery J, Leonardi M, Van Niekerk L, Druitt ML, Parker MA, Girling JE, McKinnon B, Mikocka-Walus A, Ng CHM et al. Lessons from implementing the Australian national action plan for endometriosis. *Reprod Fertil* 2022;**3**:C29–C39.
- Armour M, Sinclair J, Ng CH, Hyman MS, Lawson K, Smith CA, Abbott J. Endometriosis and chronic pelvic pain have similar impact on women, but time to diagnosis is decreasing: an Australian survey. *Sci Rep* 2020;**10**:16253.
- Bafot C, Beebejaun Y, Tomassetti C, Bosteels J, Duffy JM. Laparoscopic surgery for endometriosis. *Cochrane Database of Syst Rev* 2020;**10**:CD011031.
- Barbara G, Facchin F, Buggio L, Somigliana E, Berlanda N, Kustermann A, Vercellini P. What is known and unknown about the association between endometriosis and sexual functioning: a systematic review of the literature. *Reprod Sci* 2017;**24**:1566–1576.
- Bartula I, Sherman KA. The Female Sexual Functioning Index (FSFI): evaluation of acceptability, reliability, and validity in women with breast cancer. *Support Care Cancer* 2015;**23**:2633–2641.
- Berlanda N, Somigliana E, Frattaruolo MP, Buggio L, Dridi D, Vercellini P. Surgery versus hormonal therapy for deep endometriosis: is it a choice of the physician? *Eur J Obstet Gynecol Reprod Biol* 2017;**209**:67–71.
- Bourdel N, Chauvet P, Billone V, Douridas G, Fauconnier A, Gerbaud L, Canis M. Systematic review of quality of life measures in patients with endometriosis. *PLoS One* 2019;**14**:e0208464.
- D'Alterio MN, Saponara S, Agus M, Laganà AS, Noventa M, Loi ES, Feki A, Angioni S. Medical and surgical interventions to improve the quality of life for endometriosis patients: a systematic review. *Gynecol Surg* 2021;**18**:13.
- Della Corte L, Di Filippo C, Gabrielli O, Reppuccia S, La Rosa VL, Ragusa R, Fichera M, Commodari E, Bifulco G, Giampaolino P. The burden of endometriosis on women's lifespan: a narrative overview on quality of life and psychosocial wellbeing. *Int J Environ Res Public Health* 2020;**17**:4683.
- Devlin NJ, Shah KK, Feng Y, Mulhern B, Van Hout B. Valuing health-related quality of life: an EQ-5 D-5 L value set for England. *Health Econ* 2018;**27**:7–22.
- Dogan T, Tugut N, Golbasi Z. The relationship between sexual quality of life, happiness, and satisfaction with life in married Turkish women. *Sex Disabil* 2013;**31**:239–247.
- Evans S, Fernandez S, Olive L, Payne LA, Mikocka-Walus A. Psychological and mind-body interventions for endometriosis: a systematic review. *J Psychosom Res* 2019;**124**:109756.
- Evans S, Mikocka-Walus A, Olive L, Seidman LC, Druitt M, Payne LA. Phenotypes of women with and without endometriosis and relationship with functional pain disability. *Pain Med* 2021;**22**:1511–1521.
- Facchin F, Buggio L, Dridi D, Barbara G, Vercellini P. The subjective experience of dyspareunia in women with endometriosis: a systematic review with narrative synthesis of qualitative research. *Int J Environ Res Public Health* 2021;**18**:12112.
- Fairbanks F, Abdo CH, Barakat EC, Podgaec S. Endometriosis doubles the risk of sexual dysfunction: a cross-sectional study in a large amount of patients. *Gynecol Endocrinol* 2017;**33**:544–547.
- Faul F, Erdfelder E, Lang AG, Buchner A. G Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;**39**:175–191.
- Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual Life Res* 2021;**30**:647–673.
- Franck C, Poulsen MH, Karampas G, Giralaldi A, Rudnicki M. Questionnaire-based evaluation of sexual life after laparoscopic surgery for endometriosis: a systematic review of prospective studies. *Acta Obstet Gynecol Scand* 2018;**97**:1091–1104.
- Fritzer N, Haas D, Oppelt P, Renner S, Hornung D, Wölfler M, Ulrich U, Fischerlehner G, Sillem M, Hudelist G. More than just bad sex: sexual dysfunction and distress in patients with endometriosis. *Eur J Obstet Gynecol Reprod Biol* 2013;**169**:392–396.
- Hansen KE, Kesmodel US, Baldursson EB, Kold M, Forman A. Visceral syndrome in endometriosis patients. *Eur J Obstet Gynecol Reprod Biol* 2014;**179**:198–203.
- Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J et al.; REDCap Consortium. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform* 2019;**95**:103208.
- Healey M, Cheng C, Kaur H. To excise or ablate endometriosis? A prospective randomized double-blinded trial after 5-year follow-up. *J Minim Invasive Gynecol* 2015;**22**:511–1004.
- Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new

- five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; **20**:1727–1736.
- Jensen JT, Schlaff W, Gordon K. Use of combined hormonal contraceptives for the treatment of endometriosis-related pain: a systematic review of the evidence. *Fertil Steril* 2018; **110**:137–152.e1.
- Laguette MD, Arkerson BJ, Robinson MA, Moawad NS. Outcomes of laparoscopic management of chronic pelvic pain and endometriosis. *J Obstet Gynaecol* 2022; **42**:146–152.
- Leonardi M, Gibbons T, Armour M, Wang R, Glanville E, Hodgson R, Cave AE, Ong J, Tong YF, Jacobson TZ et al. When to do surgery and when not to do surgery for endometriosis: a systematic review and meta-analysis. *J Minim Invasive Gynecol* 2020a; **27**:390–407.e3.
- Leonardi M, Lam A, Abrão MS, Johnson NP, Condous G. Ignored because it is benign – it is time to treat endometriosis as if it were cancer. *J Obstet Gynaecol Can* 2020b; **42**:507–509.
- Martínez-Zamora MA, Coloma JL, Gracia M, Rius M, Castelo-Branco C, Carmona F. Long-term follow-up of sexual quality of life after laparoscopic surgery in patients with deep infiltrating endometriosis. *J Minim Invasive Gynecol* 2021; **28**:1912–1919.
- Mayr S, Erdfelder E, Buchner A, Paul F. A short tutorial of GPower. *Tutor Quant Methods Psychol* 2007; **3**:51–59.
- Mehedintu C, Plotogea MN, Ionescu S, Antonovici M. Endometriosis still a challenge. *J Med Life* 2014; **7**:349–357.
- Millheiser LE. Female sexual function during pregnancy and postpartum. *J Sex Med* 2012; **9**:635–636.
- Norman R, Mulhern B, Lancsar E, Lorgelly P, Ratcliffe J, Street D, Viney R. The use of a discrete choice experiment including both duration and dead for the development of an EQ-5D-5L value set for Australia. *Pharmacoeconomics* 2023; **41**:427–438.
- O'Hara R, Rowe H, Fisher J. Managing endometriosis: a cross-sectional survey of women in Australia. *J Psychosom Obstet Gynaecol* 2022; **43**:265–272.
- Oppenheimer A, Boitrelle F, Nicolas-Boluda A, Fauconnier A. Measurement properties of sexual function assessment questionnaires in women with endometriosis: a systematic review following COSMIN guidelines. *Acta Obstet Gynec Scand* 2024; **16**:1–25.
- Pessoa de Farias Rodrigues M, Lima Vilarino F, de Souza Barbeiro Munhoz A, da Silva Paiva L, de Alcantara Sousa LV, Zaia V, Barbosa P. C. Clinical aspects and the quality of life among women with endometriosis and infertility: a cross-sectional study. *BMC Women Health* 2020; **20**:1–7.
- Pluchino N, Wenger JM, Petignat P, Tal R, Bolmont M, Taylor HS, Bianchi-Demicheli F. Sexual function in endometriosis patients and their partners: effect of the disease and consequences of treatment. *Hum Reprod Update* 2016; **22**:762–774.
- Riley KA, Benton AS, Deimling TA, Kunselman AR, Harkins GJ. Surgical excision versus ablation for superficial endometriosis-associated pain: a randomized controlled trial. *J Minim Invasive Gynecol* 2019; **26**:71–77.
- Rizk B, Turki R, Lotfy H, Ranganathan S, Zahed H, Freeman AR, Shilbayeh Z, Sassy M, Shalaby M, Malik R. Surgery for endometriosis-associated infertility: do we exaggerate the magnitude of effect? *Facts Views Vis ObGyn* 2015; **7**:109.
- Rosen C, Brown J, Heiman S, Leiblum C, Meston R, Shabsigh D, Ferguson R, D'Agostino R. The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther* 2000; **26**:191–208.
- Rowlands IJ, Abbott JA, Montgomery GW, Hockey R, Rogers P, Mishra GD. Prevalence and incidence of endometriosis in Australian women: a data linkage cohort study. *BJOG* 2021; **128**:657–665.
- Schubert R, Song S, Everist R, Nesbitt-Hawes E, Abbott J. The impact of multimodal physiotherapy in an interdisciplinary setting for the management of women with persistent pelvic pain and pelvic floor tension myalgia. *Euro J Physiotherap* 2023; **15**:1–6.
- Shafir AL, Wise LA, Palmer JR, Shuaib ZO, Katuska LM, Vinayak P, Kvaskoff M, Terry KL, Missmer SA. Validity of self-reported endometriosis: a comparison across four cohorts. *Hum Reprod* 2021; **36**:1268–1278.
- Stokes R, Mikocka-Walus A, Dowding C, Druitt M, Evans S. "It's just another unfortunate part of being female": a qualitative study on dysmenorrhea severity and quality of life. *J Clin Psychol Med Settings* 2023; **30**:628–635.
- Thirlaway K, Fallowfield L, Cuzick J. The Sexual Activity Questionnaire: a measure of women's sexual functioning. *Qual Life Res* 1996; **5**:81–90.
- Van Niekerk L, Johnstone L, Matthewson M. Health-related quality of life in endometriosis: the influence of endometriosis-related symptom presence and distress. *J Health Psychol* 2022a; **27**:3121–3135.
- Van Niekerk L, Johnstone L, Matthewson M. Predictors of self-compassion in endometriosis: the role of psychological health and endometriosis symptom burden. *Hum Reprod* 2022b; **37**:264–273.
- Van Niekerk L, Weaver-Pirie B, Matthewson M. Psychological interventions for endometriosis-related symptoms: a systematic review with narrative data synthesis. *Arch Womens Ment Health* 2019; **22**:723–735.
- Van Niekerk LM, Dell B, Johnstone L, Matthewson M, Quinn M. Examining the associations between self and body compassion and health related quality of life in people diagnosed with endometriosis. *J Psychosom Res* 2023; **167**:111202.
- van Poll M, van Barneveld E, Aerts L, Maas JW, Lim AC, de Greef BT, Bongers MY, van Hanegem N. Endometriosis and sexual quality of life. *Sex Med* 2020; **8**:532–544.
- Vannuccini S, Maseroli E, Vignozzi L, Petraglia F. The challenge of endometriosis for female sexual health. *J Sex Med* 2023; **20**:240–246.
- Vercellini P, Somigliana E, Consonni D, Frattaruolo MP, De Giorgi O, Fedele L. Surgical versus medical treatment for endometriosis-associated severe deep dyspareunia: I. Effect on pain during intercourse and patient satisfaction. *Hum Reprod* 2012; **27**:3450–3459.
- Warzecha D, Szymusik I, Wielgos M, Pietrzak B. The impact of endometriosis on the quality of life and the incidence of depression—a cohort study. *Int J Environ Res Public Health* 2020; **17**:3641.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Human Reproduction, 2024, 39, 992–1002
<https://doi.org/10.1093/humrep/deae063>
Original Article

Puberty, aging and HRT

Urinary gonadotropin assay on 24-h collections as a tool to detect early central puberty onset in girls: determination of predictive thresholds

Clément Janot ^{1,2,3,*}, Pauline Perrin ¹, Véronique Raverot ¹, Patricia Bretones⁴, René Ecochard ⁵, Sarah Malburet-Testori¹, Marc Nicolino^{2,4}, Zoé Robert¹, Florence Roucher-Boulez ^{1,2,3}, Carine Villanueva^{2,4}, Kevin Perge ^{2,4}, and Ingrid Plotton ^{1,2,3}

¹Service de Biochimie et Biologie moléculaire, Centre de Biologie et de Pathologie Est, Hospices Civils de Lyon, LBMMS, Bron Cedex, France

²Faculté de Médecine Lyon Est, Université Claude Bernard Lyon 1, Lyon, France

³StemGame Platform group, UMR INSERM 1208 SBRI, Bron, France

⁴Groupe Hospitalier Est, Service d'Endocrinologie pédiatrique, Hospices Civils de Lyon, Bron Cedex, France

⁵Laboratoire Biostatistique Santé, UMR CNRS 5558 UCBL, Lyon, France

*Correspondence address. Service de Biochimie et Biologie moléculaire, Centre de Biologie et de Pathologie Est 5^e étage, Groupe Hospitalier Est, 59, Boulevard Pinel 69500 Bron, France. E-mail: clement.janot@chu-lyon.fr  <https://orcid.org/0009-0001-9336-6874>

ABSTRACT

STUDY QUESTION: Is the 24-h urinary gonadotropin assay an effective diagnostic tool in central precocious puberty (CPP) in girls?

SUMMARY ANSWER: This study is the first to provide 24-h urinary gonadotropin assay data, using an electrochemiluminescent immunoassay (CMIA), and to report its usefulness as a tool for the diagnosis of CPP.

WHAT IS KNOWN ALREADY: Data about the GnRH test in the diagnosis of CPP are variable and there is no consensus regarding its interpretation. The measurement of FSH and LH in urines was previously reported to be an alternative biological tool.

STUDY DESIGN, SIZE, DURATION: This is a retrospective two-cohort study, involving a setting and a validation cohort. A total of 516 girls, included between October 2012 and July 2015, and 632 urinary collections were analyzed in the setting cohort. In the validation cohort, 39 girls were included between January 2021 and May 2023, and 49 urinary collections were analyzed.

PARTICIPANTS/MATERIALS, SETTING, METHODS: This study included girls who consulted for an investigation of disturbed growth rate or a clinical suspicion of puberty onset in different medical centres across France (setting cohort). Girls with a suspicion of precocious puberty onset were addressed at the expert centre of paediatric endocrinology of the Groupe Hospitalier Lyon Est (validation cohort). Pelvic ultrasonography was performed and enabled their classification according to clinical and morphologic changes criteria (prepubertal or pubertal groups). The parents collected 24-h urine samples (u_{24}) according to standardized instructions. FSH and LH (urinary or plasmatic) were measured using a current and automated CMIA.

MAIN RESULTS AND THE ROLE OF CHANCE: The area under the ROC curves for CPP prediction was 0.709 for u_{24} FSH ($P < 0.001$), 0.767 for u_{24} LH ($P < 0.001$), and 0.753 for the u_{24} LH/ u_{24} FSH ratio ($P < 0.001$). We retained all possible combinations of the four thresholds in the validation cohort (u_{24} FSH = 1.1 or 2.0 IU/24 h; u_{24} LH = 0.035 or 0.08 IU/24 h). The combination of u_{24} FSH > 1.1 IU/24 h and u_{24} LH > 0.08 IU/24 h had a positive PV of 85.7% and a negative PV of 94.3%, a sensitivity of 85.7% and a specificity of 94.3%, for classifying prepubertal and pubertal girls in this cohort.

LIMITATIONS, REASONS FOR CAUTION: This is a retrospective study, in which a margin of error remains due to the inherent uncertainty regarding the clinical assessment of pubertal onset. It must be considered that the thresholds can only apply to the used reagents; measurements without extractions using other reagents are likely to show important heterogeneity.

WIDER IMPLICATIONS OF THE FINDINGS: The assay performed herein is a simple, non-invasive, and analytically robust technique meeting the criteria for an alternative to the GnRH test which could be used to supplement its lack of sensitivity.

STUDY FUNDING/COMPETING INTEREST(S): No specific funding was used. All authors declared no conflict of interest.

TRIAL REGISTRATION NUMBER: In-house #23-5214 registered study.

Keywords: precocious puberty / gonadotropins / hormones / children / urine

Introduction

Central precocious puberty (CPP) is caused by the early activation of the hypothalamic–pituitary–gonadal (HPG) axis. The prevalence was estimated to be from 1 in 5000 to 1 in 10 000 children, having a

10:1 girl:boy ratio (Carel and Léger, 2008; Maione *et al.*, 2021). The diagnosis is of crucial importance, as a search for the etiological causes, using cerebral MRI, could lead to the identification of lesions of the central nervous system, reported in <1–3% of

Received: November 14, 2023. Revised: February 15, 2024. Editorial decision: February 29, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

6–8 years old children, and up to 6–25% of <6 years old cases (Cantas-Orsdemir et al., 2018). In girls, precocious puberty onset is clinically defined by the development of secondary sexual characteristics before 8 years of age (Eckert-Lind et al., 2020; Roberts and Kaiser, 2020); breast development is the first clinical sign leading to its suspicion. As clinical examination of breast development may be difficult to confirm, a hormonal assessment is essential before introducing treatment in the case of CPP. To confirm the onset of CPP, the current guidelines (Carel et al., 2009; Bangalore Krishna et al., 2019) recommend the measurement of plasma LH using sensitive chemiluminescent or electrochemiluminescent immunoassay (CMIA or ECLIA) reagents, before or after a GnRH stimulation test (Kastin et al., 1972a,b). However, their diagnostic performances are flawed due to a significant overlap in hormone levels between the prepubertal stage, premature thelarche, and central puberty onset (Carel et al., 2009; Bangalore Krishna et al., 2019). In girls, a pelvic ultrasound can reveal morphologic evolution of the internal genitalia, and is highly specific and non-invasive, but it lacks of sensitivity (De Vries et al., 2006; Sathasivam et al., 2011).

The GnRH test is widely used by practitioners, although there are different well-defined hormone levels thresholds after stimulation (Carel and Léger, 2008; Brito et al., 2016). Nevertheless, an LH peak >5 IU/l associated with an LH/FSH peak ratio >0.66 is commonly adopted by the paediatrician community. Previous studies have reported variable protocols, using different times for blood sampling, and different analytical methods for measurements of FSH and LH (Kandemir et al., 2011; Brito et al., 2016; Ab Rahim et al., 2020; Huynh et al., 2022). The GnRH stimulation test is also invasive and stressful, involving an intravenous injection and one to six repeated samples under supervision, usually requiring day-hospitalization. The measurement of FSH and LH in urine samples was previously reported to be a useful biological tool to detect the pubertal increase of gonadotropin secretion (Buckler and Clayton, 1970; Rifkind et al., 1970; Chipman et al., 1981). Several studies have reported interesting results regarding the usefulness of first-morning voided (FMV) or random non-timed urines to detect puberty onset with good sensitivity in girls (McNeilly et al., 2012; Zung et al., 2014; Demir et al., 2016, 2021; Lucaccioni et al., 2016; Kolby et al., 2017; Shim et al., 2019; Yüce et al., 2020; Lee et al., 2021; Zhan et al., 2021). The collection of 24-h urine samples provides the complete nycthemeral activity of the HPG axis and the data correlates well with that from morning urine samples (Maesaka et al., 1990). However, few data have been reported about 24-h urinary collections for precocious puberty, except where radioimmunological techniques have been used (Morel et al., 1985).

The aim of this study was to assess 24-h urinary gonadotropin measurements as a tool for girls presenting with a clinical suspicion of CPP, by establishing and then validating the decision thresholds of FSH (u_{24} FSH) and LH (u_{24} LH).

Materials and methods

Study design and population

This is a retrospective two-cohort study, involving a setting cohort and a validation cohort.

For the setting cohort, we identified girls who consulted for an investigation of disturbed growth rate or a clinical suspicion of puberty onset in different medical centres across the Auvergne-Rhône-Alpes region and beyond (France) between October 2012 and July 2015. We included girls for whom (i) the Tanner pubertal stage for breast development was determined at the time of the consultation, prior to urine collection; and (ii) the 24-h urinary

collections were addressed for the u_{24} FSH and u_{24} LH assays to the Hormonology unit of Biochemistry and Molecular Biology department of Groupement Hospitalier Est (Hospices civils de Lyon, France). We excluded girls: (i) presenting with endocrine disease which interferes with the onset of puberty (congenital adrenal hyperplasia, peripheral puberty onset, gonadal diseases, growth hormone deficiency, thyroid hormone deficiency, syndromic pathologies, etc.); (ii) younger than 4 years old, in order to eliminate the ‘minipuberty’ due to transient late activation of the HPG axis; (iii) with urinary collections under 400 ml or without available collection volumes; or (iv) with doubtful or inconclusive determination of Tanner stages.

The Tanner classification of breast development followed the detailed clinical description previously provided and used worldwide (Marshall and Tanner, 1969; Morris and Udry, 1980; Carel and Léger, 2008). In case of unilateral right or left breast classified as S2 (and the other S1), the patient was considered S2.

In this cohort, the urine-plasma correlations as well as the distribution of urinary values regarding Tanner stages were analyzed. We established thresholds that allowed us to characterize the puberty onset using 24-h urinary gonadotropins by using S1 and S2 values. We included GnRH stimulation tests when they were performed at the time of consultation. This test consisted in an intravenous injection of 2.5 µg/kg or up to 100 µg of gonadorelin acetate. Blood samples were collected prior to the injection (T0), then 30, 60, and 90 min after injection: peak values of FSH and LH during the test were measured. The stimulation test was considered positive when the LH peak was higher than 5 IU/l, or when the LH/FSH peak ratio was higher than 0.66. It was decided to consider the peak of the LH or that of the ratio so as not to artificially reduce the sensitivity of the test.

For the validation cohort, we identified girls with a suspicion of precocious puberty onset consulted with fifteen practitioners in the Department of Pediatric Endocrinology and Metabolic Diseases of the Groupement Hospitalier Est between January 2021 and May 2023. The suspicion of precocious puberty was defined as girls for whom the first clinical signs occurred before 8 years of age, considering the delay between puberty onset and consultation. The inclusion criteria of this validation cohort were the same as those of the setting cohort. Were excluded girls: (i) with doubtful or inconclusive Tanner stages (‘unilateral’ (right or left) S2 were considered S2); (ii) with urinary collections under 600 ml; (iii) older than 8 years old at the time of clinical onset; (iv) with peripheral precocious puberty; or (v) ongoing puberty blocking treatment. As the aim in this study was to differentiate prepubertal from pubertal patients, girls with breast development stage $\geq$ S2 were merged in a S2 group. Pelvic ultrasonography (US) was performed in all girls in the S2 group in our centre. This enabled their classification into two groups: (i) pubertal S2US+ with girls who had patterns of morphologic enlargement of the uterus and ovaries due to estrogen effects, according to previously reported criteria (Table 1, derived from Stranzinger and Strouse, 2008; Asavaoie, 2014), and (ii) S2US– were girls who had no such patterns. The S1 and S2US– girls constituted the prepubertal group.

For this cohort, the aim was to validate the previously established u_{24} FSH and u_{24} LH thresholds according to the breast Tanner stages (same criteria than previously) and the US data in the prepubertal and pubertal groups. Urinary gonadotropin assays were performed and obtained as part of routine clinical practice. The clinical progression of puberty at a second consultation was included if available, in order to study the

Table 1. Ultrasonography criteria for HPG axis activation.

	Prepubertal (US-)	Pubertal (US+)
Uterus		
Shape	tubular shape	pear shape
Volume (ml)	<1.6	>1.6
Fundus/cervix size ratio	≤1:1	>1:1
Endometrial line	absent	Visible
Endometrial stripe	hypoechoic	Echogenic
Ovaries		
Volume (ml)	<2	≥2

concordance with urinary findings. Evolution of breast size was used as clinical criterion to on-course progressive precocious puberty.

Urinary collections

The parents collected the 24-h urine according to standardized instructions: patients were asked to discard the last evening's voiding before going to bed and note the time. During the following 24 h, all subsequent urine samples were collected, in a clean, dry, and empty canteen without additives, until the time recorded the previous day. The urine was homogenized at the end of the collection. The total volume was measured and a sample of the homogenate was placed in a tube without additives. The assay was performed on a non-frozen aliquot less than a week after the end of the collection.

FSH and LH measurements

FSH and LH (urinary or plasmatic) were measured by a current and automated CMIA using Architect i2000 analyzer from Abbott Laboratories® (IL 60064, USA; FSH reagent ref. 7K75; LH reagent ref. 2P40). No modification was carried out to the assays, whether they were used for urine or plasma. Both assays are calibrated against the international reference materials from World Health Organization (WHO), 92/510 and 80/552, respectively, for FSH and LH. Assays on urine were performed on samples stored at 4 °C that were never frozen, free of any previous treatment or extraction, rapidly vortexed and centrifuged just before measurement. The analytical method applied to urine in this work was validated in the hormonology laboratory, and is highly correlated with the immunoradiometric assay (IRMA) technique, notably on extraction-free urine (Supplementary Fig. S1 and Table S1). Plasma gonadotropins were collected in ethylene diamine tetraacetic acid tubes. Intra-assay coefficients of variation were <4.2% for plasma FSH, <3.6% for plasma LH (determined by Abbott Laboratories®), <4.1% for urinary FSH and <6.5% for urinary LH (determined by in-house tests on urinary samples). Inter-assay coefficients of variation were <2.8% for plasma FSH, <2.7% for plasma LH, <3.3% for urinary FSH, and <4.5% for urinary LH (determined by the long-term use of Technopath® Multichem U and Multichem IA plus Quality Controls; from Technopath Clinical Diagnostics, Ballina, Ireland). Urinary assays were related to the volume of diuresis and expressed as daily excretion ($u_{24}FSH$ and $u_{24}LH$, in IU/24h).

Statistical analyses

The median of $u_{24}FSH$ and $u_{24}LH$ were used for distribution analysis and statistical comparison using the non-parametric Kruskal–Wallis test. Post hoc Dunn's test was used for multiple comparisons. Mann–Whitney test was used for comparisons between GnRH+ and GnRH– groups. Relationships between $u_{24}FSH$, $u_{24}LH$ and plasmatic FSH and LH were analyzed using Spearman correlation. Thresholds establishment used the Youden index (YI)

($Y = \text{sensitivity} + \text{specificity} - 1$). An alpha risk of error was set at 5% for all tests and confidence intervals. All statistic tests, receiver operating characteristic (ROC) and precision–recall curves and threshold establishment were performed using the GraphPad Prism® version 6.01 (GraphPad Software, Boston, MA, USA).

Ethical approval

The present study protocol was approved by the ethics committee of our institution (23-5214). Medical and biological personal data in this research were studied anonymously. Their use followed good practices and the principles of the Declaration of Helsinki as well as the European legislation on personal data protection (MR-004 GDPR standards).

Results

Setting cohort

A total of 516 girls with 789 urinary collections were initially selected. After exclusions, 632 urinary collections were finally analyzed in this cohort (Fig. 1). Among them, 23 patients underwent the GnRH stimulation test (providing 30 stimulation and collections tests), and 159 patients had blood collection for basal plasma FSH and LH assays. There were 17 girls who had a negative test (GnRH–) and 13 who had a positive test (GnRH+). Interestingly, $u_{24}FSH$ and $u_{24}LH$ were correlated with both the LH peak and the LH/FSH peak ratio (Fig. 2A and B). However, the FSH peak alone was not correlated with either $u_{24}FSH$ or $u_{24}LH$. $u_{24}FSH$ and $u_{24}LH$ were both higher in GnRH+ patients compared with GnRH– patients ($P = 0.001$ and $P < 0.001$, respectively). Among the 13 GnRH+ girls, four had an $u_{24}LH$ and five had an $u_{24}FSH$ readings within the range of the GnRH– girls (Fig. 2C). Basal plasma FSH and LH were respectively correlated with $u_{24}FSH$ ($r^2 = 0.784$, $P < 0.001$) and $u_{24}LH$ ($r^2 = 0.715$, $P < 0.001$).

The number of patients and samples as well as the ages per Tanner stage groups are shown in Table 2. The distribution of $u_{24}FSH$ (Fig. 3A–C) showed a wide range of values for Tanner stage from 1 to 4. Although there was a significant difference between the four stages, and the S2 mean value was higher than that of S1 ($P < 0.001$), the overlap of the S1 and S2 intervals was important. One outlier point was identified in the S1 group ($u_{24}FSH = 11.62$ IU/24 h). The distribution of $u_{24}LH$ (Fig. 3B–D) showed a narrow range of values in S1 girls (up to 0.31 IU/24 h) but was higher in S2 girls (up to 2.51 IU/24 h). There was a significant difference according to Tanner stages ($P < 0.001$). The proportion of undetectable $u_{24}LH$ (less than the lower limit of detection, LLOD) was the highest in S1 girls, 65/191 (34%) of the patients. It decreased to 33/291 (11.3%), 5/119 (4.2%), and 1/31 (3.2%) in each successive Tanner group, from 2 to 4. The area under the ROC curves was 0.709 for $u_{24}FSH$ ($P < 0.001$, Fig. 3E), 0.767 for $u_{24}LH$ ($P < 0.001$), and 0.753 for the $u_{24}LH/u_{24}FSH$ ratio ($P < 0.001$). The area under the precision–recall curves (Fig. 3E) was higher for $u_{24}LH$ (0.847) than for $u_{24}FSH$ (0.772) or the $u_{24}LH/u_{24}FSH$ ratio (0.765). Several hypothetical threshold values ranged in descending order of YI. The highest index thresholds were 0.035 IU/24 h for $u_{24}LH$ (sensitivity 70.9%, specificity 70.7%, $YI = 0.416$) and 2.0 IU/24 h for $u_{24}FSH$ (sensitivity 72.3%, specificity 62.8%, $YI = 0.352$). $u_{24}LH$ did not reach more than 89% sensitivity but led to a specificity higher than 90% (91.1%) by raising the threshold to 0.08 IU/24 h (Supplementary Table S2). Similarly, $u_{24}FSH$ reached a sensitivity higher than 90% (91.5%) by lowering the threshold to 1.1 IU/24 h (Supplementary Table S2). Thus, elevated $u_{24}LH$ seemed to be more interesting to identify puberty initiation (specificity) and low $u_{24}FSH$ is better to exclude it (sensitivity). It was chosen to retain and evaluate all possible combinations of the four latter above-mentioned thresholds in the

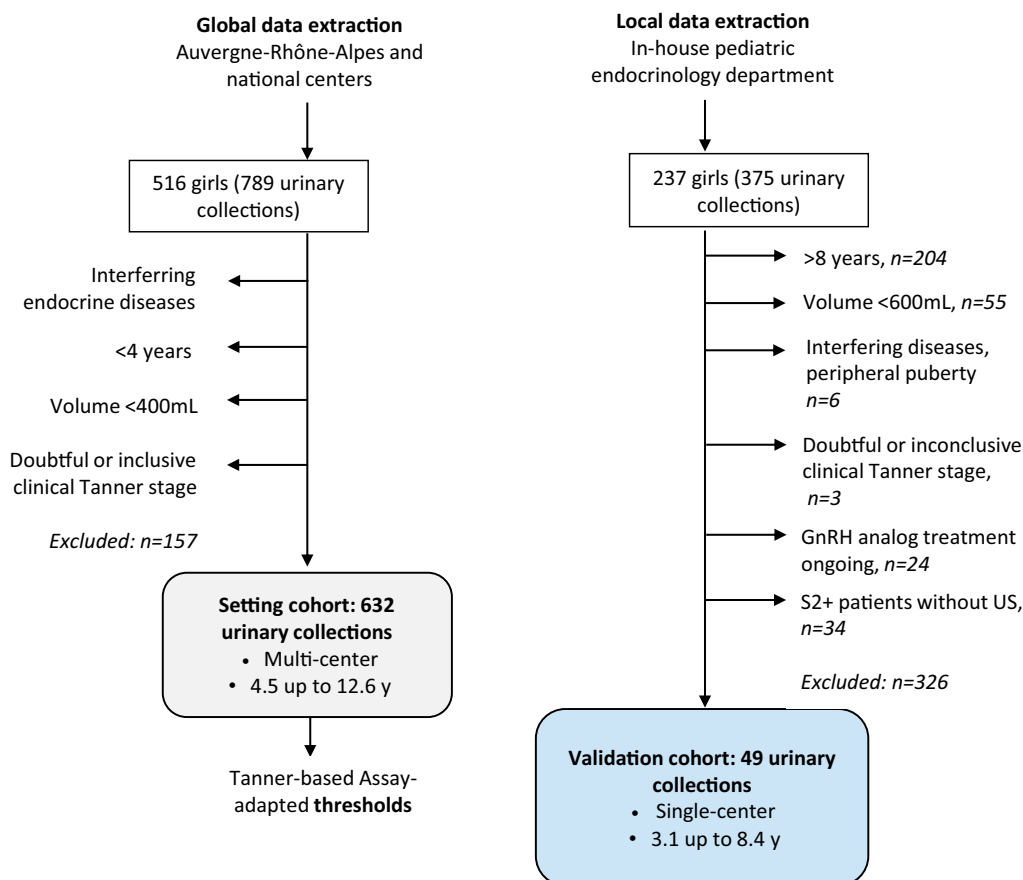


Figure 1. Flowchart of inclusion and exclusion steps of setting and validation cohorts. GnRH, gonadotropin-releasing hormone; S2+, Breast Tanner stage or more; US, ultrasonography.

validation cohort. Retrospectively, using these thresholds, the 30 GnRH tests were compared with urinary collections: 13/13 GnRH+ patients had $u_{24}\text{FSH} > 1.1 \text{ IU}/24 \text{ h}$ and 10/13 had $u_{24}\text{LH} > 0.08 \text{ IU}/24 \text{ h}$, while 4/17 GnRH- patients had $u_{24}\text{LH} > 0.08 \text{ IU}/24 \text{ h}$ (Fig. 2C). As the $u_{24}\text{LH}/u_{24}\text{FSH}$ ratio did not seem to offer a better performance than $u_{24}\text{LH}$ alone, a ratio threshold was not retained.

Validation cohort

A total of 237 girls (375 urinary collections) were initially selected, and after exclusion, 39 patients and 49 urinary collections were finally included (Fig. 1). The age of patients at the time of inclusion were comparable in each subgroup, with mean ages between 6.44 and 7.34 years (S1, S2US-, S2US+, Supplementary Table S3). There were 15 girls who were >8 years old at the time of consultation but the clinical signs occurred before this age. Four girls had breast development at S3 Tanner stage and were merged with S2 girls. No girls presented with a S4 or S5 breast development. Among the 21 girls with Tanner stage 1 (S1), only 12 had a pelvic US and among them, two showed slight signs of an onset of morphological changes; their $u_{24}\text{FSH}$ were 1.61 and $2.4 \text{ IU}/24 \text{ h}$, and their $u_{24}\text{LH}$ were 0.024 and $<0.02 \text{ IU}/24 \text{ h}$. Diagnostic performances of $u_{24}\text{FSH}$ and/or $u_{24}\text{LH}$ to identify S2US+ girls (predictive value, PV) were tested in all combinations of previously retained thresholds in the setting cohort (Table 3). The best positive PV was 85.7% (± 9.8) combining $u_{24}\text{FSH} > 1.1 \text{ IU}/24 \text{ h}$ and $u_{24}\text{LH} > 0.08 \text{ IU}/24 \text{ h}$ and the negative PV was 94.3% (± 6.5). By increasing the $u_{24}\text{FSH}$ threshold to $2.0 \text{ IU}/24 \text{ h}$, the negative PV reached 95%, but at the expense of the positive PV (46%). In the prepubertal group, 33/35 girls had

$u_{24}\text{FSH}$ or $u_{24}\text{LH}$ below 1.1 and $0.08 \text{ IU}/24 \text{ h}$. Conversely, 12/14 S2US+ girls had $u_{24}\text{FSH}$ and $u_{24}\text{LH}$ above the thresholds (Fig. 4).

Of this cohort, 27 girls were re-examined (median delay: 7 months) by the practitioners of the centre, who determined whether a progressive puberty was ongoing or not. Of the 25 girls in the prepubertal group, 17 were followed up and 15 did not have evolving puberty at the second consultation. One of the two girls in this group for whom urinary assays was found above the thresholds was still non-pubertal at follow-up. Of the 14 girls in the pubertal group (S2US+), ten were reviewed. Among these, nine with elevated urinary gonadotropin were proposed to receive puberty blockers, but in two cases the family did not choose the treatment for their child because they were close to 8 years old. Of the two out of the 14 S2US+ girls for whom urinary gonadotropins were below the thresholds, one had a non-progressive breast development (premature thelarche), the other were found to have a tumour of the sex cords. Among S2US- patients, only the one with pubertal urinary gonadotropins ($u_{24}\text{FSH} = 4.73 \text{ IU}/24 \text{ h}$ and $u_{24}\text{LH} = 0.31 \text{ IU}/24 \text{ h}$) was finally diagnosed with a slowly progressive puberty with an indication of treatment.

Overall, the combination of $u_{24}\text{FSH}$ and $u_{24}\text{LH}$ had a positive PV of 85.7% and a negative PV of 94.3% to classify prepubertal and pubertal girls in this cohort.

Discussion

This study is the first to provide 24-h urinary gonadotropin assay data, using the CMLA reagent of Abbott Laboratories®, and report its usefulness as a tool for paediatrician in the diagnosis of CPP.

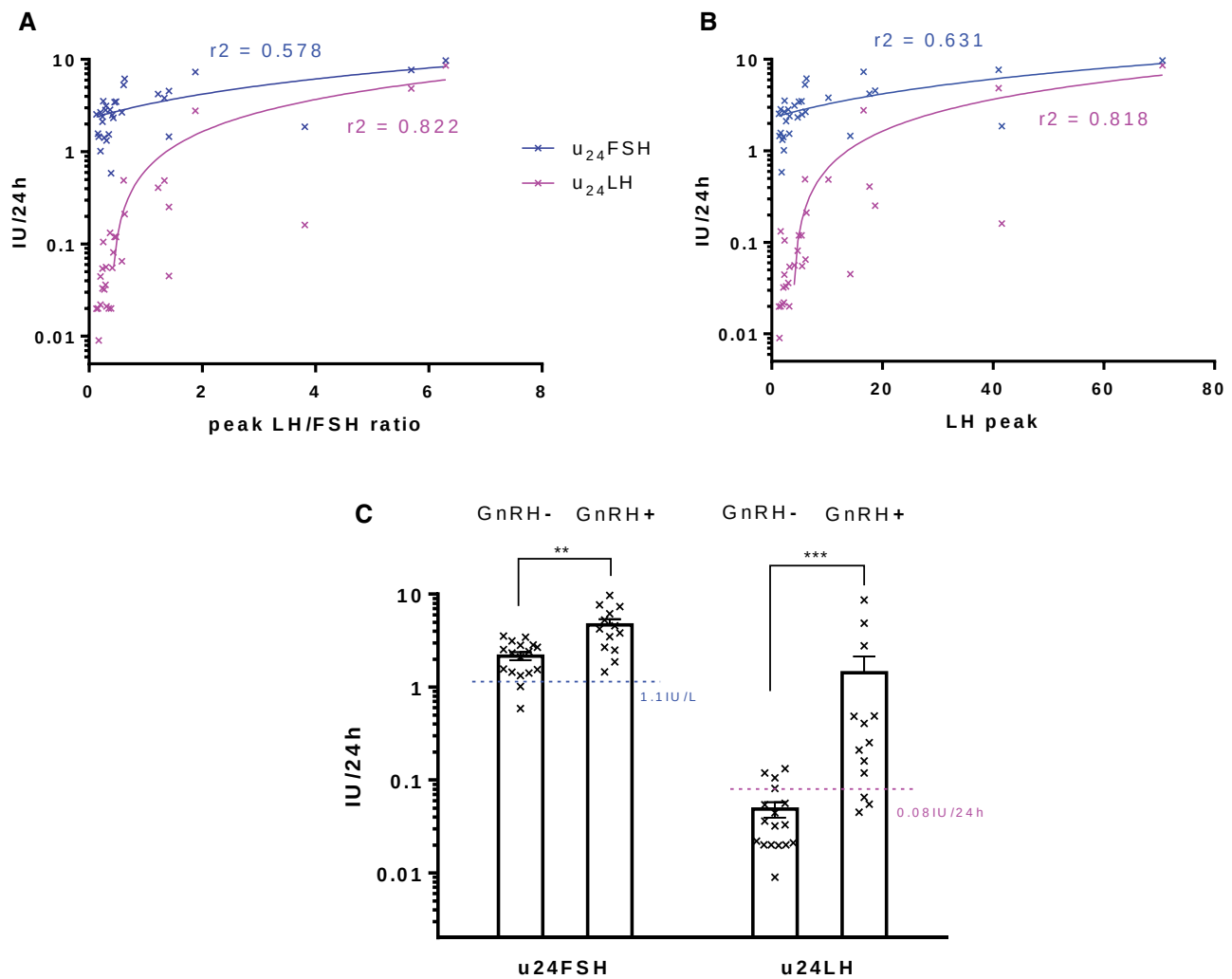


Figure 2. Correlations between 24-h urinary gonadotropins and GnRH tests. (A and B) Correlation (r = Spearman coefficient) with LH peak and LH/FSH peak ratio after GnRH test in 30 urinary collection (23 patients). (C) Distribution of $u_{24}FSH$ and $u_{24}LH$ according to GnRH test response. Blue- and purple-hashed lines respectively correspond with $u_{24}FSH$ and $u_{24}LH$ thresholds used hereafter (** $P = 0.001$; *** $P < 0.0001$; Mann-Whitney tests). FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone.

Table 2. Setting cohort characteristics.

	Samples (patients)	Age (years in mean $\pm$ SD)	Age of breast development onset (years)		
			2.5th p	Median	97.5th p
All	632 (516)	8.16 $\pm$ 1.07	4.83	7.75	9.50
S1	191 (142)	7.78 $\pm$ 1.04		–	
S2	291 (242)	8.18 $\pm$ 1.03	5.15	7.80	9.67
S3	119 (107)	8.51 $\pm$ 1.03	4.70	7.50	9.25
S4	31 (25)	8.84 $\pm$ 1.06	4.70	8.00	10.20

The present results supported that $u_{24}FSH$ and $u_{24}LH$ are able to detect an increase of the gonadotropin excretion, reflecting the activation of the HPG axis. This was defined by clinical (breast Tanner stage ≥ 2) and US (morphological changes of internal genitalia) criteria. The choice of combined thresholds, $u_{24}FSH > 1.1 IU/24 h$ and $u_{24}LH > 0.08 IU/24 h$, allowed the detection of CPP onset with 85.7% of positive PV and 94.3% of negative PV. The assay performed herein is a simple, non-invasive, and analytically robust technique meeting the criteria (sensitivity, specificity, and PVs) to become an alternative to GnRH test. The urinary collection can be achieved without difficulty at home,

thus avoiding hospitalization, care-related stress and missing school, as well as costs and inconvenience associated with the administration of GnRH and repeated venous blood samplings.

Nocturnal FSH and LH pulses are early biological signs of central pubertal onset (Boyar *et al.*, 1972), preceding the first clinical signs (Demir *et al.*, 1996). Currently, the GnRH stimulation test is considered the gold standard to confirm central pubertal onset. However, its performance in the diagnosis of central puberty has been reported to be variable by some authors (Yazdani *et al.*, 2012; Demir *et al.*, 2016). Regarding the interpretation of the test, there are disparities of the compounds used, which are not all

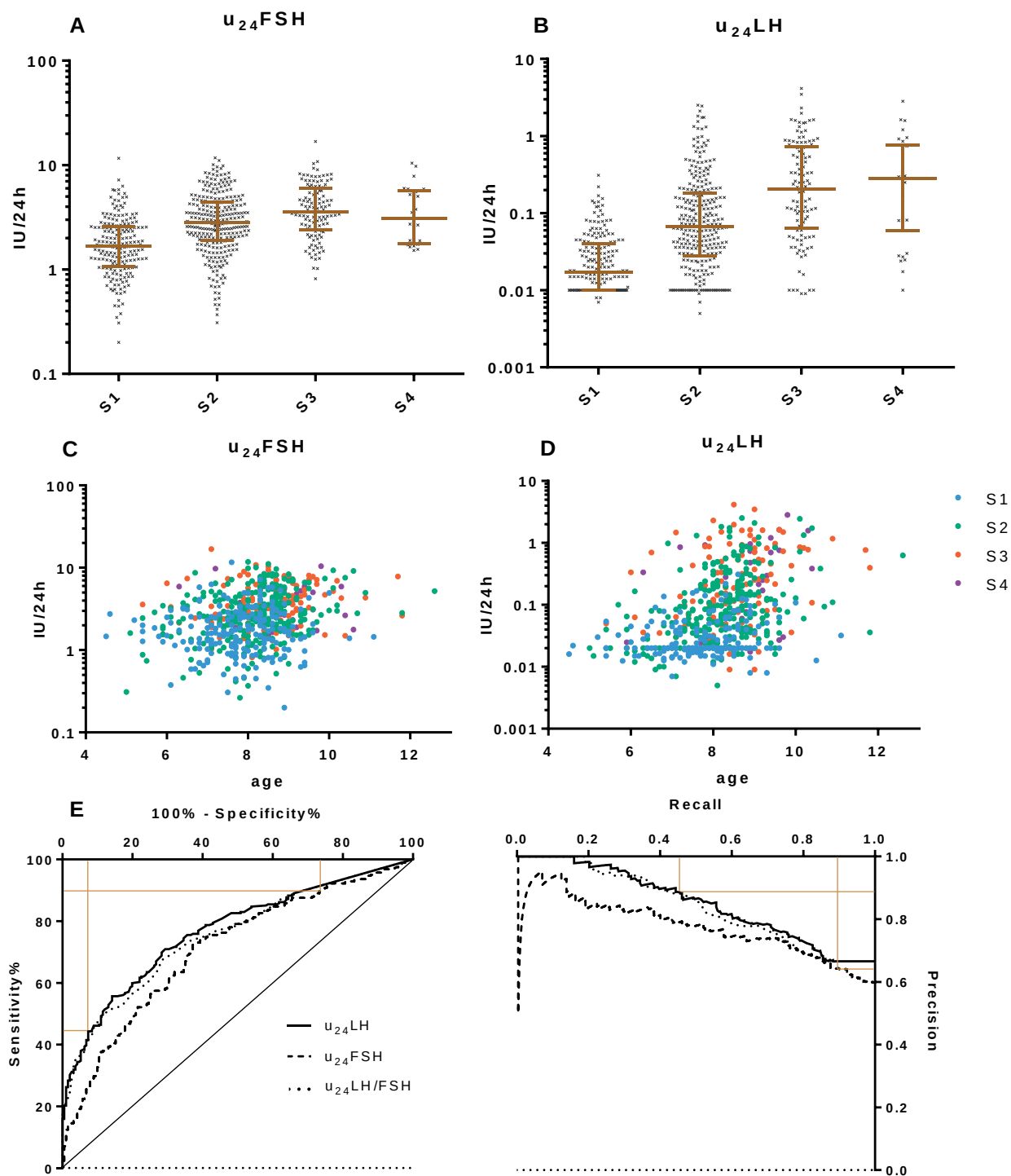


Figure 3. Distributions of values in the setting cohort and receiver operating characteristic (ROC) curves. (A and B) Distribution of values in log-10 scale per Tanner breast stages. Undetectable $u_{24}LH$ are represented as aligned dots on the $y = LLOD$ line, namely 0.01 IU (B). (C and D) Urinary gonadotropins according to Tanner breast stage and age. (E) Receiver operating (sensitivity and specificity) and Precision-Recall (precision=predictive value; recall=sensitivity) curves. Horizontal and vertical lines point finally validated thresholds. FSH, follicle-stimulating hormone; LH, luteinizing hormone; LLOD, low limit of detection.

available in every regions of the world, and the thresholds highly vary between expert teams (Brito et al., 2016). By identifying pubertal onset based on the clinic, Demir et al. (2016) showed that the GnRH test had a sensitivity of 59% and a specificity of 66% to distinguish S1 from S2 Tanner girls. Similarly, Yazdani et al. (2012) reported that the LH/FSH peak ratios exceeded 1 in only 55% of girls with central pubertal onset. In our validation cohort,

we were able to use a clinical progression criterion for some patients to confirm central pubertal onset or not. Although we think this is a good indicator, further studies would be interesting to know whether this criterion is as good or better than the GnRH test. Interestingly, it was noted herein that the use of urinary gonadotropin to confirm HPG axis activation allowed us to reclass four girls as pubertal while the GnRH test was negative. Urinary

Table 3. Thresholds validation.

$u_{24}FSH$ thresholds (IU/24 h)	$u_{24}LH$ thresholds (IU/24 h)	Combined	Positive PV	Negative PV
>1.1	>0.035	OR	32.5%	88.9%
		AND	50.0%	92.0%
>1.1	>0.08	OR	32.5%	88.9%
		AND	85.7%	94.3%
>2.0	>0.035	OR	40.6%	94.1%
		AND	57.9%	90.0%
>2.0	>0.08	OR	46.4%	95.2%
		AND	84.6%	91.7%

The results in bold correspond to the combination of thresholds offering the best predictive performance that we select in this work.

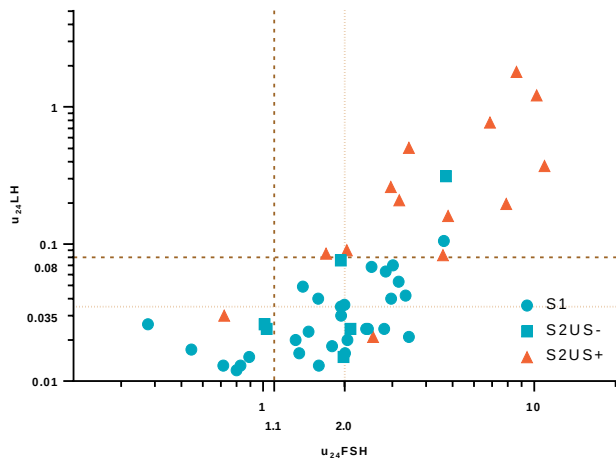


Figure 4. Predictive performance of 24-h urinary gonadotropins in the validation cohort. $u_{24}FSH$ and $u_{24}LH$ are represented in a decimal logarithmic scale. Bold dashed line stands for finally retained thresholds (1.1 IU/24 h, 0.08 IU/24 h) and slight dashed line stand for alternative thresholds evaluated (2.0 IU/24 h, 0.035 IU/24 h). FSH, follicle-stimulating hormone; LH, luteinizing hormone; S1, Breast Tanner Stage 1; S2US-, Breast Tanner stage 2 or more without ultrasonographic morphological changes; S2US+, Breast Tanner stage 2 or more with ultrasonographic morphological changes.

gonadotropins could be able to supplement for the lack of sensitivity of the GnRH test. The response of the axis to GnRH pulses depends on sex hormone effects, involving activation of kisspeptin neurons (Cortés et al., 2015; Hu et al., 2022). In some situations, such as obesity, a highly emerging issue in paediatrics, a non-dynamic assessment of the rise of gonadotropin appears more appropriate than a GnRH test given the effect of leptin on kisspeptin neurons (Itriyeve, 2022). This statement should first be confirmed in a comparative study.

Several studies have proposed various protocols based on simple emptying of the bladder before bedtime and collecting in the morning, accurate notification of the time between the last emptying and collection, or standardization of the hour of collection, at one or several times (Kolby et al., 2017; Shim et al., 2019; Lee et al., 2021; Demir et al., 2023). Currently, no study has compared these methods in order to show a potential superiority. Yet the duration and intensity of gonadotropin pulsed are variable and their time of occurrence can be delayed in children with disorders of the sleep rhythm. Moreover, the studies of Wide et al. (2010, 2022) reported that diverse sialylated forms of glycan in physiological proportions of LH and FSH impacted their elimination clearance by basically doubling. As these forms exist in varying proportions throughout life, we think that 24-h urine

collection can present interesting benefits to ensure that gonadotropin pulses are fully collected and delayed clearance are covered. The present protocol ensures that the complete circadian excretion is similarly measured for all subjects. Comparison with morning or random micturition would be of high interest. Previous studies have agreed that there is no benefit in calculating the urinary gonadotropins/urinary creatinine ratio; this does not refine correlation of values with breast Tanner stages (Zhan et al., 2021), and even impairs it (Demir et al., 2021). Moreover, Singh et al. (2015) showed in a large cohort of teenagers that an adjustment of steroid or urinary LH assays on hydration status (osmolality, specific gravity, urine creatinine) did not confer an improved relevance of the results. It was therefore chosen herein not to adjust urinary FSH and LH on creatinine.

The current study involved a first large and multicenter setting cohort to evaluate intervals and identify thresholds, and a much more restrictive single-centre validating cohort to validate the diagnostic performances of these thresholds. Good correlations of u_{24} gonadotropins with LH peaks and LH/FSH peak ratios during GnRH test, as well as with basal plasma FSH and LH levels were found. In contrast, the FSH peak after GnRH test did not correlate with any of $u_{24}FSH$ or $u_{24}LH$. A recent study showed very similar results on randomly collected urinary samples (Lee et al., 2021). Moreover, the correlation coefficients found herein were much more comparable with previous data (Demir et al., 2016; Lucaccioni et al., 2016; Kolby et al., 2017), and higher than some others (Lee et al., 2021; Zhan et al., 2021).

The retained thresholds determined in the setting cohort considered the inter-individual variability of presented phenotypes, as well as that of $u_{24}FSH$ and $u_{24}LH$ in non-pubertal girls, which is clearly enlightened in this work, particularly for $u_{24}FSH$. Yet we think that there is a possibility that the high number of paediatrician practitioners may have caused some heterogeneous over-classification of Tanner breast staging, likely to falsely downshift $u_{24}FSH$ and $u_{24}LH$ values distributions in the S2 group. This would explain the unexpected amount of $u_{24}LH$ below LLOD in S2 group, which can be explained by adipomastia or premature thelarche. Even though some patients with interfering conditions may not have been fully excluded, the number of values support the relevance of these data. We only noted one high $u_{24}FSH$ point in S1 group, which was likely to be an outlier with probable missed diagnosis of ovarian insufficiency. Two threshold values respectively for $u_{24}FSH$ and $u_{24}LH$ thresholds were chosen to be validated. The use of the YI assumes that even if CPP is a rare condition, it is not so rare in a population of girls addressed by paediatricians for clinical suspicions, and a balance between the risks of false positives or negatives is difficult to assess.

The validation cohort allowed us to confirm the usefulness of 24-h urinary gonadotropin levels. It was constructed in a sole expert centre with homogeneously qualified and restricted community of paediatrician endocrinologists and radiologists, in which a bias in Tanner staging and ultrasound examination is highly unlikely. Besides, no more than three patients among 237 were excluded because of doubtful Tanner breast staging in the entire validation cohort constitution. The pre-pubertal group was thus constituted by including S2US- girls and S1 girls which, allowing a good distinction between premature thelarche and CPP, which our urinary thresholds were able to reproduce. This study showed that $u_{24}FSH > 1.1$ IU/24 h combined with $u_{24}LH > 0.08$ IU/24 h was predictive of girls with breast development S2 or more associated with US assessment of pubertal maturation of the internal genitalia (pubertal group). Based on available progression data, one pre-pubertal S1 girl had elevated urinary gonadotropins

but was still pre-pubertal at follow-up. The interpretation of urinary assays was very consistent with clinical conclusion, as only S2 (or more) girls for whom urinary gonadotropin levels were greater than the thresholds were diagnosed with CPP and received puberty blockers. Two S1 girls showed slight signs of an onset of morphological evolution (S1US+), and although their $u_{24}FSH$ were $>1.1 IU/24 h$, $u_{24}LH$ was quite low ($<0.035 IU/24 h$ or undetectable). It is difficult to conclude whether these two girls were at a very early step of puberty initiation, in which case, internal genitalia morphologic changes would be visible before breast development. Without a detectable rise of urinary LH, this would be congruent with the limited lack of sensitivity of the $u_{24}LH$ threshold. Nevertheless, it is possible that some girls having $u_{24}LH$ comprised between 0.035 and 0.08 IU/24 h may be in a 'middle-zone' of awakening the HPG axis prior to any clinical or US signs. This should be elucidated in a prospective study. On the other hand, three S2US+ had only slightly elevated $u_{24}LH$ higher than 0.08 (0.083, 0.085, 0.09 IU/24 h). According to coefficient of variation at their raw levels (IU/l), the repeat of the assay would presumably have not classified them differently. Yet the inherent limit of a threshold must be considered, and a control collection should be advised in case of any borderline results. No patients had $u_{24}FSH < 1.1 IU/24 h$ with $u_{24}LH > 0.08 IU/24 h$, which appears to be a paradoxical hypothetical situation, that may evoke pre-analytical or analytical errors.

The present results lead to support the idea that 24-h urinary gonadotropin measurements have their place in the diagnosis of CPP onset, regarding the predictive performances presented herein. Studies from last decade reporting data on FMV urines or random urines have shown good usefulness, with sensitivity from 65% to 92% and specificity from 63% to 100% (Demir et al., 2021; Zhan et al., 2021). A meta-analysis from Xu et al. (2022) that retained six studies, in spite of heterogeneous collecting protocols (collection protocols, sample treatment prior to analysis) and analytical techniques, revealed an overall sensitivity of 79% and specificity of 84%. Comparisons of diagnostic thresholds from different studies must consider these pre-analytical and assay method differences. It must, however, be considered that the proposed and validated thresholds from this study can only apply to an Abbott Laboratories® CMIA assay performed directly on non-extracted 24-h urine; measurement without extraction using other reagents must be evaluated due to a likely high heterogeneity. This automated immunoassay kit for gonadotropins in urine is highly correlated with RIA techniques (Supplementary Fig. S1), which had demonstrated their value in clinical practice in the past (Morel et al., 1985).

However, this study has some limitations. First, related to its retrospective design, further prospective follow-up studies should be performed to describe the chronology of the increase of 24-h urinary gonadotropins and that of clinical pubertal development. Long-term clinical follow-up data on pubertal progression was only partly available for patients in both the prepubertal and the pubertal groups (27/39). Although there was a good association between urinary gonadotropin levels and the decision to treat or clinical progression in these patients, the retrospective nature of the study means that we cannot rule out the possibility that urinary gonadotropin levels may have influenced management. Also, there is a margin of error due to the inherent uncertainty of clinical assessment of pubertal onset. These limitations specific to clinical examination linked to the number of practitioners must be kept in mind, particularly in the setting cohort, in which the geographical scope of inclusion was wide. Finally, urine sampling, although requiring only basic

equipment, is sometimes subject to omission or involuntary urine loss by children, which can compromise the completeness of the collection. This is an easy sampling procedure, but imposes certain pre-analytical requirements. Assays can be easily repeated on new 24-h collections. The authors also point out that this study was made possible by the applicability of the Abbott assay kit to urine samples, and that this assay is not available to all centres treating these patients. As another strong point, this setting cohort was based on the inclusion of a large number of girls with no breast development (142 S1 girls with 191 urines); this provided us an original control group of girls with clinically non-initiated central puberty. One of the strengths of the validation cohort is that it was constructed using US criteria focused on the uterus and on the ovaries, and performed at the same time that the urinary collection. Moreover, the large number of values in the setting cohort is a strong argument for the reliability of the intervals.

In conclusion, the present results provide evidence of the reliability of 24-h urinary gonadotropins in the diagnosis of CPP in girls, as our method performs at least as well as the GnRH test and is well predictive of clinical and radiological pubertal onset. The herein presented $u_{24}FSH$ and $u_{24}LH$ thresholds must be used in further prospective studies aiming to compare with the GnRH test for diagnosis and treatment monitoring, as well as with FMV or random collection of urine samples.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

Data are available on request.

Acknowledgements

We thank Dr Régine Cartier, Dr Laurent Desmurs, Dr Caroline Richet, and Dr Fanny Zhao, from the Hospices Civils de Lyon, as well as the patients and their families for their collaboration.

Authors' roles

I.P. conceived study design, supervised, and led the study. C.J. conducted the data analysis and formatting, devised the scientific approach, and drafted the manuscript. P.P. and V.R. contributed to method development and provided biological expertise with F.R.-B. S.M.-T. and Z.R. participated in the data collection, analysis, and interpretation. M.N., C.V., P.B., and K.P. conducted and supervised the clinical management and provided medical expertise. R.E. contributed to the data interpretation. All authors contributed to the manuscript co-writing.

Funding

No specific funding was used.

Conflict of interest

All authors declared no conflict of interest.

References

- Ab Rahim SN, Omar J, Tuan Ismail TS. Gonadotropin-releasing hormone stimulation test and diagnostic cutoff in precocious puberty: a mini review. *Ann Pediatr Endocrinol Metab* 2020;**25**:152–155.
- Asavaoie C. Ovarian and uterine ultrasonography in pediatric patients. Pictorial essay. *Med Ultrason* 2014;**16**:160–167.
- Bangalore Krishna K, Fuqua JS, Rogol AD, Klein KO, Popovic J, Houk CP, Charmandari E, Lee PA, Freire AV, Ropelato MG et al. Use of gonadotropin-releasing hormone analogs in children: update by an International Consortium. *Horm Res Paediatr* 2019;**91**:357–372.
- Boyar R, Finkelstein J, Roffwarg H, Kapen S, Weitzman E, Hellman L. Synchronization of augmented luteinizing hormone secretion with sleep during puberty. *N Engl J Med* 1972;**287**:582–586.
- Brito VN, Spinola-Castro AM, Kochi C, Kopacek C, Silva PCAD, Guerra-Júnior G. Central precocious puberty: revisiting the diagnosis and therapeutic management. *Arch Endocrinol Metab* 2016;**60**:163–172.
- Buckler JMH, Clayton BE. Output of luteinizing hormone in the urine of normal children and those with advanced sexual development. *Arch Dis Child* 1970;**45**:478–484.
- Cantas-Orsdemir S, Garb JL, Allen HF. Prevalence of cranial MRI findings in girls with central precocious puberty: a systematic review and meta-analysis. *J Pediatr Endocrinol Metab* 2018;**31**:701–710.
- Carel J-C, Eugster EA, Rogol A, Ghizzoni L, Palmert MR, Antoniazzi F, Berenbaum S, Bourguignon J-P, Chrousos GP, Coste J et al.; Members of the ESPE-LWPES GnRH Analogs Consensus Conference Group. Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics* 2009;**123**:e752–e762.
- Carel J-C, Léger J. Precocious puberty. *N Engl J Med* 2008;**358**:2366–2377.
- Chipman JJ, Moore RJ, Marks JF, Fevre M, Segel T, Ramsey J, Boyar RM. Interrelationship of plasma and urinary gonadotropins: correlations for 24 hours, for sleep/wake periods, and for 3 hours after luteinizing hormone-releasing hormone stimulation. *J Clin Endocrinol Metab* 1981;**52**:225–230.
- Cortés ME, Carrera B, Riosco H, Pablo Del Río J, Vigil P. The role of kisspeptin in the onset of puberty and in the ovulatory mechanism: a mini-review. *J Pediatr Adolesc Gynecol* 2015;**28**:286–291.
- De Vries L, Horev G, Schwartz M, Phillip M. Ultrasonographic and clinical parameters for early differentiation between precocious puberty and premature thelarche. *Eur J Endocrinol* 2006;**154**:891–898.
- Demir A, Aydın A, Büyükgebiz A, Stenman U-H, Hero M. Urinary gonadotropin measurements by enhanced luminometric assays (LIA) for the evaluation of pubertal development. *J Pediatr Endocrinol Metab* 2021;**34**:859–866.
- Demir A, Büyükgebiz A, Aydın A, Hero M. Quantification of overnight urinary gonadotropin excretion predicts imminent puberty in girls: a semi-longitudinal study. *Hormones (Athens)* 2023;**23**:141–150. [Internet] Available from: <https://link.springer.com/10.1007/s42000-023-00499-7>.
- Demir A, Voutilainen R, Juul A, Dunkel L, Alfthan H, Skakkebaek NE, Stenman UH. Increase in first morning voided urinary luteinizing hormone levels precedes the physical onset of puberty. *J Clin Endocrinol Metab* 1996;**81**:2963–2967.
- Demir A, Voutilainen R, Stenman U-H, Dunkel L, Albertsson-Wikland K, Norjavaara E. First morning voided urinary gonadotropin measurements as an alternative to the GnRH test. *Horm Res Paediatr* 2016;**85**:301–308.
- Eckert-Lind C, Busch AS, Petersen JH, Biro FM, Butler G, Bräuner EV, Juul A. Worldwide secular trends in age at pubertal onset assessed by breast development among girls: a systematic review and meta-analysis. *JAMA Pediatr* 2020;**174**:e195881.
- Hu K-L, Chen Z, Li X, Cai E, Yang H, Chen Y, Wang C, Ju L, Deng W, Mu L. Advances in clinical applications of kisspeptin-GnRH pathway in female reproduction. *Reprod Biol Endocrinol* 2022;**20**:81.
- Huynh QTV, Le NQK, Huang S-Y, Ho BT, Vu TH, Pham HTM, Pham AL, Hou J-W, Nguyen NTK, Chen YC. Development and validation of clinical diagnostic model for girls with central precocious puberty: machine-learning approaches. *PLoS One* 2022;**17**:e0261965.
- Itriyeve K. The effects of obesity on the menstrual cycle. *Curr Probl Pediatr Adolesc Health Care* 2022;**52**:101241.
- Kandemir N, Demirbilek H, Özön ZA, Gönç N, Alikasıfoğlu A. GnRH stimulation test in precocious puberty: single sample is adequate for diagnosis and dose adjustment. *J Clin Res Pediatr Endocrinol* 2011;**3**:12–17.
- Kastin AJ, Schally AV, Gual C, Arimura A. Release of LH and FSH after administration of synthetic LH-releasing hormone. *J Clin Endocrinol Metab* 1972a;**34**:753–756.
- Kastin AJ, Schally AV, Schalch DS, Korenman SG, Miller MC, Gual C, Perez-Pasten E. Characterization of the hormonal responses to luteinizing hormone-releasing hormone (LH-RH) in prepubertal and adult subjects. *Pediatr Res* 1972b;**6**:481–486.
- Kolby N, Busch AS, Aksglaede L, Sørensen K, Petersen JH, Andersson A-M, Juul A. Nocturnal urinary excretion of FSH and LH in children and adolescents with normal and early puberty. *J Clin Endocrinol Metab* 2017;**102**:3830–3838.
- Lee SY, Kim JM, Kim YM, Lim HH. Single random measurement of urinary gonadotropin concentration for screening and monitoring girls with central precocious puberty. *Ann Pediatr Endocrinol Metab* 2021;**26**:178–184.
- Lucaccioni L, McNeilly J, Mason A, Giacomozzi C, Kyriakou A, Shaikh MG, Iughetti L, Ahmed SF. The measurement of urinary gonadotropins for assessment and management of pubertal disorder. *Hormones (Athens)* 2016;**15**:377–384. <http://www.hormones.gr/8646/article/the-measurement-of-urinary-gonadotropins-for%E2%80%A6.html>
- Maesaka H, Suwa S, Tachibana K, Kikuchi N. Quantitation of urinary gonadotropins in normal children. *Pediatr Res* 1990;**28**:401–404.
- Maione L, Bouvattier C, Kaiser UB. Central precocious puberty: recent advances in understanding the aetiology and in the clinical approach. *Clin Endocrinol (Oxf)* 2021;**95**:542–555.
- Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child* 1969;**44**:291–303.
- McNeilly JD, Mason A, Khanna S, Galloway PJ, Ahmed SF. Urinary gonadotrophins: a useful non-invasive marker of activation of the hypothalamic pituitary-gonadal axis. *Int J Pediatr Endocrinol* 2012;**2012**:10.
- Morel Y, La Selve H, Chatelain P, Perez J, Varenne JP, de Peretti E, Forest MG, David M, Bertrand J. Value of the assay of urinary gonadotropins in pediatric endocrinology. *Arch Fr Pediatr* 1985;**42** (Suppl 1):579–585.
- Morris NM, Udry JR. Validation of a self-administered instrument to assess stage of adolescent development. *J Youth Adolesc* 1980;**9**:271–280.
- Rifkind AB, Kulin HE, Rayford PL, Cargille CM, Ross GT. 24-hour urinary luteinizing hormone (LH) and follicle stimulating hormone (FSH) excretion in normal children. *J Clin Endocrinol Metab* 1970;**31**:517–525.

- Roberts SA, Kaiser UB. Genetics in endocrinology: genetic etiologies of central precocious puberty and the role of imprinted genes. *Eur J Endocrinol* 2020;**183**:R107–R117.
- Sathasivam A, Rosenberg HK, Shapiro S, Wang H, Rapaport R. Pelvic ultrasonography in the evaluation of central precocious puberty: comparison with leuprolide stimulation test. *J Pediatr* 2011;**159**:490–495.
- Shim YS, An SH, Lee HJ, Kang MJ, Yang S, Hwang IT. Random urinary gonadotropins as a useful initial test for girls with central precocious puberty. *Endocr J* 2019;**66**:891–903.
- Singh GK, Balzer BW, Desai R, Jimenez M, Steinbeck KS, Handelsman DJ. Requirement for specific gravity and creatinine adjustments for urinary steroids and luteinizing hormone concentrations in adolescents. *Ann Clin Biochem* 2015;**52**:665–671.
- Stranzinger E, Strouse PJ. Ultrasound of the pediatric female pelvis. *Semin Ultrasound CT MR* 2008;**29**:98–113.
- Wide L, Eriksson K, Sluss PM, Hall JE. The common genetic variant of luteinizing hormone has a longer serum half-life than the wild type in heterozygous women. *J Clin Endocrinol Metab* 2010;**95**:383–389.
- Wide L, Eriksson K, Sluss PM, Hall JE. Determination of half-lives of circulating FSH and LH glycoforms in women during GnRH receptor blockade. *J Clin Endocrinol Metab* 2022;**107**:e4058–e4062.
- Xu D, Zhou X, Wang J, Cao X, Liu T. The value of urinary gonadotropins in the diagnosis of central precocious puberty: a meta-analysis. *BMC Pediatr* 2022;**22**:453.
- Yazdani P, Lin Y, Raman V, Haymond M. A single sample GnRHa stimulation test in the diagnosis of precocious puberty. *Int J Pediatr Endocrinol* 2012;**2012**:23.
- Yüce Ö, Bideci A, Çelik N, Çamurdan O, Cinaz P. Diagnostic value of urinary luteinizing hormone levels in the monitoring of precocious puberty treatment. *Arch Endocrinol Metab* 2020;**64**:121–127. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2359-39972020005001207&lng=en&nrm=iso.
- Zhan S, Huang K, Wu W, Zhang D, Liu A, Dorazio RM, Shi J, Ullah R, Zhang L, Wang J et al. The use of morning urinary gonadotropins and sex hormones in the management of early puberty in Chinese girls. *J Clin Endocrinol Metab* 2021;**106**:e4520–e4530.
- Zung A, Burundukov E, Ulman M, Glaser T, Rosenberg M, Chen M, Zadik Z. The diagnostic value of first-voided urinary LH compared with GNRH-stimulated gonadotropins in differentiating slowly progressive from rapidly progressive precocious puberty in girls. *Eur J Endocrinol* 2014;**170**:749–758.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Human Reproduction, 2024, 39, 1003–1012
<https://doi.org/10.1093/humrep/deae055>
Original Article

Puberty, aging and HRT

Bone health in women with premature ovarian insufficiency/early menopause: a 23-year longitudinal analysis

A.R. Jones ^{1,2,*}, J. Enticott¹, P.R. Ebeling^{2,3}, G.D. Mishra ⁴, H.T. Teede ^{1,2}, and A.J. Vincent^{1,2}

¹Monash Centre for Health Research and Implementation, Monash University, Melbourne, VIC, Australia

²Department of Endocrinology, Monash Health, Melbourne, VIC, Australia

³Department of Medicine, School of Clinical Sciences, Monash University, Melbourne, VIC, Australia

⁴Australian Women and Girls' Health Research Centre, School of Public Health, University of Queensland, Brisbane, QLD, Australia

*Correspondence address. Monash University Faculty of Medicine Nursing and Health Sciences, Monash Centre for Health Research and Implementation, Locked Bag 29, Clayton, VIC 3800, Australia. E-mail: alicia.jones@monash.edu  <https://orcid.org/0000-0001-8419-7987>

ABSTRACT

STUDY QUESTION: What is the frequency of, and predictors for, osteoporosis, fractures, and osteoporosis management (investigation, treatment) in women with premature ovarian insufficiency (POI; menopause <40 years) and early menopause (EM; menopause 40–44 years)?

SUMMARY ANSWER: Over the 23-year follow-up duration, at a mean age of 68 years, women with POI/EM had higher osteoporosis/fracture risk and prevalence, higher osteoporosis screening and anti-osteoporosis medication use compared to women with usual age menopause; increasing age was predictive of increased risk of osteoporosis/fracture and menopause hormone therapy (MHT) prior to or at study entry (aged 45–50 years) was protective.

WHAT IS KNOWN ALREADY: Women with POI/EM have increased risk of osteoporosis and fractures with limited data regarding risk factors for reduced bone density and fractures. Clinical guidelines recommend screening with dual X-ray absorptiometry (DXA) and treatment with MHT for most women with POI/EM to reduce osteoporosis and fracture risk; however, studies indicate gaps in osteoporosis knowledge, guideline uptake, and management adherence by clinicians and women.

STUDY DESIGN, SIZE, DURATION: The Australian Longitudinal Study on Women's Health is a prospective longitudinal study of Australian women. This study uses the cohort of women born between 1946 and 1951, surveyed nine times between 1996 and 2019. Data from the Australian administrative health records, including hospital admissions data (fractures, osteoporosis), Medicare Benefits Schedule (DXA), and the Pharmaceutical Benefits Scheme (PBS; MHT, anti-osteoporosis medication, available only from 2002) were linked to survey data.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Survey respondents with self-reported age of menopause were included. POI/EM was defined as menopause <45 years. T-test or chi-square were used for comparisons at baseline ($P < 0.05$ indicates significance). Generalized estimating equations for panel data explored predictors for the longitudinal outcomes of osteoporosis, fractures, DXA rates, MHT use, and anti-osteoporosis medication (in women with osteoporosis/fracture, from Survey 4 onwards only). Univariable regression was performed, and variables retained where $P < 0.2$, to form the multivariable model, and bootstrapping with 100 repetitions at 95% sampling of the original dataset to ensure robustness of results.

MAIN RESULTS AND THE ROLE OF CHANCE: Eight thousand six hundred and three women were included: 610 (7.1%) with POI/EM. Mean (SD) baseline age was 47.6 (1.45) years in the entire cohort and mean (SD) age of menopause was 38.2 (7.95) and 51.3 (3.04) years in women with POI/EM and usual age menopause, respectively ($P < 0.001$). Over the 23 years, of women with POI/EM, 303 (49.7%) had osteoporosis/fractures, 421 (69.0%) had DXA screening, 474 ever used MHT (77.7%), and 116 (39.1%) of those with osteoporosis/fractures used anti-osteoporosis medication. Of women with usual age menopause, 2929 (36.6%) had osteoporosis/fractures, 4920 (61.6%) had DXA screening, 4014 (50.2%) used MHT, and 964 (33.0%) of those with osteoporosis/fractures used anti-osteoporosis medication. Compared to women with menopause at age ≥ 45 years and after adjusting for other risk factors, women with POI/EM had increased risk of osteoporosis (odds ratio [OR] 1.37; 95% CI 1.07–1.77), fractures (OR 1.45; 1.15–1.81), DXA testing (OR 1.64; 1.42–1.90), MHT use (OR 6.87; 5.68–8.30), and anti-osteoporosis medication use (OR 1.50; 1.14–1.98). In women with POI/EM women, increasing age was associated with greater risk of osteoporosis/fracture (OR 1.09; 1.08–1.11), and MHT prior to or at study entry (aged 45–50 years), was protective (OR 0.65, 0.45–0.96). In women with POI/EM, age (OR 1.11; 1.10–1.12), fractures (OR 1.80, 1.38–2.34), current smoking (OR 0.60; 0.43–0.86), and inner (OR 0.68; 0.53–0.88) or outer regional (OR 0.63; 0.46–0.87) residential location were associated with DXA screening. In women with POI/EM, increasing age (OR 1.02; 1.01–1.02), and currently consuming alcohol (OR 1.17; 1.06–1.28), was associated with having ever used MHT. In the 299 women with POI/EM and osteoporosis/fractures, only 39.1% ever received treatment with an anti-osteoporosis medication. Increasing age (OR 1.07; 1.04–1.09) and lower BMI (OR 0.95; 0.92–0.98) were associated with greater likelihood of treatment with anti-osteoporosis medication.

Received: October 9, 2023. Revised: January 22, 2024. Editorial decision: February 6, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

LIMITATIONS, REASONS FOR CAUTION: Survey data including age of menopause were self-reported by participants; fracture questions were not included in the 2001 survey, and location or level of trauma of self-reported fractures was not asked. Additional risk/protective factors such as vitamin D status, calcium intake, and exercise were not able to be included. Due to sample size, POI and EM were combined for all analyses, and we were unable to differentiate between causes of POI/EM. PBS data were only available from 2004, and hospital admissions data were state-based, with all of Australia were only available from 2007.

WIDER IMPLICATIONS OF THE FINDINGS: This study supports previous literature indicating increased risk of osteoporosis and fractures in women with POI, and adds evidence for women with POI/EM, where there was a relative paucity of data. This is the first study to analyse a variety of clinical and demographic risk factors for osteoporosis and fractures in women with POI/EM, as well as analysing investigation and treatment rates. In these women, using MHT prior to or at study entry, aged 45–50 years, was protective for osteoporosis/fractures; however, having ever used MHT was not, highlighting the importance of early treatment with MHT in these women to preserve bone strength. Although women with POI/EM and osteoporosis or fractures were more likely to use anti-osteoporosis medications than those with usual age menopause, overall treatment rates are low at <40%, demonstrating a significant treatment gap that should be addressed to reduce future fracture risk.

STUDY FUNDING/COMPETING INTEREST(S): This study was funded by The Australian NHMRC Centre of Research Excellence Women's Health in Reproductive Life (CRE-WHRL, project number APP1171592). A.R.J. is the recipient of a National Health and Medical Research Council post-graduate research scholarship (grant number 1169192). P.R.E. is supported by a National Health and Medical Research Council grant 1197958. P.R.E. reports grants paid to their institution from Amgen, Sanofi, and Alexion, honoraria from Amgen paid to their institution, and honoraria from Alexion and Kyowa-Kirin.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: premature ovarian insufficiency / early menopause / osteoporosis / fractures / longitudinal

Introduction

The usual age of menopause globally is between 48 and 52 years (InterLACE Study Team, 2019). Premature ovarian insufficiency (POI), loss of ovarian function occurring before the age of 40 years, affects around 2–4% of women, and early menopause (EM), menopause occurring between 40 and 45 years, affects up to 12% of women (Mishra et al., 2017; Golezar et al., 2019). EM can occur spontaneously, or secondary to medical or surgical therapies such as bilateral oophorectomy, radiotherapy, or chemotherapy (The ESHRE Guideline Group on POI et al., 2016).

Long-term health concerns associated with POI/EM include increase in cardiovascular disease, cognitive dysfunction, and reduced bone mineral density (BMD), as well as reduced overall quality of life and life expectancy, largely due to cardiovascular disease (Muka et al., 2016; The ESHRE Guideline Group on POI et al., 2016).

Peak bone mass is usually achieved during the third decade of life, and oestrogen deficiency during childhood to young adulthood can prevent optimal peak bone mass accrual (Biller et al., 1989; Weaver et al., 2016; Siegel et al., 2017). Bone mass remains relatively stable during mid-life, then, during the late perimenopause and postmenopausal periods, there is rapid loss of bone density, of around 2% per year (Rosenthal et al., 1989; Ahlborg et al., 2003; Finkelstein et al., 2008). Depending on the age of menopause, women with POI/EM may experience both lower peak bone mass, and earlier menopausal bone loss, predisposing to osteoporosis and fractures (Nguyen et al., 2021).

Studies consistently demonstrate lower BMD in women with POI, with a corresponding increase in fracture risk (Sullivan et al., 2015, 2017; Shea et al., 2021; Costa et al., 2023). Risk factors for lower BMD in women with POI include lower age of menopause, low vitamin D, low body weight, lack of exercise, delay in diagnosis, poor adherence to menopause hormone therapy (MHT), and low calcium intake (Popat et al., 2009). Despite affecting significantly more women than POI alone, there is a relative paucity of data on bone health in women with POI/EM, and studies examining osteoporosis and fracture risk in these women have had mixed results, with some showing increased risk (Hadjidakis et al., 2003; Paganini-Hill et al., 2005; Anagnostis et al., 2019) and others showing no difference (Papaioannou et al., 2005; Banks et al., 2009).

International guidelines recommend the use of MHT in women with POI/EM, until the usual age of menopause (The ESHRE Guideline Group on POI et al., 2016; The North American Menopause Society, 2022). MHT has been shown to improve BMD in these women, and, although evidence for fracture reduction in this population is unclear, data extrapolated from studies in postmenopausal women in general has shown significant fracture reduction (Burgos et al., 2017; Sullivan et al., 2017; Barrionuevo et al., 2019; Costa et al., 2023). However, data suggest that many women fail to receive appropriate MHT to protect their bones (Bachelot et al., 2016; Goh et al., 2019).

This study aimed to investigate the prevalence of osteoporosis and fractures in women with POI/EM and the risk factors for developing osteoporosis and fractures, as well as examining gaps in osteoporosis investigation and treatment in a country with universal healthcare and subsidized access to medication.

Materials and methods

Participants

The Australian Longitudinal Study on Women's Health (ALSWH) is an ongoing prospective study examining the health of Australian women. Full details of the study are described elsewhere (Brown et al., 1996). In brief, four birth cohorts of women were invited to participate, with random selection from the Australian government health insurance database (Medicare), which includes all Australian citizens and permanent residents, and purposeful oversampling of people living outside urban areas. Women complete surveys every 2–3 years. The current study uses data from the cohort of women born between 1946 and 1951, who completed nine surveys between 1996 and 2019.

Women were included from the first survey (baseline, 1996) if they provided an age of menopause in any survey. Age of menopause was self-reported from Survey 2 onwards. The first reported age of menopause was included. Women were excluded if no age of menopause was provided, if a hysterectomy was performed prior to a reported age of menopause, or if age of menopause reported was greater than their current age. POI/EM was defined as menopause <45 years, and usual age menopause was defined as menopause ≥45 years. POI and EM women were combined into one group to optimize sample size. POI/EM was confirmed by ensuring absence of menses in the 12 months prior to

the first survey, unless MHT or the combined oral contraceptive pill (COCP) was being used, as these questions were asked in the surveys. Bilateral oophorectomy was included within POI/EM if an age of surgery was provided. Given the age at first survey was 45–50 years, at baseline in 1996, all women with POI/EM had already gone through menopause.

Surveys were developed by the original study, and included demographic details, medical conditions, and self-reported biometry, which was used to calculate a BMI. Comorbidities known to influence osteoporosis were self-reported (hyperthyroidism, hyperparathyroidism, rheumatoid arthritis, malabsorption, inflammatory bowel disease, chronic liver disease, chronic kidney disease, diabetes, connective tissue disorders, any transplant). Residential location was defined using the Australian Statistical Geography Standard Remoteness Structure, categorized as major cities, inner regional, outer regional, remote and very remote areas, with remote and very remote locations combined due to small numbers (Australian Bureau of Statistics, July 2021–June 2026). Socioeconomic status was classified by the index of relative socioeconomic disadvantage (IRSD), which summarizes economic and social conditions of people by area and includes factors such as income and education, with lower values indicating greater disadvantage. Surveys are freely available online (Australian Longitudinal Study on Women's Health, 2022).

The survey data have been linked to administrative health datasets, including the Pharmaceutical Benefits Scheme (PBS, Australian government-subsidized medications), Medicare Benefits Schedule (MBS, Australian government-subsidized medical services), and hospital separation data. Linked PBS data are available from 2002 onwards, MBS data are available from 1996 onwards, and availability of hospital separation data differs by state, with all states data included from 2007.

Outcomes

The primary outcomes were to determine the proportion of women with POI/EM with osteoporosis and fractures and to identify predictors for developing osteoporosis/fractures. Secondary outcomes were the proportion of women with POI/EM undergoing osteoporosis investigation using dual X-ray absorptiometry (DXA), MHT use, and treatment with anti-osteoporosis medications. Secondary analyses aimed to identify predictors for these. Outcomes in women with POI/EM were identified and then compared with women with usual age menopause (age >45 years).

Osteoporosis diagnosis was collected from hospital separation data using ICD-9 and -10 codes for primary or secondary diagnoses (Supplementary Table S1) and was also self-reported in all surveys. The cumulative prevalence of osteoporosis is reported, as osteoporosis was primarily listed as a comorbidity in hospital separation data, and so incidence of osteoporosis could not be included.

Fractures were collected from hospital separation data from 1995 onwards, using ICD codes (Supplementary Table S1), and excluded fractures of the skull, hands, and feet. Fractures were also self-reported in all surveys except Survey 3 (2001). Self-reported fractures did not specify the location or number of fractures, and only reported incident fractures for the past 12 months. For this reason, fractures are reported as the cumulative prevalence of having had a fracture.

DXA rates were taken from MBS data, and self-reported in surveys from 2007 onwards. DXA rates are reported as the number who completed a DXA per survey period.

MHT use was collected from PBS data from 2002 onwards, and included transdermal or oral oestradiol, conjugated equine oestrogen ± medroxyprogesterone acetate, transdermal or oral

oestradiol/norethisterone, oestrone, oestradiol/cyproterone, oestradiol hemihydrate, oestradiol/dydrogesterone, and the COCP, but excluding vaginal oestrogen preparations. MHT was also self-reported by participants from Survey 1 (1996) onwards. MHT treatment is reported as 'current use of MHT' (per survey period, use of any MHT or COCP since the previous survey); and 'cumulative use of MHT' (current or previous MHT at any time point, and COCP use from the Survey 1 onwards). We did not include self-reported COCP use prior to Survey 1, as the time of use was not provided, and the focus of this study was oestrogen for bone protection, rather than contraception.

Anti-osteoporosis medication was examined only in women who had either osteoporosis or a fracture and was available only from Survey 4 onwards. Treatment included all PBS-listed osteoporosis therapy (alendronate, risedronate, etidronate (together grouped as oral bisphosphonates), zoledronic acid, denosumab, raloxifene, strontium, teriparatide, and calcitriol). Treatment is reported as cumulative use at any time point. The number of prescriptions was used to determine duration of use and compliance. Compliance was defined as $\geq 80\%$ medication possession index, except for denosumab and zoledronic acid. The antiresorptive effect of denosumab is rapidly reversed after missing or delaying a dose, and so compliance was defined as a subsequent dose within 7 months of the previous dose (Bone et al., 2011). In contrast, zoledronic acid has a long half-life with prolonged efficacy, and so compliance was defined as subsequent dose within 18 months (Reid et al., 2018).

Statistical methods

Categorical variables were expressed as number (%), and categorical variables at baseline were compared between women with and without POI/EM using chi-square tests. Continuous variables were expressed as mean (SD) when normally distributed, and median (quartile 1, quartile 3) when not normally distributed. Continuous variables at baseline were compared between women with and without POI/EM using t-tests if normally distributed, and Mann–Whitney *U* test if not normally distributed. At each survey, the outcomes of osteoporosis, fractures, DXA, MHT, and anti-osteoporosis medications were compared between women with and without POI/EM using chi-square tests.

Generalized estimating equations for panel data were used to analyse predictive factors for the longitudinal outcomes of osteoporosis/fractures, DXA, MHT, and anti-osteoporosis medication in women with POI/EM. Independent variables included factors known to influence the risk of osteoporosis. The IRSD was included in multivariable models in place of education or marital status as it includes these factors. Interaction testing between residential location and IRSD was performed due to a theoretical relationship and found to be significant for the outcomes of osteoporosis, DXA and treatment, and so residential location only was included in these models. Univariable regression was performed, with variables retained at a significance level of $P < 0.2$, to form the final multivariable model. Bootstrapping was performed, with 100 repetitions at 95% sampling of the original dataset, to ensure robust results. All *P*-values were calculated from two-tailed tests of statistical significance, with $P < 0.05$ considered significant. Analyses were conducted using Stata software, v15 (StataCorp, Texas, USA).

Ethical approval

Ethical approval was obtained from the University of Newcastle HREC (EC00144), the University of Queensland HREC (EC00456/7), the Australian Institute of Health and Welfare HREC (EC00103), the NSW Population and Health Services Research Ethics

Committee (EC00410, this included a Victorian-specific module to address Victorian legislative requirements), the ACT Health HREC (EC00100), the Department of Health Western Australia HREC (EC00422), and Tasmanian Health and Medical HREC (EC00337).

Results

Of the 13 714 women included in the original study in 1996, 8603 women reported an age of menopause and were included in the current study, including 610 (7.1%) with POI/EM (223 with POI) (Fig. 1). Baseline characteristics of women are listed in Table 1, the mean (SD) age at baseline was 47.6 (1.45) years, and the mean age at last follow-up was 67.8 (5.45) years. The mean (SD) age of menopause was 38.2 (7.95) and 51.3 (3.04) years in women with POI/EM and usual age menopause, respectively ($P < 0.001$).

Age, educational attainment, socioeconomic disadvantage, marital status, and smoking status differed between women with POI/EM and usual age menopause. At baseline, 432 (70.8%) of women with POI/EM were either currently using, or previously used MHT, compared to 1977 (24.7%) of women with usual age menopause. At baseline, 161 (1.9%) of women had bilateral oophorectomy, including 122 women with POI/EM, and 39 with usual age menopause; 149 women had a history of breast cancer (24 with POI/EM) and 337 had a history of endometrial, ovarian, or cervical cancer (39 with POI/EM). Comorbidities associated with an increased risk of osteoporosis, experienced over the 23-year follow-up period, varied between POI/EM and usual age menopause women (Table 2).

Osteoporosis and fractures

Overall, 1936 (22.5%) of women had a diagnosis of osteoporosis during follow-up, including 191 (31.3%) of women with POI/EM and 1745 (21.8%) of women with usual age menopause. A fracture was recorded in 2015 (23.4%) women, including 190 (31.2%) of women with POI/EM and 1825 (22.8%) of women with usual age menopause. However, only 719 (35.7%) of those with a

fracture had a known diagnosis of osteoporosis. Of those with POI/EM, 303 (49.7%) had either osteoporosis or a fracture, compared to 2929 (36.6%) of women with usual age menopause. Of the 842 fractures occurring in 738 women taken from hospital separation data, 500 of these were classified as major osteoporotic fractures (hip, spine, forearm, humerus), including 73 hip fractures in 69 women (10 women with POI/EM and 59 women with usual age menopause).

Osteoporosis and fracture prevalence was higher in women with POI/EM compared to women with usual age menopause at all surveys (Fig. 2A). Univariable and multivariable longitudinal analysis of the entire cohort found that women with POI/EM had significantly higher risk of osteoporosis (odds ratio [OR] 1.37, 95% CI 1.07–1.77) and fractures (OR 1.45, 95% CI 1.15–1.81), than women with usual age menopause (Supplementary Tables S2 and S3). In the entire cohort, increasing age, lower BMI, increasing number of comorbidities, and having had a DXA were also associated with greater likelihood of having osteoporosis, while current MHT use was protective (Supplementary Table S2). Increasing age and increasing number of comorbidities was associated with greater likelihood of having had a fracture, and current MHT use was protective (Supplementary Table S3). Bootstrapping did not alter results.

For women with POI/EM, univariable and multivariable longitudinal analysis indicated that increasing age was associated with an increased risk of osteoporosis/fracture (OR 1.09; 95% CI 1.08–1.11) (Table 3). MHT use prior to or at baseline, in 1996, but not having ever used MHT or current use of MHT, was protective for osteoporosis/fractures (OR 0.65; 0.45–0.96). BMI, smoking, alcohol, and the number of comorbidities were associated with osteoporosis/fractures on univariable analysis, but did not remain significant in the multivariable analysis. Bootstrapping did not alter results.

Investigation

Over the 23 years, 5341 (62.1%) of all women had at least one DXA (self-reported or MBS), including 421 (69.0%) of those with POI/EM and 4920 (61.6%) of those with usual age menopause. Of the 3886 who has an MBS DXA recorded, 1925 had a single DXA over the 23-year period, 839 had two DXA, and 1122 had >2 DXA done. Use of DXA was significantly higher in women with POI/EM compared with women with usual age menopause at all surveys (Fig. 2B), and longitudinal analysis also found POI/EM was associated with significantly higher likelihood of having a DXA (OR 1.64; 95% CI 1.42–1.90, Supplementary Table S4). Increasing age, lower BMI, increasing number of comorbidities, and previous fractures were also associated with a greater likelihood of having a DXA, while current smoking and living outside a major city were associated with reduced likelihood of DXA.

Of women with POI/EM, increasing age (OR 1.11; 1.10–1.12) and previous fracture (OR 1.80; 1.38, 2.34) were associated with greater likelihood of having a DXA (Supplementary Table S5). Living in an inner (OR 0.68; 0.53–0.88) or outer (0.63; 0.458–0.87) regional area, and current smoking (OR 0.60; 0.43–0.86) were associated with less likelihood of DXA. Increasing number of comorbidities increased the likelihood of DXA on univariable analysis, but after adjustment was no longer significant. Bootstrapping did not alter results.

Menopause hormone therapy

MHT was used by 4488 (52.2%) of women before or during the study, including 474 (77.7%) of women with POI/EM and 4014 (50.2%) of women with usual age menopause. At baseline, 432/610 (70.8%) of women with POI/EM currently or previously used

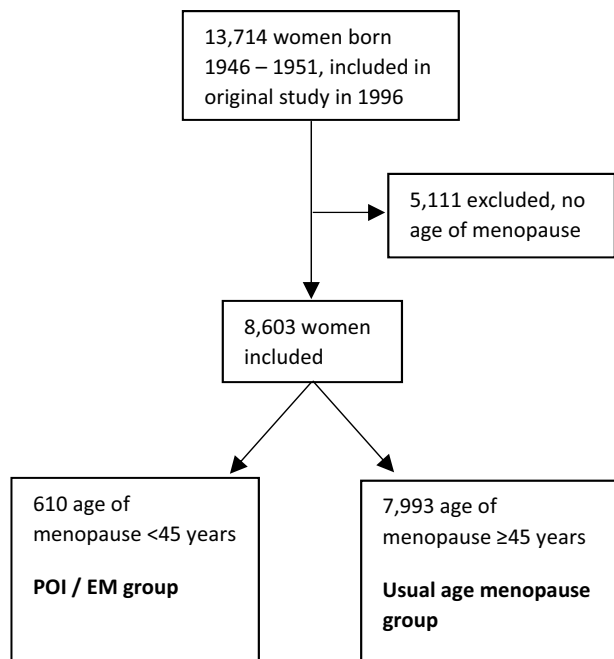


Figure 1. Flowchart of participant inclusion. POI: premature ovarian insufficiency; EM: early menopause.

Table 1. Baseline characteristics of included women.

		Premature ovarian insufficiency/early menopause n = 610	Usual age menopause n = 7993	Total n = 8603	P
Age (years; mean ± SD)		47.7 ± 1.4	47.6 ± 1.5	47.6 ± 1.5	0.03
BMI (kg/m ² ; mean ± SD)		25.5 ± 4.8	25.5 ± 4.9	25.5 ± 4.9	0.96
Education (n, %)	<High school	327 (54.5)	3583 (45.2)	3910 (45.8)	<0.001
	High school completion	109 (18.2)	1379 (17.4)	1488 (17.4)	
	Trade/apprentice/certificate/diploma	108 (18.0)	1614 (20.4)	1722 (20.2)	
	University/higher degree	56 (9.3)	1356 (17.1)	1412 (16.6)	
Marital status (n, %)	Partnered	467 (77.1)	6634 (83.4)	7101 (82.9)	<0.001
	Non-partnered	139 (22.9)	1323 (16.6)	1462 (17.1)	
Country of birth (n, %)	Australia	450 (74.9)	6052 (76.5)	6502 (76.4)	0.36
	Outside Australia	151 (25.1)	1858 (23.5)	2009 (23.6)	
IRSD (mean ± SD)		987.4 ± 94.5	1005 ± 88.2	1004.1 ± 88.8	<0.001
Residential location (n, %)	Major city	225 (36.9)	2920 (36.5)	3145 (36.6)	0.69
	Inner regional	227 (37.2)	3097 (38.8)	3324 (38.6)	
	Outer regional	128 (21.0)	1604 (20.1)	1732 (20.1)	
	Remote/very remote	30 (4.9)	372 (4.7)	402 (4.7)	
Smoking status (n, %)	Never	241 (40.8)	4262 (55.0)	4503 (54.0)	<0.001
	Former	156 (26.4)	2270 (29.3)	2426 (29.1)	
	Current	194 (32.8)	1213 (15.7)	1407 (16.9)	
Alcohol intake (n, %)	No	89 (14.8)	1064 (13.4)	1153 (13.5)	0.35
	Yes	514 (85.2)	6868 (86.6)	7382 (86.5)	
Menopause status (n, %)	Pre-menopausal	0	6886 (86.8)	6886 (80.6)	<0.001
	Post-menopausal	610 (100)	1045 (13.2)	1655 (19.4)	
Exogenous oestrogen (n, %)	Current use (MHT or COCP)	337 (55.3)	1457 (18.2)	1794 (20.9)	<0.001
	Past use (MHT only)	95 (15.6)	520 (6.5)	615 (7.2)	
	Never use	178 (29.2)	6016 (75.3)	6194 (72.0)	

Significant P-values are shown in bold.

COCP: combined oral contraceptive pill; IRSD: index of relative socioeconomic disadvantage; MHT: menopause hormone therapy (oral, transdermal, nasal).

Table 2. Cumulative comorbidities during follow-up, by menopause group.

		Premature ovarian insufficiency/early menopause	Usual age menopause	Total	P
Hyperthyroidism		3 (0.5)	36 (0.5)	39 (0.5)	0.76
Hyperparathyroidism		1 (0.2)	6 (0.08)	7 (0.1)	0.40
Rheumatoid arthritis		85 (13.9)	751 (9.4)	836 (9.7)	<0.001
Malabsorption		1 (0.2)	25 (0.3)	26 (0.3)	1.00
Inflammatory bowel disease		0	37 (0.5)	37 (0.4)	0.11
Chronic liver disease		4 (0.7)	20 (0.3)	24 (0.3)	0.09
Chronic kidney disease		12 (2.0)	126 (1.6)	138 (1.6)	0.41
Diabetes		99 (16.2)	1086 (13.6)	1185 (13.8)	0.08
Connective tissue disease		12 (2.0)	79 (1.0)	91 (1.1)	0.04
Any transplant (solid organ and bone marrow)		6 (1.0)	9 (0.1)	15 (0.2)	<0.001
Any comorbidity		177 (29.0)	1849 (23.1)	2026 (23.6)	0.001
Number of comorbidities	1	133 (21.8)	1562 (19.5)	1695 (19.7)	<0.001
	2	43 (7.1)	257 (3.2)	300 (3.5)	
	3	1 (0.2)	28 (0.4)	29 (0.3)	
	4	0	2 (0.03)	2 (0.02)	

Data expressed as n (%). Significant P-values are shown in bold.

MHT, and of these, 22% used MHT for <1 year, 39% for 1–4 years, 28% for 5–10 years, and 6% for >10 years. Of women with osteoporosis or a fracture, 1820 (56.3%) had used MHT, including 232 (77.6%) of those with POI/EM and 1588 (54.2%) of those with usual age menopause. At all surveys, more women with POI/EM currently used MHT, and had ever used MHT, compared with women with usual age menopause (Fig. 2C). On longitudinal analysis, women with POI/EM had higher likelihood of having ever used MHT (OR 6.87, 95% CI 5.68–8.30, Supplementary Table S6). Increasing age and BMI were also associated with ever using MHT, while smoking, alcohol, number of comorbidities, and prevalent osteoporosis or fractures, was not.

In women with POI/EM, increasing age (OR 1.02; 1.01–1.02), and currently consuming alcohol (OR 1.17; 1.06–1.28), was

associated with having ever used MHT (Supplementary Table S7). BMI, smoking, number of comorbidities, and prevalent osteoporosis or fracture were predictive of MHT use in univariable analysis but did not remain significant in the multivariable analysis. Bootstrapping did not alter results.

Types of MHT used, from PBS data from 2002 onwards, are in Supplementary Table S8. Of the 1180 women who used PBS MHT, the median (Q1, Q3) duration of use was 12.0 (4, 37) months, and 420 (35.6%) were compliant with therapy.

Anti-osteoporosis medications

Analysis of treatment included 3215 participants with either osteoporosis or a fracture (297 with POI/EM, 2918 with usual age menopause), from Survey 4 (2004) onwards. A total of 1080

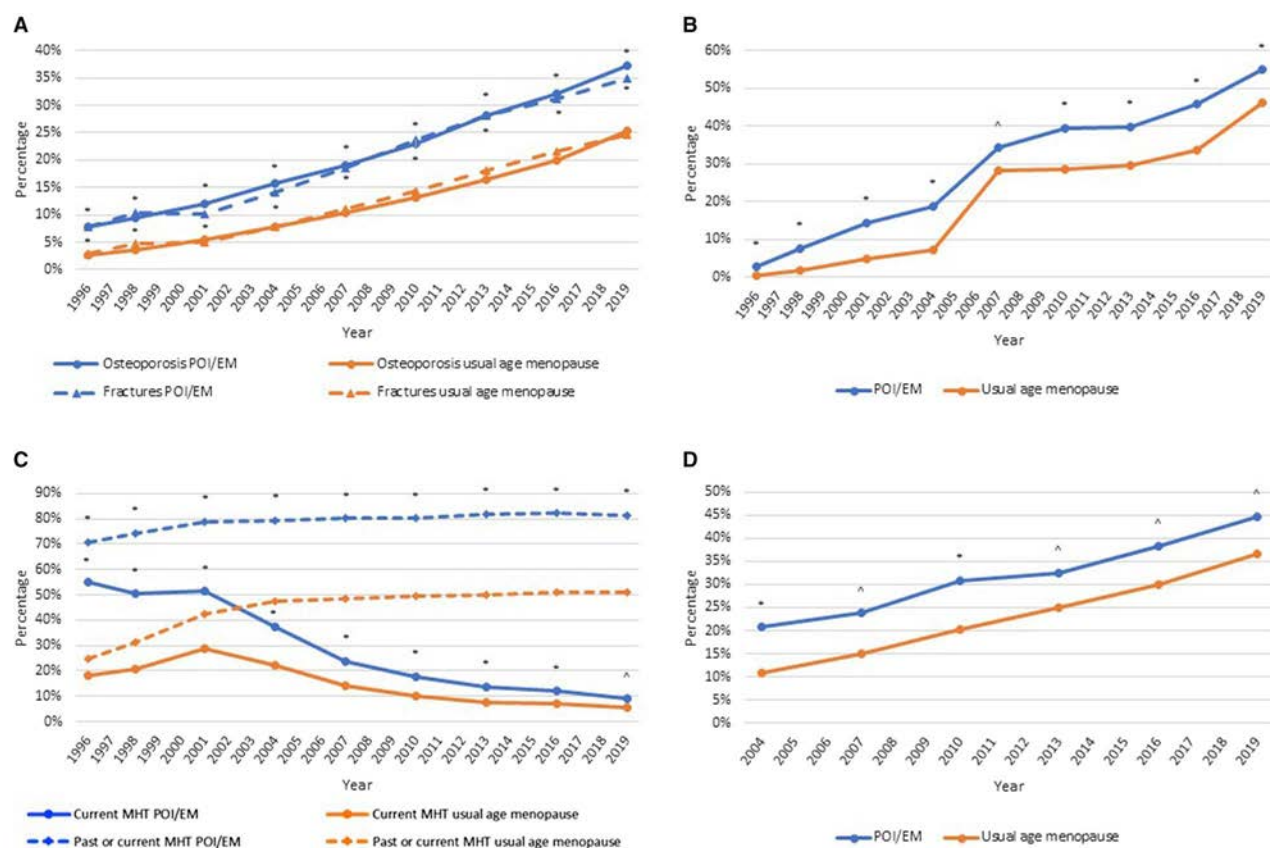


Figure 2. Frequency of outcomes at each survey, by menopausal group. (A) Cumulative frequency of osteoporosis and all fractures; (B) Dual X-ray absorptiometry use per survey period; (C) cumulative or current use of menopause hormone therapy; and (D) cumulative treatment with anti-osteoporosis medication in women with osteoporosis or fractures. Unadjusted frequency at each survey is compared by chi-square tests. * $P < 0.001$; $^{\wedge}P < 0.05$. POI: premature ovarian insufficiency; EM: early menopause; MHT: menopause hormone therapy.

Table 3. Predictors of developing osteoporosis and/or fractures in women with premature ovarian insufficiency/early menopause.

Factor		Univariable model				Multivariable model			
		OR	95% CI		P-value	OR	95%		P-value
Age		1.10	1.08	1.11	<0.001	1.09	1.08	1.11	<0.001
BMI		1.07	1.04	1.10	<0.001	1.01	0.99	1.04	0.291
Smoking	Never	Ref							
	Former	0.94	0.75	1.17	0.550	0.94	0.72	1.23	0.664
	Current	0.48	0.35	0.66	<0.001	0.94	0.69	1.29	0.695
Alcohol	No	Ref							
	Yes	0.65	0.52	0.81	<0.001	0.85	0.68	1.06	0.147
Number of comorbidities		2.37	1.92	2.91	<0.001	1.12	0.88	1.42	0.373
Country of birth	Australia	Ref							
	Outside Australia	1.35	0.98	1.86	0.069	1.14	0.77	1.70	0.515
Residential location	Major city	Ref							
	Inner regional	0.85	0.66	1.10	0.212	1.06	0.82	1.36	0.657
	Outer regional	0.79	0.54	1.16	0.226	1.12	0.79	1.58	0.540
	Remote	0.70	0.45	1.09	0.117	1.57	0.96	2.58	0.074
IRSD		1.00	1.00	1.00	0.107	1.00	1.00	1.00	0.820
MHT use ever		1.41	1.03	1.94	0.035				
MHT use current		0.41	0.34	0.49	<0.001				
MHT prior to or at study entry		0.79	0.58	1.07	0.134	0.66	0.45	0.96	0.032

IRSD: index of relative socioeconomic disadvantage; MHT: menopause hormone therapy; OR: odds ratio. Significant P-values are shown in bold.

(33.6%) received anti-osteoporosis medications at some stage, including 116 (39.1%) of those with POI/EM and 964 (33.0%) of those with usual age menopause. Of the 1925 with OP, 938 (48.7%) received treatment, including 99/186 (53.2%) with POI/EM and 839/1739 (48.3%) with usual age menopause. Of those with a fracture,

596 (29.7%) received treatment, including 71/188 (37.8%) with POI/EM and 525/1818 (28.9%) with usual age menopause. At all surveys, women with POI/EM were significantly more likely to have used anti-osteoporosis medication, than women with usual age menopause (Fig. 2D), and longitudinal analysis confirmed

women with POI/EM were more likely to use anti-osteoporosis medication (OR 1.50, 95% CI 1.14–1.98, [Supplementary Table S9](#)). Increasing age and lower BMI were also associated with using anti-osteoporosis medications, while current smoking was associated with lower likelihood of anti-osteoporosis medication use.

In the 299 women with POI/EM, increasing age (OR 1.07; 1.04–1.09) and lower BMI (OR 0.95; 0.92–0.98) were associated with greater likelihood of using anti-osteoporosis medications ([Supplementary Table S10](#)). There was no effect of smoking, alcohol, or the number of comorbidities. Bootstrapping did not alter results.

The types, duration, and compliance with anti-osteoporosis medications are listed in [Supplementary Table S11](#). Compliance did not differ between women with and without POI/EM.

Discussion

In this 23-year longitudinal study, by the age of 68 years, half of the women with POI/EM developed osteoporosis or a fracture. Increasing age was associated with an increased risk of osteoporosis/fractures in these women, whereas MHT use prior to or at study entry was protective. The risk of osteoporosis and fracture risk were 37% and 45% higher, respectively, in women with POI/EM compared with those with usual age menopause. Women with POI/EM were also more likely to have had a DXA, use MHT, and use anti-osteoporosis medications; however, the overall rates of use of the latter among women with osteoporosis or fractures was low at under 40%, highlighting a significant treatment gap in this population.

Our findings expand upon previous research of a subset of this cohort, which demonstrated that women with spontaneous menopause aged <40 years, and 41–45 years, had increased risk (OR 2.00 and 1.30, respectively) of incident osteoporosis over 6 years, compared with those undergoing menopause at 50–51 years ([Xu et al., 2020](#)). Here, we provide longer follow-up duration, linked data and fracture data, and adjustment for confounding factors. Few other longitudinal studies include both women with POI and EM. A prospective longitudinal study of Swedish women over 34 years found 56% of women with POI/EM had osteoporosis by the age of 77 years, with an 83% increased risk of osteoporosis compared with women with later menopause ([Svejme et al., 2012](#)). However, POI/EM was defined as menopause prior to the age of 47 years, rather than the conventional definition of prior to 45 years (The ESHRE Guideline Group on POI et al., 2016). The current study provides convincing evidence for increased osteoporosis risk in women with POI/EM after accounting for other risk factors.

One-third of women with POI/EM experienced an incident fracture during this study. The current findings of increased fracture risk concur with a systematic review including 14 studies of women with POI/EM, which found a 36% increased risk of fracture compared with women with menopause >45 years, although there was heterogeneity between included studies, and possible differences between ethnic groups ([Anagnostis et al., 2019](#)). Whether there is site-specific fracture risk remains unclear. A longitudinal study including 49 000 women with POI/EM who had never used MHT found little if any increased risk for hip fracture ([Banks et al., 2009](#)). Mixed findings are reported for vertebral fractures, with no increase in self-reported vertebral fractures in Canadian and Japanese cohorts with POI/EM ([Papaioannou et al., 2005](#); [Shimizu et al., 2018](#)). However, a study of Dutch women found a 90% increased risk for vertebral fractures confirmed on spinal X-rays in women with POI/EM compared to those with menopause age >50 years ([Kliff et al., 2004](#)). Unfortunately, fracture site was not collected in ALSWH survey data, and the low

number of fractures for each site recorded in linked hospital data prevented further exploration. Ongoing follow-up of the current cohort may enable such comparisons in the future.

In women with POI/EM, increasing age was associated with an increased risk of osteoporosis or fracture. This is consistent with a previous review, which also found no effect of BMI, but which did not examine other potential risk factors ([Anagnostis et al., 2019](#)). Studies limited to women with spontaneous POI, reported increased risk of low BMD associated with earlier onset of menstrual irregularity, delay in diagnosis and commencement of MHT, poor adherence to MHT, low vitamin D, low body weight, low calcium intake, lack of regular exercise, and increasing age ([Leite-Silva et al., 2009](#); [Popat et al., 2009](#)). However, these cross-sectional studies only examined women at a younger age (mean age ~30 years), and risk factors at only one time point. A strength of the current study was that risk factors were assessed at multiple time points, which takes into consideration changes in smoking status, body weight, and other dynamic risk factors over time.

Using MHT ever, at the start of the study (aged 45–50 years), was protective for osteoporosis/fractures, whereas having ever used MHT later was not protective. This highlights the importance of early treatment with MHT in women who undergo POI/EM. International guidelines recommend MHT in women with POI/EM, without contraindications, until the age of usual menopause, to prevent osteoporosis (The ESHRE Guideline Group on POI et al., 2016; [The North American Menopause Society, 2022](#)). Delay in diagnosis/treatment, non-adherence to MHT, dose, and type of MHT appear to influence the BMD response to MHT ([Burgos et al., 2017](#); [Gonçalves et al., 2022](#); [Samad et al., 2022](#); [Costa et al., 2023](#)), and there is as yet no definite evidence for fracture reduction ([Gonçalves et al., 2022](#)). Although our findings suggest a protective effect of early MHT on fractures, being observational, we cannot determine causality. At study entry, 70% of women with POI/EM had previously or currently used MHT. Of concern is that 20% of women with POI/EM did not receive appropriate MHT (after excluding women with possible oestrogen-responsive cancers). Sub-optimal MHT use by women with POI/EM has been reported in other studies with the proportion of women using MHT varying between 22 and 50% ([Shea et al., 2021](#); [Svejme et al., 2012](#)). Additionally, 30–40% of women with POI are non-adherent to MHT, due to a perceived lack of benefit, side effects, or fears of breast cancer ([Popat et al., 2009](#); [Bachelot et al., 2016](#)). Unfortunately, due to the nature of the current study, it was not possible to examine the type, dose, adherence, and duration of MHT in earlier life, limiting the ability to examine these factors.

DXA is recommended for all women with POI at baseline, with follow-up imaging dependent on BMD ([Kiriakova et al., 2019](#); [The ESHRE Guideline Group on POI et al., 2016](#)). In our cohort, during the follow-up period, 69% of women with POI/EM had one or more DXA performed, and women with POI/EM were significantly more likely to have a DXA than women with usual age menopause. DXA utilization was similar to that in cross-sectional studies of Australian women with POI ([Gibson-Helm et al., 2014](#); [Jones et al., 2020](#)). However, DXA performed prior to study entry or at diagnosis of POI/EM was not captured in our study, and thus the total number of women with POI/EM who completed screening at any time is likely to be higher. Although women with POI/EM were more likely than women with usual age menopause to receive osteoporosis treatment, management of osteoporosis overall was sub-optimal, with <40% of women with osteoporosis/fractures receiving an anti-osteoporosis medication at any stage.

This reflects a known global gap in osteoporosis care with low rates of antiresorptive use (20–30%) and failure to prevent re-fractures, and highlights an important practice point for clinicians (Kanis et al., 2021). There are no specific guidelines for using antiresorptive agents in women with POI/EM, although studies in younger women have demonstrated increased BMD with bisphosphonate therapy (Tauchmanova et al., 2006; Shapiro et al., 2011; Kiriakova et al., 2019).

This study has several strengths. It is the largest study of bone health in Australian women with POI/EM, one of few longitudinal studies assessing osteoporosis and fractures in this population, and the first to report on longitudinal risk factors, investigation, and anti-osteoporosis medication. Although POI/EM was self-reported, it was validated by ensuring all women with POI/EM had amenorrhoea at study entry. Outcomes are both self-reported and derived from administrative records. There are also several limitations to our study. Survey data were self-reported, and given data collection commenced from age 45 to 50 years, limited data were available on earlier risk factors, fractures, screening, treatment, or MHT use. We were unable to include additional risk/protective factors such as vitamin D use, calcium intake, or exercise. Although the prevalence of POI/EM was similar to that reported previously (Golezar et al., 2019), we were unable to distinguish between different causes of POI/EM. Self-reported fractures only included those in the prior 12 months, rather than the 3-year survey period included for hospital separation data fractures, and vertebral fractures will be underestimated without performing routine X-rays. Linked administrative databases were not available for the full study period, with some states' hospital separation data not available for earlier surveys, and PBS medications were only available from 2002. Compliance estimations assumed that women took all medications dispensed. Finally, the cohort had greater Australian-born women, with higher educational levels, than the broader Australian population (Dobson et al., 2015).

In this large prospective cohort of Australian women over 23 years, 7% experienced POI/EM, and half of women with POI/EM had osteoporosis or a fracture by the age of 68 years. In women with POI/EM, MHT use prior to or at the age of 45–50 years was protective for osteoporosis/fractures, highlighting the importance of early treatment in managing women with POI/EM. Although DXA utilization was high, one-fifth of women did not receive appropriate MHT and use of anti-osteoporosis medications were low, highlighting a significant practice gap in preventing future fractures.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

ALSWH survey data are owned by the Australian Government Department of Health and Aged Care and due to the personal nature of the data collected, release by ALSWH is subject to strict contractual and ethical restrictions. Ethical review of ALSWH is by the Human Research Ethics Committees at The University of Queensland and The University of Newcastle. De-identified data are available to collaborating researchers where a formal request to make use of the material has been approved by the ALSWH Data Access Committee and by relevant data custodians (if applicable). The committee is receptive of requests for datasets required to replicate results. Information on applying for ALSWH

data is available from <https://alswh.org.au/for-data-users/applying-for-data/>

- The Centre for Health Record Linkage (CHeReL), NSW Ministry of Health and ACT Health, for the NSW Admitted Patients Collection; and the ACT Admitted Patient Care Collection.
- Queensland Health as the source for Queensland Hospital Admitted Patient Data Collection; and the Statistical Analysis and Linkage Unit (Queensland Health) for the provision of data linkage.
- The Department of Health Western Australia, including the Data Linkage Branch, and the WA Hospital Morbidity Data Collection.
- SA NT Datalink, SA Health, and Northern Territory Department of Health, for the SA Public Hospital Separations, and NT Public Hospital Inpatient Activity Data Collections.
- The Department of Health Tasmania, and the Tasmanian Data Linkage Unit, for the Public Hospital Admitted Patient Episodes Data Collection.
- Victorian Department of Health as the source of the Victorian Admitted Episodes Dataset; and the Centre for Victorian Data Linkage (Victorian Department of Health) for the provision of data linkage.

Acknowledgements

The authors would like to thank Mr Sanjeeva Ranasinha and Dr Raja Ram Dhungana for additional statistical support. The research on which this paper is based was conducted as part of the Australian Longitudinal Study on Women's Health by the University of Queensland and the University of Newcastle. We are grateful to the Australian Government Department of Health and Aged Care for funding and to the women who provided the survey data. The authors acknowledge the Australian Government Department of Health and Aged Care, and Veterans' Affairs, for providing MBS and PBS data; and the Australian Institute of Health and Welfare (AIHW) as the integrating authority. The authors acknowledge the following:

Authors' roles

A.R.J. was involved in study design, execution, manuscript drafting, and critical discussion. P.R.E. was involved in manuscript drafting and critical discussion. G.D.M. was involved in study design, execution, and manuscript drafting. H.T.T. was involved in study design, obtaining funding, manuscript drafting, and critical analysis. J.E. was involved in data analysis and manuscript drafting. A.J.V. was involved in study design, manuscript drafting, and critical discussion.

Funding

This study was funded by The Australian NHMRC Centre of Research Excellence Women's Health in Reproductive Life (CRE-WHIREL, project number APP1171592). A.R.J. is the recipient of a National Health and Medical Research Council post-graduate research scholarship (grant number 1169192). P.R.E. is supported by a National Health and Medical Research Council grant 1197958.

Conflict of interest

P.R.E. reports grants paid to their institution from Amgen, Sanofi, and Alexion, honoraria from Amgen paid to their institution, and honoraria from Alexion and Kyowa-Kirin.

References

- Ahlborg HG, Johnell O, Turner CH, Rannevik G, Karlsson MK. Bone loss and bone size after menopause. *N Engl J Med* 2003;**349**:327–334.
- Anagnostis P, Siolos P, Gkekas NK, Kosmidou N, Artzouchaltzi A-M, Christou K, Paschou SA, Potoupnis M, Kenanidis E, Tsidiris E et al. Association between age at menopause and fracture risk: a systematic review and meta-analysis. *Endocrine* 2019;**63**:213–224.
- Australian Bureau of Statistics. *Remoteness Structure*. Canberra: ABS, July 2021–June 2026. <https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/remoteness-structure/remoteness-areas> (18 September 2023, date last accessed).
- Australian Longitudinal Study on Women's Health. *Australian Longitudinal Study on Women's Health Surveys*, 2022. <https://alswh.org.au/for-data-users/data-documentation/surveys> (1 October 2023, date last accessed).
- Bachelot A, Nicolas C, Gricourt S, Dulon J, Leban M, Golmard JL, Touraine P. Poor compliance to hormone therapy and decreased bone mineral density in women with premature ovarian insufficiency. *PLoS One* 2016;**11**:e0164638.
- Banks E, Reeves GK, Beral V, Balkwill A, Liu B, Roddam A; Million Women Study Collaborators. Hip fracture incidence in relation to age, menopausal status, and age at menopause: prospective analysis. *PLoS Med* 2009;**6**:e1000181.
- Barrionuevo P, Kapoor E, Asi N, Alahdab F, Mohammed K, Benkhadra K, Almasri J, Farah W, Sarigianni M, Muthusamy K et al. Efficacy of pharmacological therapies for the prevention of fractures in postmenopausal women: a network meta-analysis. *J Clin Endocrinol Metab* 2019;**104**:1623–1630.
- Billir BM, Saxe V, Herzog DB, Rosenthal DI, Holzman S, Klibanski A. Mechanisms of osteoporosis in adult and adolescent women with anorexia nervosa. *J Clin Endocrinol Metab* 1989;**68**:548–554.
- Bone HG, Bolognese MA, Yuen CK, Kendler DL, Miller PD, Yang YC, Grazette L, San Martin J, Gallagher JC. Effects of denosumab treatment and discontinuation on bone mineral density and bone turnover markers in postmenopausal women with low bone mass. *J Clin Endocrinol Metab* 2011;**96**:972–980.
- Brown W, Bryson L, Byles J, Dobson A, Manderson L, Schofield M, Williams G. Women's Health Australia: establishment of the Australian longitudinal study on women's health. *J Womens Health* 1996;**5**:467–472.
- Burgos N, Cintron D, Latortue-Albino P, Serrano V, Rodriguez Gutierrez R, Faubion S, Spencer-Bonilla G, Erwin PJ, Murad MH. Estrogen-based hormone therapy in women with primary ovarian insufficiency: a systematic review. *Endocrine* 2017;**58**:413–425.
- Costa GPO, Ferreira-Filho ES, Simoes R, Soares-Junior JM, Baracat EC, Maciel GAR. Impact of hormone therapy on the bone density of women with premature ovarian insufficiency: a systematic review. *Maturitas* 2023;**167**:105–112.
- Dobson AJ, Hockey R, Brown WJ, Byles JE, Loxton DJ, McLaughlin D, Tooth LR, Mishra GD. Cohort profile update: Australian Longitudinal Study on Women's Health. *Int J Epidemiol* 2015;**44**:1547, 1547a–1547f.
- Finkelstein JS, Brockwell SE, Mehta V, Greendale GA, Sowers MR, Ettinger B, Lo JC, Johnston JM, Cauley JA, Danielson ME et al. Bone mineral density changes during the menopause transition in a multiethnic cohort of women. *J Clin Endocrinol Metab* 2008;**93**:861–868.
- Gibson-Helm M, Teede H, Vincent A. Symptoms, health behavior and understanding of menopause therapy in women with premature menopause. *Climacteric* 2014;**17**:666–673.
- Goh M, Nguyen HH, Khan NN, Milat F, Boyle JA, Vincent AJ. Identifying and addressing osteoporosis knowledge gaps in women with premature ovarian insufficiency and early menopause: a mixed-methods study. *Clin Endocrinol (Oxf)* 2019;**91**:498–507.
- Golezar S, Ramezani Tehrani F, Khazaei S, Ebadi A, Keshavarz Z. The global prevalence of primary ovarian insufficiency and early menopause: a meta-analysis. *Climacteric* 2019;**22**:403–411.
- Gonçalves CR, Vasconcellos AS, Rodrigues TR, Comin FV, Reis FM. Hormone therapy in women with premature ovarian insufficiency: a systematic review and meta-analysis. *Reprod Biomed Online* 2022;**44**:1143–1157.
- Hadjidakis DJ, Kokkinakis EP, Sfakianakis ME, Raptis SA. Bone density patterns after normal and premature menopause. *Maturitas* 2003;**44**:279–286.
- InterLACE Study Team. Variations in reproductive events across life: a pooled analysis of data from 505 147 women across 10 countries. *Hum Reprod* 2019;**34**:881–893.
- Jones AR, Goh M, Langham R, Boyle J, Milat F, Ebeling PR, Teede H, Vincent AJ. Osteoporosis and premature ovarian insufficiency: geographic variation in clinicians' and consumers' knowledge gaps and barriers to care. *Arch Osteoporos* 2020;**15**:38.
- Kanis JA, Norton N, Harvey NC, Jacobson T, Johansson H, Lorentzon M, McCloskey EV, Willers C, Borgström F. SCOPE 2021: a new scorecard for osteoporosis in Europe. *Arch Osteoporos* 2021;**16**:82.
- Kiriakova V, Cooray SD, Yeganeh L, Somarajah G, Milat F, Vincent AJ. Management of bone health in women with premature ovarian insufficiency: systematic appraisal of clinical practice guidelines and algorithm development. *Maturitas* 2019;**128**:70–80.
- Klift M, de Laet C, McCloskey EV, Johnell O, Kanis JA, Hofman B, Pols H. Risk factors for incident vertebral fractures in men and women: the Rotterdam Study. *J Bone Miner Res* 2004;**19**:1172–1180.
- Leite-Silva P, Bedone A, Pinto-Neto AM, Costa JV, Costa-Paiva L. Factors associated with bone density in young women with karyotypically normal spontaneous premature ovarian failure. *Arch Gynecol Obstet* 2009;**280**:177–181.
- Mishra GD, Pandeya N, Dobson AJ, Chung HF, Anderson D, Kuh D, Sandin S, Giles GG, Bruinsma F, Hayashi K et al. Early menarche, nulliparity and the risk for premature and early natural menopause. *Hum Reprod* 2017;**32**:679–686.
- Muka T, Oliver-Williams C, Kunutsors S, Laven JSE, Fauser BCJM, Chowdhury R, Kavousi M, Franco OH. Association of age at onset of menopause and time since onset of menopause with cardiovascular outcomes, intermediate vascular traits, and all-cause mortality: a systematic review and meta-analysis. *JAMA Cardiol* 2016;**1**:767–776.
- Nguyen HH, Milat F, Vincent AJ. New insights into the diagnosis and management of bone health in premature ovarian insufficiency. *Climacteric* 2021;**24**:481–490.
- Paganini-Hill A, Atchison KA, Gornbein JA, Nattiv A, Service SK, White SC. Menstrual and reproductive factors and fracture risk: the Leisure World Cohort Study. *J Women's Health (Larchmt)* 2005;**14**:808–819.
- Papaioannou A, Joseph L, Ioannidis G, Berger C, Anastassiades T, Brown JP, Hanley DA, Hopman W, Josse RG, Kirkland S et al. Risk factors associated with incident clinical vertebral and nonvertebral fractures in postmenopausal women: the Canadian Multicentre Osteoporosis Study (CaMos). *Osteoporos Int* 2005;**16**:568–578.

- Popat BV, Calis AK, Vanderhoof HV, Cizza CG, Reynolds FJ, Sebring MN, Troendle MJ, Nelson ML. Bone mineral density in estrogen-deficient young women. *J Clin Endocrinol Metab* 2009;**94**:2277–2283.
- Reid IR, Horne AM, Mihov B, Stewart A, Garratt E, Wong S, Wiessing KR, Bolland MJ, Bastin S, Gamble GD. Fracture prevention with zoledronate in older women with osteopenia. *N Engl J Med* 2018;**379**:2407–2416.
- Rosenthal DI, Mayo-Smith W, Hayes CW, Khurana JS, Biller BMK, Neer RM, Klibanski A. Age and bone mass in premenopausal women. *J Bone Miner Res* 1989;**4**:533–538.
- Samad N, Nguyen HH, Hashimura H, Pasco J, Kotowicz M, Strauss BJ, Ebeling PR, Milat F, Vincent AJ. Abnormal trabecular bone score, lower bone mineral density and lean mass in young women with premature ovarian insufficiency are prevented by oestrogen replacement. *Front Endocrinol (Lausanne)* 2022;**13**:860853.
- Shapiro CL, Halabi S, Hars V, Archer L, Weckstein D, Kirshner J, Sikov W, Winer E, Burstein HJ, Hudis C et al. Zoledronic acid preserves bone mineral density in premenopausal women who develop ovarian failure due to adjuvant chemotherapy: final results from CALGB trial 79809. *Eur J Cancer* 2011;**47**:683–689.
- Shea AK, Buwembo A, Mayhew A, Sohel N, Griffith LE, Raina P. The association between primary ovarian insufficiency and osteoporosis in the Canadian Longitudinal Study on Aging. *Menopause* 2021;**28**:693–698.
- Shimizu Y, Sawada N, Nakamura K, Watanabe Y, Kitamura K, Iwasaki M, Tsugane S; JPHC Study group. Menstrual and reproductive factors and risk of vertebral fractures in Japanese women: the Japan Public Health Center-based prospective (JPHC) study. *Osteoporos Int* 2018;**29**:2791–2801.
- Siegel DA, Claridy M, Mertens A, George E, Vangile K, Simoneaux SF, Meacham LR, Wasilewski-Masker K. Risk factors and surveillance for reduced bone mineral density in pediatric cancer survivors. *Pediatr Blood Cancer* 2017;**64**:e26488.
- Sullivan SD, Lehman A, Nathan NK, Thomson CA, Howard BV. Age of menopause and fracture risk in postmenopausal women randomized to calcium + vitamin D, hormone therapy, or the combination: results from the Women's Health Initiative Clinical Trials. *Menopause* 2017;**24**:371–378.
- Sullivan SD, Lehman A, Thomas F, Johnson KC, Jackson R, Wactawski-Wende J, Ko M, Chen Z, Curb JD, Howard BV. Effects of self-reported age at nonsurgical menopause on time to first fracture and bone mineral density in the Women's Health Initiative Observational Study. *Menopause* 2015;**22**:1035–1044.
- Svejme O, Ahlborg HG, Nilsson JA, Karlsson MK. Early menopause and risk of osteoporosis, fracture and mortality: a 34-year prospective observational study in 390 women. *BJOG* 2012;**119**:810–816.
- Tauchmanova L, De Simone G, Musella T, Orio F, Ricci P, Nappi C, Lombardi G, Colao A, Rotoli B, Selleri C. Effects of various anti-reabsorptive treatments on bone mineral density in hypogonadal young women after allogeneic stem cell transplantation. *Bone Marrow Transplant* 2006;**37**:81–88.
- The ESHRE Guideline Group on POI; Webber L, Davies M, Anderson R, Bartlett J, Braat D, Cartwright B, Cifkova R, de Muinck Keizer-Schrama S, Hogervorst E et al. ESHRE Guideline: management of women with premature ovarian insufficiency. *Hum Reprod* 2016;**31**:926–937.
- The North American Menopause Society. The 2022 hormone therapy position statement of The North American Menopause Society. *Menopause* 2022;**29**:767–794.
- Weaver CM, Gordon CM, Janz KF, Kalkwarf HJ, Lappe JM, Lewis R, O'Karma M, Wallace TC, Zemel BS. The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. *Osteoporos Int* 2016;**27**:1281–1386.
- Xu X, Jones M, Mishra GD. Age at natural menopause and development of chronic conditions and multimorbidity: results from an Australian prospective cohort. *Hum Reprod* 2020;**35**:203–211.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Human Reproduction, 2024, 39, 1013–1022
<https://doi.org/10.1093/humrep/deae037>
Original Article

Reproductive biology

The impact of ovarian stimulation on the human endometrial microenvironment

Anat Chemerinski^{1,*,†}, May Shen^{2,†}, Nuriban Valero-Pacheco^{3,†}, Qingshi Zhao^{1,†}, Trystn Murphy¹, Lea George¹, Alex Lemenze^{3,4}, Lauren Sherman⁵, Debra Heller⁶, Xiaowei Chen⁷, Tracy Wu¹, Peter G. McGovern¹, Sara S. Morelli¹, Ripla Arora¹, Aimee M. Beaulieu^{3,8,†}, and Nataki C. Douglas^{1,3,†}

¹Department of Obstetrics, Gynecology and Reproductive Health, Rutgers Biomedical and Health Sciences, Newark, NJ, USA

²Department of Obstetrics, Gynecology and Reproductive Biology, Michigan State University, East Lansing, MI, USA

³Center for Immunity and Inflammation, Rutgers Biomedical and Health Sciences, Newark, NJ, USA

⁴Molecular and Genomics Informatics Core, Rutgers Biomedical and Health Sciences, Newark, NJ, USA

⁵Department of Medicine-Hematology/Oncology, Rutgers Biomedical and Health Sciences, Newark, NJ, USA

⁶Department of Pathology, Rutgers Biomedical and Health Sciences, Newark, NJ, USA

⁷Department of Pathology, Columbia University Medical Center, New York, NY, USA

⁸Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers Biomedical and Health Sciences, Newark, NJ, USA

*Correspondence address. Department of Obstetrics, Gynecology and Reproductive Health, Rutgers Biomedical and Health Sciences, 185 S Orange Avenue, Newark, NJ 07103, USA. Tel: +1-973-972-5554; E-mail: ac2059@njms.rutgers.edu <https://orcid.org/0000-0002-8745-9409>

[†]These authors contributed equally to this work.

[‡]These are co-senior authors with equal contribution.

ABSTRACT

STUDY QUESTION: How does ovarian stimulation (OS), which is used to mature multiple oocytes for ART procedures, impact the principal cellular compartments and transcriptome of the human endometrium in the periovulatory and mid-secretory phases?

SUMMARY ANSWER: During the mid-secretory window of implantation, OS alters the abundance of endometrial immune cells, whereas during the periovulatory period, OS substantially changes the endometrial transcriptome and impacts both endometrial glandular and immune cells.

WHAT IS KNOWN ALREADY: Pregnancies conceived in an OS cycle are at risk of complications reflective of abnormal placentation and placental function. OS can alter endometrial gene expression and immune cell populations. How OS impacts the glandular, stromal, immune, and vascular compartments of the endometrium, in the periovulatory period as compared to the window of implantation, is unknown.

STUDY DESIGN, SIZE, DURATION: This prospective cohort study carried out between 2020 and 2022 included 25 subjects undergoing OS and 25 subjects in natural menstrual cycles. Endometrial biopsies were performed in the proliferative, periovulatory, and mid-secretory phases.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Blood samples were processed to determine serum estradiol and progesterone levels. Both the endometrial transcriptome and the principal cellular compartments of the endometrium, including glands, stroma, immune, and vasculature, were evaluated by examining endometrial dating, differential gene expression, protein expression, cell populations, and the three-dimensional structure in endometrial tissue. Mann–Whitney *U* tests, unpaired *t*-tests or one-way ANOVA and pairwise multiple comparison tests were used to statistically evaluate differences.

MAIN RESULTS AND THE ROLE OF CHANCE: In the periovulatory period, OS induced high levels of differential gene expression, glandular-stromal dyssynchrony, and an increase in both glandular epithelial volume and the frequency of endometrial monocytes/macrophages. In the window of implantation during the mid-secretory phase, OS induced changes in endometrial immune cells, with a greater frequency of B cells and a lower frequency of CD4 effector T cells.

LARGE SCALE DATA: The data underlying this article have been uploaded to the Genome Expression Omnibus/National Center for Biotechnology Information with accession number GSE220044.

LIMITATIONS, REASONS FOR CAUTION: A limited number of subjects were included in this study, although the subjects within each group, natural cycle or OS, were homogenous in their clinical characteristics. The number of subjects utilized was sufficient to identify significant differences; however, with a larger number of subjects and additional power, we may detect additional differences. Another limitation of the study is that proliferative phase biopsies were collected in natural cycles, but not in OS cycles. Given that the OS cycle subjects did not have known endometrial factor infertility, and the comparisons involved subjects who had a similar and robust response to stimulation, the findings are generalizable to women with a normal response to OS.

WIDER IMPLICATIONS OF THE FINDINGS: OS substantially altered the periovulatory phase endometrium, with fewer transcriptomic and cell type-specific changes in the mid-secretory phase. Our findings show that after OS, the endometrial microenvironment

Received: August 2, 2023. Revised: February 19, 2024. Editorial decision: February 22, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved. For permissions, please email: journals.permissions@oup.com

in the window of implantation possesses many more similarities to that of a natural cycle than does the periovulatory endometrium. Further investigation of the immune compartment and the functional significance of this cellular compartment under OS conditions is warranted.

STUDY FUNDING/COMPETING INTERESTS: Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases (R01AI148695 to A.M.B. and N.C.D.), Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD109152 to R.A.), and the March of Dimes (5-FY20-209 to R.A.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or March of Dimes. All authors declare no conflict of interest.

Keywords: endometrium / ovarian stimulation / ART / RNA sequencing / endometrial glands / endometrial stroma / immune cells

Introduction

Nearly 80,000 infants are born each year from ART cycles (Centers for Disease Control and Prevention, 2020). Pregnancies conceived in an ovarian stimulation (OS) cycle are at risk for higher rates of preterm delivery, low birth weight, and small for gestational age infants (Maheshwari et al., 2018), as well as complications reflective of abnormal placentation and placental function. While significant advances have been made in optimizing embryo handling and selection, leading to improved live birth rates with single embryo transfers, less is understood about how the endometrial environment contributes to normal or abnormal pregnancies arising from ART. The endometrium is a complex tissue composed of many cell types; in this manuscript, we investigate the impact of OS on these individual cell types, including stromal, glandular epithelial, vascular, and immune cells. As 50,000 fresh embryo transfers are performed each year, it is imperative to better understand how OS could negatively impact endometrial development in each of these compartments, resulting in a tissue that is less receptive to embryo implantation and that does not support proper placentation.

Since the 1970s, dating of endometrial tissue has been used to assess endometrial maturation during the menstrual cycle (Noyes et al., 1975). Histologic changes in the endometrial glands and stroma correlate with ovarian-derived estradiol (E2) and progesterone (P4) hormones. Thus, endometrial dating can identify endometrial pathology linked to aberrant E2 and/or P4 production. When dating the endometrium, the pathologist assigns a separate date to the glands and stroma (Noyes et al., 1975). In unstimulated natural menstrual cycles, endometrial glands and stroma possess a similar histologic date and are synchronized in their development. OS, resulting in supraphysiologic circulating levels of E2 and P4, has been associated with aberrant development and differentiation of the epithelium and stroma (Bourgain and Devroey, 2003). Some, but not all studies, have concluded that OS results in glandular-stromal dyssynchrony (Noci et al., 1997; Meyer et al., 1999; Basir et al., 2001). The lack of consistency in these studies likely arises from the heterogeneity in OS protocols, oocyte maturation triggers, biopsy timepoints, and subject characteristics.

More recently, studies have sought to shed light on the effects of OS on the endometrium through gene expression profiling. The initial studies used microarray technology (Horcajadas et al., 2005, 2007, 2008; Martinez-Conejero et al., 2007; Macklon et al., 2008; Liu et al., 2008, 2010; Haouzi et al., 2009, 2010; Zhao et al., 2010; Kodithuwakku et al., 2011; Bermejo et al., 2014) and were analyzed alongside other similar studies in a recent systematic review (Walker et al., 2023). Recently, three studies have utilized RNA sequencing to determine how OS changes the global endometrial transcriptome (Li et al., 2020; Wu et al., 2020; Han et al., 2022). Li et al. compared gene expression in OS and natural cycles (NCs) in the mid-secretory phase. Wu et al. evaluated the effects of different OS protocols (minimal vs conventional stimulation)

on gene expression in the periovulatory period at the time of oocyte retrieval. In the third study by Han et al., all subjects were treated with similar OS protocols and were subsequently divided based on peak serum LH levels during OS. Five days after oocyte retrieval, which is the start of the ‘window of implantation’ and two days prior to mid-secretory phase, endometrial gene expression was compared in subjects with low, medium, and high peak serum LH levels. All studies found aberrations in endometrial gene expression induced by OS. In the periovulatory period, Wu et al. noted that conventional and minimal OS protocols had differential effects on endometrial expression of Wnt family genes. Up-regulation of the canonical Wnt/ β -catenin signaling pathway, a pathway involved in endometrial gland formation and endometrial decidualization (Tulac et al., 2003), was noted in subjects receiving minimal OS. Han et al. found the highest number of differentially expressed endometrial genes, as well as alterations in genes comprising the endometrial receptivity assay (Diaz-Gimeno et al., 2011) in the low vs medium LH comparison (Han et al., 2022). In the mid-secretory period, Li et al. found differential expression of genes related to endometrial receptivity in OS vs NC subjects. Generalizability of the findings of the earlier studies is limited by the study design; Li et al. included infertile women in the unstimulated control group and Wu et al. compared the effects of different OS protocols without an unstimulated control group. The study by Han et al. is unique in its goal of understanding how changes in LH induced by OS may impact the endometrium, however, the small sample size, as well as the inclusion of both poor and normal responders in the analyses limits its generalizability.

The immune compartment, though less well-studied, was shown in recently published studies in mice and humans to undergo cell type-specific changes following OS. In mice, OS decreases leukocyte and endothelial cell density (Libby et al., 2021), as well as uterine natural killer (uNK) cell density (Ma et al., 2023), in early pregnancy prior to placentation. In humans, a recent analysis of the transcriptome of uNK cells isolated from endometrium showed that OS alters signaling pathways associated with immune cell movement and immune cell trafficking (Kanter et al., 2023). These studies shed light on some of the changes observed in the endometrium following OS; however, a clinically validated approach does not yet exist for optimizing implantation and pregnancy outcomes after fresh embryo transfer. Thus, significant gaps remain in our understanding of the impact of OS on different cellular compartments of the endometrium.

To address some of the limitations associated with previous studies and improve our understanding of the unique impact of OS on different cell types in the human endometrium, we performed a comprehensive analysis of the endometrium under unstimulated and stimulated conditions. The theory of dysregulated decidualization leading to ‘endometrial spectrum disorders’ has been proposed as an explanation for the adverse pregnancy outcomes associated with ART (Conrad, 2020) and we sought to

elucidate the underpinnings of this dysregulation. We prospectively recruited subjects with proven fertility in unstimulated ovulatory cycles (henceforth referred to as 'natural cycles') from a general gynecology clinic. To isolate the effect of OS on an endometrium with full potential for implantation, we simultaneously recruited women, without evidence of endometrial factor infertility, undergoing OS cycles in an ART clinic. We performed endometrial biopsies in both groups and characterized the cellular compartments of the endometrium both longitudinally across natural and OS cycles, as well as compared natural and OS cycles in order to determine the impact of OS cycle conditions on the periovulatory and mid-secretory endometrium. We hypothesized that there would be significant differences between the natural cycle (NC) and OS cycle groups in both the periovulatory and mid-secretory phases.

In this study, we present the results of a four-way comparative examination of the endometrium under stimulated and unstimulated conditions, which has not previously been performed. We evaluate both the endometrial transcriptome and its principal cellular compartments (glands, stroma, immune and vasculature) at three critical time points, the proliferative, periovulatory, and mid-secretory phases, by examining endometrial dating, differential gene expression, protein expression, cell populations, and the three-dimensional structure. We found that many cell types, including glandular and immune, are altered by OS during the periovulatory period. In contrast, the endometrial microenvironment in the mid-secretory phase is more similar in NC and OS subjects than previously thought, with the notable exception that endometrial B cells are markedly expanded after OS.

Materials and methods

Human subjects

Endometrial samples were collected from women with normal, 21- to 35-day menstrual cycles and from women in OS cycles. Using an Institutional Review Board (IRB)-approved protocol at Rutgers New Jersey Medical School and a university-affiliated fertility clinic, 35 women with normal ovulation and proven fertility and 34 women utilizing ART for treatment of infertility or oocyte freezing were prospectively enrolled between July 2020 and August 2022 (Fig. 1A). All study participants gave their signed informed consent for the review of their medical records, collection of blood, and endometrial sampling. Women ages 21–45 who were not pregnant were eligible for the study. For the NC group, women with normal ovulation and proven fertility who were not actively trying to conceive and had not used hormonal treatment in the 3 months prior to recruitment were included. For the OS group, women undergoing OS followed by in vitro fertilization and cryopreservation of all viable embryos, as a treatment approach for the diagnoses of unexplained, male factor or tubal factor infertility, and those undergoing OS with oocyte cryopreservation were enrolled. Women with the following disorders were excluded: (i) endocrine or autoimmune disorders, including antiphospholipid syndrome and polycystic ovary syndrome (PCOS), (ii) anatomic disorders of the reproductive tract including hydrosalpinges or submucosal fibroids, (iii) recurrent pregnancy loss, (iv) recurrent implantation failure, (v) endometrial factor contributing to infertility diagnosis, or (vi) history of preeclampsia or intrauterine growth restriction in a previous pregnancy. Once enrolled, study participants were instructed to use condoms during sexual intercourse for the duration of their enrollment.

In the NC group, 33 study participants who reported regular menstrual cycles and at least one prior pregnancy were confirmed to be ovulatory based on urine LH testing and a serum P4 level >3 ng/ml in the secretory phase. For these participants, endometrial biopsies were performed in the subsequent menstrual cycle with the Endocell[®], a disposable endometrial cell sampler (Wallach Surgical Devices, Trumbull, CT, USA). Biopsies were performed in the proliferative phase between cycle days (CDs) 10–13 for 7 women, in the periovulatory phase 12–24 h after the first detection of the LH surge with urine ovulation predictor kits for 13 women, and in the mid-secretory phase 8–9 days after detection of the LH surge for 13 women.

Study participants undergoing OS used a GnRH antagonist (Cetrotide, Ganirelix, or Fyremadel), gonadotropins (Follistim or Gonal-F in combination with Menopur), and hCG 10,000 IU to induce final oocyte maturation. The starting gonadotropin dose was based on age, BMI, serum anti-Müllerian hormone level, and antral follicle count. hCG was administered when at least two follicles measured ≥18 mm and the majority were ≥13 mm, after considering serum E2 and days of stimulation. Oocyte retrieval was performed 34–36 h after hCG was administered. OS clinical treatment protocols were determined and managed by the treating physicians, and endometrial biopsies were performed on 33 women by physicians on the study team. Biopsies were performed at the periovulatory stage two days after hCG was administered, at the time of oocyte retrieval for 24 women, while mid-secretory biopsies were performed nine days after hCG was administered for 9 women. Characteristics of study participants are summarized in Fig. 1, Table 1, and Supplementary Tables S1 and S2.

Serum

Blood (5 ml) was collected in nonheparinized, serum separator tubes, allowed to clot, and centrifuged at 4°C. For study participants in NCs, blood samples were collected at the first study visit and on the day of the endometrial biopsy. For study participants in OS cycles, blood samples were collected as per the clinical treatment protocol, and on the day of the endometrial biopsy. Serum E2 and P4 levels were determined (LabCorp, Raritan, NJ).

Tissue collection

Endometrial biopsies were collected under sterile conditions from the uterine cavity. For each biopsy, the endometrial tissue was placed in a tissue culture dish and separated into multiple pieces with fine forceps. For all 50 samples included for analysis (Fig. 1A), ~100 mg endometrial tissue was fixed in formalin and stored at room temperature and ~100 mg endometrial tissue was preserved in RNAlater (Qiagen, 76160) for RNA extraction. Nineteen samples were used for RNA-Seq studies. Endometrial tissue (~500 mg) from 42 samples was fixed in dimethylsulfoxide (DMSO):methanol for wholemount immunostaining. For 34 samples, all remaining endometrial tissue was placed on ice, and processing for flow cytometry was initiated within 1 hour of tissue collection.

Endometrial dating

Formalin-fixed and paraffin-embedded tissue was sectioned at 5 µm, stained with hematoxylin and eosin (H&E), and reviewed independently by two gynecological pathologists (D.H. and X.C.) who were blinded to the study participant group. Endometrial dating was performed according to the histopathological criteria set forth by Noyes *et al.* (1975). Subjects were excluded if samples had inconclusive endometrial dating or if histologic dating between the two pathologists differed by more than 3 days.

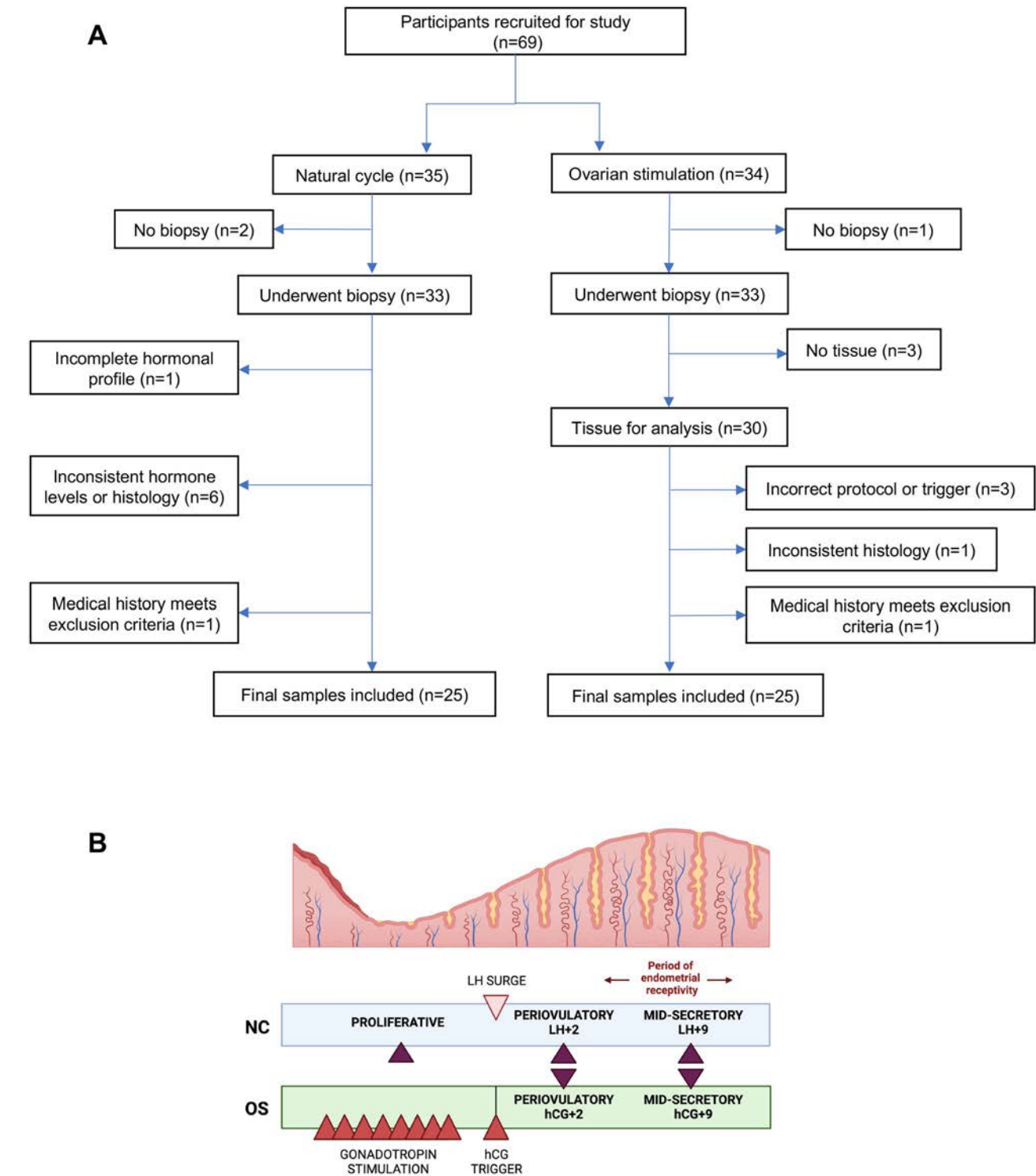


Figure 1. Study design. (A) Recruitment flowchart of enrolled subjects. (B) Schematic representation of natural and ovarian stimulation (OS) cycles. Purple arrowheads indicate timing of endometrial biopsies. For natural cycles (NC), detection of the luteinizing hormone (LH) surge (pink arrowhead) was used to time endometrial biopsies. For OS cycles, the daily gonadotropin stimulation and human chorionic gonadotropin (hCG) trigger are indicated by orange arrowheads.

($n = 1$, Fig. 1A). For NCs, subjects were excluded if histologic dating and presumed phase of the menstrual cycle were inconsistent ($n = 6$, Fig. 1A).

RNA extraction and RNA-Seq analysis

Nineteen samples were included in the RNA sequencing analysis (NC: 6 periovulatory and 4 mid-secretory; OS: 5 periovulatory and

4 mid-secretory). Total RNA was extracted from the endometrial tissue using the RNeasy Plus Mini Kit (Qiagen, 74136) according to the manufacturer's protocol. Following extraction, RNA concentration and integrity were determined. For analysis of the endometrial transcriptome, RNA was sent to the Computational Genomics Core at Albert Einstein College of Medicine, Bronx, NYC, for next-generation mRNA sequencing. The raw

Table 1. Demographic characteristics of study subjects in natural and ovarian stimulation cycles.

	Proliferative		Periovulatory		Mid-secretory		
	NC (n = 6)	NC (n = 10)	OS* (n = 13)	P-value	NC (n = 9)	OS (n = 9)	P-value
Age	28.8 ± 5.8	29.4 ± 4.8	33.0 ± 1.9	0.02	30.4 ± 5.4	36.1 ± 4.3	0.03
BMI (kg/m ²)	31.6 (26.6, 39.9)	28.4 (26.2, 38.4)	25.5 (23.0, 28.1)	0.08	28.9 (26.8, 38.2)	27.3 (22.7, 33.8)	0.36
Gravidity	3.0 (3.0, 4.0)	2.5 (1.8, 4.5)	0.0 (0.0, 0.8)	<0.0001	3.5 (1.5, 4.0)	1.0 (0.0, 1.0)	0.0009
Parity	2.0 (2.0, 3.0)	2.0 (1.0, 3.3)	0.0 (0.0, 0.0)	<0.0001	2.5 (1.3, 3.8)	0.0 (0.0, 1.0)	<0.0001
E2 (pg/ml)	170.7 ± 88.8	182.5 ± 85.4	978.8 ± 477.3	<0.0001	160.5 ± 114.6	682.2 ± 557.7	0.01
P4 (ng/ml)	0.4 ± 0.3	1.0 ± 0.7	12.1 ± 5.3	<0.0001	11.0 ± 3.6	19.0 ± 15.1	0.0005

* Three subjects with poor response to ovarian stimulation (OS-P) were removed. Total NC participants = 25, total OS participants = 22. Data are shown as mean ± SD or median (IQR). Age, E2, and P4 levels compared with unpaired t-test; BMI, gravidity, parity compared with Mann-Whitney U test. E2, estradiol; IQR, interquartile range; NC, natural cycle; OS, ovarian stimulation; P4, progesterone.

transcriptome reads were assessed for quality control (FASTQC v 0.11.8) (Andrews, 2010) and trimmed for quality/adaptor contamination (cutadapt v2.5) (Martin, 2011). Then, 75 base-pair single-end reads were used. The average read count was 62,568,930. Trimmed reads were aligned to the human genome (GRCh37) using STAR (v2.6.1) (Dobin et al., 2013), followed by transcript abundance calculations and hit count extraction with StringTie (v2.0) (Pertea et al., 2015) and featureCounts (v1.6.4) (Love et al., 2014), respectively. Hit count normalization and differential gene expression group cross-comparisons were performed using DESeq2 (v1.26.0) (Liao et al., 2014). Significant differentially expressed gene thresholds were set to FDR-adjusted $P < 0.05$. Pathway term enrichment was performed with Ingenuity Pathway Analysis.

Whole mount immunofluorescent staining, imaging, and 3D reconstruction

Whole mount immunofluorescent staining was performed on endometrial biopsy tissue that was fixed in DMSO:methanol (1:4 solution) and stored at -20°C . Endometrial biopsy tissue was removed from fixative and rehydrated in (1:1) methanol:PBST (PBS, 1% triton) for 15 min followed by a 15-min wash in 100% PBST. Samples were put in a blocking solution (PBS, 1% triton, 2% powdered milk) for 1 h at room temperature. Samples were then stained with primary antibodies rabbit anti-cytokeratin 8 (Abcam, ab53280, 1:500) and mouse anti-CD31 (Abcam, ab9498, 1:500) in blocking solution for 7–14 days at 4°C . Samples were then subjected to two 15-min and three 45-min washes of PBST. Samples were then incubated with secondary antibodies, donkey anti-rabbit IgG Alexa Fluor 555 (Invitrogen, A31572, 1:300), donkey anti-mouse IgG Alexa Fluor 647 (Invitrogen, A31571, 1:300), and Hoechst (B2261, Sigma Aldrich, 1:500) for 7 days at 4°C . Biopsy samples were then washed with PBST for 15 min followed by three 45-min washes, transferred to a methanol:PBST (1:1) solution for 5 min, and washed in 100% methanol for 15 min. Samples were incubated at 4°C overnight in a 3% H_2O_2 solution prepared in Methanol. Biopsy samples were then washed with 100% methanol for 15 min once and then for an hour. Samples were then cleared in benzyl alcohol:benzyl benzoate (1:4) (Sigma Aldrich, 108006, B6630) overnight.

Samples were imaged using a Leica SP8 TCS white light laser confocal microscope utilizing 10× air objective and a 7.0- μm Z stack (Flores et al., 2020). Imaris v9.2.1 (Bitplane) commercial software was used to analyze confocal image files and create 3D renderings. The surface function in Imaris was used to create 3D renderings of uterine glands (cytokeratin 8), vasculature (PECAM), and whole sample (Hoechst). Volumetric data of the structures were taken from the surface statistics function. To normalize for different sizes of endometrial tissue, the gland

volume and the vasculature volume were normalized using the total volume of endometrial tissue from the Hoechst staining.

Immunostaining

Immunofluorescence (IF) and immunohistochemistry (IHC) staining were performed on formalin-fixed and paraffin-embedded tissue, sectioned at 5 μm , using commercially available antibodies to detect Ki-67 (Abcam, Cat. # ab15580) and the B-cell marker, CD20 (Sigma Aldrich, Cat. # HPA014341). Sections were deparaffinated and rehydrated in a graded alcohol bath. Heat-mediated antigen retrieval was performed using a pressure cooker. For IF, the sections were blocked with bovine serum albumin and then incubated overnight in a humidified chamber at 4°C with the primary antibody. On the second day, the sections were incubated with a secondary antibody (AF594 conjugated anti-rabbit antibody, 1:1000, Invitrogen Cat. # A21207), counterstained with DAPI, and mounted. For colorimetric IHC, biotinylated anti-rabbit (Vector Laboratories Cat. # BA-1000, 1:400), avidin/biotin blocking kit (Vector SP-2001), Vectastain ABC kit, and DAB substrate kit (Vector SK-4100) were used. The specificity of primary antibodies was previously determined (Yeaman et al., 2001; Hu et al., 2018). Slides stained with secondary antibody only served as negative controls for IF and IHC staining.

Cell counts for Ki-67+ gland epithelial and stromal cells

Two-dimensional sections of endometrial tissue samples were stained for nuclei (DAPI) and proliferating cells (Ki-67) and imaged. Three sections $\sim 100 \mu\text{m}$ apart were analyzed from each biopsy. For quantitation, images were imported into Imaris v9.2.1 (Bitplane) with Matlab (XT) module. To isolate glandular epithelium, a surface was made from the DAPI channel using Background Subtraction (Local Contrast) with the default Surfaces Detail value. Because gland epithelial cells are situated closer together than stromal cells and stain brighter with DAPI, this step filters out many stromal cells. If the staining was suboptimal, then multiple surfaces were created for smaller regions of interest surrounding the glands before merging the surfaces into one. Once the surface for glandular cells was created, any remaining non-glandular cells were removed manually. The clean gland surface was then masked under the DAPI channel. A Gaussian blur (of 20) was applied, and then, using Channel Arithmetics function, the channel intensity was amplified up to 30-fold to create a new channel. A surface was made from this new glands channel while using the function Split Touching Objects (Region Growing) to count all individual gland epithelial cells.

To isolate the stroma, a new channel was created by subtracting the newly made glands channel (see above) from the DAPI

channel using the Channel Arithmetic function. Similar to above, a Gaussian blur (of 20) was applied, and channel intensity was amplified up to 30-fold to create a new stroma channel. A surface was made from the new channel while using the function Split Touching Objects (Region Growing) to count all stromal cells.

To obtain the number of proliferating gland cells, Channel Arithmetics was used to multiply the Ki-67 channel with the new glands channel. Similarly, to obtain the number of proliferating stromal cells, Channel Arithmetics was used to multiply the Ki-67 channel with the new stromal channel. A surface was then created for each using Split Touching Objects. The number of separated objects was then exported from Imaris for statistical analysis and to calculate the proportion of proliferating gland cells and stroma cells in each section.

Flow cytometry and cell sorting

For flow cytometry and cell sorting, endometrial tissues were finely minced with scissors and then enzymatically digested in pre-warmed RPMI-1640 containing 1mg/ml Collagenase A (Roche, 11088793001) and 0.1mg/ml DNase I (Roche, 10104159001) at 37°C in a benchtop shaker. After 15 min of shaking, tissues were disrupted by repeated passage through an 18-G needle attached to a syringe, and then shaken for an additional 15 min at 37°C. Cells were then washed with RPMI-1640 containing 3% FBS (RPMI-3) and filtered through a 100-µm cell strainer. Red blood cells (RBCs) were lysed with RBC Lysis Buffer (Tonbo Biosciences, TNB-4300-L100). Cells were stained with LIVE/Dead Fixable Aqua (Invitrogen, L34965A), according to the manufacturer's instructions. Staining for surface antigens (including isotype controls) was performed by incubating cells with fluorophore-conjugated antibodies for 30 min on ice. In some cases, surface staining was followed by intranuclear staining using the Foxp3 Transcription Factor Staining Buffer Kit (eBioscience, 00-5523-00), as per the manufacturer's instructions.

The following antibodies were used for flow cytometry or fluorescence-activated cell sorting (FACS): from BD Biosciences: CD45-BUV395 (clone HI30); from BioLegend: CD14-FITC, BV421 and APC-Fire 750 (M5E2), CD3-FITC and PerCP-Cy5.5 (UCHT1), CD19-APC (SJ25C1), CD16-FITC (3G8), CD4-PE-Cy7 (OKT4), TCRgd-APC (B1), IgD-BV421 (IA6-2), IgM-APC-Fire 750 (MHM-88), CD86-PerCP-Cy5.5 (IT2.2), CD21-PE-Cy7 (Bu32), BAFFR-PE (11C1), CD27-BV711 (O323), CD38-BV785 (HIT2), CD24-BV605 (ML5), EpCAM-PerCP-Cy5.5 (9C4) and CD309 (VEGFR2)-BV421 (A16085H); and from Thermo Fisher Scientific (eBioscience): CD56-APC (CMSSB), CD3-APC-eFluor 780 (UCHT1), CD19-APC-eFluor 780 (SJ25C1), CD8-FITC (RPA-T8), Foxp3-eFluor 450 (236A/E7), and CD31-APC-eFluor 780 (WM59). Flow cytometric analyses were performed using a Fortessa X20 cytometer (BD) and FlowJo v10 software (BD).

For analyses on immune cells, cells were first gated on FSC and SSC parameters, then on singlets, live cells, and CD45⁺ cells. Among live CD45⁺ cells, monocytes/macrophages were identified as CD3[−]CD19[−]CD14⁺ cells, eosinophils as CD3[−]CD19[−]CD14[−]SSC^{hi} cells, NK cells as CD56^{hi/dim}CD3[−]CD14[−]CD19[−] cells, T cells as CD3⁺CD19[−] cells, with further identification of effector CD4⁺ T cells (CD4⁺Foxp3[−]), CD8⁺ T cells, γδ T cells (TCRγδ⁺), and Tregs (CD4⁺Foxp3⁺), and B cells were identified as CD19⁺CD3[−]CD14[−] cells.

For analysis of endometrial B-cell subsets, mononuclear cells were first enriched from enzymatically dissociated endometrial tissue using Lymphoprep (STEMCELL, 20144) and then endometrial B cells were further enriched using the EasySep Human Pan-B Cell Enrichment Kit (STEMCELL, 19554), according to manufacturer's instructions, and then cells were surface stained as above except DAPI was used for dead cell exclusion. B cells were

gated as live CD45⁺CD19⁺CD3[−]CD14[−] cells, then memory B cells were identified as CD38[−]CD24⁺ cells, regulatory B cells as CD38⁺CD24⁺ cells, plasma cells as CD38⁺CD24[−] cells, and naive B cells as CD38[−]CD24[−].

For analyses on endothelial cells, cells were first gated on FSC and SSC parameters, then as singlets, live cells, and CD45[−] cells. Among live CD45[−] cells, endothelial cells were identified as live EpCAM[−]CD31⁺ or EpCAM[−]VEGFR2⁺ cells.

Imaging

H&E, IF, and IHC images were captured using the 20x objective of the Keyence BZ-X710 All-in-One fluorescence microscope with Keyence BZ-X Viewer version 01.03 software (Keyence, Osaka, Japan). Standard filters were used to image DAPI, Alexa Fluor 488, and Alexa Fluor 594.

Statistical analysis

Data were analyzed to identify outliers, and normality was determined using the Shapiro-Wilk test. For non-parametric data, medians were compared using the Mann-Whitney U test or Kruskal-Wallis with post hoc multiple comparisons tests and data are expressed as medians with interquartile range (IQR). For normally distributed data, means were compared using unpaired t-tests or ordinary one-way ANOVA with Tukey's multiple comparisons tests and data are expressed as mean ± SD. Statistical analyses were performed using Prism Version 9.0 (GraphPad). Statistical significance was defined as *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

Results

Demographic data

Study participants with regular, ovulatory menstrual cycles and proven fertility were enrolled in one of three groups that corresponded to endometrial biopsy collection in the proliferative, periovulatory, or mid-secretory phase of a NC (Fig. 1B). The demographics of participants in the three NC groups were similar with respect to age, BMI, gravidity (number of prior pregnancies), and parity (number of live births) (Supplementary Table S1).

Women undergoing ART were enrolled in one of two groups that corresponded to endometrial biopsy collection in the periovulatory or mid-secretory phase of an OS cycle. Within the periovulatory OS cycle group, 3 out of 16 subjects had significantly fewer oocytes collected (2.0 vs 16.0, P = 0.008) and significantly lower peak serum E2 levels (E2 1036 ± 254.2 vs 1806 ± 802.6 pg/ml, P = 0.04) than the rest of the periovulatory cohort (Supplementary Table S2), consistent with poor response to OS (Blumenfeld, 2020). These three subjects also received a higher total dose of gonadotropins compared to the periovulatory OS group (6458 ± 437 vs 3612 ± 1359 IU, P = 0.002) and the mid-secretory OS group (6458 ± 437 vs 5050 ± 907 IU, P = 0.18). Having identified a unique group of poor responders (OS-P), these subjects were analyzed as a separate group. When comparing ages across the three OS groups (OS-P, periovulatory OS, and mid-secretory OS), there was a statistically significant difference (P = 0.03) detected. However, after adjusting for multiple comparisons, the post hoc pairwise tests of significance only showed a trend toward statistical significance (OS periovulatory vs OS mid-secretory: 33.0 ± 1.9 vs 36.1 ± 4.3, P = 0.07; OS periovulatory vs OS-P periovulatory: 33.0 ± 1.9 vs 37.0 ± 1.0, P = 0.09). After removing the OS-P subjects, comparison of oocyte yields and peak E2 levels was similar in the OS periovulatory and OS mid-secretory phase groups (Supplementary Table S2), suggesting similar individual responses to the ART clinical protocols used for OS.

When comparing subjects in OS cycles (with OS-P removed) vs NCs, we found that the mean age of the OS cycle subjects was significantly higher than the NC subjects in both the periovulatory phase (33.0 vs 29.4, $P=0.02$) and the mid-secretory phase (36.1 vs 30.4, $P=0.03$) (Table 1). The median BMIs of OS and NC subjects were similar in the periovulatory (25.5 vs 28.4, $P=0.08$) and mid-secretory (27.3 vs 28.9, $P=0.36$) phases (Table 1).

OS induces higher serum E2 and P4 in both the periovulatory and mid-secretory phases

Given that circulating E2 and P4 levels reflect ovarian function and menstrual cycle stage in NCs and the response to exogenous gonadotropin stimulation in OS cycles, we measured serum E2 and P4 at the time of endometrial biopsy to validate the timing of sample collection and to identify differences in the natural and OS cycle groups. One subject was excluded because her cycle phases and serum hormones were inconsistent (Fig. 1A). We compared serum E2 and P4 levels from the NC, OS-P, and OS groups in the periovulatory phase and those from NC and OS groups in the mid-secretory phase. Serum E2 and P4 levels were significantly higher in the OS group compared to the NC group in the periovulatory phase (E2: 978.8 ± 477.3 vs 182.5 ± 85.4 pg/ml, $P < 0.0001$; P4: 12.1 ± 5.3 vs 1.0 ± 0.7 ng/ml, $P < 0.0001$) and the mid-secretory phase (E2: 682.2 ± 557.7 vs 160.5 ± 114.6 pg/ml, $P = 0.01$; P4: 19.0 ± 15.1 vs 11.0 ± 3.6 ng/ml, $P = 0.0005$) (Table 1).

OS induces glandular-stromal dyssynchrony in the periovulatory phase

We next performed blinded histological dating on endometrial samples from each group to validate timing of sample collection, confirm sample integrity, and compare relative stromal and glandular dating in NC and OS conditions. Glands and stroma with features of pre-ovulatory endometrium are called 'proliferative'. Post-ovulation, glands and stroma are assigned a date, starting with 'CD 15' when the characteristic features of ovulation are first observed, and ending with 'CD 28' when decidualization is complete and breakdown becomes visible in the absence of a pregnancy (Noyes et al., 1975). For our analyses, we converted the conventional CD nomenclature to post-ovulatory day (POD), whereby CD 15 is equivalent to POD 1, CD 16 is equivalent to POD 2, and all proliferative samples were assigned POD -1.

Among the NC subjects, we found that some periovulatory samples were dated as proliferative, with tubular endometrial glands containing columnar cells and dense stroma, whereas others possessed characteristic features of the endometrium in the first few PODs, such as subnuclear and supranuclear vacuoles in the glandular epithelium (Fig. 2A). Among the NC subjects in the proliferative phase, glandular and stromal dating was synchronous in 6/6 subjects and is designated as POD -1 (Fig. 2B). Among NC subjects in the periovulatory phase, glandular dating ranged from POD -1 to POD 4 and stromal dating ranged from POD -1 to POD 4 (Fig. 2B). As expected, samples

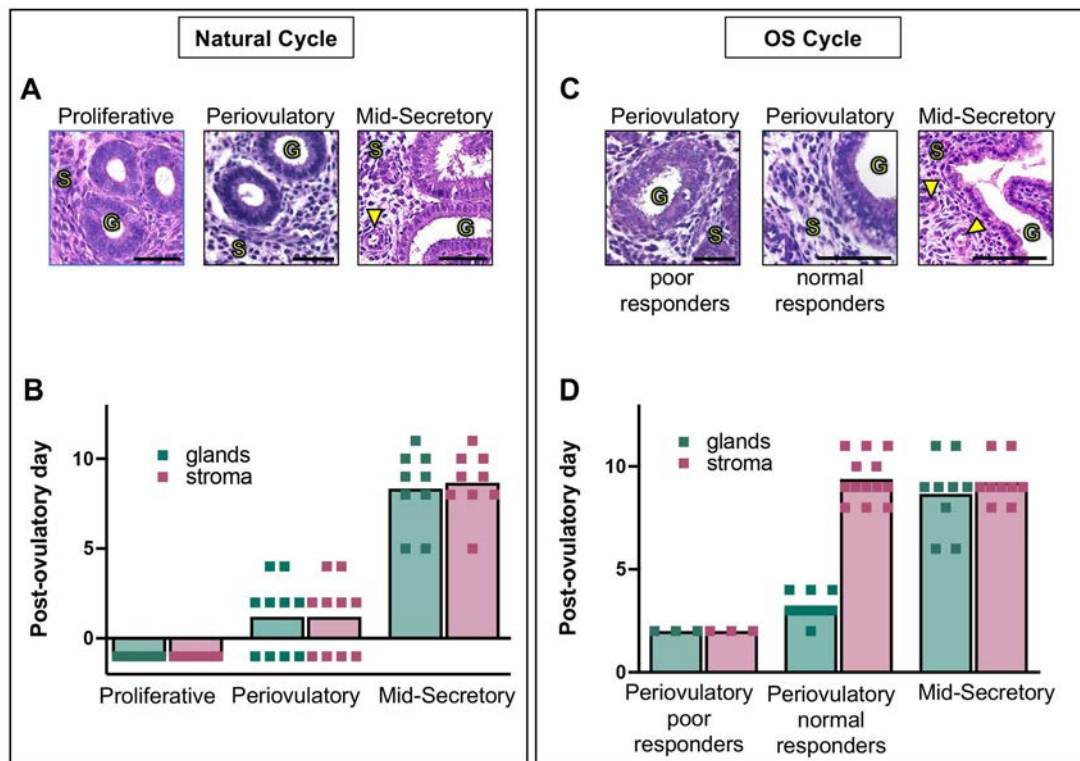


Figure 2. Glandular-stromal dyssynchrony in the periovulatory phase endometrium of subjects with normal response to ovarian stimulation. (A) Representative images of H&E-stained endometrial biopsy tissue sections from natural, unstimulated cycles in the proliferative phase showing tubular glands and dense stroma, periovulatory phase showing tubular glands, and mid-secretory phase showing tortuous glands, and stromal edema. (B) Conventional cycle day was converted from endometrial dating nomenclature to POD, whereby cycle day 15 is equivalent to POD 1, cycle day 16 is equivalent to POD 2, and all proliferative samples were assigned POD -1. Histologic dating of glands and stroma demonstrates glandular-stromal synchrony in the proliferative, periovulatory, and mid-secretory phases of natural cycles. (C) Representative images of H&E-stained endometrial biopsy tissue sections from OS cycles at the periovulatory phase (poor responders showing subnuclear and supranuclear vacuoles, normal responders showing subnuclear and supranuclear vacuoles and stromal edema) and mid-secretory phase showing tortuous glands, stromal edema, and spiral arterioles. (D) Histologic dating of glands and stroma demonstrates glandular-stromal dyssynchrony in the periovulatory phase of normal responders, but not poor responders. Each dot represents one subject. Scale bars = 50 μ m; yellow arrowhead: spiral arteriole. G, glands; H&E, hematoxylin and eosin; OS, ovarian stimulation; POD, post-ovulatory day; S, stroma.

from NC subjects in the mid-secretory phase were characterized by tortuous glands, stromal edema, and the presence of spiral arterioles (Fig. 2A), with glandular dating that ranged from POD 5 to 11 and stromal dating that ranged from POD 5 to 11 (Fig. 2B). We found that glandular and stromal dates were synchronous in 10/10 NC subjects in the periovulatory phase and in 8/9 NC subjects in the mid-secretory phase.

In contrast to findings in NC subjects, samples from OS subjects exhibited dyssynchrony in glandular-stromal dating, with stromal advancement relative to glandular dating in the periovulatory phase. Specifically, glandular dating ranged from POD 3 to 14, whereas stromal dating ranged from POD 8 to 11, for 13/16 OS subjects sampled in the periovulatory phase (Fig. 2C and D, periovulatory normal responders). However, for 3/16 OS subjects sampled in the periovulatory phase (Fig. 2C and D, periovulatory poor responders), both glandular and stromal dating ranged from POD 2–3 and thus did not show dyssynchrony. In the mid-secretory phase of OS cycles, glandular dating ranged from POD 6 to 12 and stromal dating ranged from POD 8 to 12, which corresponded with synchronous glandular-stromal dates in 8/9 OS subjects (Fig. 2D). Taken together, these data show that OS causes glandular-stromal dyssynchrony only in the periovulatory phase, likely resulting from endometrial exposure to supraphysiologic levels of E2 that accompany normal, but not poor responses, to OS. Following group validation by serum hormone levels and histology, the OS-P group was not included in the subsequent analyses.

OS induces greatest transcriptomic differences in the periovulatory phase

To determine global changes in endometrial gene expression, we performed bulk RNA sequencing on 19 endometrial tissue samples, including all validated NC and OS samples and excluding poor responders (OS-P). Our objectives were to determine how the endometrial transcriptome changes within natural and OS cycles and how OS impacts endometrial gene expression at both the periovulatory and mid-secretory phases.

Principal component analysis (PCA) shows clustering of both natural and OS cycle samples into distinct groups that reflect endometrial tissue collection at the periovulatory or mid-secretory phases (Fig. 3A). As shown in the volcano plot (Fig. 3Bi), comparison of the mid-secretory to periovulatory phase of NCs identified 2236 up- and 2495 down-regulated differentially expressed genes (DEGs) with adjusted $P < 0.05$ and $\log_2FC \geq |1|$. Similarly, comparison of the mid-secretory to periovulatory phase of OS cycles identified 2114 up- and 2198 down-regulated DEGs with adjusted $P < 0.05$ and $\log_2FC \geq |1|$ (Fig. 3Bii). When comparing OS and NCs, PCA shows a greater difference in gene expression in the periovulatory compared to the mid-secretory phase (Fig. 3A). Consistent with the PCA, volcano plots show a larger number of up- and down-regulated DEGs with adjusted $P < 0.05$ and $\log_2FC \geq |1|$ in the periovulatory phase (857 up and 1119 down in OS vs. NC) (Fig. 3Biii) as compared to the mid-secretory phase (6 up and 13 down in OS vs. NC) (Fig. 3Biv). Thus, the greatest differences in endometrial gene expression in OS vs NC subjects occur in the periovulatory phase, consistent with the findings in the histologic dating of the glandular epithelium and stroma.

Biological pathways associated with the transition from the periovulatory phase to mid-secretory phase in natural and OS cycles

We used IPA to perform a comparative analysis to identify the pathways that were significantly altered (with adjusted $P < 0.05$

and $\log_2FC \geq |1.2|$) in the transition from the periovulatory phase to mid-secretory phase of natural and OS cycles. The heatmap (Fig. 3C) reveals similar up- and down-regulation in natural and OS cycles for most pathways, including significant up-regulation of the Pathogen Induced Cytokine Storm Signaling Pathway and Acute Phase Response Signaling, and down-regulation of Cyclins and Cell Cycle Regulation, and the Kinetochore Metaphase Signaling Pathway.

We then used IPA to identify the top five differentially-regulated pathways and linked functions for the following categories: canonical pathways, upstream regulators, diseases and disorders, molecular and cellular functions, and physiologic system development and function (Supplementary Fig. S1A and B and Fig. 4A and B). In the NC, the mid-secretory vs periovulatory phase comparison revealed alterations in pathways related to the cell cycle and injury repair, as well as differential regulation of pathways or molecular and cellular functions associated with beta-estradiol and P4 (Supplementary Fig. S1A). In the OS cycle mid-secretory vs periovulatory phase comparison, many of the top five pathways were identical to those in the NC comparison (Supplementary Fig. S1B). Notably, differential regulation of the immune cell trafficking canonical pathway was unique to OS. Though beta-estradiol was differentially regulated in the OS cycle comparison, P4 was not.

When comparing the natural and OS cycles in the periovulatory phase, Wnt/ β -catenin signaling was differentially regulated with OS, as were molecular and cellular functions associated with dopamine receptor DRD1, beta-estradiol, and EOMES (Fig. 4A). When comparing the natural and OS cycles in the mid-secretory phase, the inflammatory responses and estrogen biosynthesis pathways were differentially regulated (Fig. 4B). Taken together, our data demonstrate that OS alters many pathways related to processes that are important for endometrial differentiation, such as cell proliferation, cellular differentiation, and immune cell function.

Glandular proliferation decreases over the course of natural, but not OS, cycles

Given our findings that cell cycle pathways are altered by OS, we sought to examine cellular proliferation in the endometrium. To determine the pattern of proliferation in natural menstrual cycles, we performed IF staining using the proliferation marker, Ki-67, a protein expressed in the nucleus of cells undergoing active proliferation (Fig. 5A–C). The proportion of Ki-67+ cells was quantified in the proliferative, periovulatory, and mid-secretory phases in both the glandular epithelium and stroma, selecting three subjects per group that were similar to one another in terms of histology and serum hormone levels. For each subject, three slides at least 100 μm apart were scored. In NC subjects, the proportion of Ki-67+ cells in both the glandular epithelium and stroma decreased from the proliferative phase to the mid-secretory phase (Fig. 5A–C and F). In contrast, in OS cycles, the proportion of Ki-67+ cells in the glandular epithelium and stroma did not decrease from the periovulatory to mid-secretory phase (Fig. 5D, E, and G). When the NC and OS groups were compared at the periovulatory and mid-secretory timepoints, there was no difference in the proportion of Ki-67+ cells between these two groups in either the glandular epithelium or stroma (Fig. 5H). Compared with NC, OS was associated with greater within-subject variance in Ki67 expression in both the glandular epithelium and stroma in the periovulatory phase, but less so in the mid-secretory, phase (Supplementary Fig. S2A and B). Taken together, our data suggest that the decrease in glandular epithelial proliferation seen in the NC between the periovulatory and

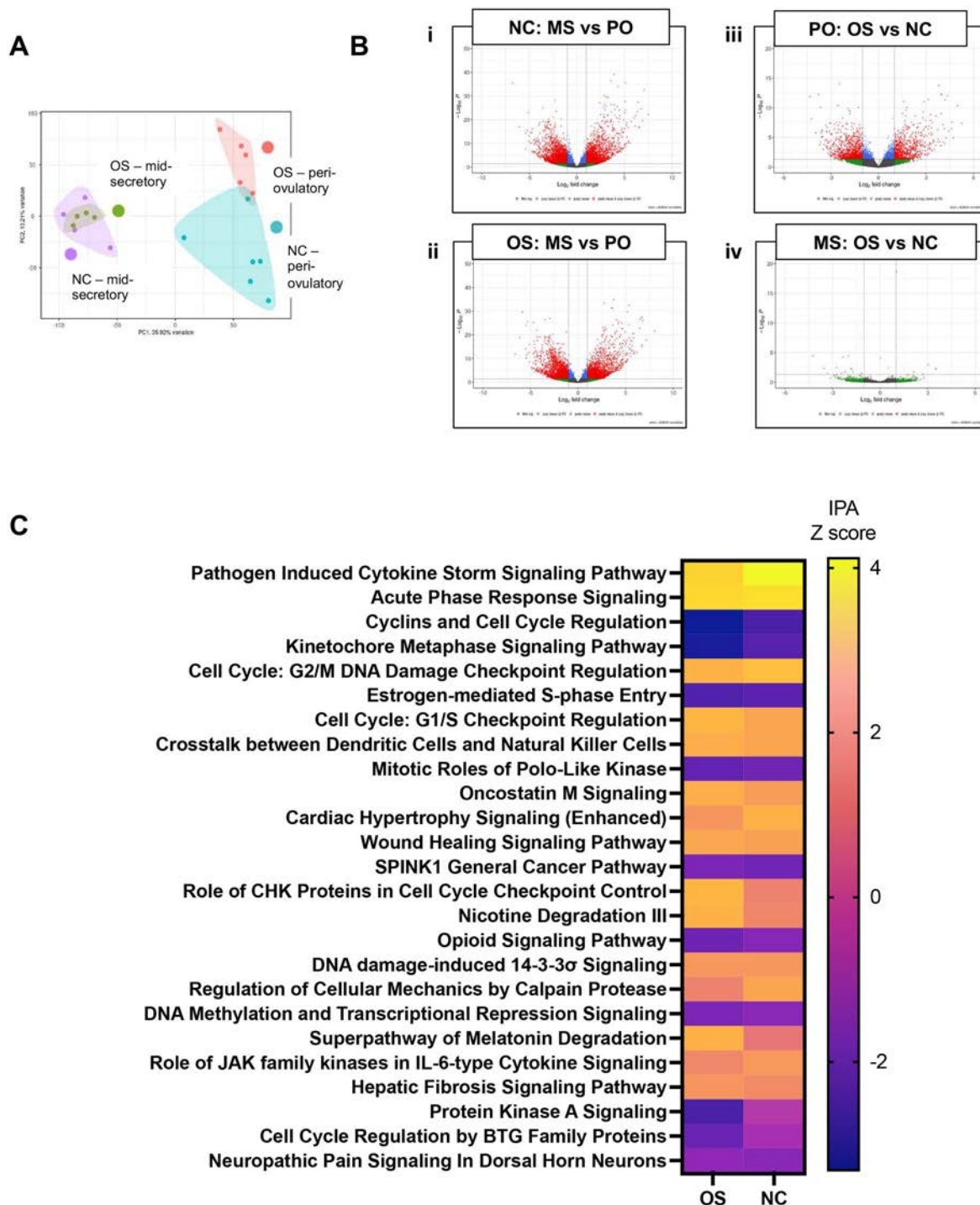


Figure 3. Differentially expressed genes and altered pathways in natural and ovarian stimulation cycles. (A) PCA plot demonstrating clustering of study groups. (B) Volcano plots demonstrating differentially expressed genes in each of the four group comparisons (i–iv). Genes with $-\log_{10}P$ -value < 0.05 and $\log_2$ fold change $> |1|$ are represented in red. (C) Heat maps based on ingenuity pathway analysis (IPA) Z-score of the top 25 altered pathways in the comparison of peri-ovulatory vs mid-secretory phases of ovarian stimulation and natural cycles. MS, mid-secretory; NC, natural cycle; OS, ovarian stimulation; PCA, principal component analysis; PO, peri-ovulatory.

mid-secretory phases is less pronounced in OS subjects; this may be due to increased variance in OS samples.

OS does not alter the frequency of endometrial endothelial cells

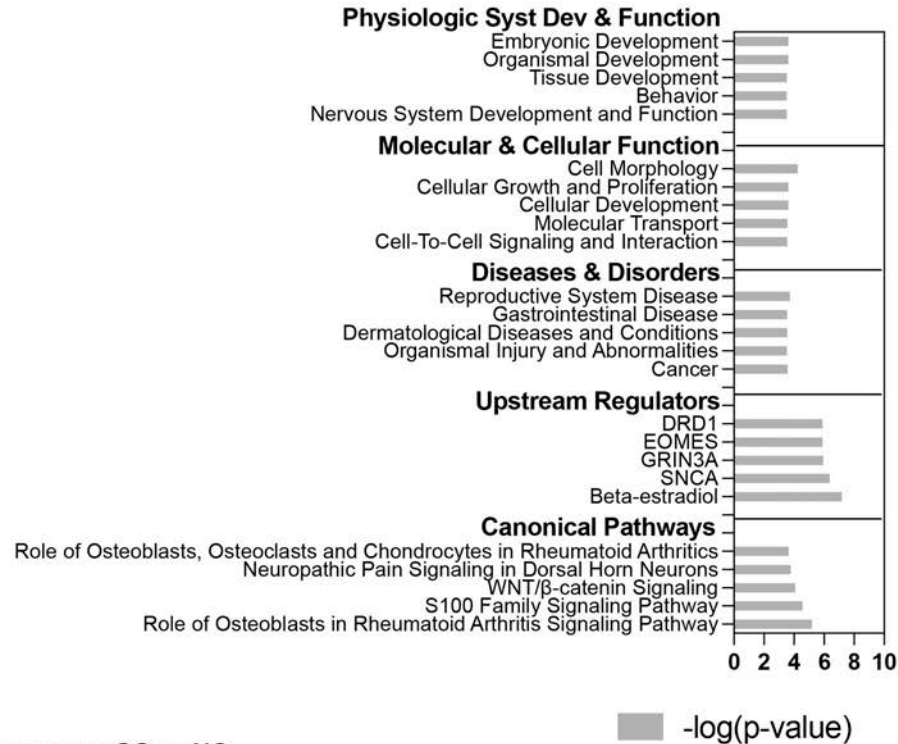
Changes in the endometrial vasculature during the menstrual cycle are thought to be directly regulated by estrogens and P4 (Massri et al., 2023). Previous studies have found conflicting

results with respect to changes in genes and proteins associated with the vasculature during OS cycles, with some reporting a reduction in the expression of certain vascular genes (Junovich et al., 2011) and some reporting an increase in others (Lee et al., 2008). In order to determine whether OS causes alterations in the vascular compartment on a cellular level, the frequency of CD31 $^{+}$ and VEGFR2 $^{+}$ cells was assessed by flow cytometry. CD31 (PECAM-1), a protein expressed on endothelial cells, can be used

as a marker of vasculature differentiation. VEGFR2 (KDR) is also expressed on endothelial cells and is activated by binding VEGFA, -C, and -D to mediate angiogenic processes (Wang et al., 2020). Most endothelial cells in the endometrium express these two markers, and they can be used to quantify vasculature (Tawia et al., 1993; Mints et al., 2002).

We found that in natural menstrual cycles, the frequency of CD31⁺ cells remained unchanged from the proliferative to periovulatory phase (1.3 vs 0.83%, $P=0.07$), but decreased from the proliferative to mid-secretory phase (0.83 vs 0.3%, $P=0.002$) (Fig. 6A). Similarly, the frequency of VEGFR2⁺ cells decreased from the proliferative to mid-secretory phase (1.09 vs 0.34%, $P=0.006$).

A Periovulatory: OS vs NC



B Mid-Secretory: OS vs NC

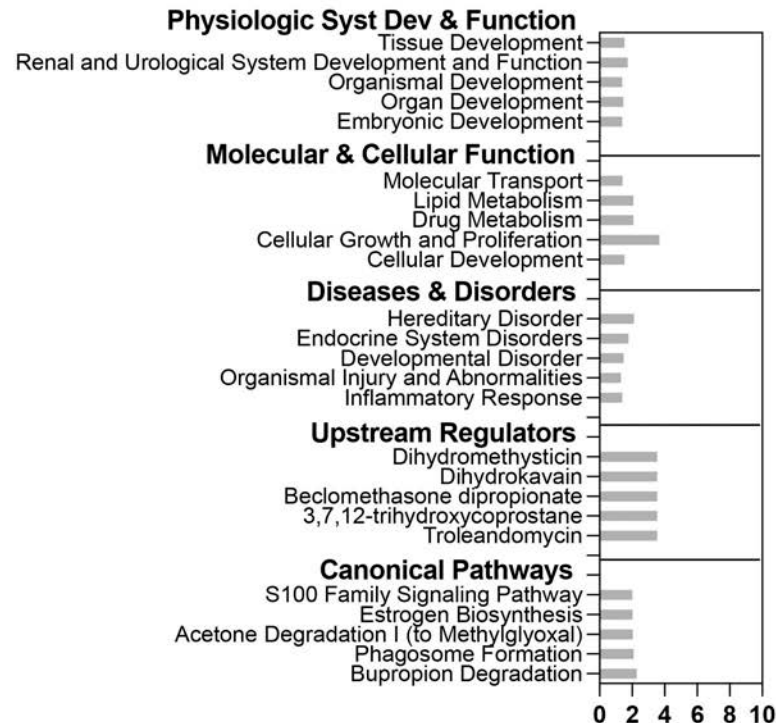


Figure 4. Ovarian stimulation differentially regulates pathways in the periovulatory and mid-secretory phase endometrium. IPA was used to identify the top five differentially regulated pathways for the following categories: physiologic system development and function, molecular and cellular function, diseases and disorders, upstream regulators, and canonical pathways. (A) Comparisons between OS cycles and NCs in the periovulatory phase. (B) Comparisons between OS cycles and NCs in the mid-secretory phase. IPA, ingenuity pathway analysis; NC, natural cycle; OS, ovarian stimulation.

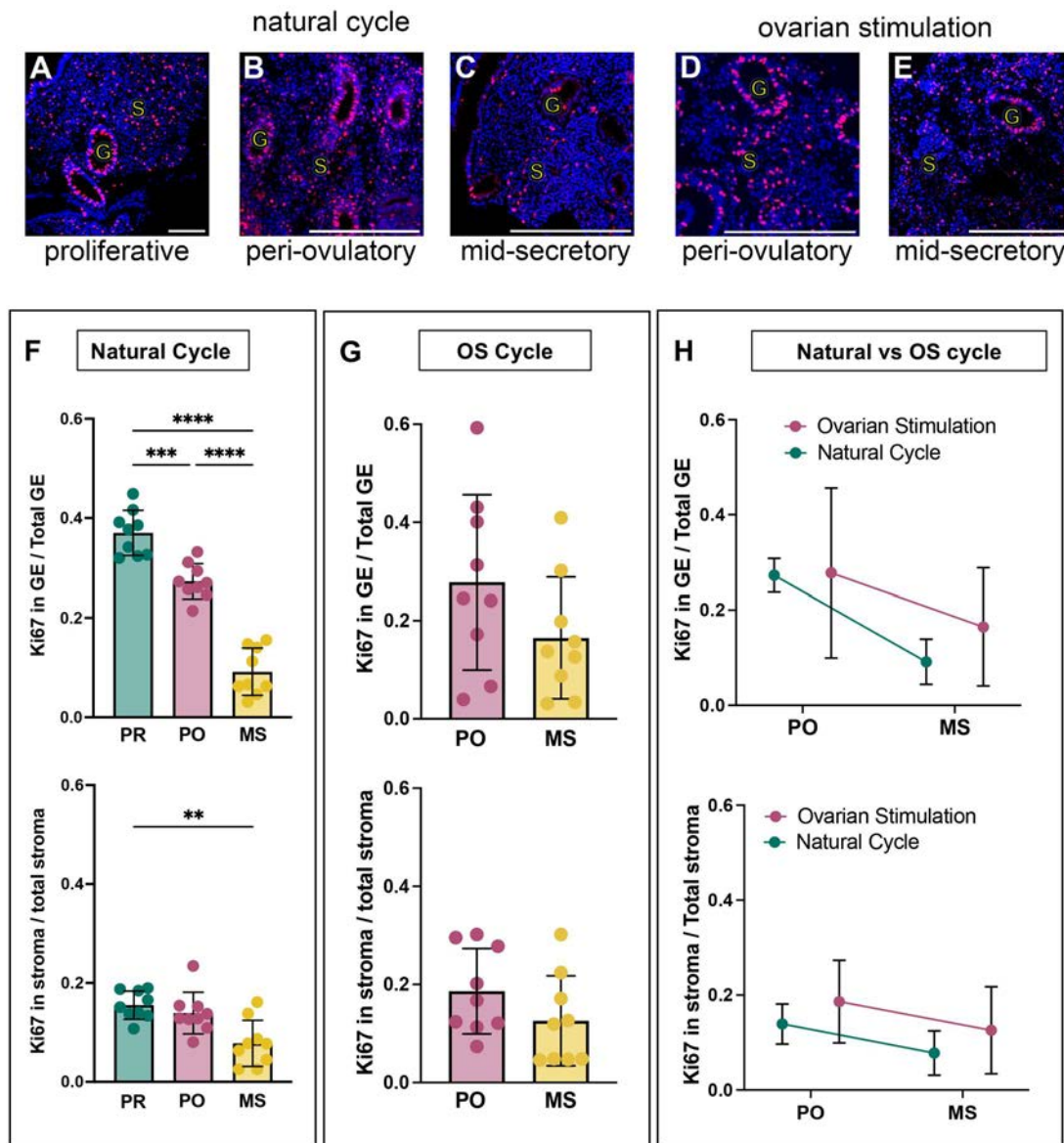


Figure 5. Cellular proliferation in the glandular epithelium and stroma in natural and ovarian stimulation cycles. Representative images of Ki67 staining in tissue sections from endometrial biopsies collected in the proliferative, peri-ovulatory, and mid-secretory phases of natural (A–C) and ovarian stimulation (D–E) cycles. (F and G) Quantification of Ki67 expression in the glandular epithelium and stroma. Each dot represents one of three slides scored for each subject. (F) For natural cycles, data were compared with an ordinary one-way ANOVA and Tukey's multiple comparisons tests and are shown as mean ± SD. (G) For ovarian stimulation cycles, data were compared with an unpaired t-test and are shown as mean ± SD. (H) Comparison of natural cycle and ovarian stimulation groups in the peri-ovulatory and mid-secretory phases, in both the glandular epithelium and stroma. Data were compared with an unpaired t-test and are shown as mean ± SD. G, glands; GE, glandular epithelium; MS, mid-secretory; OS, ovarian stimulation; PO, peri-ovulatory; PR, proliferative; S, stroma. ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. Scale bar (A) = 100 μm ; scale bar (B–E) = 250 μm .

(Fig. 6C). In OS cycles, the frequency of CD31⁺ cells decreased significantly from the peri-ovulatory to mid-secretory phase (1.09 vs 0.3%, $P = 0.003$) (Fig. 6B), whereas the decreased frequency of VEGFR2⁺ cells did not reach statistical significance (0.91 vs 0.23%, $P = 0.08$) (Fig. 6D).

No differences were seen in the frequency of CD31⁺ cells between NC and OS groups in the peri-ovulatory (1.09 vs 0.83%, $P = 0.43$) or mid-secretory (0.3 vs 0.3%, $P = 0.51$) phases (Fig. 6E). Similarly, no differences were seen in frequency of VEGFR2⁺ cells between NC and OS cycle groups in the peri-ovulatory (0.91 vs 0.58%, $P = 0.20$) or mid-secretory (0.23 vs 0.34%, $P = 0.79$) phases (Fig. 6F). These data suggest that the frequency of endothelial cells in the endometrium is similar in NC and OS cycles at both time points.

OS increases endometrial B-cell frequency at the mid-secretory phase

To understand how the composition of the endometrial immune cell compartment varies throughout the natural menstrual cycle, we used flow cytometry to quantify the frequency of the following immune cell types in the endometrium: CD8⁺ T cells, effector and regulatory CD4⁺ T cells, CD56^{hi} and CD56^{dim} NK cells, eosinophils, CD14⁺ monocytes/macrophages, B cells, and $\gamma\delta$ T cells (Fig. 7A–D). Among NC subjects, we found that the frequency of all immune cell types measured was relatively stable across the proliferative, peri-ovulatory, and mid-secretory phases (data not shown), except for CD14⁺ monocytes/macrophages and B cells (Fig. 7G and I). Specifically, the frequency of CD14⁺ monocytes/macrophages was significantly higher in the mid-secretory phase

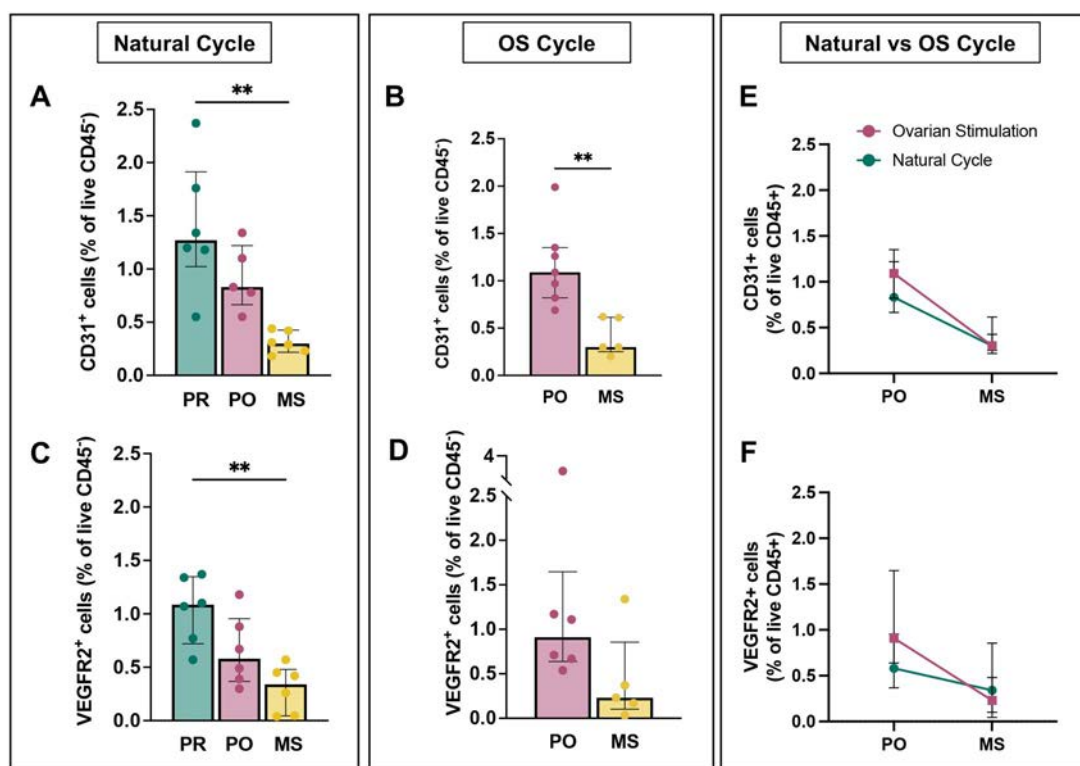


Figure 6. Frequency of endometrial endothelial cells in natural and ovarian stimulation cycles in the periovulatory and mid-secretory phases, measured by flow cytometry. Frequency of CD31+ cells among total live cells in natural (A) and ovarian stimulation (B) cycles. Frequency of VEGFR2+ cells among total live cells in natural (C) and ovarian stimulation (D) cycles. Comparison of the frequency of CD31+ cells (E) and VEGFR2+ cells (F) in natural and ovarian stimulation cycles in the periovulatory and mid-secretory phases. Data were compared with Kruskal–Wallis and Dunn’s multiple comparisons tests (A and C) or Mann–Whitney U tests (B, D, E, and F) with $^{**}P < 0.01$. Data in (A)–(F) are shown as median + IQR. CD31, cluster of differentiation 31; IQR, interquartile range; MS, mid-secretory; OS, ovarian stimulation; PO, periovulatory; PR, proliferative; VEGFR2, vascular endothelial growth factor receptor 2.

compared to the proliferative phase among NC subjects (8.4 vs 3.5%, $P = 0.005$) (Fig. 7G). The frequency of B cells was significantly higher in the mid-secretory phase compared to both the proliferative (3.6 vs 0.8%, $P = 0.001$) and periovulatory phases (3.6 vs 1.04%, $P = 0.02$) among NC subjects (Fig. 7I).

We next sought to understand how OS impacts the composition of the endometrial immune cell compartment. The frequency of most immune cell subsets was similar in OS and NC subjects, with the notable exception of B cells and to a much lesser extent CD14⁺ monocytes/macrophages and T cells (Fig. 7A–J). Specifically, the frequency of endometrial CD14⁺ monocytes/macrophages was modestly higher in OS vs NC subjects in the periovulatory (11.8 vs 7.0%, $P = 0.03$) but not the mid-secretory phase (7.5 vs 8.3%, $P = 0.11$), and the frequency of endometrial CD4⁺ effector T cells was modestly lower in OS vs NC subjects in the mid-secretory (14.2 vs 19.8%, $P = 0.02$) but not the periovulatory phase (15.2 vs 20.5%, $P = 0.08$) (Fig. 7E and F). In contrast, OS led to a striking ~8-fold increase in the frequency of endometrial B cells in the mid-secretory phase, as compared to the NC subjects (26.6 vs 1.1%, $P = 0.001$) (Fig. 7F, I, and J). IHC staining of mid-secretory phase endometrial tissue sections shows B cells scattered throughout the stroma with increased density in OS vs NC subjects (Fig. 7L). Taken together, our results suggest that OS selectively alters the frequency of monocytes/macrophages in the periovulatory phase, and of endometrial B cells and CD4⁺ T effector cells in the mid-secretory phase.

Because OS dramatically increased the frequency of endometrial B cells at the mid-secretory stage, we used flow cytometry to determine whether OS affected any B-cell subset specifically.

Analysis of endometrial B cells from OS showed a similar level of surface expression of the B-cell markers IgD, IgM, CD27, BAFFR, CD21, and CD86 compared to NC controls (data not shown). Similarly, analysis of endometrial B-cell subsets based on CD38 and CD24 expression as previously described (Blair et al., 2010; Zaimoku et al., 2020), revealed a similar distribution of naive, plasma, and memory cells within the B-cell compartment of NC and OS cycle samples (Fig. 7K). However, OS cycle samples had a higher relative frequency of CD38⁺CD24⁺ B cells, which have been previously identified as regulatory B cells or Bregs (Blair et al., 2010; Zaimoku et al., 2020). Overall, these results show that at the mid-secretory stage, OS leads to an increase in all B-cell subsets, with a slight relative increase in the frequency of CD38⁺CD24⁺ B cells with a regulatory phenotype (Fig. 7K).

OS alters glandular volume, but not vascular volume

Mouse studies using 3D reconstruction technology have demonstrated an increase in epithelial density, an increase in the ratio of glandular epithelium to luminal epithelium, and a decrease in gland diameter in mice undergoing OS compared with unstimulated controls (Qu et al., 2022). Using novel approaches to whole tissue analysis with 3D imaging (Arora et al., 2016), we sought to determine if gland volume and vasculature volume in intact, non-dissociated endometrial tissue reveals differences between natural and OS conditions that cannot be appreciated through analysis of dissociated tissue (Fig. 8A–D). To our knowledge, this is the first time this type of analysis has been performed on human endometrial tissue.

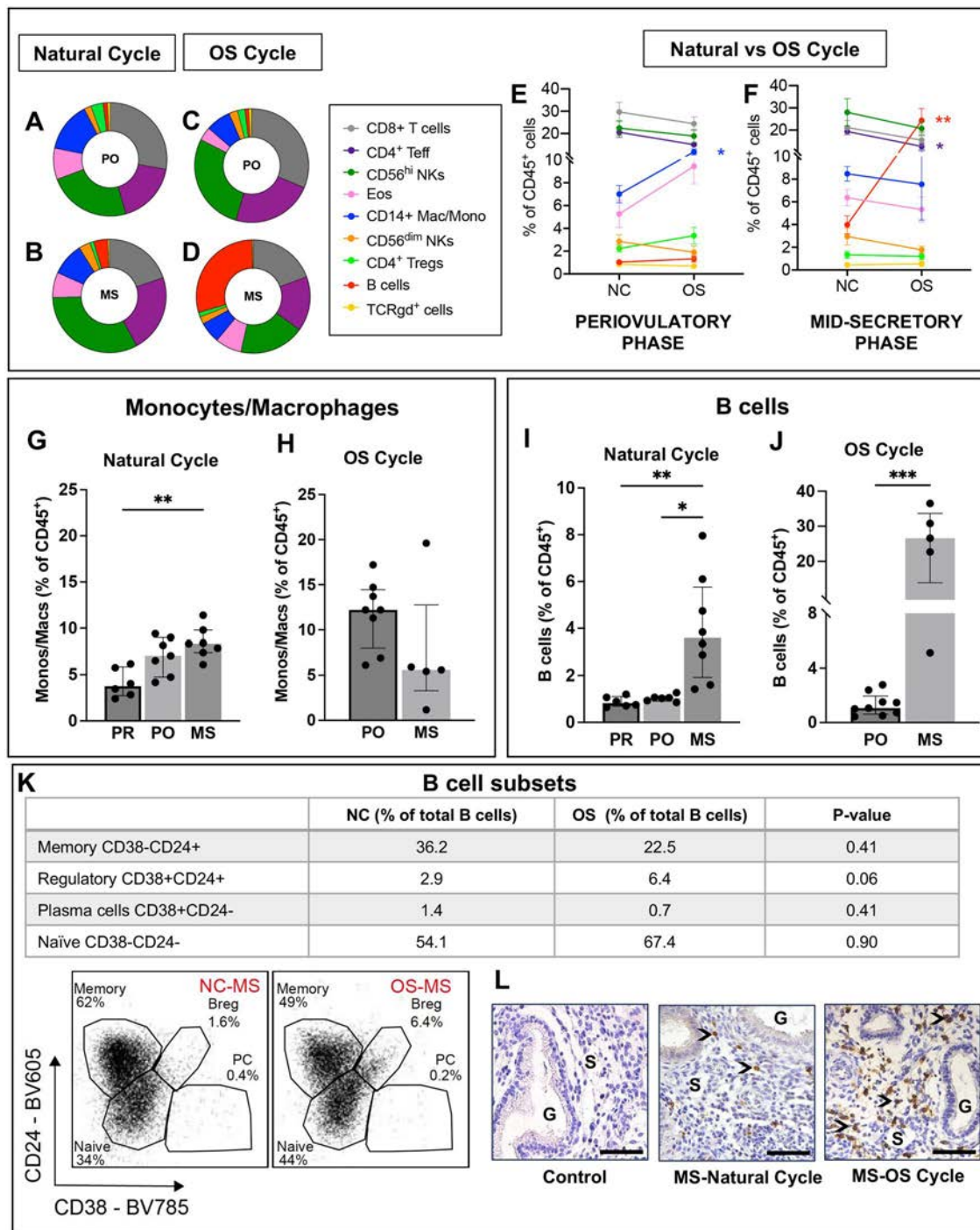


Figure 7. Frequency of endometrial immune cells in natural and ovarian stimulation cycles in the periovulatory and mid-secretory phases, measured by flow cytometry. Pie charts showing endometrial immune cell distribution among CD45⁺ cells in NC periovulatory (A) and mid-secretory (B) phases and ovarian stimulation cycle periovulatory (C) and mid-secretory (D) phases. Comparison of endometrial immune cell frequency among CD45⁺ cells across subjects in natural and OS cycles in the periovulatory (E) and mid-secretory (F) phases. Longitudinal comparison of CD14⁺ monocyte/macrophage frequency in natural (G) and OS cycles (H). Longitudinal comparison of B cell frequency in natural (I) and OS cycles (J). (K) Table showing frequency of indicated B cell subsets among total B cells in the mid-secretory phase of natural and OS cycles, with representative flow cytometry plot. (L) Representative images of endometrial tissue sections stained to detect expression of CD20, showing B cells in the endometrial stroma in the mid-secretory phase of natural and OS cycles. Data in (A)–(D) shown as median percentage, Mann–Whitney test. Data in (E), (F), (H), and (J) shown as median + IQR, Mann–Whitney test. Data in (G) and (I) shown as median + IQR, Kruskal–Wallis test. Data in (K) compared with Mann–Whitney test and shown as median percentage. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. G, glands; IQR, interquartile range; MS, mid-secretory; NC, natural cycle; OS, ovarian stimulation; PO, periovulatory; PR, proliferative; S, stroma. Black arrowheads = B cells. Scale bars = 50 μ m.

We first characterized changes in vascular volume as determined by three-dimensional analysis. There was no difference in vascular volume between the proliferative, periovulatory and mid-secretory phases of NCs (Fig. 8E) or between the

periovulatory and mid-secretory phases of OS cycles (Fig. 8F). When comparing natural to OS cycles, we found no difference in vascular volume at either the periovulatory or mid-secretory phases (Fig. 8G).

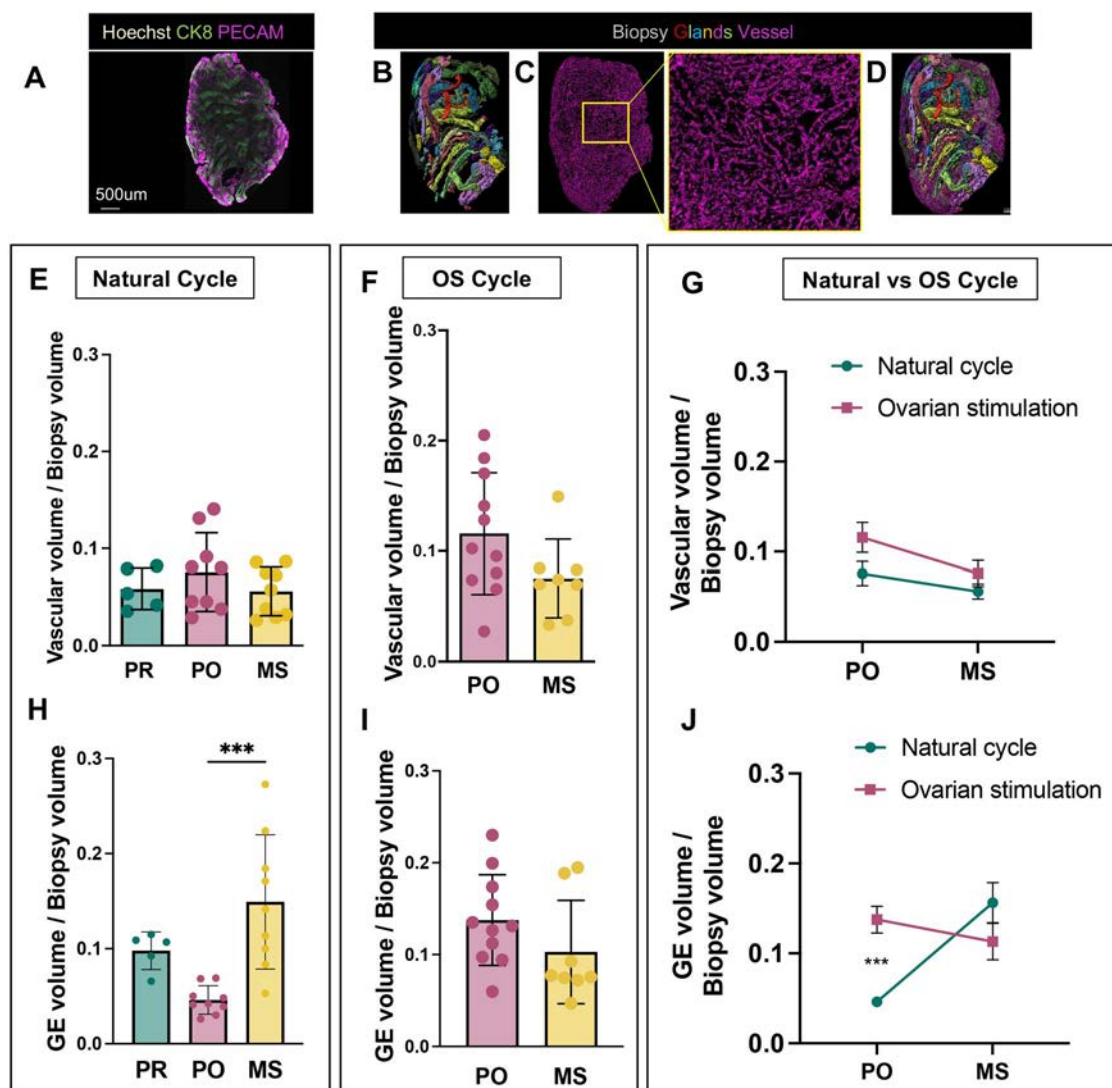


Figure 8. 3D analysis of glandular epithelium and vasculature. (A) Two-dimensional surface rendering of glandular epithelium (CK8) and vasculature (PECAM) in a natural cycle, periovulatory endometrial biopsy sample. (B) Three-dimensional rendering of glandular epithelium, with each gland depicted in a separate color. (C) Three-dimensional rendering of vasculature, inset showing detailed structure. (D) Three-dimensional overlay of glandular epithelium and vasculature. Vascular volume in natural (E) and ovarian stimulation (F) cycles. (G) Vascular volume from the PO to MS phases of NC and OS subjects. Glandular epithelial volume in natural (H) and ovarian stimulation (I) cycles. (J) Glandular epithelial volume from the PO to MS phases of NC and OS subjects. Data in (F), (G), (I), and (J) shown as mean $\pm$ SD, unpaired t-test. Data in (E) and (H) shown as mean $\pm$ SD, ANOVA with Tukey's multiple comparisons tests and *** $P < 0.001$. GE, glandular epithelium; MS, mid-secretory; OS, ovarian stimulation; PO, periovulatory; PR, proliferative.

We then characterized the changes in glandular volume in natural menstrual cycles by measuring volume in the proliferative, periovulatory, and mid-secretory phases of the menstrual cycle. We found a 3.2-fold increase in glandular volume from the periovulatory to mid-secretory phase in natural menstrual cycles (Fig. 8H), but no significant change from the periovulatory to mid-secretory phase in OS cycles (Fig. 8I). Next, we sought to determine whether there are differences in glandular volume between OS and NC groups. In OS compared with NCs, three-dimensional imaging analysis demonstrated a significant increase in gland volume in the periovulatory period, with no difference in the mid-secretory phase (Fig. 8J). In contrast to the histologic analysis which did not suggest glandular changes in the periovulatory phase, these data revealed an increase in gland volume at that time point, which may either be due to glandular enlargement or to an increase in the number of glandular epithelial cells.

Genes associated with development and differentiation of cellular compartments of the endometrium

To further understand the molecular impact of OS, we determined how expression of genes involved in the development and differentiation of the glands, stroma, vascular and immune compartments changed in the transition from the periovulatory phase to mid-secretory phase in both unstimulated natural and OS cycles. When examining genes associated with glandular development and differentiation, mRNA expression patterns were similar; there was significantly increased expression of *SPP1* and *PAEP* and decreased expression of *GLI1* and *IHH* in the mid-secretory compared to periovulatory phase of both natural and OS cycles (Fig. 9A). mRNA expression of markers of decidualization, *IGFBP1*, *PRL*, and *LEFTY2*, was more abundant in the mid-secretory phase vs periovulatory phase of natural menstrual cycles. In contrast, expression of *IGFBP1* and *PRL*, but not *LEFTY2*,

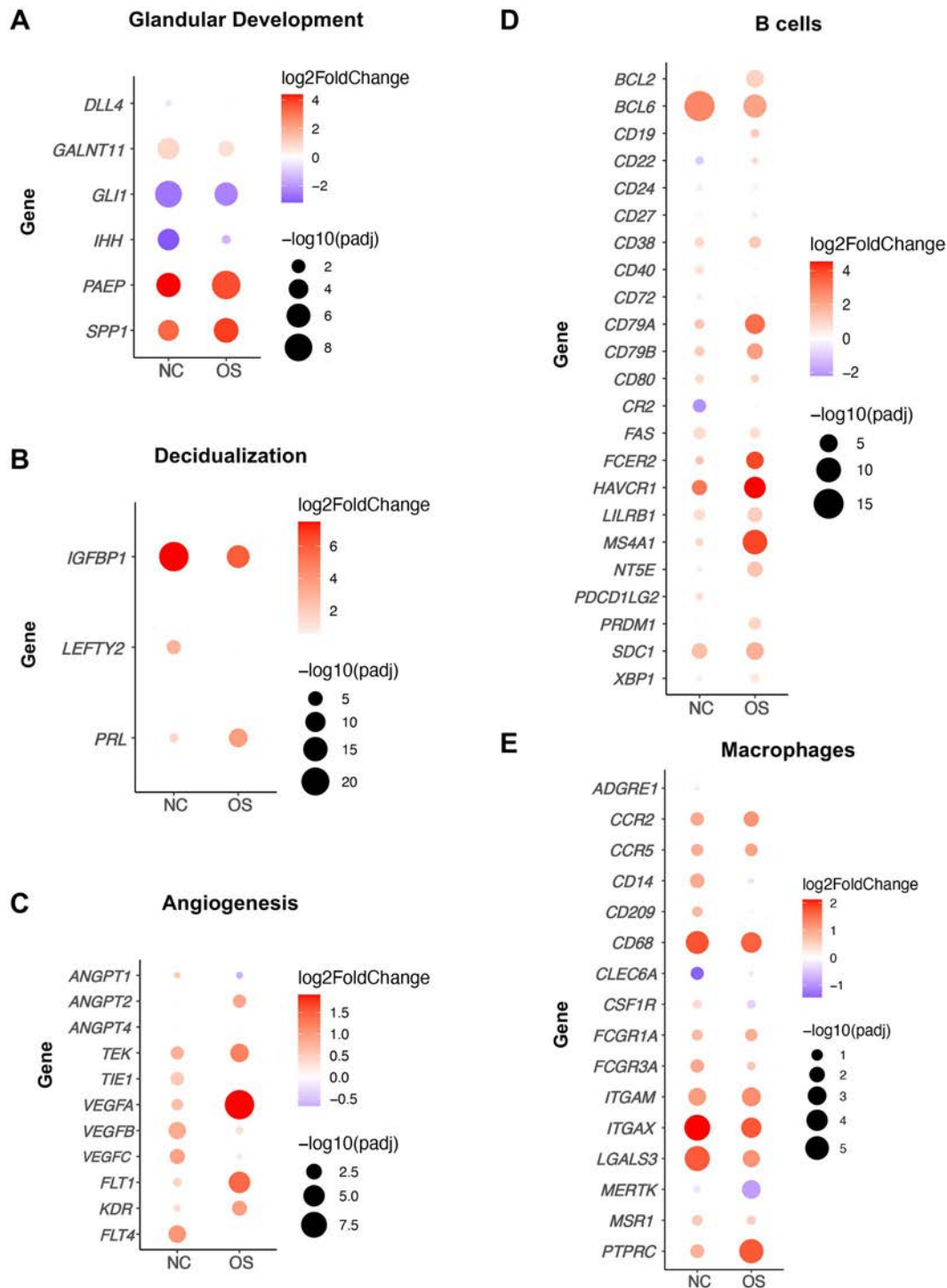


Figure 9. Differentially expressed genes in natural and ovarian stimulation cycles in periovulatory to mid-secretory phases. Dot plots showing differentially expressed genes in ovarian stimulation (periovulatory vs mid-secretory) and NCs (periovulatory vs mid-secretory) for genes that mediate glandular differentiation (A), stromal decidualization (B), angiogenesis (C), B cells (D), and macrophages (E). Log₂fold change represented by dot color, -log₁₀(adjusted P) represented by dot size. NC: natural cycle; OS, ovarian stimulation.

was higher in the mid-secretory vs periovulatory phase of OS cycles (Fig. 9B). mRNA expression of markers of angiogenesis, such as VEGFA and its receptors FLT1 (VEGFR1) and KDR (VEGFR2), was higher in the mid-secretory vs periovulatory phase in both OS and NCs (Fig. 9C). When examining genes associated with B cells, there was significantly increased expression of BCL2 and BCL6, as well as FCER2, HAVCR1, and MS4A1 in the mid-secretory vs periovulatory phase of OS cycles. Expression of

CD22, a regulator of B-cell receptor signaling, was increased in OS and decreased in NCs when we compared changes from the mid-secretory vs periovulatory phases (Fig. 9D). Macrophage marker CD68 was more abundant in the mid-secretory vs periovulatory phase of both natural and OS cycles; expression of CLEC6A was decreased in the mid-secretory phase of NCs and CSFR1 expression was decreased in the mid-secretory phase of OS cycles (Fig. 9E).

Discussion

In this study, we considered the endometrium as both an intact tissue, as well as a sum of its principal compartments: glandular, stromal, vascular, and immune. We characterized the endometrium through bulk RNA sequencing and 3D structure analyses and examined the principal compartments at a cellular and molecular level through histologic assessment, immunostaining, and flow cytometry. Herein, we demonstrated that the endometrium in OS cycles differs significantly from the endometrium in natural menstrual cycles at the periovulatory phase, with glandular-stromal dyssynchrony, many differentially expressed genes, increased glandular volume, and an increased frequency of monocytes/macrophages. Notably, although the mid-secretory endometrium in OS cycles is more similar to that in NCs, as evidenced by glandular-stromal synchrony and the small number of differentially expressed genes, there is a dramatically increased frequency of B cells and a modestly decreased frequency of effector CD4⁺ T cells in the mid-secretory phase endometrium of the OS subjects. Our finding that pronounced differences in the mid-secretory phase were limited to the immune compartment provokes the hypothesis that immune perturbations may be important contributors to the adverse outcomes, such as failed implantation and poor placentation, that are associated with fresh embryo transfers in OS cycles.

To our knowledge, this longitudinal assessment of the endometrium in natural and OS cycles with concurrent comparisons across groups has not yet been reported and represents the first comprehensive study of the changes in various cellular compartments under both natural and stimulated conditions. This study has many additional strengths, including the comprehensive examination of unstimulated, natural menstrual cycles and OS cycles at multiple time points, the concurrent comparison of unstimulated and stimulated cycles in both the periovulatory and mid-secretory phases in women without known endometrial factor infertility, and homogeneity in OS protocols. One limitation of this study is that proliferative phase biopsies were not collected in women undergoing OS. Additionally, the study included a small number of subjects, though the subjects within the NC and OS groups were homogenous in their characteristics. Although the total number of subjects was small, subjects in the OS group studied at either the periovulatory or mid-secretory timepoint did not differ in terms of their baseline characteristics: they were all treated with similar OS protocols, had a similar response to stimulation (except the 3 OS-P subjects, who were removed or separated out for comparative analyses), and importantly, did not have known endometrial factor infertility. Therefore, findings in each of these groups are generalizable to women with a normal response to OS.

Published reports of glandular-stromal changes following OS have yielded inconsistent results, likely due to heterogeneity in subjects and OS protocols used (Bourgain and Devroey, 2003). In an effort to address some of the limitations of the published reports, we restricted our enrollment criteria to subjects of similar age, without endometrial factor infertility and those utilizing similar gonadotropin stimulation protocols and the same oocyte maturation trigger. During our analysis of glandular-stromal histology in periovulatory phase subjects, we identified a subset of individuals with poor response to stimulation and serum hormone levels more closely resembling those in natural menstrual cycles. Glandular-stromal development was synchronous in these subjects. In contrast, we found glandular-stromal dyssynchrony in the periovulatory phase in all OS subjects with normal

response to OS. Therefore, a lack of distinction between poor and normal responders could explain inconsistencies in findings and may have diluted the observed impact of OS in prior studies (Noci et al., 1997). One study that sought to characterize endometrial histology and changes in cellular proliferation following OS found a decrease in glandular and stromal proliferation in the periovulatory period without substantial glandular-stromal dyssynchrony (Bourgain et al., 2002). Our two-dimensional analyses showed that OS did not alter glandular histology or cellular proliferation in the periovulatory period. In contrast, our three-dimensional analysis (which measures glandular volume as a proportion of the biopsy volume) showed that OS increased glandular volume in the periovulatory period. Reasons for these seemingly discrepant results could be differences in analyses; two-dimensional assessments of proliferating glandular epithelium and stroma do not incorporate the total size of the biopsy tissue and the three-dimensional analysis cannot differentiate between an increase in the number of glandular epithelial cells vs an increase in the volume of individual epithelial cells. A three-dimensional analysis of cellular proliferation in the glands and stroma would be the best approach.

Studies in mice are often used to investigate implantation events and early pregnancy, but in contrast to decidualization in the human endometrium, which is triggered by ovulation, decidualization in mice is triggered by the presence of an embryo (Massri et al., 2023). Studies in mice demonstrate an increase in sprouting angiogenesis and non-sprouting angiogenesis accompanying decidualization (Massri et al., 2023). While a direct human comparison cannot be made, it is reasonable to hypothesize a similar change in the decidualized human endometrium even in the absence of an embryo. Interestingly, there are inconsistent findings with respect to changes in vasculature during natural and OS cycles. While some studies have found a reduction in the expression of vascular markers in the late secretory phase (Senapati et al., 2018), others have found increased levels of angiogenic factors (Junovich et al., 2011). We found that the frequency of CD31⁺ and VEGFR2⁺ cells was lowest in the mid-secretory phase of both natural and OS cycles and that there were no differences in expression of these vascular markers and 3D vascular volume when comparing natural and OS cycles at the periovulatory and mid-secretory phases. Angiogenesis in the secretory phase human endometrium is thought to be dominated by non-sprouting intussusceptive angiogenesis (Massri et al., 2023), a process by which new blood vessels are formed by the splitting of an existing vessel. Vascular remodeling that results from this type of angiogenesis may not be captured by measuring frequency of vascular markers or vascular volume by our 3D analysis, as these metrics would not necessarily change with intussusceptive angiogenesis. Some researchers consider endometrial differentiation that begins in the secretory phase to be a 'pre-decidualization' that continues to decidualization only after embryo implantation (Conrad et al., 2017). Human studies to assess embryo-driven endometrial vascular remodeling will require development of non-invasive techniques, and only then could we make a direct comparison to angiogenesis in early mouse pregnancy.

Few studies have examined the effects of OS on immune cell populations in the human endometrium. To our knowledge, we are the first to provide a global profile of endometrial immune cells throughout unstimulated menstrual cycles and under OS conditions at both the periovulatory and mid-secretory phases in a single report. One recent study found a decrease in total uNK

cells with a specific decrease in CD103⁺ NK cells, and an increase in B cells, in the mid-secretory phase of OS compared with NCs (Kanter *et al.*, 2023). We also observed a striking increase in B cells in the mid-secretory phase of OS cycles. However, our data differ from the study by Kanter *et al.* in that we did not find that OS changes the frequency of endometrial CD57^{dim} or CD57^{hi} NK populations. Instead, we observed modest but significant changes in the frequency of monocytes/macrophages in the periovulatory phase, and in effector CD4⁺ T cells in the mid-secretory phase of OS subjects, which have not previously been reported. Differences in subject inclusion, OS protocols and oocyte maturation triggers, and biopsy timepoints, may have contributed to these discrepancies. The dramatic increase in endometrial B cells after OS observed in our study and the study by Kanter *et al.* suggest an important and previously unappreciated role for B cells, and immune dysregulation more broadly, during the critical window of implantation.

In considering the importance of the periovulatory and mid-secretory phases of the endometrium, we expected that higher circulating levels of both E2 and P4 at these critical timepoints during OS cycles would result in similarly large differences in the endometrial microenvironment in OS subjects when compared to NC subjects. That this was not observed suggests that, after perturbation, the endometrium exposed to OS conditions can return to homeostasis over the course of the secretory phase to support embryo implantation during the period of receptivity. Several possibilities exist to explain the significant differences in the periovulatory phase and to contextualize the mid-secretory return to homeostasis. In considering the periovulatory phase differences, it is notable that in comparison with the periovulatory period of an OS cycle, which is abruptly and precisely timed by the administration of the oocyte maturation trigger, the periovulatory phase of an NC is a more variable and protracted event lasting 48–50 h (Fritz and Speroff, 2011) with a lower peak P4 level (Andersen *et al.*, 2020). In addition to differences in timing and magnitude, it is also possible that the hCG oocyte maturation trigger causes short-lived changes to the periovulatory endometrium that are apparent soon after its administration (Andersen *et al.*, 2020). An early rise in P4 under OS conditions may account for some of the variation in this period as well (Kalakota *et al.*, 2022). In considering the mid-secretory return to homeostasis, it is important to recall that studying the mid-secretory window of receptivity in the absence of an embryo does not provide actual information about an endometrium in which an embryo has implanted. Additionally, the fecundability rate in an NC is ~25% (Yen *et al.*, 2014), while the implantation rate following an embryo transfer in young patients approaches 50% (Centers for Disease Control and Prevention (2020); with embryo selection, this differential may be appropriate and the attempt to reach an implantation rate of 100% may not be rational.

Taken together, our findings suggest that the endometrial microenvironment in the mid-secretory phase after OS is much more similar to a NC than we previously thought. Consistent with our findings, pregnancy rates following OS are not substantially lower than those in NCs. Given our findings of changes in B cell, effector CD4⁺ T-cell, and monocyte/macrophage populations, further investigation of the immune compartment and the functional significance of these cell types under OS conditions is warranted. Non-invasive studies of the endometrium at the time of implantation and in the presence of an embryo, with a focus on the immune compartment, will provide a more complete picture of the changes associated with OS.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article have been uploaded to the Genome Expression Omnibus/National Center for Biotechnology Information with accession number GSE220044.

Acknowledgements

The authors would like to acknowledge Shahina B Maqbool, PhD, from the Epigenomics Shared Facility of the Albert Einstein College of Medicine; Xusheng Zhang and Robert Dubin, PhD, from the Epigenomics/Computational Genomics Core of the Albert Einstein College of Medicine Center; and Patricia Greenberg, MS, from Rutgers University Biostatistics & Epidemiology Services.

Authors' roles

A.C., M.S., N.V.-P., Q.Z., P.G.M., S.S.M., R.A., A.M.B., and N.C.D. conceived and planned the data analyses and experiments. A.C., M.S., N.V.-P., Q.Z., T.M., L.G., A.L., L.S., D.H., X.C., and T.W. carried out the experiments. A.C., M.S., N.V.-P., L.S., and Q.Z. performed the data analysis. All authors contributed to the interpretation of the results. A.C., M.S., N.V.-P., R.A., A.M.B., and N.C.D. wrote the manuscript with all authors critically editing the manuscript.

Funding

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases (R01AI148695 to A.M.B. and N.C.D.), Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD109152 to R.A.), and the March of Dimes (5-FY20-209 to R.A.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or March of Dimes.

Conflict of interest

All authors declare no conflict of interest.

References

- Andersen CY, Kelsey T, Mamsen LS, Vuong LN. Shortcomings of an unphysiological triggering of oocyte maturation using human chorionic gonadotropin. *Fertil Steril* 2020;**114**:200–208.
- Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham, England: Babraham Institute, 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (11 March 2024, date last accessed).
- Arora R, Fries A, Oelerich K, Marchuk K, Sabeur K, Giudice LC, Laird DJ. Insights from imaging the implanting embryo and the uterine environment in three dimensions. *Development* 2016;**143**:4749–4754.
- Basir GS, O WS, Ng EH, Ho PC. Morphometric analysis of peri-implantation endometrium in patients having excessively high oestradiol concentrations after ovarian stimulation. *Hum Reprod* 2001;**16**:435–440.

- Bermejo A, Cerrillo M, Ruiz-Alonso M, Blesa D, Simon C, Pellicer A, Garcia-Velasco JA. Impact of final oocyte maturation using gonadotropin-releasing hormone agonist triggering and different luteal support protocols on endometrial gene expression. *Fertil Steril* 2014;**101**:138–146 e133.
- Blair PA, Norena LY, Flores-Borja F, Rawlings DJ, Isenberg DA, Ehrenstein MR, Mauri C. CD19(+)CD24(hi)CD38(hi) B cells exhibit regulatory capacity in healthy individuals but are functionally impaired in systemic Lupus Erythematosus patients. *Immunity* 2010;**32**:129–140.
- Blumenfeld Z. What is the best regimen for ovarian stimulation of poor responders in ART/IVF? *Front Endocrinol (Lausanne)* 2020;**11**:192.
- Bourgain C, Devroey P. The endometrium in stimulated cycles for IVF. *Hum Reprod Update* 2003;**9**:515–522.
- Bourgain C, Ubaldi F, Tavaniotou A, Smits J, Van Steirteghem AC, Devroey P. Endometrial hormone receptors and proliferation index in the periovulatory phase of stimulated embryo transfer cycles in comparison with natural cycles and relation to clinical pregnancy outcome. *Fertil Steril* 2002;**78**:237–244.
- Centers for Disease Control and Prevention. *National ART Summary*. U.S. Department of Health and Human Services. 2020. <https://www.cdc.gov/art/reports/2020/summary.html> (11 March 2024, date last accessed).
- Conrad KP. Evidence for corpus luteal and endometrial origins of adverse pregnancy outcomes in women conceiving with or without assisted reproduction. *Obstet Gynecol Clin North Am* 2020;**47**:163–181.
- Conrad KP, Rabaglino MB, Post Uiterweer ED. Emerging role for dysregulated decidualization in the genesis of preeclampsia. *Placenta* 2017;**60**:119–129.
- Diaz-Gimeno P, Horcadas JA, Martinez-Conejero JA, Esteban FJ, Alama P, Pellicer A, Simon C. A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature. *Fertil Steril* 2011;**95**:50–60, 60.e1–15.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.
- Flores D, Madhavan M, Wright S, Arora R. Mechanical and signaling mechanisms that guide pre-implantation embryo movement. *Development* 2020;**147**:dev193490.
- Fritz MA, Speroff L. *Clinical Gynecologic Endocrinology and Infertility*, 8th edn. Philadelphia, PA, USA: Wolters Kluwer, 2011.
- Han S, Liu MH, Lv YS, Ren HY, Guo J, Li Y, Liu S. Effects of low luteinizing hormone during ovarian stimulation on endometrial gene expression and function—transcriptome analysis during the implantation window. *Reprod Sci* 2022;**29**:1908–1920.
- Haouzi D, Assou S, Dechanet C, Anahory T, Dechaud H, De Vos J, Hamamah S. Controlled ovarian hyperstimulation for in vitro fertilization alters endometrial receptivity in humans: protocol effects. *Biol Reprod* 2010;**82**:679–686.
- Haouzi D, Assou S, Mahmoud K, Tondeur S, Reme T, Hedon B, De Vos J, Hamamah S. Gene expression profile of human endometrial receptivity: comparison between natural and stimulated cycles for the same patients. *Hum Reprod* 2009;**24**:1436–1445.
- Horcadas JA, Diaz-Gimeno P, Pellicer A, Simon C. Uterine receptivity and the ramifications of ovarian stimulation on endometrial function. *Semin Reprod Med* 2007;**25**:454–460.
- Horcadas JA, Minguez P, Dopazo J, Esteban FJ, Dominguez F, Giudice LC, Pellicer A, Simon C. Controlled ovarian stimulation induces a functional genomic delay of the endometrium with potential clinical implications. *J Clin Endocrinol Metab* 2008;**93**:4500–4510.
- Horcadas JA, Riesewijk A, Polman J, van Os R, Pellicer A, Mosselman S, Simon C. Effect of controlled ovarian hyperstimulation in IVF on endometrial gene expression profiles. *Mol Hum Reprod* 2005;**11**:195–205.
- Hu M, Li J, Zhang Y, Li X, Brannstrom M, Shao LR, Billig H. Endometrial progesterone receptor isoforms in women with polycystic ovary syndrome. *Am J Transl Res* 2018;**10**:2696–2705.
- Junovich G, Mayer Y, Azpiroz A, Daher S, Iglesias A, Zylverstein C, Gentile T, Pasqualini S, Markert UR, Gutierrez G. Ovarian stimulation affects the levels of regulatory endometrial NK cells and angiogenic cytokine VEGF. *Am J Reprod Immunol* 2011;**65**:146–153.
- Kalakota NR, George LC, Morelli SS, Douglas NC, Babwah AV. Towards an improved understanding of the effects of elevated progesterone levels on human endometrial receptivity and oocyte/embryo quality during assisted reproductive technologies. *Cells* 2022;**11**:11.
- Kanter J, Gordon SM, Mani S, Sokalska A, Park JY, Senapati S, Huh DD, Mainigi M. Hormonal stimulation reduces numbers and impairs function of human uterine natural killer cells during implantation. *Hum Reprod* 2023;**38**:1047–1059.
- Kodithuwakku SP, Ng PY, Liu Y, Ng EH, Yeung WS, Ho PC, Lee KF. Hormonal regulation of endometrial olfactomedin expression and its suppressive effect on spheroid attachment onto endometrial epithelial cells. *Hum Reprod* 2011;**26**:167–175.
- Lee YL, Liu Y, Ng PY, Lee KF, Au CL, Ng EH, Ho PC, Yeung WS. Aberrant expression of angiopoietins-1 and -2 and vascular endothelial growth factor-A in peri-implantation endometrium after gonadotrophin stimulation. *Hum Reprod* 2008;**23**:894–903.
- Li L, Wang P, Liu S, Bai X, Zou B, Li Y. Transcriptome sequencing of endometrium revealed alterations in mRNAs and lncRNAs after ovarian stimulation. *J Assist Reprod Genet* 2020;**37**:21–32.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;**30**:923–930.
- Libby VR, Wilson R, Kresak A, Cameron C, Redline R, Mesiano S, Weinerman R. Superovulation with gonadotropin-releasing hormone agonist or chorionic gonadotropin for ovulation trigger differentially affects leukocyte populations in the peri-implantation mouse uterus. *FS Sci* 2021;**2**:198–206.
- Liu Y, Kodithuwakku SP, Ng PY, Chai J, Ng EH, Yeung WS, Ho PC, Lee KF. Excessive ovarian stimulation up-regulates the Wnt-signaling molecule DKK1 in human endometrium and may affect implantation: an in vitro co-culture study. *Hum Reprod* 2010;**25**:479–490.
- Liu Y, Lee KF, Ng EH, Yeung WS, Ho PC. Gene expression profiling of human peri-implantation endometria between natural and stimulated cycles. *Fertil Steril* 2008;**90**:2152–2164.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.
- Ma R, Jin N, Lei H, Dong J, Xiong Y, Qian C, Chen S, Wang X. Ovarian stimulation in mice resulted in abnormal placentation through its effects on proliferation and cytokine production of uterine NK cells. *Int J Mol Sci* 2023;**24**:5907.
- Macklon NS, van der Gaast MH, Hamilton A, Fauser BC, Giudice LC. The impact of ovarian stimulation with recombinant FSH in combination with GnRH antagonist on the endometrial transcriptome in the window of implantation. *Reprod Sci* 2008;**15**:357–365.
- Maheshwari A, Pandey S, Amalraj Raja E, Shetty A, Hamilton M, Bhattacharya S. Is frozen embryo transfer better for mothers and babies? Can cumulative meta-analysis provide a definitive answer? *Hum Reprod Update* 2018;**24**:35–58.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J* 2011;**17**:10–12.

- Martinez-Conejero JA, Simon C, Pellicer A, Horcajadas JA. Is ovarian stimulation detrimental to the endometrium? *Reprod Biomed Online* 2007;**15**:45–50.
- Massri N, Loia R, Sones JL, Arora R, Douglas NC. Vascular changes in the cycling and early pregnant uterus. *JCI Insight* 2023;**8**:e163422.
- Meyer WR, Novotny DB, Fritz MA, Beyler SA, Wolf LJ, Lessey BA. Effect of exogenous gonadotropins on endometrial maturation in oocyte donors. *Fertil Steril* 1999;**71**:109–114.
- Mints M, Blomgren B, Falconer C, Palmblad J. Expression of the vascular endothelial growth factor (VEGF) family in human endometrial blood vessels. *Scand J Clin Lab Invest* 2002;**62**:167–175.
- Noci I, Borri P, Coccia ME, Criscuoli L, Scarselli G, Messeri G, Paglierani M, Moncini D, Taddei G. Hormonal patterns, steroid receptors and morphological pictures of endometrium in hyperstimulated IVF cycles. *Eur J Obstet Gynecol Reprod Biol* 1997;**75**:215–220.
- Noyes RW, Hertig AT, Rock J. Dating the endometrial biopsy. *Am J Obstet Gynecol* 1975;**122**:262–263.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;**33**:290–295.
- Qu Y, Zhang J, Guo S, Zhang L, Qian J, Zhu X, Duan E, Zhang Y. Three-dimensional visualization of mouse endometrial remodeling after superovulation. *Front Cell Dev Biol* 2022;**10**:933852.
- Senapati S, Wang F, Ord T, Coutifaris C, Feng R, Mainigi M. Superovulation alters the expression of endometrial genes critical to tissue remodeling and placentation. *J Assist Reprod Genet* 2018;**35**:1799–1808.
- Tawia SA, Beaton LA, Rogers PA. Immunolocalization of the cellular adhesion molecules, intercellular adhesion molecule-1 (ICAM-1) and platelet endothelial cell adhesion molecule (PECAM), in human endometrium throughout the menstrual cycle. *Hum Reprod* 1993;**8**:175–181.
- Tulac S, Nayak NR, Kao LC, Van Waes M, Huang J, Lobo S, Germeyer A, Lessey BA, Taylor RN, Suchanek E et al. Identification, characterization, and regulation of the canonical Wnt signaling pathway in human endometrium. *J Clin Endocrinol Metab* 2003;**88**:3860–3866.
- Walker ER, McGrane M, Aplin JD, Brison DR, Ruane PT. A systematic review of transcriptomic studies of the human endometrium reveals inconsistently reported differentially expressed genes. *Reprod Fertil* 2023;**4**:e220115.
- Wang X, Bove AM, Simone G, Ma B. Molecular bases of VEGFR-2-mediated physiological function and pathological role. *Front Cell Dev Biol* 2020;**8**:599281.
- Wu JL, Keller P, Kanchwala M, Xing C, Babayev SN, Carr BR, Bukulmez O, Word RA. Controlled ovarian stimulation protocols alter endometrial histomorphology and gene expression profiles. *Reprod Sci* 2020;**27**:895–904.
- Yeaman GR, Collins JE, Fanger MW, Wira CR, Lydyard PM. CD8+ T cells in human uterine endometrial lymphoid aggregates: evidence for accumulation of cells by trafficking. *Immunology* 2001;**102**:434–440.
- Yen SSC, Strauss JF, Barbieri RL. *Yen & Jaffe's Reproductive Endocrinology: physiology, Pathophysiology, and Clinical Management*, 7th edn. Philadelphia, PA: Elsevier/Saunders, 2014.
- Zaimoku Y, Patel BA, Kajigaya S, Feng X, Alemu L, Quinones Raffo D, Groarke EM, Young NS. Deficit of circulating CD19(+) CD24(hi) CD38(hi) regulatory B cells in severe aplastic anaemia. *Br J Haematol* 2020;**190**:610–617.
- Zhao Y, Garcia J, Kolp L, Cheadle C, Rodriguez A, Vlahos NF. The impact of luteal phase support on gene expression of extracellular matrix protein and adhesion molecules in the human endometrium during the window of implantation following controlled ovarian stimulation with a GnRH antagonist protocol. *Fertil Steril* 2010;**94**:2264–2271.

Reproductive biology

Deciphering the role of PGRMC2 in the human endometrium during the menstrual cycle and *in vitro* decidualization using an *in vitro* approach

Yassmin Medina-Laver ^{1,†}, Roberto Gonzalez-Martin ^{1,†}, Pedro de Castro ¹, Indra Diaz-Hernandez¹, Pilar Alama^{1,2}, Alicia Quiñonero¹, Andrea Palomar ¹, and Francisco Dominguez ^{1,*}

¹Research Group in Reproductive Medicine, IVIRMA Global Research Alliance, IVI Foundation, Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia, Comunitat Valenciana, Spain

²Gynecology, IVIRMA Global Research Alliance, IVI-RMA Valencia, Valencia, Comunitat Valenciana, Spain

*Correspondence address. Fundación Instituto Valenciano de Infertilidad, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril, Martorell, 106—Tower A, 1st Floor, Comunitat Valenciana, 46026 Valencia, Spain. Tel: +34-963903305; E-mail: francisco.dominguez@ivirma.com  <https://orcid.org/0000-0002-7620-8081>

[†]These authors contributed equally to this work.

ABSTRACT

STUDY QUESTION: What is the human endometrial non-classical progesterone receptor (PGR) membrane component 2 (PGRMC2) expression pattern throughout the menstrual cycle and what role does it play during decidualization?

SUMMARY ANSWER: Endometrial PGRMC2 expression fluctuates during the human menstrual cycle and is abundantly expressed in human endometrial stromal cells (hEnSCs) during *in vitro* decidualization, process where PGRMC2 is involved in embryo implantation-related pathways.

WHAT IS KNOWN ALREADY: The endometrial response to progesterone is mediated by the classical and non-classical PGRs. We previously demonstrated that PGR membrane component 1 (PGRMC1) is critical for endometrial function, embryo implantation, and future placentation, however, the role(s) of PGRMC2, which is structurally similar to PGRMC1, have not been studied in the human endometrium.

STUDY DESIGN, SIZE, DURATION: This prospective study comprehensively evaluated the endometrial expression of PGRMC2 throughout the human menstrual cycle and during *in vitro* decidualization of hEnSCs (isolated from 77 endometrial biopsies that were collected from 66 oocyte donors), using immunohistochemistry, RT-qPCR, western blot, transcriptomic, and proteomic analyses. In addition, functional analysis was carried out to validate the implication of PGRMC2 in hEnSCs during embryo invasion using an *in vitro* outgrowth model.

PARTICIPANTS/MATERIALS, SETTING, METHODS: *In vitro* decidualization of hEnSCs was induced using co-treatment with cAMP and medroxyprogesterone 17-acetate progestin, and evaluated by measuring prolactin by ELISA and F-actin immunostaining. RT-qPCR was employed to compare expression with other PGRs. To reveal the function of PGRMC2 during the decidualization process, we specifically knocked down PGRMC2 with siRNAs and performed RNA-seq and quantitative proteomics techniques (SWATH-MS). The common differentially expressed genes (DEGs) and proteins (DEPs) were considered for downstream functional enrichment analysis. Finally, to verify its implication in the trophoblast invasion, an outgrowth model was carried out where hEnSCs with silenced PGRMC2 were co-cultured with human trophoblastic spheroids (JEG-3) following *in vitro* decidualization.

MAIN RESULTS AND THE ROLE OF CHANCE: In contrast to PGRMC1 and classical PGRs, endometrial PGRMC2 gene expression was significantly lower during the late- versus mid-secretory phase ($P < 0.05$). Accordingly, the elevated PGRMC2 protein abundance observed in the endometrial epithelial glands throughout the menstrual cycle dropped in the late secretory phase, when abundance decreased in all endometrial compartments. Nevertheless, PGRMC2 protein increased during the mid-secretory phase in stromal and glandular cells, and PGRMC2 mRNA ($P < 0.0001$) and protein ($P < 0.001$) levels were significantly enhanced in the membranes/organelles of decidualized hEnSCs, compared to non-decidualized hEnSCs. Notably, PGRMC1 and PGRMC2 mRNA were significantly more abundant than classical PGRs throughout menstrual cycle phases and in decidualized and non-decidualized hEnSCs ($P < 0.05$). RNA-seq and proteomics data revealed 4687 DEGs and 28 DEPs, respectively, in decidualized hEnSCs after PGRMC2 silencing. While functional enrichment analysis showed that the 2420 upregulated genes were mainly associated with endoplasmic reticulum function, vesicular transport, morphogenesis, angiogenesis, cell migration, and cell adhesion, the 2267 downregulated genes were associated with aerobic respiration and protein biosynthesis. The protein enrichment analysis showed that 4 upregulated and 24 downregulated proteins were related to aerobic respiration, cellular response, metabolism, localization of endoplasmic reticulum proteins, and ribonucleoside biosynthesis routes. Finally, PGRMC2 knockdown significantly compromised the ability of the decidualized hEnSCs to support trophoblast expansion in an outgrowth model ($P < 0.05$).

LARGE-SCALE DATA: Transcriptomic data are available via NCBI's Gene Expression Omnibus (GEO) under GEO Series accession number GSE251843 and proteomic data via ProteomeXchange with identifier PXD048494.

Received: February 4, 2023. Revised: January 11, 2024. Editorial decision: February 18, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

LIMITATIONS, REASONS FOR CAUTION: The functional analyses were limited by the discrete number of human endometrial biopsies. A larger sample size is required to further investigate the potential role(s) of PGRMC2 during embryo implantation and maintenance of pregnancy. Further, the results obtained in the present work should be taken with caution, as the use of a pure primary endometrial stromal population differentiated *in vitro* does not fully represent the heterogeneity of the endometrium *in vivo*, nor the paracrine communications occurring between the distinct endometrial cell types.

WIDER IMPLICATIONS OF THE FINDINGS: The repression of endometrial PGRMC2 during the late- versus mid-secretory phase, together with its overexpression during decidualization and multiple implications with embryo implantation not only highlighted the unknown roles of PGRMC2 in female reproduction but also the potential to exploit PGRMC2 signaling pathways to improve assisted reproduction treatments in the future.

STUDY FUNDING/COMPETING INTEREST(S): This research was funded by Instituto de Salud Carlos III (ISCIII) granted to F.D. (PI20/00405 and PI23/00860), co-funded by the European Union. Y.M.-L. was supported by a predoctoral research grant from Generalitat Valenciana (ACIF/2019/262). R.G.-M. was supported by Generalitat Valenciana (CIAPOT/2022/15). P.d.C. was supported by a predoctoral grant for training in research into health (PFIS FI20/00086) from the Instituto de Salud Carlos III. I.D.-H. was supported by the Spanish Ministry of Science, Innovation and Universities (FPU18/01550). A.P. was supported by the Instituto de Salud Carlos III (PFIS FI18/00009). This research was also supported by IVI Foundation-RMA Global (1911-FIVI-103-FD). The authors declare no conflict of interest.

Keywords: PGRMC2 / human endometrium / menstrual cycle / endometrial stromal cells / decidualization / progesterone receptor / progesterone / transcriptomics / proteomics

Introduction

The hormonally regulated human endometrium plays a critical role in embryo implantation and pregnancy maintenance (Salazar and Calzada, 2007; Loutradis *et al.*, 2008; Morel *et al.*, 2016). In response to steroid hormones (mainly progesterone (P4)), endometrial tissues undergo physiological changes to enhance receptivity to embryo implantation, a process commonly referred to as the 'window of implantation' (WOI) (Graham and Clarke, 1997; Ruiz-Alonso *et al.*, 2013; Yoshinaga, 2014). One P4-dependent critical event is decidualization, which is characterized by profound morphological and biochemical differentiation of human endometrial stromal cells (hEnSCs) into decidual cells (Sakai *et al.*, 2003), causing an increase of secreted prolactin (PRL) protein, commonly used as a marker of decidualization.

This P4 response is predominately regulated by classical P4 receptor (classical progesterone receptor (PGR)) isoforms, PGR α and PGR β . Despite the essential role that classical PGRs play in female reproduction (Conneely *et al.*, 2002; Large and DeMayo, 2012; Wetendorf *et al.*, 2017; Medina-Laver *et al.*, 2021), studies in knockout mice lacking these classical PGRs demonstrate that cells are still able to respond to P4 (Frye and Vongher, 1999) through a faster mechanism via non-classical progesterone membrane receptors (non-classical PGRs) (Kimura *et al.*, 2012). Non-classical PGRs are important contributors in reproduction (Gellersen *et al.*, 2009; Medina-Laver *et al.*, 2021) regulating gene transcription (Karteris *et al.*, 2006; Salsano *et al.*, 2021), the acrosomal reaction in human sperm (Foresta *et al.*, 1993), sexual behavior in mice (Frye and Vongher, 1999; Frye *et al.*, 2006), and meiotic maturation in *Xenopus* oocytes (Finidori-Lepicard *et al.*, 1981). We showed that one of these non-classical PGRs, P4 receptor membrane component 1 (PGRMC1), is critical for endometrial function and embryo implantation (Garrido-Gómez *et al.*, 2014; Salsano *et al.*, 2017, 2020, 2021).

A non-classical PGR with structural similarity to PGRMC1 (Cahill, 2007), PGRMC2, is also expressed in many reproductive tissues, including the uterus, ovaries, and placenta (Kimura *et al.*, 2012). PGRMC2 regulates sterol metabolism (Kimura *et al.*, 2012; Wu *et al.*, 2019), contributes to cell proliferation through extracellular-signal-regulated kinase (ERK) activation (Liu *et al.*, 2009), is associated with ovarian reserve status in young women (Skiadas *et al.*, 2012), participates in cell cycle regulation (Griffin *et al.*, 2014; Peluso *et al.*, 2014, 2018, 2019; Li *et al.*, 2019; Wu *et al.*, 2019), and is involved in the progression of endometriosis (Keator *et al.*, 2012; Bunch *et al.*, 2013). Yet, the functions of PGRMC2 in

the human endometrium and its specific role in decidualization remain unclear. Its importance is evident, in fact, PGRMC2 deficiency causes premature uterine senescence (unlike normal senescence that occurs in the physiological decidualization process; Deryabin *et al.*, 2020; Lucas *et al.*, 2020) and leads to post-implantation failure (Clark *et al.*, 2016; Mccallum *et al.*, 2016). A similar scenario has been observed in PGRMC1- and PGRMC2-knockout zebrafish (Wu *et al.*, 2019; Wu and Zhu, 2020). Additionally, there is limited knowledge regarding PGRMC2 expression across the human menstrual cycle and its function in human endometrial receptivity.

Understanding how P4 signaling occurs in endometrial cells is important to guide the development of more effective therapeutic approaches to the numerous reproductive diseases caused by disrupted P4 signaling, resulting in subfertility or infertility (Bunch *et al.*, 2013; Kim *et al.*, 2013; Zhang and Murphy, 2014; Szekeres-Bartho, 2018; Hirota, 2019; Liao *et al.*, 2019). Here, to fill these knowledge gaps, we characterized endometrial PGRMC2 expression throughout the human menstrual cycle and in hEnSCs subjected to an *in vitro* model of human endometrial decidualization. Additionally, we describe the functional implications of PGRMC2 with a silencing approach assay, integrating RNA-seq, and proteomic techniques. We evaluated the functional implication of PGRMC2 in the embryo implantation process with an outgrowth *in vitro* model. Finally, we also compared endometrial PGRMC2 expression with PGRMC1 and PGR.

Materials and methods

Ethical approval

The study was approved by the ethical committee for clinical research at the Instituto Universitario, Instituto Valenciano de Infertilidad (IU-IVIRMA Valencia), Spain (protocol number 1911-FIVI-103-FD). All participants provided written informed consent before undergoing surgical procedures and sample collection.

Endometrial tissue collection

A total of 77 endometrial tissue samples ($n=14$ for immunohistochemistry, $n=38$ for *in vivo* mRNA quantification, $n=35$ for *in vivo* protein quantification, $n=22$ for *in vitro* experiments, $n=4$ for RNA-seq, $n=5$ for sequential window acquisition of all theoretical fragment ion mass spectra (SWATH-MS), and $n=6$ for the outgrowth model), were obtained from 66 healthy fertile egg donors who fulfilled the criteria of the IVIRMA Valencia oocyte

donation program (Fig. 1). Donors included individuals aged 18–35 years with a body mass index of 18.0–28.0 kg/m² with regular menstrual cycles (24–35 days) checked by the first day of the menstrual cycle and ultrasonographic evaluation of follicle

growth, normal ovarian reserve (antral follicle count >8 and anti-Müllerian hormone (AMH) measurement), and proven fertility. Ultrasonographic and gynecological examinations were performed on all participants to exclude those with uterine

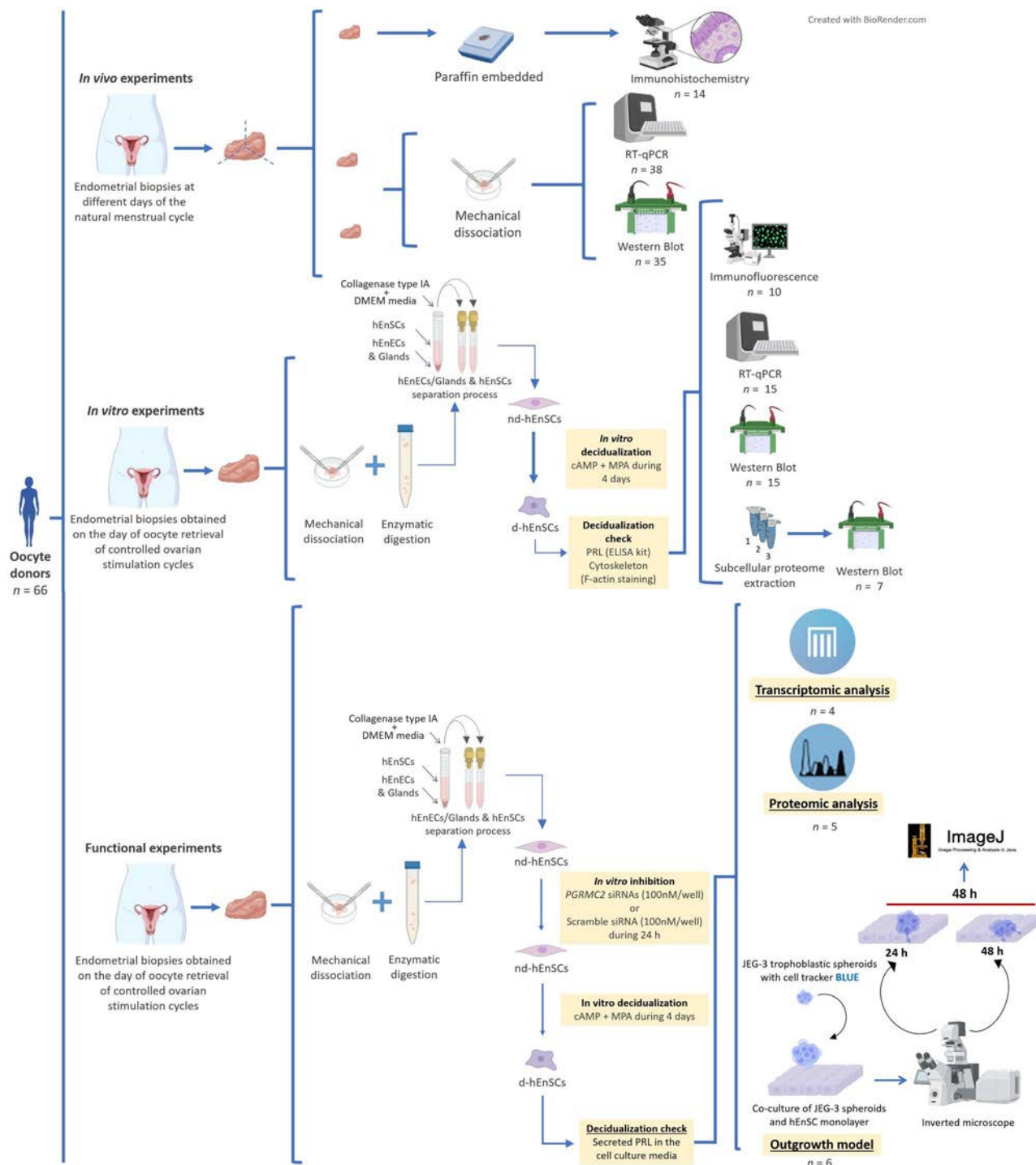


Figure 1. Study design. Endometrial biopsies were collected from 66 donors at different days of the natural menstrual cycle (for *in vivo* experiments) or undergoing oocyte donation cycles (for *in vitro* experiments and functional experiments). For *in vivo* experiments, endometrial biopsies were cut in three sections. One of said sections was used to perform immunohistochemistry. The other two sections were used to analyze PGRMC2 expression by RT-qPCR and western blot. For *in vitro* and functional experiments, hEnSCs were isolated from endometrial biopsies to analyze PGRMC2 localization, mRNA expression, and protein abundance during *in vitro* decidualization and to study transcriptomically and proteomically the differences in d-hEnSCs when PGRMC2 was silenced. Finally, an outgrowth model with the use of JEG-3 spheroids was carried out silencing PGRMC2 in d-hEnSCs. d-hEnSCs, decidualized human endometrial stromal cells; DMEM, Dulbecco's modified Eagle medium; hEnSCs, human endometrial stromal cells; hEnECs, human endometrial epithelial cells; MPA, medroxyprogesterone 17-acetate progestin; nd-hEnSCs, non-decidualized human endometrial stromal cells; PGRMC2, progesterone receptor membrane component 2; PRL, prolactin; RT-qPCR, quantitative reverse transcription polymerase chain reaction. Figure created with BioRender.com.

pathologies (e.g. adenomyosis, endometriosis, endometritis, pelvic inflammatory disease), abnormal bleeding or irregular menstrual cycles (<24 or >35 days), hormone treatment and/or contraceptives, ovarian stimulation (for *in vivo* experiments), stress, pharmacotherapy, and pregnancy within 3 months of sample collection. One single endometrial biopsy was obtained from 59 donors, two endometrial biopsies were obtained from six donors, and three endometrial biopsies were obtained from one donor throughout various menstrual phases. All of them were obtained with a Pipelle catheter (CCD Laboratoire, France). For *in vivo* experiments, endometrial samples were collected at different days of the natural menstrual cycle and were distributed into five groups: early proliferative (EP; *n* = 9, Days 3–8), late proliferative (LP; *n* = 8, Days 9–14), early secretory (ES; *n* = 8, Days 15–18), mid-secretory (MS; *n* = 8, Days 19–22), and late secretory (LS; *n* = 7, Days 23–28). For *in vitro* experiments, endometrial samples were obtained on the day of oocyte retrieval of controlled ovarian stimulation cycles (Fig. 1).

PGRMC2 immunohistochemistry in human endometrial tissue

For immunohistochemical analysis (*n* = 14), endometrial tissue at different phases of the menstrual cycle was immediately fixed in 4% (vol/vol) paraformaldehyde, embedded in paraffin, cut in 6 mm sections (5 µm thickness), and placed on microscope slides (Thermo Scientific, USA). After deparaffinization using xylene and rehydration with decreasing concentrations of ethanol and distilled water, sections were rinsed with 1× phosphate-buffered saline (1× PBS) and processed as previously described (Salsano *et al.*, 2017).

Slides were incubated overnight at 4°C in a humid box with rabbit polyclonal anti-human PGRMC2 (PA5-59465, Invitrogen, USA), tested in previous studies (Richardson *et al.*, 2020; Lozovyy *et al.*, 2021), diluted 1:300 with 3% bovine serum albumin (BSA; Sigma-Aldrich, Spain). We included negative (incubated with 1× PBS and 3% BSA in the absence of primary antibodies) and positive controls (human intestine tissue; TTR008-25EA, Sigma-Aldrich, Spain). Secondary antibodies were included in the peroxidase/diaminobenzidine (DAB) kit (DAKO; Agilent Technologies, USA), and immunoreactivity was detected using DAB. After counterstaining with hematoxylin and washing with distilled water, slides were mounted with SafeMount mounting medium (Sigma-Aldrich, Spain). Slides without antibodies were analyzed by hematoxylin-eosin staining to check the biopsy condition. Slides were examined with a Nikon Eclipse 80i microscope equipped with a Nikon digital sight DS-L1 camera.

hEnSC isolation, cell culture, and *in vitro* decidualization

To isolate hEnSCs, endometrial samples were minced into smaller pieces ($\leq 1 \text{ mm}^3$) and digested with a 1:10 dilution of collagenase type IA in phenol red-free Dulbecco's modified Eagle's medium (Sigma-Aldrich, Spain) at 4°C overnight. hEnSCs were isolated from the digested endometrial sample by gravity sedimentation as described previously (Dominguez *et al.*, 2010). Primary hEnSCs were *in vitro* decidualized with a 0.5-mM stable cyclic adenosine 3':5' monophosphate analog bromo-cAMP (Sigma-Aldrich, Spain) and 1-µM medroxyprogesterone 17-acetate progestin (Sigma-Aldrich, Spain) treatment over 4 days. Control hEnSCs were cultured in parallel without hormone treatment (non-decidualized hEnSCs (nd-hEnSCs)).

Decidualization was checked by measuring secreted human PRL (*n* = 15) in conditioned medium by ELISA (Novus Biologicals, USA) and assessment of cytoskeletal reorganization (*n* = 10) by

F-actin staining in decidualized hEnSCs (d-hEnSCs) and nd-hEnSCs. ELISA detection limit was 2 ng/ml PRL, and colorimetric values were determined using a standard curve. Results are expressed as µU/ml PRL (mean ± SD) (Supplementary Fig. S1).

For F-actin staining, cells were fixed with 4% paraformaldehyde for 15 min, washed, and incubated with 1:1000 rhodamine phalloidin (ab235138, Abcam, UK) at room temperature (RT). The structure of the cytoskeleton was visualized and photographed using a fluorescence inverted AXIO microscope (Zeiss, Germany).

PGRMC2 immunofluorescence in hEnSCs

To evaluate PGRMC2 localization in hEnSCs, after *in vitro* decidualization treatment, confluent nd-hEnSCs (*n* = 10) and d-hEnSCs (*n* = 10) were fixed with 4% paraformaldehyde for 15 min, washed and, permeabilized in 0.2% Triton X-100. Cells were blocked with 5% BSA and incubated with 0.4 µg/ml rabbit polyclonal anti-human PGRMC2 (PA5-59465, Invitrogen, USA) in 3% BSA at 4°C overnight. Secondary Alexa Fluor 488 (Santa Cruz Biotechnology, Germany) was incubated at a 1:1000 dilution in 3% BSA for 1 h at RT in the dark and mounted with 6-diamidino-2-phenylindole (DAPI; P36931, Thermo Scientific, USA). Experiments were performed in duplicate. The cytoskeleton was visualized and photographed using a fluorescence inverted AXIO microscope (Zeiss, Germany).

Subcellular proteome extraction in hEnSCs

To analyze PGRMC2 subcellular localization, a ProteoExtract Subcellular Proteome Extraction kit (539790, Calbiochem, Millipore, France) was used to separate the cytoplasmic, membrane/organelle, and nuclear protein fractions of nd-hEnSCs and d-hEnSCs, according to the manufacturer's instructions (*n* = 7 per subcellular fraction). Purity of each enriched fraction was assessed by western blotting (WB) with mouse monoclonal anti-human Hsp70 (1:1000 dilution; sc373867, Santa Cruz Biotechnology, Germany) for the cytosolic fraction, rabbit polyclonal anti-human CD98 (1:1000 dilution; sc9160, Santa Cruz Biotechnology, Germany) for the membrane/organelle fraction, and rabbit polyclonal anti-human histone H3 (1:1000 dilution; PA5-16183, Invitrogen, USA) for the nuclear fraction. PGRMC2 protein expression in all subcellular compartments was assessed by WB, and β-actin was used as a loading control (1:1000 dilution; sc47778, Santa Cruz Biotechnology, Germany). Protein quantification was performed using ImageJ software, and protein abundance was normalized to β-actin. Results were expressed as protein relative abundance (PGRMC2/β-actin) ± SD.

PGRMC2 silencing in hEnSCs

Cells were plated, as described previously, one day before transfection. A pool of three different siRNAs specific to PGRMC2 (Origene, USA; Supplementary Table S1) were used to knockdown PGRMC2 gene in d-hEnSCs (iPGRMC2 d-hEnSCs). Trilencer-27 Universal Scrambled Negative Control siRNA Duplex (Origene, USA) was used as internal control (Sc d-hEnSCs) at 100 nM/well. A nd-hEnSCs and d-hEnSCs without transfection were also included as negative controls (control nd-hEnSCs or control d-hEnSCs, respectively). The PGRMC2 siRNAs pool (100 nM/well) was transfected into cells using Lipofectamine 3000 Transfection Reagent (Invitrogen, USA) for 24 h. Then, *in vitro* decidualization was induced for 4 days. PGRMC2 silencing was checked by RT-qPCR and WB.

Protein extraction and WB

We performed WB to assess endometrial PGRMC2 protein (≈28 kDa) abundance across the human menstrual cycle *in vivo*

and in hEnSCs decidualized *in vitro*. Endometrial biopsies obtained across the human menstrual cycle were washed with 1× PBS and minced into smaller pieces ($\leq 1 \text{ mm}^3$) with scalpels at 4°C. Protein was extracted from total endometrial samples (total $n = 35$; EP, $n = 7$; LP, $n = 8$; ES, $n = 6$; MS, $n = 8$; LS, $n = 6$) and from hEnSCs ($n = 15$) using RIPA lysis buffer containing 1% Triton X-100, 1% sodium dodecyl sulfate (SDS), 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% NP-40, 1× protease inhibitor (Roche complete 7×), and 1× phosphatase inhibitor mixture (Sigma-Aldrich, Spain). After 20–30 min of incubation on ice, lysed cells were centrifuged at 14 000 g for 20 min at 4°C. The supernatant was recovered, and protein content was determined by Bradford protein assay following the manufacturer's protocol. Protein was denatured for 5 min at 95°C, and 15 µg of protein per sample was separated on 8–12% SDS-polyacrylamide gels and transferred to polyvinylidene difluoride (PVDF) membranes with a Trans-Blot Turbo Transfer System (Bio-Rad Laboratories, USA). PGRMC2 protein expression was determined and compared to the expression of PGRMC1 (goat monoclonal anti-human PGRMC1, 0.33 µg/ml; ab48012, Abcam, UK) and classical PGR (PR α and PR β isoforms; mouse monoclonal anti-human PGR, 1:500 dilution; MA1-410, Invitrogen, USA). β -Actin (mouse monoclonal anti-human β -actin; 1:2000 dilution; Santa Cruz Biotechnology, Germany) was used as a loading control. Bands were developed using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific, USA) and visualized on an Amersham Imager 680 system. Protein quantitation was performed using ImageJ software, and protein abundance was normalized to β -actin (protein relative abundance). Results are expressed as protein relative abundance mean $\pm$ SD and as fold change in *in vitro* experiments. PVDF membranes were scanned using Fujifilm LAS-300.

RNA extraction, reverse transcription, and real-time qPCR

Relative expression of PGRMC2 mRNA at different phases of the menstrual cycle (total $n = 38$; EP, $n = 8$; LP, $n = 8$; ES, $n = 8$; MS, $n = 7$; LS, $n = 7$) and in hEnSCs following *in vitro* decidualization ($n = 15$) was determined by RT-qPCR and compared to the expression of PGRMC1 and classical PGR. RNA was isolated from total endometrial samples and hEnSCs with Trizol and a RNeasy Mini kit (Qiagen, USA), respectively. RNA concentration was quantified using a NanoDrop 2000c spectrophotometer (Thermo Scientific, USA), and cDNA was synthesized using the PrimeScript RT reagent kit (Takara, Japan) from 0.5 µg of RNA per sample.

RT-qPCR was performed on a StepOnePlus qPCR System (Applied Biosystems, Spain) with PowerUp SYBR Green Master Mix (Applied Biosystems, Spain). Primers for PGRMC2, PGRMC1, total PGR (PGR-t), PGR β isoform (PGR- β), and GAPDH are listed in [Supplementary Table S2](#). Reactions were run in duplicate. Relative mRNA expression was calculated by the $2^{-\Delta\Delta C_T}$ method and normalized to the expression of GAPDH. Results are expressed as fold change mean $\pm$ SD in gene expression during menstrual cycle and in hEnSCs. Analysis of PGRMC2 gene expression between different phases of the menstrual cycle was performed comparing the ΔC_T of each sample with the ΔC_T mean of EP-phase samples. To compare classical and non-classical PGR mRNA expression across the menstrual cycle stages and in hEnSCs, the inverse of ΔC_T ($1/\Delta C_T$) of each sample was compared between genes in each phase or decidual condition, respectively.

Transcriptomic analysis by RNA-seq

Raw count matrix derived from RNA-seq data libraries between Sc d-hEnSCs and iPGRMC2 d-hEnSCs was processed and

subjected to statistical analysis ($n = 4$). Normalized read counts were obtained by dividing raw read counts by the scaling factor associated with the sample they belonged to. Differential expressed genes (DEGs) were assessed between Sc d-hEnSCs and iPGRMC2 d-hEnSCs using DESeq2 (DESeq2 v1.34.0). Genes were considered relevant when adjusted $P < 0.01$ and fold change was > 2 or < 0.5 . Common DEGs (false discovery rate (FDR)-adjusted $P < 0.05$) were used for downstream functional enrichment analysis with g:Profiler (g_SCS < 0.05 , term size up to 2000 genes). Plugin EnrichmentMap in Cytoscape 3.9.1 software was employed to generate functional enrichment map of DEGs-related Biological Processes ([Reimand et al., 2019](#)).

RNA-seq validation was performed in d-hEnSCs from the same donors used in RNA-seq analyses ($n = 4$). RNA expression levels between Sc d-hEnSCs and iPGRMC2 d-hEnSCs of the selected genes (C-X-C motif chemokine ligand 14 (CXCL14), DEPP autophagy regulator 1 (DEPP1), PGRMC2, tenascin C (TNC), hepatocyte growth factor (HGF), collagen type I alpha 1 chain (COL1A1)) were validated by RT-qPCR, using the same methodology to that previously described (See 'RNA extraction, reverse transcription, and real-time qPCR' section) and normalized to the expression of GAPDH. Validation primers are listed in [Supplementary Table S2](#). Results are expressed as fold change mean $\pm$ SD.

Proteomic analysis by SWATH-MS

SWATH-MS procedure was carried out as previously described ([Perez-Patiño et al., 2016](#)). Briefly, all samples ($n = 5$) were pooled to build the spectral library using an 12% precast gel (Bio-Rad Laboratories, USA) and an Eksport nanoLC 425 nanoflow system (Eksigent Technologies, ABSCIEX) coupled to a nanoESI qTOF MS (6600 plus TripleTOF, ABSCIEX). Resulting peptides were analyzed in a mass spectrometer nanoESI qTOF (6600 plus TripleTOF, ABSCIEX). Data were read with ProteinPilot v5.0 to generate peak list directly from 6600 TripleTOF wiff files. For individual SWATH analysis 20 µg of total protein extracted was loaded in a 1-dimensional SDS-polyacrylamide gel electrophoresis (1D-SDS-PAGE) gel to clean and concentrate samples. The wiff files obtained from SWATH experiment were analyzed by Peak View 2.2 and Marker View (Sciex) (FDR $< 1\%$). Protein areas were normalized by the total sum of the areas of all the quantified proteins.

Proteins with P -value < 0.05 were described as differentially expressed proteins (DEPs) between control d-hEnSCs and iPGRMC2 d-hEnSCs. Common DEPs were used for downstream functional enrichment analysis with g:Profiler (Benjamini-Hochberg FDR < 0.05). Plugin EnrichmentMap in Cytoscape 3.9.1 software was employed to generate functional enrichment map of DEPs-related Biological Processes ([Reimand et al., 2019](#)).

Ki-67 immunofluorescence in d-hEnSCs

To evaluate the proliferation of d-hEnSCs after PGRMC2 silencing, control d-hEnSCs ($n = 3$), Sc d-hEnSCs ($n = 3$), and iPGRMC2 d-hEnSCs ($n = 3$) were fixed with 4% paraformaldehyde for 15 min, washed and, permeabilized in 0.2% Triton X-100. Cells were blocked with 5% BSA and incubated with 0.5 µg/ml rabbit polyclonal anti-human Ki-67 (ab15580, Abcam, UK) in 3% BSA at 4°C overnight. Secondary Alexa Fluor 488 (Santa Cruz Biotechnology, Germany) was incubated at a 1:1000 dilution in 3% BSA for 1 h at RT in the dark and mounted with DAPI. Experiments were performed in duplicate. The % of proliferative cells (Ki-67 positive) was observed by a fluorescent inverted AXIO microscope (Zeiss, Germany) and measured by ImageProPlus software. Results are expressed as % of proliferative cells (Ki-67)/total cells (DAPI) $\pm$ SD.

Trophoblast invasion in d-hEnSCs

An outgrowth model was carried out where control, Sc, and iPGRMC2 d-hEnSCs ($n=6$) were co-cultured *in vitro* with human trophoblastic spheroids (JEG-3 cell line (cat. HTB-36, The American Type Culture Collection (ATCC), USA)). Trophoblast cells were stained using blue CellTracker fluorescent probes (1:1000; Invitrogen, USA) and cultured for 48 h in low-attachment plate to facilitate spheroids formation. Then, 48-h spheroids were co-cultured with the corresponding d-hEnSCs with three to four spheroids per well. Experiments were performed in duplicate. The trophoblast expanded area (EA) was observed by a fluorescent inverted AXIO microscope (Zeiss, Germany) at 24 and 48 h and measured by ImageJ software. Data is represented as normalized EA, being $EA = 48\text{ h area} - 24\text{ h area}$ and normalizing by dividing the corresponding EA of each spheroid by the mean EA of spheroids cultured in control d-hEnSCs. All decidualization procedures were checked by measuring secreted PRL by ELISA.

Statistical analysis

GraphPad Prism 8.3.0 (San Diego, CA) was used for statistical analyses and graphics generation. Normality and logarithmic tests (Anderson–Darling, D’Agostino & Pearson, Shapiro–Wilk, and Kolmogorov–Smirnov test) were performed to analyze the distribution of our results. Data are presented as mean $\pm$ SD. Gene expression of PGRMC2, PGRMC1, and PGR-t across the menstrual cycle, the differences of gene expression between genes in EP, LP, ES, MS, and LS phases and comparisons of PGRMC2 protein abundance between control, Sc, and iPGRMC2 d-hEnSCs were assessed by a one-way analysis of variance test. A Kruskal–Wallis test was used to assess differences in PGR- β mRNA expression across the menstrual cycle, for WB analysis of PGRMC2 expression across the menstrual cycle and for outgrowth model analysis between control, Sc, and iPGRMC2 d-hEnSCs. A Wilcoxon test for paired samples was performed to analyze PGRMC2 gene expression and protein abundance in hEnSCs.

A paired t-test was performed to assess the expression of PGRMC2, in the genes expressions used to validate RNA-seq results, and in PRL levels between nd- and d-hEnSCs. A Friedman test was used to compare PGRMC2, PGRMC1, PGR-t, and PGR- β mRNA expression in d-hEnSCs and nd-hEnSCs, to evaluate the significant differences between control, Sc, and iPGRMC2 d-hEnSCs PGRMC2 gene expression, and to compare the percentage of Ki-67 protein staining between control, Sc, and iPGRMC2 d-hEnSCs. A two-way ANOVA test was performed to assess PGRMC2 protein abundance in all subcellular fractions.

All results were normalized and compared with those of the EP phase (in menstrual cycle analyses) or endogenous control expression. In all cases, a P -value of <0.05 was considered statistically significant.

Results

PGRMC2 localization, mRNA expression, and protein abundance in the human endometrium across the menstrual cycle

We identified PGRMC2 localization and abundance by imaging human endometrial tissues across the menstrual cycle (Fig. 2A–H). Further, PGRMC2 staining was more intense during the MS phase in endometrial epithelial glands and hEnSCs, in contrast to the LS phases, where PGRMC2 protein expression decreased in all

endometrial cellular types. Luminal epithelial cells showed consistent PGRMC2 protein expression from proliferative phases until the MS phase. In line with this, RT-qPCR analysis revealed a significant decrease in PGRMC2 mRNA expression in the LS phase compared with the MS phase ($P < 0.05$; Fig. 2I). We identified a 28-kDa protein corresponding to PGRMC2 in all the endometrial samples. However, we did not detect significant differences in PGRMC2 protein abundance across the menstrual cycle by WB (Fig. 2J).

Differences between classical and non-classical PGRs mRNA levels in the human endometrium across the menstrual cycle

We next assessed mRNA expression of classical and non-classical PGRs to elucidate the role of PGRMC2 in the human endometrium across the menstrual cycle. Expression of PGRMC1 ($P < 0.05$) and classical PGR ($P < 0.05$) were significantly reduced during the secretory phases compared to the proliferative phases (Fig. 3A–C), in contrast to PGRMC2 (Fig. 2I). To determine the quantity of each receptor present in each menstrual phase, we compared the expression of genes encoding the non-classical and classical PGRs (Fig. 3D–H). Non-classical PGRs had greater expression in all phases ($P < 0.05$). PGRMC1 was the most expressed receptor in the proliferative phases and ES phase, while PGRMC2 was the most expressed receptor in LS phase. During the MS phase (Fig. 3G), the expression of non-classical PGRMC1 and PGRMC2 were significantly increased compared with classical PGRs in endometrial tissue ($P < 0.05$).

PGRMC2 localization, mRNA, and protein analysis in hEnSCs during *in vitro* decidualization

To determine if the behavior of PGRMC2 after *in vitro* decidualization correlated with the observed decrease in PGRMC2 expression during the LS phase compared to the MS phase, we assessed the expression of PGRMC2 in decidualized and non-decidualized stromal cells. PGRMC2 gene expression was significantly increased ($P < 0.0001$) in d-hEnSCs compared with nd-hEnSCs (Fig. 4A). Accordingly, PGRMC2 protein levels were significantly higher in d-hEnSCs ($P < 0.001$; Fig. 4B, C).

To assess if PGRMC2 localization changes during decidualization, we induced *in vitro* decidualization in primary hEnSCs and performed immunofluorescence staining. PGRMC2 was localized mainly in membranes of d-hEnSCs (Fig. 4E, G); in nd-hEnSCs, PGRMC2 was less concentrated and more difficult to define in specific cellular compartments (Fig. 4D, F).

To verify these results, we isolated cytoplasmic, membranes/organelles, and nuclear fractions from nd-hEnSCs and d-hEnSCs. WB analysis confirmed the presence of PGRMC2 in hEnSCs, and PGRMC2 abundance was significantly greater in the membranes/organelles fraction ($P < 0.05$) versus the other subcellular fractions (Fig. 4H, I).

Differences between classical and non-classical PGRs mRNA levels after *in vitro* decidualization in hEnSCs

To confirm that the *in vitro* behavior of non-classical and classical PGRs corresponded to endometrial tissue observations, the expression of genes encoding PGRs was compared in hEnSCs after decidualization in cell culture (Supplementary Fig. S2). In non-decidualized cells, the expression of non-classical PGRs was significantly higher than classical PGRs ($P < 0.001$; Supplementary Fig. S2A). After *in vitro* decidualization, a significant increase in non-classical PGRs mRNA expression was also

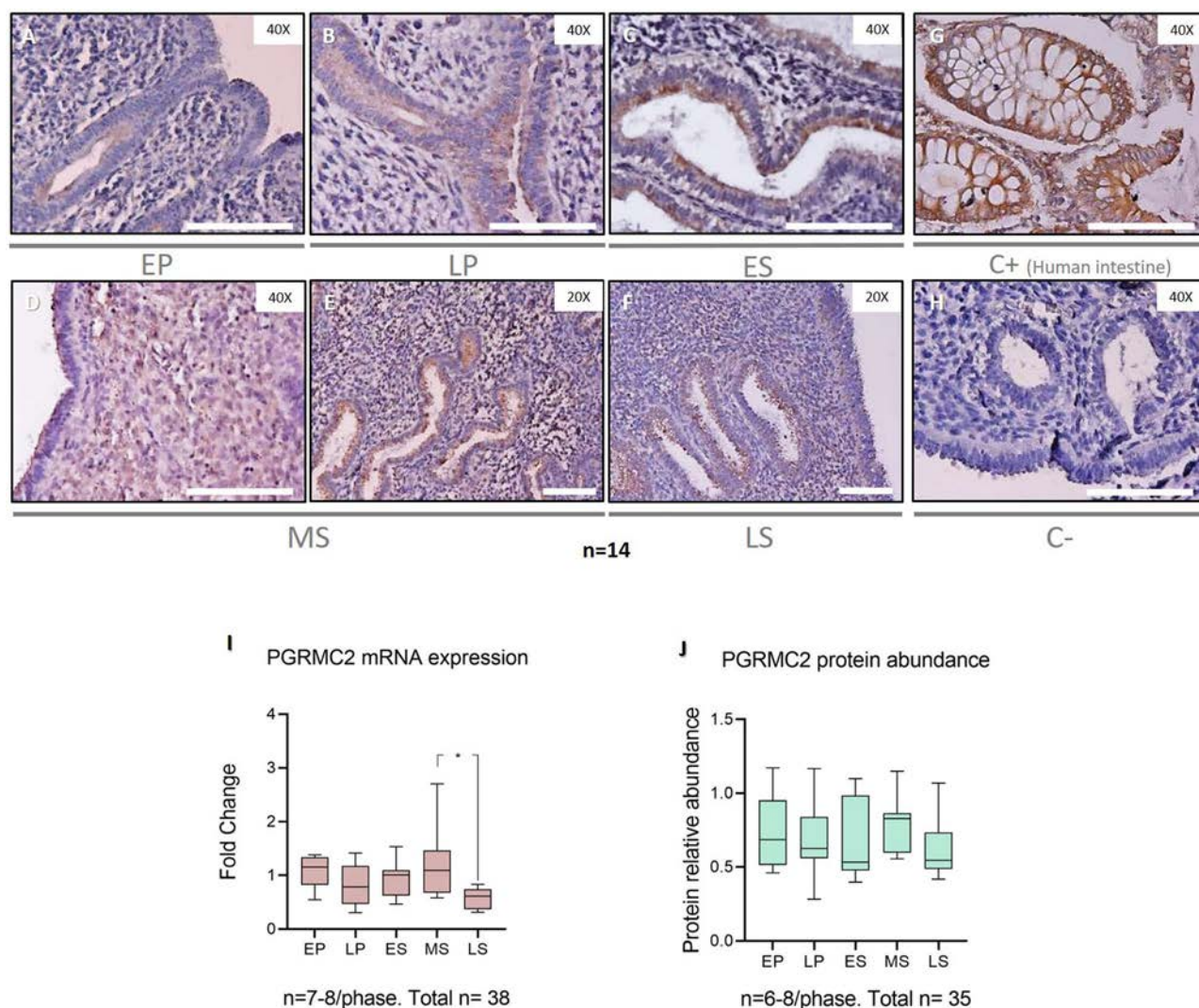


Figure 2. Progesterone receptor membrane component 2 (PGRMC2) localization, mRNA expression, and protein abundance in the human endometrium across the menstrual cycle. (A–F) Immunostaining of PGRMC2 in human endometrium at different menstrual cycle stages. Positive control tissue (human intestine) probed with anti-PGRMC2 (G) and a negative control (H) are also shown. Images were acquired at $\times 20$ and $\times 40$ magnifications. Scale bars: 100 μm . PGRMC2 mRNA and protein abundance were quantified by RT-qPCR (I) and western blot (J), respectively, in human endometrium throughout the menstrual cycle. Box and whiskers show PGRMC2 (I) gene expression, represented as mRNA fold change, and (J) protein abundance, represented as protein relative abundance, both in the different endometrial phases. In the plots, central bar denotes the median value, the box contains the 25th to 75th percentiles, and the whiskers mark the 5th and 95th percentiles; * $P < 0.05$; EP, early proliferative; LP, late proliferative; ES, early secretory; MS, mid-secretory; LS, late secretory; C+, positive control; C–, negative control.

observed ($P < 0.05$), being PGRMC2 gene expression the highest (Supplementary Fig. S2B).

Transcriptomic impact of PGRMC2 silencing during decidualization

To unravel the function of PGRMC2 during the decidualization process, we performed a PGRMC2 silencing in hEnSCs. Then, we compared iPGRMC2 d-hEnSCs with Scramble d-hEnSCs using RNA-seq.

The silencing approach assay was validated by RT-qPCR showing an $\approx 80\%$ reduction of PGRMC2 expression in d-hEnSCs transfected with PGRMC2 siRNAs when compared to control d-hEnSCs ($P < 0.01$; Supplementary Fig. S3A).

RNA-seq data showed 4687 DEGs between Sc d-hEnSCs and iPGRMC2 d-hEnSCs ($n = 4$), 2420 being upregulated and 2267 downregulated (FDR-adjusted $P < 0.05$) (Fig. 5A and Supplementary Table S3). We subsequently performed a functional enrichment

analysis in which we observed that genes overexpressed in iPGRMC2 d-hEnSCs participate in biological processes related to endoplasmic reticulum (ER) function, vesicular transport, lipid biosynthesis, cell signaling, phosphorylation, morphogenesis, angiogenesis, cell migration, and cell adhesion. In contrast, genes under-expressed in iPGRMC2 d-hEnSCs are involved in biological processes related to aerobic respiration, RNA translation, and ribosome biosynthesis (Fig. 5B and Supplementary Table S4).

Regarding cellular components, the most expressed genes in iPGRMC2 d-hEnSCs were mainly associated to the nucleus, the ER, the Golgi apparatus, cell vesicles and vacuoles, the plasma membrane, cell junctions, and the extracellular matrix, while the least expressed genes in iPGRMC2 d-hEnSCs were mainly associated with the mitochondria, ribosomes, and the catalytic complex (Supplementary Table S4).

Regarding molecular functions, the most expressed genes in iPGRMC2 d-hEnSCs mainly had kinase function, binding to

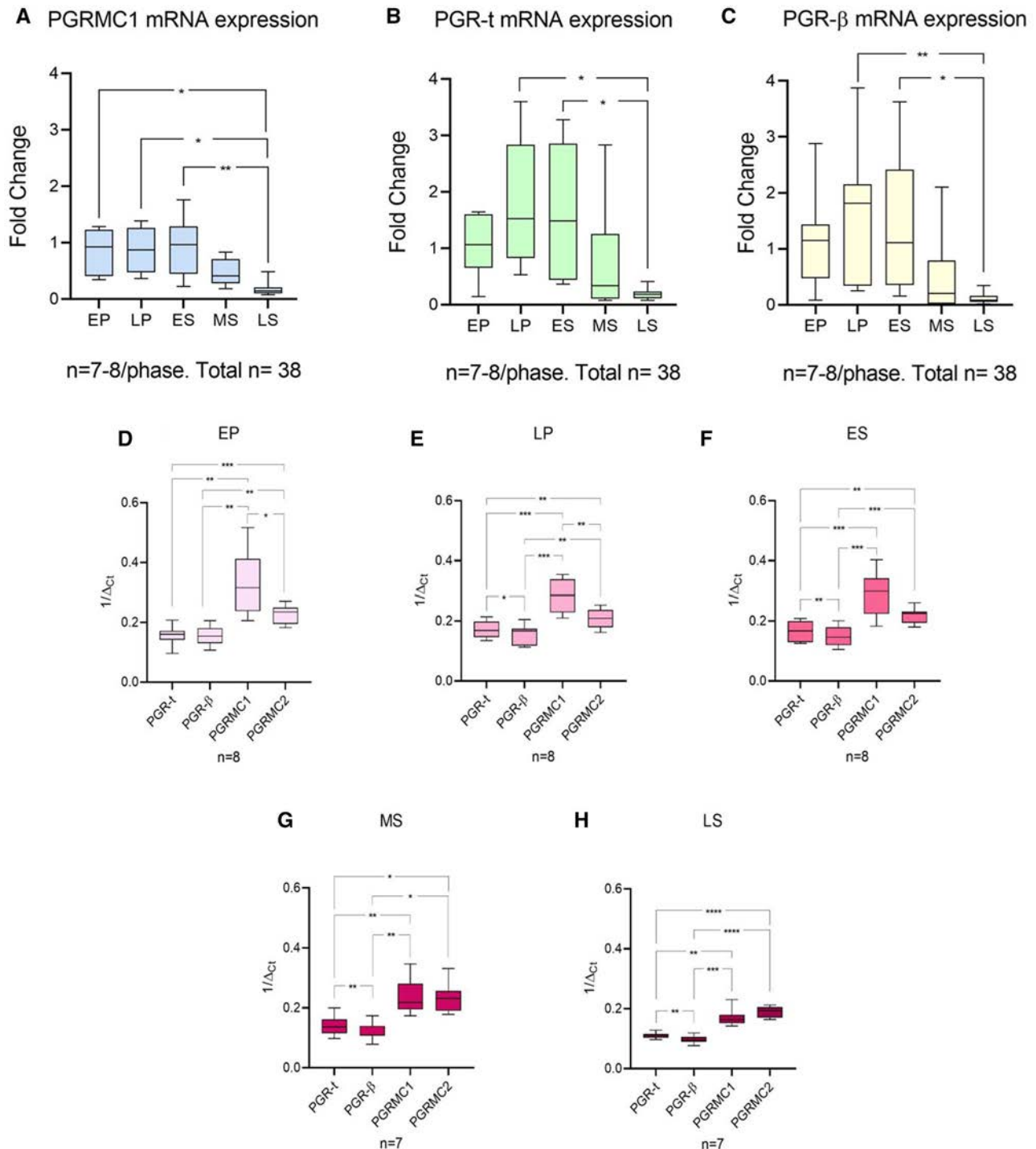


Figure 3. Differences in mRNA expression between classical and non-classical progesterone receptors (PGRs) across the human menstrual cycle. (A–C) Box and whiskers show the fold change expression of progesterone receptor membrane component 1 (PGRMC1), total progesterone receptors (PGR-t), and beta progesterone receptor (PGR-β) mRNA across the human menstrual cycle. (D–H) Box and whiskers show the comparison of classical and non-classical PGRs mRNA expression in the human endometrium during each phase of the menstrual cycle. mRNA gene expression was represented by fold change (A–C) and inverse of ΔCt ($1/\Delta\text{Ct}$) (D–H). In the plots, central bar denotes the median value, the box contains the 25th to 75th percentiles, and the whiskers mark the 5th and 95th percentiles; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; EP, early proliferative; LP, late proliferative; ES, early secretory; MS, mid-secretory; LS, late secretory.

different substrates and extracellular matrix constituents. On the other hand, the least expressed genes had oxidoreductase function, proton transport, and structural constituents of ribosomes (Supplementary Table S4).

To validate RNA-seq data by RT-qPCR, we selected six DEGs (HGF, COL1A1, TNC, CXCL14, DEPP1, and PGRMC2) based on their

potential significance in female reproductive pathology. We confirmed significant PGRMC2 ($P < 0.0001$), DEPP1 ($P < 0.05$), and CXCL14 ($P < 0.05$) downregulation and TNC ($P < 0.05$) upregulation in iPGRMC2 d-hEnSCs compared with Sc d-hEnSCs (Supplementary Fig. S4A–D). However, we did not detect significant differences in COL1A1 and HGF mRNA expression (Supplementary Fig. S4E, F).

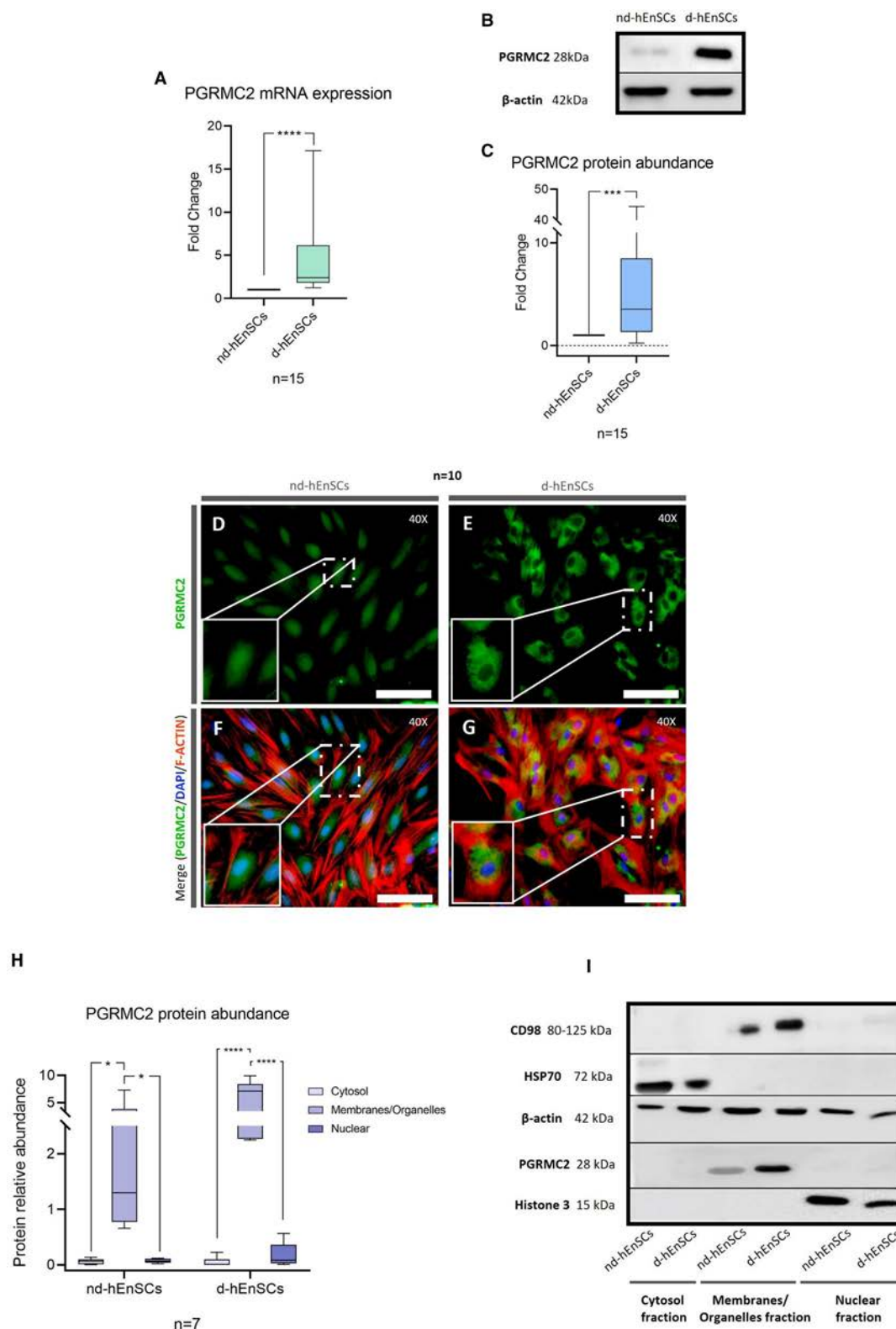


Figure 4. Progesterone receptor membrane component 2 (PGRMC2) mRNA expression, protein abundance, and subcellular localization in human endometrial stromal cells (hEnSCs) during *in vitro* decidualization. (A–C) Box and whiskers show the expression of PGRMC2 in non-decidualized human endometrial stromal cells (nd-hEnSCs) and decidualized human endometrial stromal cells (d-hEnSCs) by quantitative reverse transcription polymerase chain reaction (RT-qPCR) (A) and western blot (B and C). (D–G) PGRMC2 (green) localization in nd-hEnSCs and d-hEnSCs by immunofluorescence; F-actin, rhodamine phalloidin (red); nuclei, DAPI (blue). Merged images are presented in F and G. Images were acquired at $\times 40$ magnifications. Scale bars: 100 μ m. (H, I) Box and whiskers show the subcellular protein abundance of PGRMC2 in nd-hEnSCs and d-hEnSCs. In the plots, central bar denotes the median value, the box contains the 25th to 75th percentiles, and the whiskers mark the 5th and 95th percentiles. Specific markers for membrane/organelles (CD98), cytosol (HSP70), and nuclear fraction (His3) were used. Gene expression and protein abundance are represented as fold change during *in vitro* decidualization, and subcellular protein abundance is represented as protein relative abundance; * $P < 0.05$; *** $P < 0.001$; **** $P < 0.0001$.

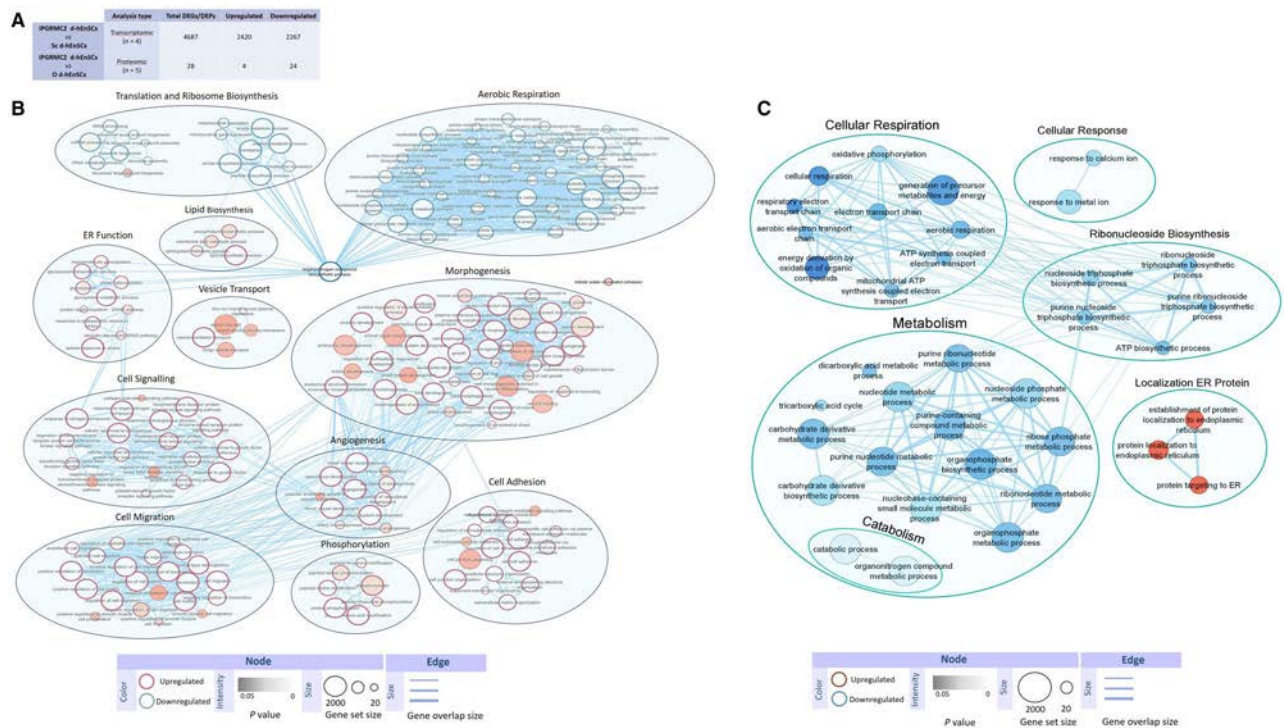


Figure 5. Transcriptomic and proteomic analyses of decidualized human endometrial stromal cells (d-hEnSCs) when progesterone receptor membrane component 2 (PGRMC2) was silenced. **(A)** Differentially expressed genes (DEGs; false discovery rate (FDR)-adjusted $P < 0.05$) and proteins (DEPs) (total, upregulated and downregulated) in pairwise comparison of Scramble control d-hEnSCs (Sc d-hEnSCs) and d-hEnSCs with PGRMC2 silenced (iPGRMC2 d-hEnSCs) in transcriptomic ($n = 4$) and negative control d-hEnSCs (O d-hEnSCs) and d-hEnSCs with PGRMC2 silenced (iPGRMC2 d-hEnSCs) in proteomic analyses ($n = 5$). **(B)** Transcriptomic enrichment map for gene ontology biological processes analyzed by g:Profiler in pairwise comparison of Sc d-hEnSCs and iPGRMC2 d-hEnSCs. Node size represents the number of genes associated with a specific term. Node color represents the FDR-adjusted P-value. **(C)** Proteomic enrichment map for biological processes analyzed by g:Profiler (Benjamini-Hochberg FDR < 0.05) in pairwise comparison of O d-hEnSCs and iPGRMC2 d-hEnSCs. Node size represents the number of proteins associated with a specific term. Node color represents the FDR-adjusted P-value. ER, endometrial reticulum. Both enrichment maps were created with Cytoscape 3.9.1.

Proteomic impact of PGRMC2 silencing during decidualization

To evaluate the changes in the global protein profile during the decidualization process in the absence of PGRMC2, we performed a PGRMC2 silencing in d-hEnSCs. Then, we compared iPGRMC2 d-hEnSCs with control d-hEnSCs using SWATH-MS.

The silencing approach assay was validated by WB, showing a $\approx 60\%$ of reduction of PGRMC2 abundance in hEnSCs transfected with PGRMC2 siRNAs when compared to control d-hEnSCs ($P < 0.001$; [Supplementary Fig. S3B](#)). The SWATH-MS analysis of control d-hEnSCs compared to iPGRMC2 d-hEnSCs ($n = 5$) identified the presence of 2070 different proteins, with a total of 28 DEPs when PGRMC2 protein was silenced, 4 being upregulated and 24 downregulated ($P < 0.05$) ([Fig. 5A](#) and [Supplementary Table S5](#)). Of the 28 DEPs found, the functional enrichment analysis identified 40 biological processes significantly deregulated in d-hEnSCs, mainly related with cellular respiration, cellular response, metabolism, localization of ER proteins, and ribonucleoside biosynthesis ([Fig. 5C](#) and [Supplementary Table S6](#)). Sixty-five cellular components were also found, mainly related to organelle membranes, vesicles, extra-cellular matrix, ER, and mitochondria components ([Supplementary Table S6](#)). Finally, molecules bonding, transporter activity, and steroid biosynthesis are the main molecular functions altered in iPGRMC2 d-hEnSCs ([Supplementary Table S6](#)).

PGRMC2 silencing impact in cell proliferation in d-hEnSCs

To evaluate the possible impact of PGRMC2 silencing on the proliferation of hEnSCs during *in vitro* decidualization process, we

evaluated proliferation marker Ki-67 by immunofluorescence. Fluorescence microscopy quantification showed no statistically significant difference in the number of Ki-67 positive cells between control d-hEnSCs, Sc d-hEnSCs, and iPGRMC2 d-hEnSCs ([Supplementary Fig. S5](#)).

PGRMC2 silencing affects trophoblast invasion

To verify PGRMC2 implication during trophoblast invasion, an *in vitro* invasion model was carried out where iPGRMC2 d-hEnSCs were co-cultured with JEG-3 trophoblast spheroids. The trophoblast spheroids invaded the hEnSC monolayer at different expansion rates between control, Sc, and iPGRMC2 d-hEnSCs at 24 and 48 h of co-culture. Optical and fluorescence measures revealed a significant decrease in trophoblast expansion in iPGRMC2 d-hEnSCs, compared to control ($P < 0.001$) and Sc ($P < 0.01$) d-hEnSCs ([Fig. 6D](#)). In fact, a similar area of the labeled spheroids between 24 and 48 h can be appreciated in iPGRMC2 d-hEnSCs ([Fig. 6C](#)).

Discussion

In the uterus, classical and non-classical PGR expressions oscillate throughout the menstrual cycle, regulating critical events for endometrial receptivity and pregnancy establishment ([Medina-Laver et al., 2021](#)). Despite classical PGRs contributing to the majority of female reproductive processes ([Luetjens et al., 2006](#); [Kowalik et al., 2013](#)), cells completely devoid of classical PGRs also respond to P4 ([Frye and Vongher, 1999](#)).

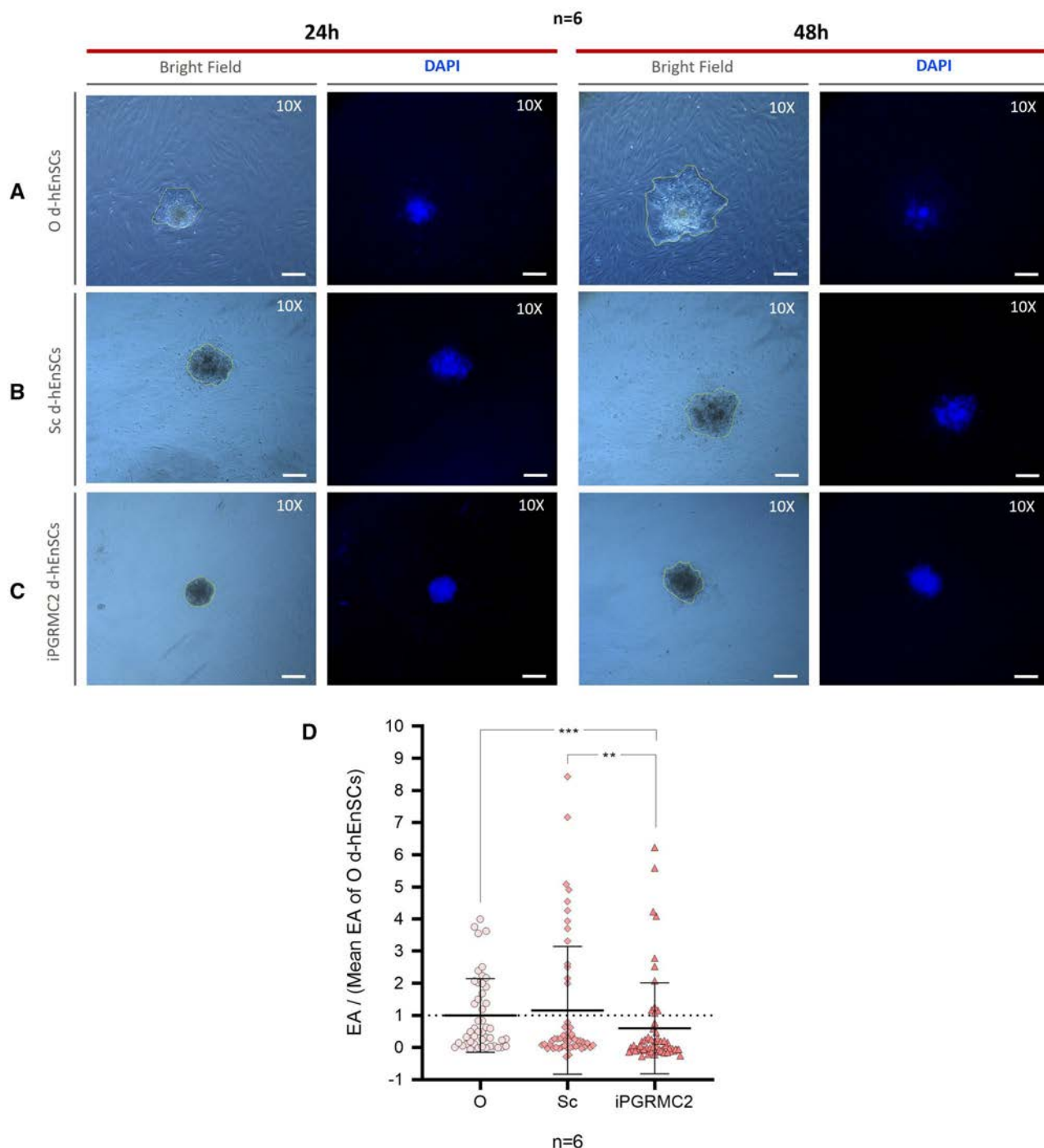


Figure 6. Differences in the trophoblast outgrowth area at 24 and 48 h of co-culture when endometrial stromal progesterone receptor membrane component 2 (PGRMC2) is silenced. Optical and fluorescence photographs showed JEG-3 trophoblastic spheroids labeled in blue invading the stromal monolayer at different expansion area (EA) rates between decidual negative control (O d-hEnSCs; **A**), decidual scramble control (Sc d-hEnSCs; **B**), and decidual-inhibited PGRMC2 stromal cells (iPGRMC2 d-hEnSCs; **C**). Photographs were acquired at 24 and 48 h of co-culture and at $\times 10$ magnifications. Scale bars: 100 μm . **(D)** Quantification of expanded area (EA) differences between negative control (O; $n = 47$ spheroids), scramble control (Sc; $n = 53$ spheroids), and inhibited PGRMC2 (iPGRMC2; $n = 56$ spheroids) decidual endometrial stromal cells. A total of $n = 6$ decidual human endometrial stromal cells (d-hEnSCs) were used. Data are represented as normalized EA $\pm$ SD; ** $P < 0.01$; *** $P < 0.001$.

PGRMC2, a non-classical PGR, inhibits antral follicle development (Griffin et al., 2014; Will et al., 2017) and participates in meiotic spindle assembly (Jühlen et al., 2016). Low PGRMC2 expression was observed in women with poor ovarian reserve, endometrial hyperplasia, and endometriosis (Skiadas et al., 2012; Bunch et al., 2013; Vázquez-Martínez et al., 2020) whereas PGRMC2 absence altered the expression of vascular endothelial growth factor A (VEGF-A), Kit ligand, and AMH (Peluso et al.,

2018). However, the temporal expression and roles of PGRMC2 remain elusive in the human endometrium, and particularly, during decidualization.

To fill this knowledge gap, we first investigated PGRMC2 mRNA and protein expression throughout the menstrual cycle in healthy human endometrial tissue.

We found that endometrial epithelial glands showed the strongest expression of PGRMC2 across the menstrual cycle, with

an evident decrease during the LS phase, aligning with previous findings in macaques and patients with endometriosis (Keator et al., 2012; Bunch et al., 2013). Further, we observed oscillating expression of PGRMC2 in the hEnSCs throughout the menstrual cycle, with a decrease in the LS phase compared to the MS phase, suggesting a potential function of PGRMC2 during embryo implantation.

In contrast to PGRMC1 (Garrido-Gómez et al., 2014; Salsano et al., 2017), PGRMC2 mRNA expression was considerably higher in the endometrium during the MS phase with respect to the LS phase, corroborating that PGRMC1 expression patterns mimic those of classical PGRs (Wang et al., 1998). The complementary expression profiles of endometrial PGRMC1/2 and classical PGRs during the WOI suggests PGRMC2 plays a key role in embryo implantation, as was previously postulated in macaques (Keator et al., 2012).

In the traditional paradigm, classical PGRs drive decidualization, however, our findings evidence additional contributions of non-classical PGRs. We previously reported that PGRMC1 expression was altered in the receptive endometrium with respect to the non-receptive endometrium (Garrido-Gómez et al., 2014), PGRMC1 overexpression during the secretory phases inhibited decidualization (Salsano et al., 2017), and PGRMC1 inhibition in d-hEnSCs dysregulates a myriad of decidualization-related genes (Salsano et al., 2021). In view of these findings, we reasoned PGRMC2 might contribute to endometrial decidualization through similar molecular mechanisms. In our *in vitro* model of human endometrial decidualization, PGRMC2 expression was enhanced in decidualized cells suggesting PGRMC2 deficiency during human decidualization or embryo attachment may lead to reproductive failure, as documented in mice (Clark et al., 2016; McCallum et al., 2016) and zebrafish (Wu et al., 2019).

Our transcriptomic and proteomic findings supported this theory, revealing that targeted PGRMC2 knockdown substantially altered d-hEnSCs expression patterns. Notably, we found the downregulated genes were related to mitochondrial energy production, and protein synthesis (particularly ribosome function, translation processes, and protein localization in the ER) while repressed proteins were related to cellular respiration, biosynthesis, and metabolic response. Conversely, upregulated genes were related to the endomembrane system (mainly the ER and the Golgi apparatus), plasma membrane, and cell junctions. The proteins encoded by the upregulated genes are mainly involved in biological processes related to morphogenesis, angiogenesis, cell migration, and cell adhesion, as well as cellular processes such as ER function, vesicle transport, cell signaling, phosphorylation, and lipid biosynthesis. Proteomically, we detected an overexpression of proteins located in the ER of d-hEnSCs after PGRMC2 silencing.

Taken together, the differential gene expression in the d-hEnSC organelles and membranous systems corroborate the role of PGRMC2 in the extensive secretory remodeling occurring during decidualization (Anelli et al., 2022). Our observations of PGRMC2 localization in membranes and organelles confirm suggested associations for PGRMC1/2 (Albrecht et al., 2012; Bunch et al., 2013; Griffin et al., 2014; Peluso et al., 2014; Jühlen et al., 2016).

Inhibition of PGRMC1, which has complementary expression to PGRMC2, upregulated genes related to ATP generation and transport, protein and lipid biosynthesis, post-transcriptional processing, vesicle trafficking, and oxidative stress alleviation (Salsano et al., 2020, 2021), indicating both non-classical receptors

are involved in the extensive cytoplasmic remodeling during decidualization.

On the other hand, inhibiting classical PGRs during decidualization upregulates genes related to proliferation, angiogenesis, apoptosis, cell migration, adhesion, cytoskeletal organization, cytokine, and chemokine secretion (Kaya et al., 2015; Mazur et al., 2015). While some of these decidualization processes (such as proliferation and apoptosis) are regulated exclusively by classical PGRs, angiogenesis, cell migration, and adhesion appear to be regulated by both non-classical (PGRMC2) and classical PGRs.

PGRMC2 seems to have key ER-related functions, given its expression in the endomembrane system and protection against ER stress. Based on observations from the knockout of other transcription factors regulating decidualization and required for ER and Golgi apparatus remodeling (Zhao et al., 2020; Anelli et al., 2022; Pittari et al., 2022), PGRMC2 silencing appears to impede protein biosynthesis and cellular energy production. PGRMC2 silencing enhanced ER stress, extracellular matrix-receptor interactions, and focal adhesion signaling pathways through post-transcriptional regulation in either the ER or Golgi apparatus (Zhao et al., 2020; Pittari et al., 2022).

A successful trophoblast invasion, necessary for both initiating and maintaining pregnancy, depends on optimal endometrial receptivity (Singh et al., 2011). Invasion is characterized by morphological, biochemical, and vascular changes in the endometrium, along with adequate decidualization of hEnSCs. Our transcriptomic findings showed PGRMC2 silencing upregulates genes but not proteins related to angiogenesis, cell migration, and adhesion which are essential for trophoblastic invasion (Muter et al., 2023). Thus, we hypothesize the RNA overexpression may be a compensatory response to the insufficient protein biosynthesis. A similar response was previously reported in SKOV-3 ovarian cancer cells with limited migration (Albrecht et al., 2012). Our *in vitro* outgrowth model supports our transcriptomic findings, suggesting that PGRMC2 regulates cell migration, but not proliferation, during embryo invasion into the d-hEnSCs. Given the elevated ER stress in d-hEnSCs associated with early pregnancy loss (Liu et al., 2011; Gao et al., 2012; Soczewski et al., 2020), this mechanism of action could explain why the lack of PGRMC2 is associated with post-implantation failure (Clark et al., 2016; McCallum et al., 2016).

Finally, the differential overexpression of non-classical versus classical PGRs in various menstrual phases and in hEnSCs highlights the prominent role of non-classical PGRs in reproductive functions. Moreover, the complementary expression patterns of PGRMC1 and PGRMC2 in both proliferative and secretory phases, as well as in non-decidualized and decidualized cells, suggests a parallel role in embryo implantation, and potentially, in maintaining pregnancy.

Given the limited sample sizes in both transcriptomic and proteomic analyses, our conclusions remain preliminary and should be taken with caution. Using a primary endometrial stromal population that was differentiated *in vitro* cannot fully represent the heterogeneity of the endometrium *in vivo*, nor does it account for the paracrine communications between endometrial cell types.

In conclusion, this study presents the first *in vivo* characterization of PGRMC2 expression throughout the human menstrual cycle and *in vitro* decidualization of hEnSCs. We provide a comparative analysis between the classical and non-classical PGRs systems, along with a novel transcriptomic and proteomic approach to assess the functional implications of targeted PGRMC2 knockdown in hEnSCs. Nevertheless, additional studies are needed to validate PGRMC2's mechanism of action and its

intertwining relationship(s) with other non-classical and classical PGRs during decidualization, to fully understand the role of these proteins in female reproduction.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

All raw sequencing data are available through NCBI's Gene Expression Omnibus (GEO) under GEO Series accession number GSE251843 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE251843>). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository under accession number PXD048494.

Acknowledgements

We thank the professionals at the IVIRMA Valencia clinic for their involvement and assistance in sample collection and to the Genomic and Proteomic Services of the University of Valencia, especially to Vicente Sentandreu and Oreto Antúñez.

Authors' roles

F.D. conceived and designed the experiments. Y.M.-L. performed the experiments. Y.M.-L. and R.G.-M. analyzed the results, and wrote the manuscript. P.A. selected the oocyte donors and collected the samples. R.G.-M., P.d.C., I.D.-H., A.P., and A.Q. supported the experiments and data interpretation, and provided critical discussions. All authors read and approved the final manuscript.

Funding

This research was funded by Instituto de Salud Carlos III (ISCIII) granted to F.D. (PI20/00405 and PI23/00860), co-funded by the European Union. Y.M.-L. was supported by a predoctoral research grant from Generalitat Valenciana (ACIF/2019/262). R.G.-M. was supported by Generalitat Valenciana (CIAPOT/2022/15). P. d.C. was supported by a predoctoral grant for training in research into health (PFIS FI20/00086) from the Instituto de Salud Carlos III. I.D.-H. was supported by the Spanish Ministry of Science, Innovation and Universities (FPU18/01550). A.P. was supported by the Instituto de Salud Carlos III (PFIS FI18/00009). This research was also supported by IVI Foundation-RMA Global (1911-FIVI-103-FD).

Conflict of interest

The authors declare no conflict of interest.

References

- Albrecht C, Huck V, Wehling M, Wendler A. In vitro inhibition of SKOV-3 cell migration as a distinctive feature of progesterone receptor membrane component type 2 versus type 1. *Steroids* 2012; **77**:1543–1550.
- Anelli T, Dalla Torre M, Borini E, Mangini E, Ullisse A, Semino C, Sitia R, Panina-Bordignon P. Profound architectural and functional readjustments of the secretory pathway in decidualization of endometrial stromal cells. *Traffic* 2022; **23**:4–20.
- Bunch K, Tinnemore D, Huff S, Hoffer ZS, Burney RO, Stallings JD. Expression patterns of progesterone receptor membrane components 1 and 2 in endometria from women with and without endometriosis. *Reprod Sci* 2013; **21**:190–197.
- Cahill MA. Progesterone receptor membrane component 1: an integrative review. *J Steroid Biochem Mol Biol* 2007; **105**:16–36.
- Clark N, Pru CA, Yee S, Lydon JP, Peluso JJ, Pru JK. Conditional ablation of progesterone receptor membrane component 2 causes female premature reproductive senescence. *Endocr Soc* 2016; **157**:3309–3319.
- Conneely OM, Mulac-Jericevic B, DeMayo F, Lydon JP, O'Malley BW. Reproductive functions of progesterone receptors. *Recent Prog Horm Res* 2002; **57**:339–355.
- Deryabin P, Griukova A, Nikolsky N, Borodkina A. The link between endometrial stromal cell senescence and decidualization in female fertility: the art of balance. *Cell Mol Life Sci* 2020; **77**:1357–1370.
- Dominguez F, Gadea B, Mercader A, Esteban FJ, Pellicer A, Simón C. Embryologic outcome and secretome profile of implanted blastocysts obtained after coculture in human endometrial epithelial cells versus the sequential system. *Fertil Steril* 2010; **93**:774–782.e1.
- Finidori-Lepicard J, Schorderet-Slatkine S, Hanoune J, Baulieu E-E. Progesterone inhibits membrane-bound adenylate cyclase in *Xenopus laevis* oocytes. *Nat* 1981; **292**:255–257.
- Foresta C, Rossato M, Di Virgilio F. Ion fluxes through the progesterone-activated channel of the sperm plasma membrane. *Biochem J* 1993; **294** (Pt 1):279–283.
- Frye CA, Sumida K, Dudek BC, Harney JP, Lydon JP, O'Malley BW, Pfaff DW, Rhodes ME. Progesterone's effects to reduce anxiety behavior of aged mice do not require actions via intracellular progesterone receptors. *Psychopharmacology (Berl)* 2006; **186**:312–322.
- Frye CA, Vongher JM. Progesterone has rapid and membrane effects in the facilitation of female mouse sexual behavior. *Brain Res* 1999; **815**:259–269.
- Gao HJ, Zhu YM, He WH, Liu AX, Dong MY, Jin M, Sheng JZ, Huang HF. Endoplasmic reticulum stress induced by oxidative stress in decidual cells: a possible mechanism of early pregnancy loss. *Mol Biol Rep* 2012; **39**:9179–9186.
- Garrido-Gómez T, Quiñero A, Antúñez O, Díaz-Gimeno P, Bellver J, Simón C, Domínguez F. Deciphering the proteomic signature of human endometrial receptivity. *Hum Reprod* 2014; **29**:1957–1967.
- Gellersen B, Fernandes MS, Brosens JJ. Non-genomic progesterone actions in female reproduction. *Hum Reprod Update* 2009; **15**:119–138.
- Graham J, Clarke C. Physiological action of progesterone in target tissues. *Endocr Rev* 1997; **18**:502–519.
- Griffin D, Liu X, Pru C, Pru JK, Peluso JJ, Al GET. Expression of progesterone receptor membrane component-2 within the immature rat ovary and its role in regulating mitosis and apoptosis of spontaneously. *Biol Reprod* 2014; **91**:1–11.
- Hirota Y. Progesterone governs endometrial proliferation differentiation switching and blastocyst implantation. *Endocr J* 2019; **66**:199–206.
- Jühlen R, Landgraf D, Huebner A, Koehler K. Identification of a novel putative interaction partner of the nucleoporin ALADIN. *Biol Open* 2016; **5**:1697–1705.
- Karteris E, Zervou S, Pang Y, Dong J, Hillhouse EW, Randeva HS, Thomas P. Progesterone signaling in human myometrium through two novel membrane G protein-coupled receptors: potential role in functional progesterone withdrawal at term. *Mol Endocrinol* 2006; **20**:1519–1534.
- Kaya HS, Hantak AM, Stubbs LJ, Taylor RN, Bagchi IC, Bagchi MK. Roles of progesterone receptor A and B isoforms during human endometrial decidualization. *Mol Endocrinol* 2015; **29**:882–895.

- Keator CS, Mah K, Slayden OD. Alterations in progesterone receptor membrane component 2 (PGRMC2) in the endometrium of macaques afflicted with advanced endometriosis. *Mol Hum Reprod* 2012;**18**:308–319.
- Kim JJ, Kurita T, Bulun SE. Progesterone action in endometrial cancer, endometriosis, uterine fibroids, and breast cancer. *Endocr Rev* 2013;**34**:130–162.
- Kimura I, Nakayama Y, Konishi M, Terasawa K, Ohta M, Itoh N, Fujimoto M. Functions of MAPR (membrane-associated progesterone receptor) family members as heme/steroid-binding proteins. *Curr Protein Pept Sci* 2012;**13**:687–696.
- Kowalik M, Rekawiecki R, Kotwica J. The putative roles of nuclear and membrane-bound progesterone receptors in the female reproductive tract. *Reprod Biol* 2013;**13**:279–289.
- Large M, DeMayo FJ. The regulation of embryo implantation and endometrial decidualization by progesterone receptor signaling. *Mol Cell Endocrinol* 2012;**358**:155–165.
- Li R, Zhong C, Yu Y, Liu H, Sakurai M, Yu L, Min Z, Shi L, Wei Y, Takahashi Y et al. Generation of blastocyst-like structures from mouse embryonic and adult cell cultures. *Cell* 2019;**179**:687–702.e18.
- Liao S, Mi H, Chai L, Wang H. Effects of progesterone receptor on proliferation of uterine leiomyoma cells. *J Biol Regul Homeost Agents* 2019;**33**:1685–1693.
- Liu AX, He WH, Yin LJ, Lv PP, Zhang Y, Sheng JZ, Leung PCK, Huang HF. Sustained endoplasmic reticulum stress as a cofactor of oxidative stress in decidual cells from patients with early pregnancy loss. *J Clin Endocrinol Metab* 2011;**96**:E493–E497.
- Liu L, Wang J, Zhao L, Nilsen J, McClure K, Wong K, Brinton RD. Progesterone increases rat neural progenitor cell cycle gene expression and proliferation via extracellularly regulated kinase and progesterone receptor membrane components 1 and 2. *Endocrinology [Internet]* 2009;**150**:3186–3196.
- Loutradis D, Beretsos P, Arabatzis E, Anagnostou E, Drakakis P. The role of steroid hormones in ART. *J Steroid Biochem Mol Biol* 2008;**112**:1–4.
- Lozovyy V, Richardson L, Saade G, Menon R. Progesterone receptor membrane components: key regulators of fetal membrane integrity. *Biol Reprod* 2021;**104**:445–456.
- Lucas ES, Vrljicak P, Muter J, Diniz-da-Costa MM, Brighton PJ, Kong C-S, Lipecki J, Fishwick KJ, Odendaal J, Ewington LJ et al. Recurrent pregnancy loss is associated with a pro-senescent decidual response during the peri-implantation window. *Commun Biol* 2020;**3**:37.
- Luetjens CM, Didolkar A, Kliesch S, Paulus W, Jeibmann A, Böcker W, Nieschlag E, Simoni M. Tissue expression of the nuclear progesterone receptor in male non-human primates and men. *J Endocrinol* 2006;**189**:529–539.
- Mazur EC, Vasquez YM, Li X, Kommagani R, Jiang L, Chen R, Lanz RB, Kovanci E, Gibbons WE, DeMayo FJ. Progesterone receptor transcriptome and cistrome in decidualized human endometrial stromal cells. *Endocrinology* 2015;**156**:2239–2253.
- Mccallum ML, Pru CA, Niikura Y, Yee SP, Lydon JP, Peluso JJ, Pru JK. Conditional ablation of progesterone receptor membrane component 1 results in subfertility in the female and development of endometrial cysts. *Endocrinology* 2016;**157**:3309–3319.
- Medina-Laver Y, Rodríguez-Varela C, Salsano S, Labarta E, Domínguez F. What do we know about classical and non-classical progesterone receptors in the human female reproductive tract? A review. *Int J Mol Sci* 2021;**22**:11278.
- Morel Y, Roucher F, Plotton I, Goursaud C, Tardy V, Mallet D. Evolution of steroids during pregnancy: maternal, placental and fetal synthesis. *Ann Endocrinol (Paris)* 2016;**77**:82–89.
- Muter J, Lynch VJ, McCoy RC, Brosens JJ. Human embryo implantation. *Development* 2023;**150**:dev201507.
- Peluso J, Liu X, Uliasz T, Pru C, Kelp N, Pru J. PGRMC1/2 promotes luteal vascularization and maintains the primordial follicles of mice. *Reproduction* 2018;**156**:365–373.
- Peluso JJ, Griffin D, Liu X, Horne M. Progesterone receptor membrane component-1 (PGRMC1) and PGRMC-2 interact to suppress entry into the cell cycle in spontaneously immortalized rat granulosa cells1. *Biol Reprod* 2014;**91**:1–12.
- Peluso JJ, Pru C, Liu X, Kelp N, Pru JK. Progesterone receptor membrane component 1 and 2 regulate granulosa cell mitosis and survival through a NFkB-dependent mechanism. *Biol Reprod* 2019;**100**:1571–1580.
- Perez-Patiño C, Barranco I, Parrilla I, Valero ML, Martínez EA, Rodríguez-Martínez H, Roca J. Characterization of the porcine seminal plasma proteome comparing ejaculate portions. *J Proteomics* 2016;**142**:15–23.
- Pittari D, Dalla Torre M, Borini E, Hummel B, Sawarkar R, Semino C, van Anken E, Panina-Bordignon P, Sitia R, Anelli T. CREB3L1 and CREB3L2 control Golgi remodelling during decidualization of endometrial stromal cells. *Front Cell Dev Biol* 2022;**10**:986997.
- Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* 2019;**14**:482–517.
- Richardson LS, Taylor RN, Menon R. Reversible EMT and MET mediate amnion remodeling during pregnancy and labor. *Reprod Biol* 2020;**13**:1–16.
- Ruiz-Alonso M, Blesa D, Díaz-Gimeno P, Gómez E, Fernández-Sánchez M, Carranza F, Carrera J, Vilella F, Pellicer A, Simón C. The endometrial receptivity array for diagnosis and personalized embryo transfer as a treatment for patients with repeated implantation failure. *Fertil Steril* 2013;**100**:818–824.
- Sakai N, Maruyama T, Sakurai R, Masuda H, Yamamoto Y, Shimizu A, Kishi I, Asada H, Yamagoe S, Yoshimura Y. Involvement of histone acetylation in ovarian steroid-induced decidualization of human endometrial stromal cells. *J Biol Chem* 2003;**278**:16675–16682.
- Salazar E, Calzada L. The role of progesterone in endometrial estradiol- and progesterone-receptor synthesis in women with menstrual disorders and habitual abortion. *Gynecol Endocrinol* 2007;**23**:222–225.
- Salsano S, González-Martín R, Quiñonero A, López-Martín S, Gómez-Escribano A, Pérez-Debén S, Yañez-Mo M, Domínguez F. Novel nonclassic progesterone receptor PGRMC1 pulldown-precipitated proteins reveal a key role during human decidualization. *Fertil Steril* 2020;**113**:1050–1066.e7.
- Salsano S, González-Martín R, Quiñonero A, Pérez-Debén S, Domínguez F. Deciphering the role of PGRMC1 during human decidualization using an in vitro approach. *J Clin Endocrinol Metab* 2021;**106**:2313–2327.
- Salsano S, Quiñonero A, Pérez S, Garrido Gómez T, Simón C, Domínguez F. Dynamic expression of PGRMC1 and SERBP1 in human endometrium: an implication in the human decidualization process. *Fertil Steril* 2017;**108**:832–842.e1.
- Singh M, Chaudhry P, Asselin E. Bridging endometrial receptivity and implantation: network of hormones, cytokines, and growth factors. *J Endocrinol* 2011;**210**:5–14.
- Skiadas C, Duan S, Correll M, Rubio R, Karaca N, Ginsburg E, Quackenbush J, Racowsky C. Ovarian reserve status in young women is associated with altered gene expression in membrana granulosa cells. *Mol Hum Reprod* 2012;**18**:362–371.

- Soczewski E, Grasso E, Gallino L, Hauk V, Fernández L, Gori S, Papparini D, Leirós CP, Ramhorst R. Immunoregulation of the decidualization program: focus on the endoplasmic reticulum stress. *Reproduction* 2020;**159**:R203–R211.
- Szekeres-Bartho J. The role of progesterone in feto-maternal immunological cross talk. *Med Princ Pract* 2018;**27**:301–307.
- Vázquez-Martínez E, Bello-Alvarez C, Hermenegildo-Molina A, Solís-Paredes M, Parra-Hernández S, Cruz-Orozco O, Silvestri-Tomassoni J, Escobar-Ponce L, Hernández-López L, Reyes-Mayoral C et al. Expression of membrane progesterone receptors in eutopic and ectopic endometrium of women with endometriosis. *Biomed Res Int* 2020;**2020**:2196024.
- Wang H, Critchley HOD, Kelly RW, Shen D, Baird DT. Progesterone receptor subtype B is differentially regulated in human endometrial stroma. *Mol Hum Reprod* 1998;**4**:407–412.
- Wetendorf M, Wu S-P, Wang X, Creighton CJ, Wang T, Lanz RB, Blok L, Tsai SY, Tsai M-J, Lydon JP et al. Decreased epithelial progesterone receptor A at the window of receptivity is required for preparation of the endometrium for embryo attachment. *†. Biol Reprod* 2017;**96**:313–326.
- Will E, Liu X, Peluso J. AG 205, a progesterone receptor membrane component 1 antagonist, ablates progesterone's ability to block oxidative stress-induced apoptosis of human granulosa/luteal cells†. *Biol Reprod* 2017;**96**:843–854.
- Wu X, Williams M, Patel P, Kew K, Zhu Y. Subfertility and reduced progesterone synthesis in PGRMC2 knockout zebrafish. *Gen Comp Endocrinol* 2019;**282**:113218.
- Wu X-J, Zhu Y. Downregulation of nuclear progesterone receptor (PGR) and subfertility in double knockouts of progesterone receptor membrane component 1 (pgrmc1) and pgrmc2 in zebrafish. *Gen Comp Endocrinol* 2020;**285**:113275.
- Yoshinaga K. Progesterone and its downstream molecules as blastocyst implantation essential factors. *Am J Reprod Immunol* 2014;**72**:117–128.
- Zhang C, Murphy BD. Progesterone is critical for the development of mouse embryos. *Endocrine* 2014;**46**:615–623.
- Zhao F, Liu H, Wang N, Yu L, Wang A, Yi Y, Jin Y. Exploring the role of Luman/CREB3 in regulating decidualization of mice endometrial stromal cells by comparative transcriptomics. *BMC Genomics* 2020;**21**:103–114.

Reproductive biology

Targeting osteopontin alleviates endometriosis and inflammation by inhibiting the RhoA/ROS axis and achieves non-invasive *in vitro* detection via menstrual blood

Han Wang^{1,2}, Binming Wang^{1,2}, Meiling Wu^{1,2}, Jiefang Lu¹, and Ping Duan^{1,2,*}

¹Department of Obstetrics and Gynecology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, China

²Oncology Discipline Group, The Second Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

*Corresponding address. Department of Obstetrics and Gynecology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang 325027, China. E-mail: dppdpp@126.com <https://orcid.org/0000-0001-7926-3839>

ABSTRACT

STUDY QUESTION: How does osteopontin (OPN) in endometriosis ectopic stromal cells (EESCs) participate in the pathogenesis of endometriosis and achieve non-invasive detection *in vitro*?

SUMMARY ANSWER: Targeted OPN regulates endometriosis's necroptosis and inflammatory state by inhibiting the RhoA/reactive oxygen species (ROS) axis, thereby alleviating endometriosis and enabling non-invasive detection of menstrual blood *in vitro*.

WHAT IS KNOWN ALREADY: Endometriosis is a chronic inflammatory disease. Recent studies have shown that OPN plays an important role in disease progression by regulating cell death and inflammation.

STUDY DESIGN, SIZE, DURATION: The study included 20 patients diagnosed with endometriosis (confirmed by laparoscopy and histology) and 10 controls without endometriosis. Endometriotic stromal cells were isolated from endometrial samples, while menstrual blood endometrial cells (MESCs) were isolated from menstrual blood. These cells were then cultured *in vitro* and utilized in subsequent experiments.

PARTICIPANTS/MATERIALS, SETTING, METHODS: OPN expression in EESCs was assessed using inflammatory factor sequencing, immunohistochemical staining (IHC), quantitative real-time PCR (qRT-PCR) analysis, and Western blotting (WB). The biological behavior of OPN and its effects on inflammatory factors were examined using EdU, wound-healing, Transwell, and ELISA assays. Necroptosis in EESCs and its impact on inflammatory factors were detected through qRT-PCR, WB, and Calcein-AM/PI fluorescence assays. The examination of mitochondrial stress in EESCs involved the use of the Mitochondrial Membrane Potential ($\Delta\Psi$ m) Assay, ROS detection, and Calcein-AM Loading/cobalt chloride Quenching. qRT-PCR, WB, and other experiments were conducted to verify the regulation of necroptosis and inflammatory factor levels in EESCs by OPN through the RhoA/ROS axis. Knockdown of OPN and its inhibitory effect on endometriosis lesion size were confirmed using AAV9 virus, IHC, qRT-PCR, WB, and other experiments. Additionally, OPN expression in MESCs was detected using transcriptome sequencing, RT-PCR, WB, and other experiments.

MAIN RESULTS AND THE ROLE OF CHANCE: *In vitro* assays demonstrated a significant upregulation of OPN in EESCs, and the knockdown of OPN effectively inhibited necroptosis and the release of inflammatory factors. OPN inhibited necroptosis and inflammatory factor release by mediating RhoA-dependent ROS production and blocking mixed lineage kinase domain-like protein phosphorylation at the cell membrane. *In vivo*, targeting of OPN can inhibit the growth of endometriosis lesions. Clinically, OPN was also significantly upregulated in the menstrual blood of patients with endometriosis.

LARGE SCALE DATA: N/A.

LIMITATIONS, REASONS FOR CAUTION: Due to limitations in obtaining surgical specimens, our study primarily involved collecting endometriosis tissues from women during the proliferative and secretory phases of the menstrual cycle. We observed a significant overexpression of OPN in the samples used for our investigation. However, the expression of OPN in endometriosis tissues during the intermenstrual phase remains unknown.

WIDER IMPLICATIONS OF THE FINDINGS: Our findings highlight the pivotal role of the OPN/RhoA/ROS axis in the regulation of necroptosis and the release of inflammatory factors. OPN knockdown exerts a therapeutic effect *in vivo*, and the high expression detection of OPN in menstrual blood *in vitro*. In summary, targeting OPN provides possibilities for the treatment and detection of endometriosis.

STUDY FUNDING/COMPETING INTEREST(S): This study was supported by the National Natural Science Foundation of China (82071626), the Zhejiang Province Public Welfare Technology Application Research Project (LGF21H040010), and the Clinical Research project of the Second Affiliated Hospital of Wenzhou Medical University (1010293). The authors have no conflicts of interest.

Keywords: endometriosis / OPN / necroptosis / inflammatory factors / menstrual blood

Introduction

Endometriosis is a gynecological inflammatory disease characterized by the presence of endometrial tissue outside the uterine cavity, leading to chronic pelvic pain and infertility (Zondervan et al., 2020). The development of endometriosis involves complex regulatory factors such as immunosuppression, genetic alterations, and pelvic microenvironmental stimulation (Wang et al., 2017; Bischoff and Simpson, 2000; Khan et al., 2005). However, the exact etiology of endometriosis remains unclear, making the exploration of the pathogenesis of endometriosis of significant value.

Currently, there is a deficiency of clinical diagnostic markers specifically for endometriosis. While carbohydrate antigen 125, human epididymis protein 4, and carbohydrate antigen 199 are commonly used as additional diagnostic markers (Wang et al., 2023), their sensitivity and specificity are low, and they often do not differentiate endometriosis from other gynecologic diseases like endometrial, cervical, and ovarian cancers. Therefore, the lack of reliable detection markers poses a challenge in diagnosing endometriosis. Since endometriosis is not a strictly defined cancer and its changes do not significantly affect systemic circulation, peripheral blood tests do not accurately reflect the precise changes in the endometrium of endometriosis *in vivo*. Menstrual blood, on the other hand, is regarded as a non-invasive source of endometrial tissue that can be obtained *in vitro* (Gargett et al., 2016). Therefore, we aim to explore the mechanisms of inflammation and identify early diagnostic biomarkers for endometriosis, as well as investigate the potential of menstrual blood testing as a non-invasive diagnostic method.

In our tissue sequencing results, we observed aberrant expression of inflammatory factors, including osteopontin (OPN), in endometriosis. OPN, also known as secreted phosphoprotein 1, is a critical regulator of cell death and immune responses (Huang et al., 2006; Wen et al., 2016). It plays a crucial role in various diseases, including cancer, cardiovascular disease, and inflammation (Qin et al., 2018; Lok and Lyle, 2019; Song et al., 2021). Silencing OPN has been shown to effectively reduce inflammatory response and disease progression (Shirakawa et al., 2018; Zhu et al., 2019). Although endometriosis is categorized as an inflammatory disease, whether OPN plays a regulatory role in endometriosis needs further investigation.

Necroptosis, a regulated form of cell death, triggers an innate immune response by releasing intracellular components upon cell burst (Dondelinger et al., 2016). Kinase activity is crucial for promoting inflammation, and receptor-interacting protein kinase 1 (RIP1) has been recognized as a crucial regulator of cell survival, inflammation, and disease (de Almagro et al., 2017). Z-DNA-binding protein 1 triggers receptor-interacting protein kinase 3 (RIP3)-dependent necroptosis and inflammation by inducing endogenous Z-type nucleic acid, contributing to the development of chronic inflammatory states (Zhang et al., 2020). Deficiency in RIP3 or mixed lineage kinase domain-like protein (MLKL) can prevent airway inflammation in chronic obstructive pulmonary disease after acute exposure to cigarette smoke (Mizumura et al., 2014; Oikonomou et al., 2021). Although

necroptosis is indispensable in inflammation, its role in endometriosis inflammation remains unexplored.

Mitochondria play a crucial role in cell function by providing the energy needed for growth and reproduction through various chemical reactions (Picard and Shirihi, 2022). Alterations in mitochondrial chemical reactions can contribute to the development of diseases, such as increased susceptibility to inflammation and oxidative stress with mutations in mitochondria-related genes (Andrieux et al., 2021). Elevated mitochondrial reactive oxygen species (mROS) triggers the binding of gasdermin D, a protein that forms pores in the mitochondrial membrane. These pores release mROS, promoting the occurrence of necroptosis mediated by RIP1/RIP3/MLKL in Parkinson's disease (Weindel et al., 2022). Our previous studies have shown mitochondrial dysfunction in endometriosis (Wang et al., 2022), along with exposure to high levels of oxidative stress. Therefore, investigating the interplay between mitochondrial function, necroptosis, and inflammation in endometriosis is crucial.

Our study demonstrates that OPN is highly expressed in ectopic endometriotic stromal cells (EESCs) and endometrial cells derived from menstrual blood of endometriosis patients (E-MESCs), acting as a key inhibitor of necroptosis in endometriosis. By down-regulating OPN, we can effectively impede necroptosis by restraining ROS levels, thereby mitigating the inflammatory state in endometriosis. Based on these findings, OPN shows promise in the treatment and detection of endometriosis.

Materials and methods

Clinical specimens

This study adhered to ethical guidelines from the Second Affiliated Hospital of Wenzhou Medical University. The study recruited 20 women diagnosed with endometriosis by laparoscopy and histological analysis at the Second Affiliated Hospital of Wenzhou Medical University. The presence of endometrial glands in the tissue of the lesion is a definitive diagnostic criterion for endometriosis. Tissue samples were collected during the procedure and immediately sent to the laboratory. All tissues were divided into two samples. We performed *in vitro* primary cell culture on one sample and the remaining samples were fixed in formalin and prepared for paraffin embedding. To maintain the study's precision, the patients who participated in this study did not receive any hormone therapy before surgery. Ten women without endometriosis or adenomyosis were considered the control group who were undergoing surgery for other benign gynecologic conditions, such as uterine smooth muscle tumors. The control group consisted of patients diagnosed with uterine smooth muscle tumors who sought medical attention primarily for symptoms such as excessive menstrual flow, prolonged menstrual periods, and the presence of lower abdominal masses. Ultrasound imaging confirmed the presence of larger tumors (>5 cm) requiring myomectomy, with hysteroscopic resection being the main surgical procedure performed. Additionally, endometrial sampling and biopsy were conducted during curettage to exclude the possibility of other endometrial diseases. All samples were collected during the proliferative phase of the menstrual cycle, histologically confirmed according to established criteria,

Table 1. Clinical characteristics of women with and without (control) endometriosis.

	Control (n = 10)	Endometriosis (n = 20)	P value
Age (years)	39.40 ± 2.91	37.20 ± 3.71	0.113
BMI (kg/m ²)	23.68 ± 0.85	23.73 ± 0.76	0.867
Menstrual average cycle (days)	29.00 ± 1.94	29.85 ± 1.53	0.200
Menstrual duration (days)	5.70 ± 0.67	6.15 ± 0.67	0.095
ASRM stage			
Stage 1		0/20	
Stage 2		2/20	
Stage 3		5/20	
Stage 4		13/20	

Statistical analysis was performed using Student's t-test. Data are mean ± SD. ASRM: American Society for Reproductive Medicine.

and transported to the laboratory on ice in Dulbecco's modified Eagle's medium (DMEM) (Biological Industries, Israel) to isolate and culture the endometriotic stromal cells (ESCs) for *in vitro* assays. In addition, menstrual blood was collected from the corresponding patient using a sterile menstrual cup on the second day of the last menstrual period before the procedure. The control group consisted of 6 samples, while the endometriosis group consisted of 13 samples. Detailed patient information is shown in Table 1. All tissue samples were obtained with the informed consent of the patients.

Cell culture

The lesion tissues were minced and treated with 4% Type IV collagenase (C5138, Sigma, USA) at 37°C for 45–60 min. Using the tissue sample sources described above (Clinical Specimens), we used stainless steel mesh sieves to isolate the EESCs from endometriotic tissues and the normal endometrial stromal cells (NESCs) from the eutopic endometrial tissue of patients with uterine smooth muscle tumors. After centrifugation at 1200g for 3 min, the lower cells were washed with DMEM, repeating this process three times. The cells were then resuspended in DMEM supplemented with 10% fetal bovine serum (FBS; Biological Industries, Israel) and incubated at 37°C with 5% CO₂. Menstrual blood endometrial cells (MESCs) were isolated from menstrual blood using Percoll density gradient centrifugation. First, the collected menstrual blood was mixed with PBS and centrifuged at 1500g for 5 min. The lower blood was washed with PBS, and this step was repeated three times. The washed blood was resuspended in DMEM and slowly added to a centrifuge tube containing Percoll in a ratio equivalent to DMEM. After centrifugation at 1800g for 25 min, the white membrane layer in the centrifuge tube was aspirated and washed three times with PBS. Finally, normal MESCs (N-MESCs) and MESCs of endometriosis patients (E-MESCs) were resuspended in DMEM supplemented with 10% FBS and incubated at 37°C with 5% CO₂. When the cell density reached 70–90%, cells were detached and passaged by 0.25% trypsin–EDTA (0.25%, 25200056, Gibco, USA). ESCs were verified through positive immunofluorescence staining of vimentin and CD10 and negative staining of E-cadherin, cytokeratin 7, and FSH receptor. Validated ESCs were used for subsequent experiments. Cells isolated from each patient were used for one experiment at a time in triplicate.

Immunohistochemical staining

Normal endometrial (EN) tissues (n = 6) and ectopic endometrial (EC) tissues (n = 6) were paraffin-embedded, sectioned, and deparaffinized. Antigen repair was conducted in a solution of 3% hydrogen peroxide and citric acid, followed by incubation with primary antibodies [OPN (22952-1-AP; Proteintech 1:200), RIP1 (73271; Cell Signaling Technology 1:200), RIP3 (13526; Cell

Signaling Technology 1:200), MLKL (14993; Cell Signaling Technology 1:200), Pro-caspase 8 (108333; Abcam 1:200), RhoA (54835; Abcam 1:400)], and secondary antibodies (ZSGB-BIO, China). The sections were analyzed using ImageJ. Microscopic observation of the results was facilitated by a microscope (Leica, Wetzlar, Germany), and analysis was done using ImageJ.

qRT-PCR analysis

Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, USA), followed by reverse transcription to produce cDNA. Quantitative real-time PCR (qRT-PCR) analysis was performed using SYBR Green PCR Master Mix (TOYOBO, Shanghai, China), with ACTB, GAPDH, and Tublin serving as the reference genes. The mean relative gene expression was determined, and differences were calculated using the 2^{−ΔΔCt} method. Specific sequence primers were designed as follows: forward, 5'-GCCG AGGTGATAGTGTGGTTAT-3' and reverse, 5'-ATTCAACTCC TCGCTTTCATGT-3' for OPN; forward, 5'-GCTGTAGTACTCTGGC GATATT-3' and reverse, 5'-CTCTTGGGCAGTACTCACTGATG-3' for RIP1; forward, 5'-GGTAACAACCCGAAAGTAGACT-3' and reverse, 5'-CATAACTTGACGCACGACATCAG-3' for RIP3; forward, 5'-GCATGCCTGTTTACCCATAAG-3' and reverse, 5'-CTGCTCC TTCTTGATCTCCTTGA-3' for MLKL; forward, 5'-TCAACAAGAGC CTGCTGAAGATA-3' and reverse, 5'-GGAGAGTCCGAGATTGTCAT TAC-3' for Pro-caspase8; forward, 5'-GCAAGAGAGACCCAA CAGAAGAA-3' and reverse, 5'-TCCAGCAATTCACCTCTTTCAT-3' for RhoA; forward, 5'-ACTCTTCCAGCCTTCCTTCC-3' and reverse, 5'-CAATGCCAGGGTACATGGTG-3' for ACTB; forward, 5'-CTGGG CTACACTGAGCACC-3' and reverse, 5'-AAGTGGTCGTTGAGGGC AATG-3' for GAPDH; forward, 5'-GGCCAAGGGTCACTACACG-3' and reverse, 5'-GCAGTCGCAGTTTTCACACTC-3' for Tublin.

Western blotting

Whole-cell and nuclear proteins were extracted, and the proteins were transferred to polyvinylidene fluoride membrane and incubated with primary antibody [OPN (22952-1-AP; Proteintech 1:1000), RIP1 (73271; Cell Signaling Technology 1:1000), RIP3 (13526; Cell Signaling Technology 1:1000), MLKL (14993; Cell Signaling Technology 1:1000), Pro-caspase 8 (108333; Abcam 1:1000), RhoA (54835; Abcam 1:1000), p-MLKL (196436; Abcam 1:1000), ACTB (20536-1-AP; Proteintech 1:10000), GAPDH (10494-1-AP; Proteintech 1:10000), Na⁺/K⁺-ATPase (14418-1-AP; Proteintech 1:10000)] and the secondary antibodies. The membrane was treated with an enhanced chemiluminescence solution, and the results were analyzed using ImageJ (National Institutes of Health, USA). Membrane and cytosolic proteins were extracted using a membrane and cytosol protein extraction kit (Cat# 89842, Thermo Fisher Scientific).

Cell transfection

siRNA against OPN (siOPN) and its negative control were transfected using Lipofectamine 3000 (Life Technologies, Gaithersburg, MD, USA). The target sequences of the siRNAs were as follows: siOPN-1 3'-GTCTCACCATTCTGATGAA-5', and siOPN-2 3'-CCAAGTAAGTCCAACGAAA-5'. The LV4 lentiviral vectors were used to construct the OPN knockdown and RhoA knockdown plasmids. Lentivirus was generated in HEK293T cells and the supernatant was collected for transfection of EESCs. The lentivirus supernatant was added to the complete culture medium at a multiplicity of infection of 100 along with 5 µg/ml polybrene. Transfection efficiency was evaluated after 72 h using qRT-PCR and Western blotting (WB). For cell transfection experiments, only EESCs from the third to sixth generations were used, and cells from passages beyond this range were excluded from the experiments.

Cell proliferation assay

Proliferation of EESCs was assessed using the EdU and cell counting kit-8 (CCK-8) assays. For the EdU assays, EESCs were inoculated in 24-well plates (2×10^4 /ml) and divided into three groups with three parallel wells each based on whether or not they were treated with siOPN after attachment. After treatment for 48 h, cells were labeled with EdU following the manufacturer's instructions (Beyotime, C0071S), and observed under a fluorescence microscope. The following formula was applied to calculate the proliferative capability of cells: EdU incorporation rate (%) = Edu-positive cells (green)/Hoechst positive cells (blue). For the CCK-8 assays, six replicate wells per group were used in EESCs of OPN knockdown and RhoA knockdown with or without 1.5 µM ROS inhibitor (DPI) (MCE, Shanghai, China) 10 µl of CCK-8 (MCE, Shanghai, China) was injected and incubated for 2 h after 0, 12, 24, and 36 h of treatment. Absorbance at 450 nm was detected with a microplate reader, and then a proliferation curve was plotted based on the optical density value. Experiments were conducted in triplicate.

Cell migration assay

Migration of EESCs was evaluated using the wound-healing and Transwell assays. For the wound-healing assays, EESCs were inoculated in six-well plates at 2×10^5 /well and divided into three groups with three parallel wells each based on whether or not they were treated with siOPN after attachment. Once a confluent monolayer was formed, scratches were made in the wells using a 200 µl pipet tip. The scratch distances were measured at 0, 24, and 48 h to calculate the wound closure rate (%): (scratch distance at 0 h – scratch distance at 24 h)/scratch distance at 0 h. For the Transwell assays, the upper chamber of a 24-well Transwell plate (Corning, NY, USA) was coated with diluted Matrigel (Corning, NY, USA) and allowed to solidify for 30 min at 37°C. The cell suspension was prepared at a concentration of 1.5×10^6 cells/ml, and 0.2 ml of the cell suspension was added to each upper chamber. In the lower chamber, 0.8 ml of medium containing 5% FBS was added. After 24 h of incubation, the upper chamber was fixed with formaldehyde, and the cells were stained with 6% crystalline violet. The invasive migrating cells were visualized and counted using SPOT imaging software (Nikon, Tokyo, Japan) and counted by ImageJ software.

ELISA

The levels of inflammatory factors and ROS were measured using an ELISA kit (MultiSciences, China). To assess inflammatory factor levels at the cellular level, we collected and centrifuged the

cell culture medium supernatants of NESCs, EESCs treated with or without 30 µM necroptosis inhibitor (Nec-1) (MCE, Shanghai, China), N-MESCs, and E-MESCs at 1000g for 20 min to remove any cellular debris. The concentrations of IL-1β, tumor necrosis factor-α (TNF-α), IL-6, IL-8, and IL-15 in the supernatants were then determined. Additionally, to measure inflammatory factors and ROS in mouse blood, we collected blood samples from each group of mice and centrifuged them at 1000g for 20 min. The serum concentrations of IL-1β, TNF-α, IL-6, IL-8, IL-15, and ROS were subsequently analyzed. After completing all steps of the ELISA according to the manufacturer's instructions, including sample addition, incubation, washing, and coloration, absorbance at 450 nm was detected within 15 min. The concentration of inflammatory factors was calculated based on the standard curve. Three replicate holes per group and experiments were performed in triplicate.

Lactate dehydrogenase concentration

To determine the occurrence of necroptosis in NESCs and EESCs treated with or without OPN knockdown, the levels of lactate dehydrogenase (LDH) were measured. A kit from Sigma (MAK066) was utilized for this purpose, following the manufacturer's instructions. Experiments were conducted in triplicate.

Calcein-AM/PI fluorescence

The viability of EESCs was assessed using Calcein-AM/PI staining (Solarbio, Beijing, China) following the manufacturer's instructions. Calcein-AM (2 µmol/l) and propidium iodide (4.5 µmol/l) were added to the culture wells of OPN knockdown-EESCs treated with or without TNF-α (10 µM, 24 h) and incubated for 15 min at 37°C in the dark. Living cells exhibited green cytoplasmic fluorescence, while dead cells exhibited red nuclei. Cell viability was expressed as the percentage of viable cells. Three replicate holes per group and experiments were performed in triplicate.

Mitochondrial membrane potential assay

Mitochondrial membrane potential ($\Delta\Psi_m$) in EESCs was assessed using the JC-1 fluorescence assay (Invitrogen, USA) following the manufacturer's instructions. EESCs treated with OPN knockdown, DPI (1.5 µM, 24 h), or cyclosporine A (CyA) (7 nM, 24 h) (MCE, Shanghai, China) were seeded at a density of 2×10^5 cells/well in a six-well cell culture plate. They were then incubated with 1 mg/ml JC-1 at 37°C for 20 min and washed three times with PBS. The ratio of the JC-1 aggregates (red) to the JC-1 monomers (green) fluorescence was calculated using a microscope. Three replicate holes per group and experiments were performed in triplicate.

Glycolytic rate assay

The extracellular acidification rate (ECAR), oxygen consumption rate (OCR), baseline glycation rates, and compensatory glycation rates of EESCs treated with or without OPN knockdown were measured using a Seahorse XF24 analyzer to assess cellular metabolism. In brief, 20 000 cells were plated in each well of an XF24 cell culture microplate and cultured in XF DMEM measurement solution. After washing, the cells were incubated at 37°C in a non-CO₂ incubator for 60 min. The cells were treated with Rot/AA and 2-deoxy-D-glucose according to the instructions and the above indicators were measured. All the machines and substances used in the experiment were obtained from Agilent Technologies. Three replicate holes per group and experiments were performed in triplicate.

Detection of cytosolic ROS and mitochondrial ROS

Cytosolic and mitochondrial ROS levels of NESCs and EESCs treated with OPN knockdown, DPI, or CyA were detected using DCFH-DA (Sigma-Aldrich, USA) and MitoSOX (Invitrogen, USA) staining, respectively. In the experiment, ESCs were plated at a density of 2×10^5 cells/well in a 6-well cell culture plate. They were incubated with either 10 mM DCFH-DA or 5 mM MitoSOX working solution at 37°C for 20 min in darkness. After incubation, the cells were washed with PBS and stained with 1 mg/ml Hoechst 33342 (Invitrogen, USA). Fluorescent signals were analyzed using a microscope to quantify cytosolic and mitochondrial ROS levels in the cells. Three replicate holes per group and experiments were performed in triplicate.

Calcein-AM loading/cobalt chloride quenching

Mitochondrial permeability transition pore (mPTP) opening was evaluated using the Calcein-AM loading/cobalt chloride (CoCl_2) quenching assay (Thermo Fisher Scientific, USA). EESCs treated with OPN knockdown, DPI, or CyA were treated with 1 mM Calcein-AM in the presence of 5 mM CoCl_2 at 37°C in darkness for 15 min to quench cytoplasmic and nuclear signals. Calcein was excited at 494 nm and emission was recorded at 517 nm. Before evaluation under a microscope, cells were washed twice with Hank's balanced salt solution containing 10 mM HEPES (pH 7.2). Three replicate wells per group and experiments were performed in triplicate.

Co-immunoprecipitation assay

Co-immunoprecipitation (Co-IP) assay was performed to investigate protein-protein interactions. EESCs were cultured in 10-cm dishes and harvested at a minimum cell count of 60 000 per group. Cell lysates were divided and treated with primary antibodies (OPN, RhoA, RIP3, and p-MLKL) overnight at 4°C with rotation. After washing the magnetic beads three times, the lysate with the primary antibody was combined with the magnetic beads and incubated for 4 h at 4°C. The supernatant was discarded, and the remaining lysate was boiled in a sample buffer. Protein levels of OPN, RhoA, RIP3, and p-MLKL were determined by WB analysis, using 10% of the total protein as input. The specificity of the antibodies used for immunoprecipitation was confirmed with negative controls, including nonimmune IgG.

Animal studies

All animal experiments were performed according to the Guide for the Care and Use of Laboratory Animals, which was published by the US National Institute of Health. The Animal Welfare and Research Ethics Committee at Wenzhou Medical University reviewed and approved this study. We established a mouse model of endometriosis based on a previous report (Zhang *et al.*, 2018), in which 10-week-old female C57BL/6 mice were randomized into donor and recipient groups. In the donor group, uterine horns were dissected and sliced in saline solution to fully expose the endometrial surface. Endometriosis lesions were formed in the effective abdominal cavity of recipient group mice after intraperitoneal inoculation with endometrial fragments from donor group mice. In the recipient group, the AAV9 virus (1.5×10^{11} vg/mouse; ViGene Biosciences, Shandong, China) was injected into the tail vein on the second day. On the 14th day, the mice were euthanized, and the endometrial lesions were recorded. Lesion measurements included length, width, and volume, calculated using the prolate ellipsoid geometric model: $(\text{length} \times \text{width}^2)/2$. Transfection efficiency was evaluated using immunohistochemical staining (IHC), qRT-PCR,

and WB. Animal experiments were approved by the Institutional Animal Care and Use Committee.

Statistical analysis

The data are presented as the mean $\pm$ SD from at least three independent experiments. All statistical analyses were performed using GraphPad Prism 8 (San Diego, CA, USA) software. Statistical significance was measured using Student's t-test, one-way ANOVA, and two-way ANOVA. In all figures, differences were considered significant at $P < 0.05$. NS is nonsignificant ($P \geq 0.05$), * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Results

OPN is significantly upregulated in endometriosis

We analyzed inflammation-related gene sequencing data to investigate changes in inflammatory factors in endometriosis. Our findings revealed a significant increase in OPN protein levels in EC tissues compared to that of the EN tissues (Fig. 1A). Consistently, analysis of the GSE51981 dataset showed that OPN levels correlated with the stage of endometriosis, with higher levels in Stage III/IV compared to Stage I/II (Fig. 1B). IHC experiments provided further evidence of increased levels of OPN protein in tissues from the proliferative phase of EC tissues (Fig. 1C). Given the high expression status of OPN in both ectopic endometrial epithelium and mesenchyme, we investigated the differences in malignant biological behaviors, such as cell viability and cell invasion, between endometriosis ectopic endometrial epithelial cells (EEECs) and EESCs. CCK8 assays and Transwell assays showed that EESCs had more cell proliferation and cell invasion ability compared with EEECs, so we chose EESCs as the focus of the subsequent investigation on endometriosis (Supplementary Fig. S1A and B). Considering the potential for significant changes in gene and protein expression in the endometrium throughout the menstrual cycle, we specifically examined OPN expression in EC tissues during the proliferative and secretory phases. IHC assays revealed elevated OPN expression in both the proliferative and secretory phases of EC tissues compared to EN tissues, with no significant difference observed between the two phases (Supplementary Fig. S2). Based on this finding, we can conclude that sampling time was not a sensitive point in the study setting, and therefore we chose the proliferation period as the time point for subsequent studies. qRT-PCR and WB assays were performed to demonstrate the higher OPN mRNA and protein levels in EESCs compared to NESCs (Fig. 1D and E). Together, these results validate the significant upregulation of OPN in endometriosis, suggesting its potential role in disease progression.

Knockdown of OPN inhibits EESC proliferation, migration, invasion, and inflammatory factor release

To investigate the impact of OPN on the biological behavior of EESCs, we transfected EESCs with siRNA targeting OPN (siOPN) to knock down its expression, and subsequently a significant decrease in OPN mRNA and protein levels was observed (Fig. 2A and B). EdU assays revealed that OPN knockdown significantly inhibited EESC proliferation (Fig. 2C). Wound-healing and Transwell assays demonstrated suppression of EESC migratory and invasive abilities upon OPN knockdown (Fig. 2D and E). Consistent with these findings, analysis of inflammation-related gene sequencing data revealed significant changes in the expression of inflammatory factors in endometriosis (Fig. 2F). To investigate the impact of OPN on the levels of inflammatory factors in EESCs, we used ELISA assays. These revealed that OPN

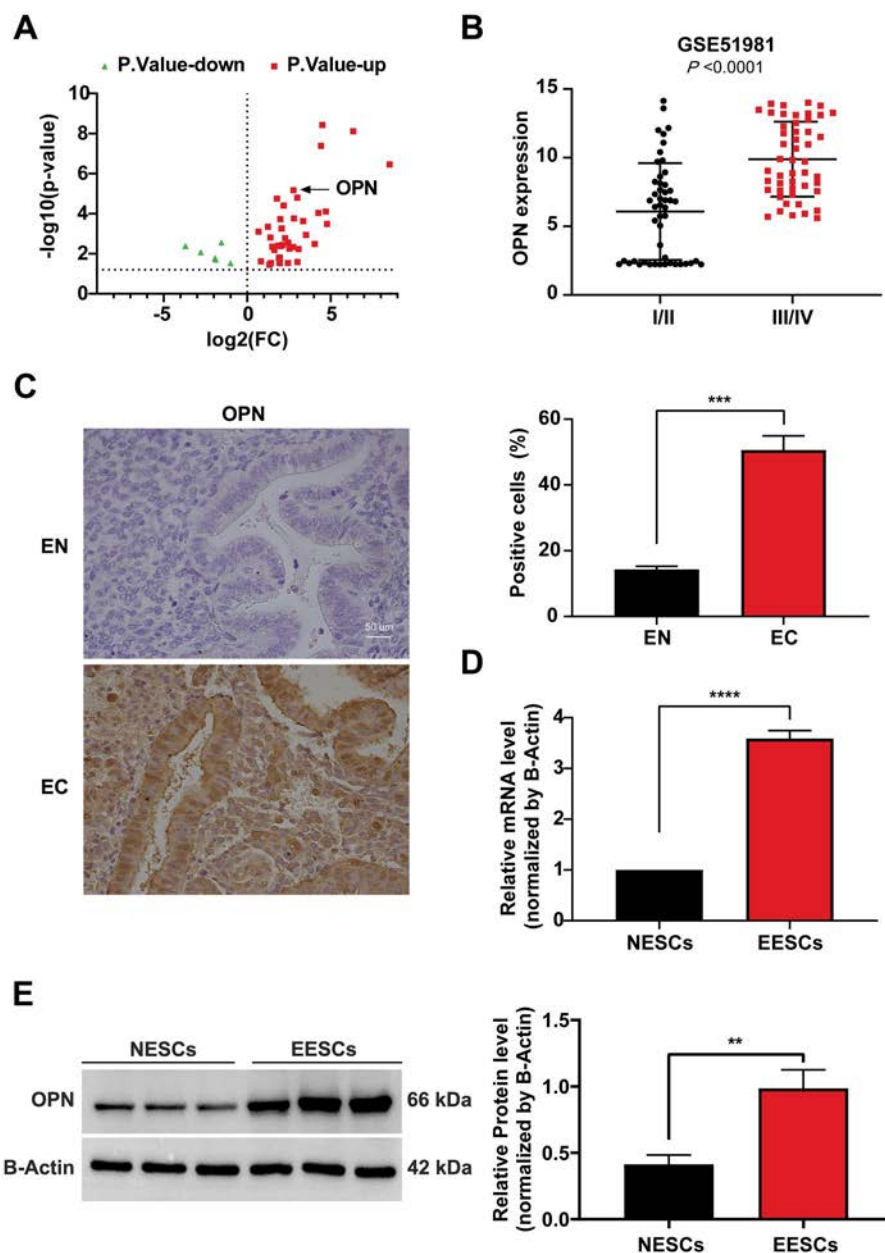


Figure 1. OPN is significantly upregulated in endometriosis. (A) OPN expression was significantly higher, as indicated by cytokine microarray analysis of inflammatory factors. The differentially expressed (fold change (FC) > 2 , $P < 0.05$ by Student's t-test) inflammatory factors are shown. Red and green colors represent upregulation and downregulation, respectively. (B) OPN expression levels in different stages of endometriosis in the GEO database (GSE51981) (black, I/II stages; red, III/IV stages). (C) Representative photographs of OPN protein expression and the percentage of positive OPN protein expression in EN (black) tissues and EC (red) tissues by IHC assays (OPN is stained dark brown, original magnification $\times 400$). (D) qRT-PCR assays confirmed OPN mRNA levels in NESCs (black) and EESCs (red). (E) WB assays revealed OPN levels in NESCs (black) and EESCs (red). All data are shown as the mean $\pm$ SD of three independent experiments ($**P < 0.01$, $***P < 0.001$, $****P < 0.0001$). EC: ectopic endometrial; EESCs: endometriosis ectopic stromal cells; EN: normal endometrial; IHC: immunohistochemical staining; NESCs: normal endometrial stromal cells; OPN: osteopontin; qRT-PCR: quantitative real-time PCR analysis; WB: Western blotting.

knockdown inhibited the levels of pro-inflammatory factors including IL-1 β , TNF- α , IL-6, IL-8, and IL-15, while promoting the levels of the anti-inflammatory factor IL-10 (Fig. 2G). These findings suggested that OPN promotes proliferation, migration, invasion, and release of inflammatory factors in EESCs. Thus, we confirmed that the knockdown of OPN could inhibit these cellular processes and may have beneficial effects on endometriosis.

Increased necroptosis in endometriosis and its modulation by OPN knockdown

Considering that necroptosis promotes inflammation (Pasparakis and Vandenabeele, 2015; Newton and Manning, 2016), we

investigated whether necroptosis is present in the inflammatory disease of endometriosis. Significant alterations in necroptosis marker proteins, including RIP1, RIP3, MLKL, and Pro-caspase8, were observed in EESCs compared with NESCs by qRT-PCR and WB. Specifically, RIP1, RIP3, and MLKL were enhanced in EESCs, whereas Pro-caspase8 was attenuated (Fig. 3A and B). LDH can be released from cells through cell membrane pores that are triggered by necroptosis (Cheng et al., 2021). Thus, we evaluated the LDH levels and found them to be increased in EESCs, indicating that necroptosis occurs in EESCs (Fig. 3C). To determine the impact of necroptosis on the release of inflammatory factors in endometriosis, we inhibited necroptosis using Nec-1, regarded as a

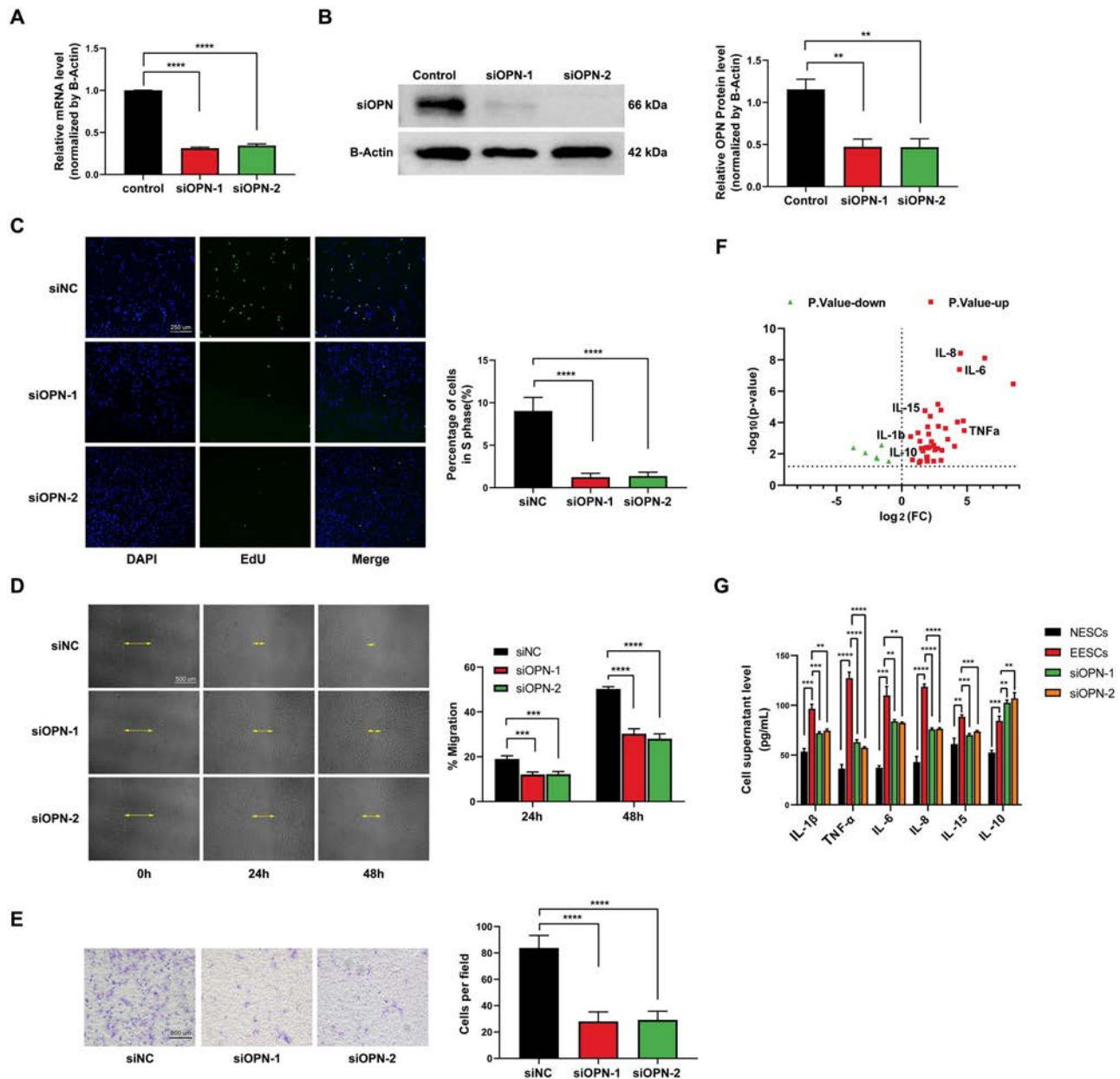


Figure 2. Knockdown of OPN inhibits EESCs' proliferation, migration, invasion, and inflammatory factor release. (A and B) qRT-PCR and WB assays showed the knockdown efficiency of siOPN in EESCs (black, control; red, EESCs treated with siOPN-1; green, EESCs treated with siOPN-2). (C) Proliferative capability of EESCs treated with or without siOPN was assessed by EdU assay and the EdU incorporation rate was quantified (green, EdU; blue, DAPI; original magnification 200×). (D) Migration capability of EESCs treated with or without siOPN was assessed by wound-healing assay and the wound closure rate was quantified (yellow arrow, wound-healing area; original magnification 40×). (E) Migration capability of EESCs treated with or without siOPN was assessed by Transwell assay and the migration cells were quantified (original magnification 200×). (F) Inflammatory factor microarray analysis showed that the expression of inflammatory factors was upregulated. The differentially expressed (fold change >2, $P < 0.05$ by Student's *t*-test) inflammatory factors are shown. Red and green colors represent upregulation and downregulation, respectively. (G) ELISA assays measured the levels of inflammatory factors (IL-1β, TNF-α, IL-6, IL-8, IL-15, and IL-10) in EESCs with or without siOPN treatment (black, NESC; red, EESC; green, EESCs treated with siOPN-1; orange, EESCs treated with siOPN-2). All data are shown as the mean ± SD of three independent experiments (** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). EESCs: endometriosis ectopic stromal cells; ESCs: endometrial stromal cells; NESC: normal endometrial stromal cells; OPN: osteopontin; qRT-PCR: quantitative real-time PCR; siOPN: siRNA of OPN; TNF-α: tumor necrosis factor-α; WB: Western blotting.

Nec-1, and found that Nec-1 suppressed the release of pro-inflammatory factors and promoted the release of anti-inflammatory factors, indicating necroptosis affects the release of inflammatory factors in endometriosis (Fig. 3D). To assess the impact of OPN on necroptosis in endometriosis, we introduced a lentivirus-mediated OPN knockdown vector into EESCs. We confirmed the efficiency of transfection through qRT-PCR and WB experiments, which revealed a significant decrease in OPN

expression at both the mRNA and protein levels upon OPN knockdown (Fig. 3E and F). OPN knockdown reduced the level of necroptosis marker protein in EESCs (Fig. 3G and H) and hindered the release of LDH, indicating that OPN promoted necroptosis (Fig. 3I). Intriguingly, live-dead assays indicated that OPN knockdown suppressed cell death by inhibiting necroptosis, but OPN knockdown alone also facilitated cell death, suggesting that necroptosis is not the sole mechanism through which OPN promoted

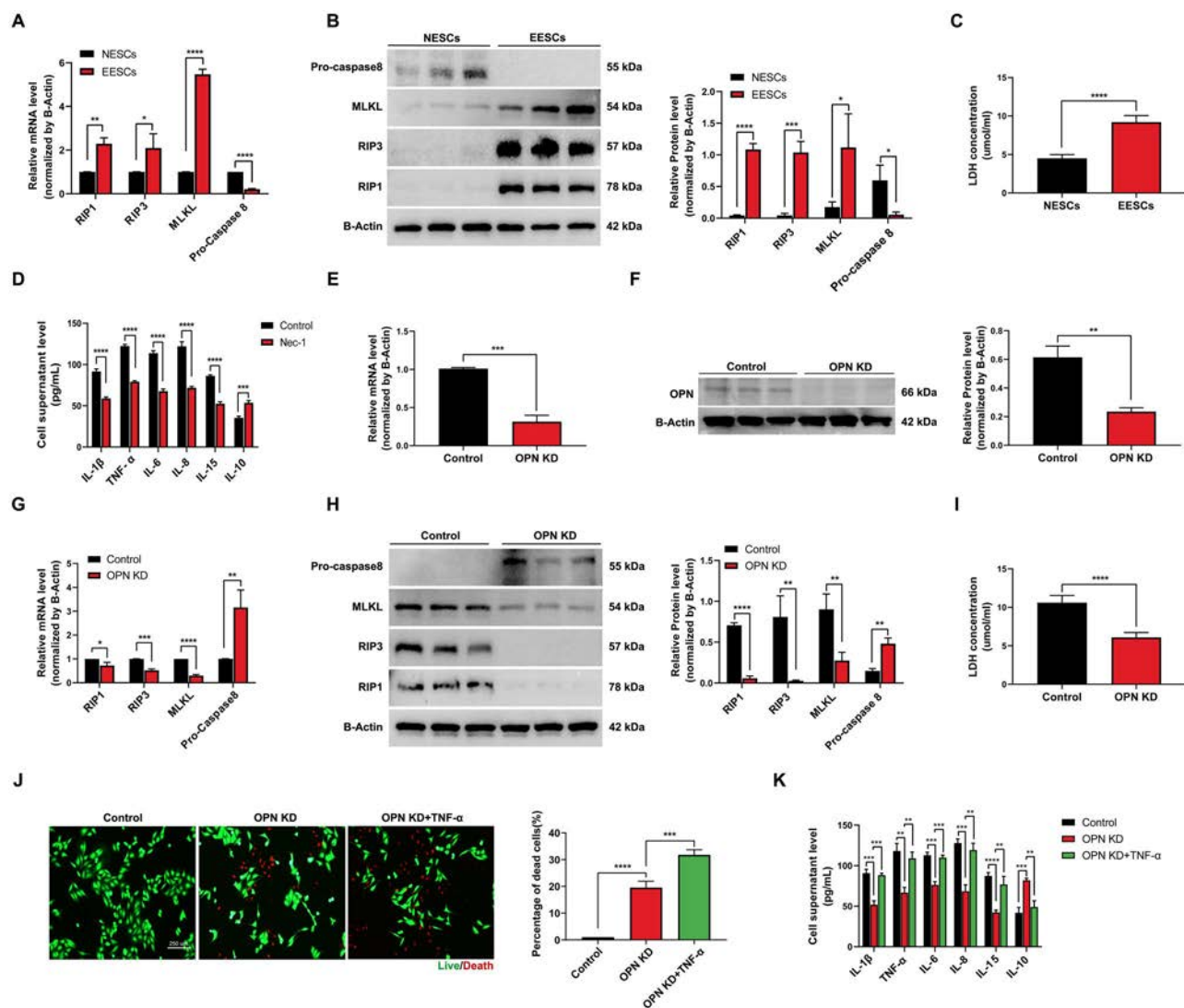


Figure 3. Increased necroptosis in endometriosis and its modulation by OPN knockdown. (A and B) qRT-PCR and Western blot (WB) assays confirmed RIP1, RIP3, MLKL, and Pro-caspase8 mRNA and protein levels in NESCs (black) and EESCs (red). (C) LDH production in NESCs (black) and EESCs (red). (D) ELISA assays measured inflammatory factor levels in EESCs treated with or without (black) Nec-1 (30 μM, 24 h). (E and F) qRT-PCR and WB assays demonstrated efficient OPN knockdown in EESCs (black, control; red, EESCs treated with OPN knockdown). (G and H) qRT-PCR and WB assays showing RIP1, RIP3, MLKL, and Pro-caspase8 mRNA and protein levels in EESCs transfected with control (black) or OPN knockdown (red). (I) LDH production in EESCs treated with control (black) or OPN knockdown (red). (J) Live and dead cells were stained with a Calcein-AM/PI double staining assay (green, live cell; red, dead cell; original magnification 200×), and cell viability was expressed as the percentage of surviving cells (black, control; red, EESCs treated with OPN knockdown; green, EESCs treated with OPN knockdown plus TNF-α). (K) ELISA assays measured inflammatory factor levels in EESCs treated with (red) or without (black) OPN knockdown and OPN knockdown plus TNF-α (green). All data are shown as the mean ± SD of three independent experiments (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). EESCs: endometriosis ectopic stromal cells; LDH: lactate dehydrogenase; MLKL: mixed lineage kinase domain-like protein; Nec-1: necroptosis inhibitor; NESCs: normal endometrial stromal cells; OPN: oseopontin; qRT-PCR: quantitative real-time PCR; RIP1: receptor-interacting protein kinase 1; RIP3: receptor-interacting protein kinase 3; TNF-α: tumor necrosis factor-α.

cell death (Fig. 3J). Finally, ELISA assays revealed that TNF-α, an activator of necroptosis, stimulated the secretion of pro-inflammatory factors and inhibited the secretion of anti-inflammatory factors, which were both counteracted by OPN knockdown (Fig. 3K). Collectively, these findings indicated that necroptosis promotes the release of inflammatory factors from EESCs, and OPN knockdown inhibits inflammatory factor release by suppressing necroptosis.

OPN knockdown suppresses secretion of inflammatory factors by inhibiting mitochondrial stress and ROS release

Mitochondria play a crucial role in the activation of inflammation (Zhou et al., 2020). In our previous study, impaired

mitochondrial respiration and increased mitochondrial stress have been carefully explored in endometriosis. To further investigate the mechanism of mitochondrial stress, we measured the $\Delta\Psi_m$ using JC-1 fluorescence and found a significant reduction in EESCs (Fig. 4A). Additionally, mitochondrial stress often increases ROS (Zhao et al., 2021). Thus, we assessed the levels of mROS and cytosolic ROS (cROS) using fluorescent dyes Mito-SOX and DCFH-DA, respectively. mROS and cROS levels were elevated in EESCs, which were consistent with mitochondrial stress (Fig. 4B). Despite the necessity of ROS for optimal activation of the inflammatory response in endometriosis (Donnez et al., 2016), it remains unclear whether and how OPN affects ROS release. To this end, Seahorse metabolic analysis showed that OCR, ECAR, basal, and compensatory glycolytic rates were

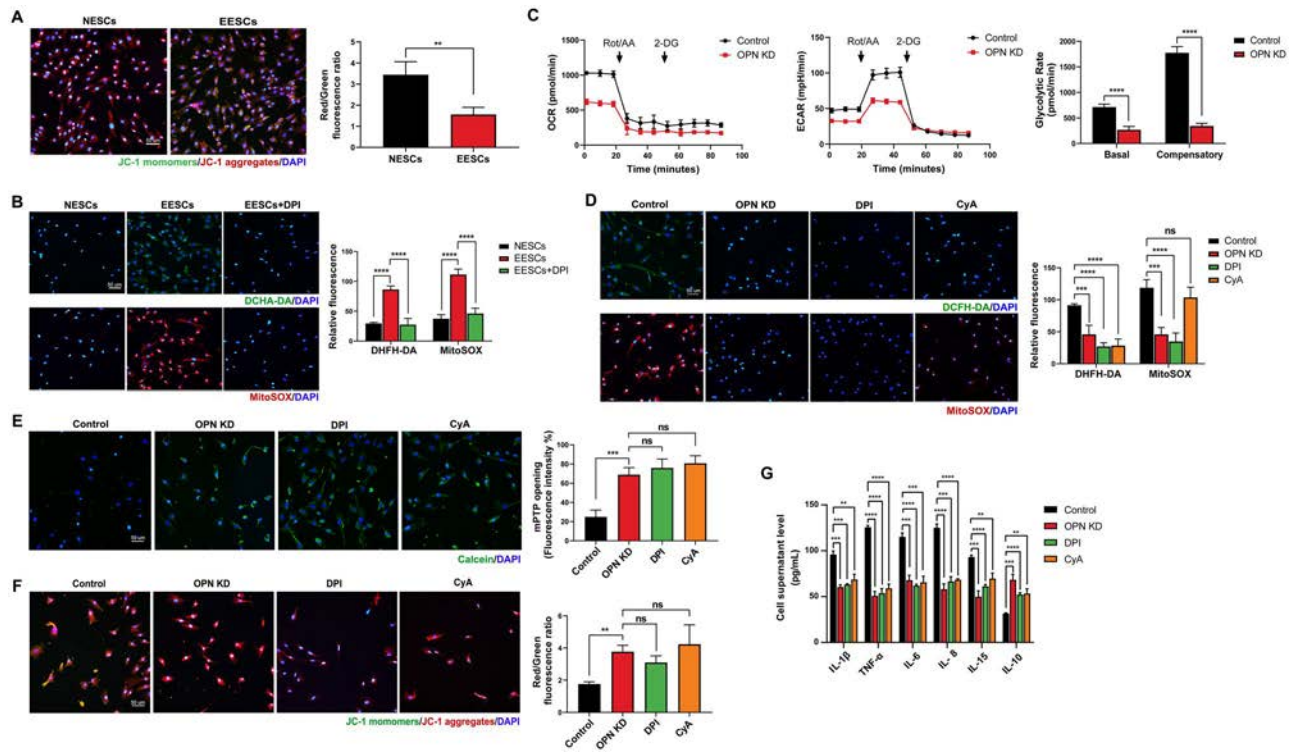


Figure 4. OPN knockdown suppresses inflammatory factors secretion by inhibiting mitochondrial stress and ROS release. (A) Assessment of $\Delta\Psi_m$ using JC-1 staining in NESCs and EESCs (green, JC-1 monomers; red, JC-1 aggregates; original magnification 400 $\times$), and fluorescence intensity was quantified (black, NESCs; red, EESCs). (B) MitoSOX and DCFH-DA staining images in NESCs and EESCs with or without DPI (1.5 μ M, 24 h) (green, DCFH-DA; red, MitoSOX; blue, DAPI; original magnification 400 $\times$), and fluorescence intensity was quantified (black, NESCs; red, EESCs; green, EESCs treated with DPI). (C) Analysis of ECAR, OCR, and glycolytic rate in EESCs treated with or without OPN knockdown using a Seahorse XF24 analyzer (black, control; red, EESCs treated with OPN knockdown). (D) MitoSOX and DCFH-DA staining images in EESCs treated with OPN knockdown, DPI, or CyA (7 nM, 24 h) and fluorescence intensity was quantified (black, control; red, EESCs treated with OPN knockdown, green, EESCs treated with DPI; orange, EESCs treated with CyA; original magnification 400 $\times$). (E) Measurement of mPTP opening in EESCs treated with OPN knockdown, DPI, or CyA using Calcein loading/CoCl₂ quenching (green, Calcein; original magnification 400 $\times$) and fluorescence intensity was quantified. (F) Assessment of $\Delta\Psi_m$ using JC-1 staining in EESCs treated with OPN knockdown (red), DPI (green), or CyA (orange) (original magnification 400 $\times$). (G) ELISA assays measured inflammatory factor levels in EESCs treated with or without OPN knockdown (red), DPI (green), and CyA (orange). All data are shown as the mean $\pm$ SD of three independent experiments (NS is nonsignificant, ** P < 0.01, *** P < 0.001, **** P < 0.0001). CoCl₂: cobalt chloride; CyA: cyclosporine A; DPI: ROS inhibitor; ECAR: extracellular acidification rate; EESCs: endometriosis ectopic stromal cells; mPTP: mitochondrial permeability transition pore; NESCs: normal endometrial stromal cells; OCR: oxygen consumption rate; OPN: osteopontin; ROS: reactive oxygen species; $\Delta\Psi_m$: mitochondrial membrane potential.

inhibited in OPN knockdown cells, respectively, indicating a shift from mitochondrial oxidative phosphorylation to glycolysis (Fig. 4C). Moreover, OPN knockdown reduced both mROS and cROS levels, similar to the effects of a ROS inhibitor, DPI. It is well-known that mROS can be released into the cytosol through the mPTP, which is considered a major source of cROS (Kwong and Molkenin, 2015; Dey et al., 2018; Zhao et al., 2020). Consistently, the administration of mPTP inhibitors (CyA), significantly decreased the cROS levels in EESCs (Fig. 4D). Finally, Calcein release and JC-1 fluorescence assays revealed that OPN knockdown inhibited mPTP opening, as indicated by the increase in Calcein fluorescence (Fig. 4E) and increased $\Delta\Psi_m$ (Fig. 4F). ELISA assays showed that OPN knockdown reduced inflammation levels similarly to DPI and CyA (Fig. 4G). Together, these findings suggested that OPN knockdown suppresses inflammatory factor secretion by inhibiting mitochondrial stress and ROS release.

OPN knockdown inhibits release of ROS and inflammatory factors by promoting RhoA

As RhoA signaling counteracts oxidative stress by inhibiting the mitochondrial death pathway (Ren et al., 2022). We

conducted qRT-PCR and WB assays and showed that OPN knockdown increased RhoA expression at both the mRNA and protein levels (Fig. 5A and B). To further explore the role of RhoA in OPN-mediated ROS release, we used a lentivirus to knock down the expression of RhoA. qRT-PCR and WB assays showed that RhoA was successfully knocked down (Fig. 5C and D). Knockdown of RhoA reversed the decrease in mROS and cROS levels caused by OPN knockdown (Fig. 5E), indicating that RhoA is involved in the regulation of ROS release by OPN. Subsequently, Calcein release and JC-1 fluorescence assays showed that RhoA knockdown promoted mPTP opening and $\Delta\Psi_m$ suppression (Fig. 5F and G). Finally, ELISA assays showed that RhoA knockdown increased and inhibited the release of pro-inflammatory and anti-inflammatory factors, respectively (Fig. 5H), indicating RhoA knockdown reversed the effect of OPN knockdown. Potential interactions between OPN and RhoA predicted by the STRING website (Fig. 5I) and Co-IP assays confirmed the binding of OPN to RhoA in EESCs (Fig. 5J). Collectively, these results showed that RhoA is involved in OPN-regulated mitochondrial stress, and OPN knockdown inhibits ROS and inflammatory factors released by promoting RhoA.

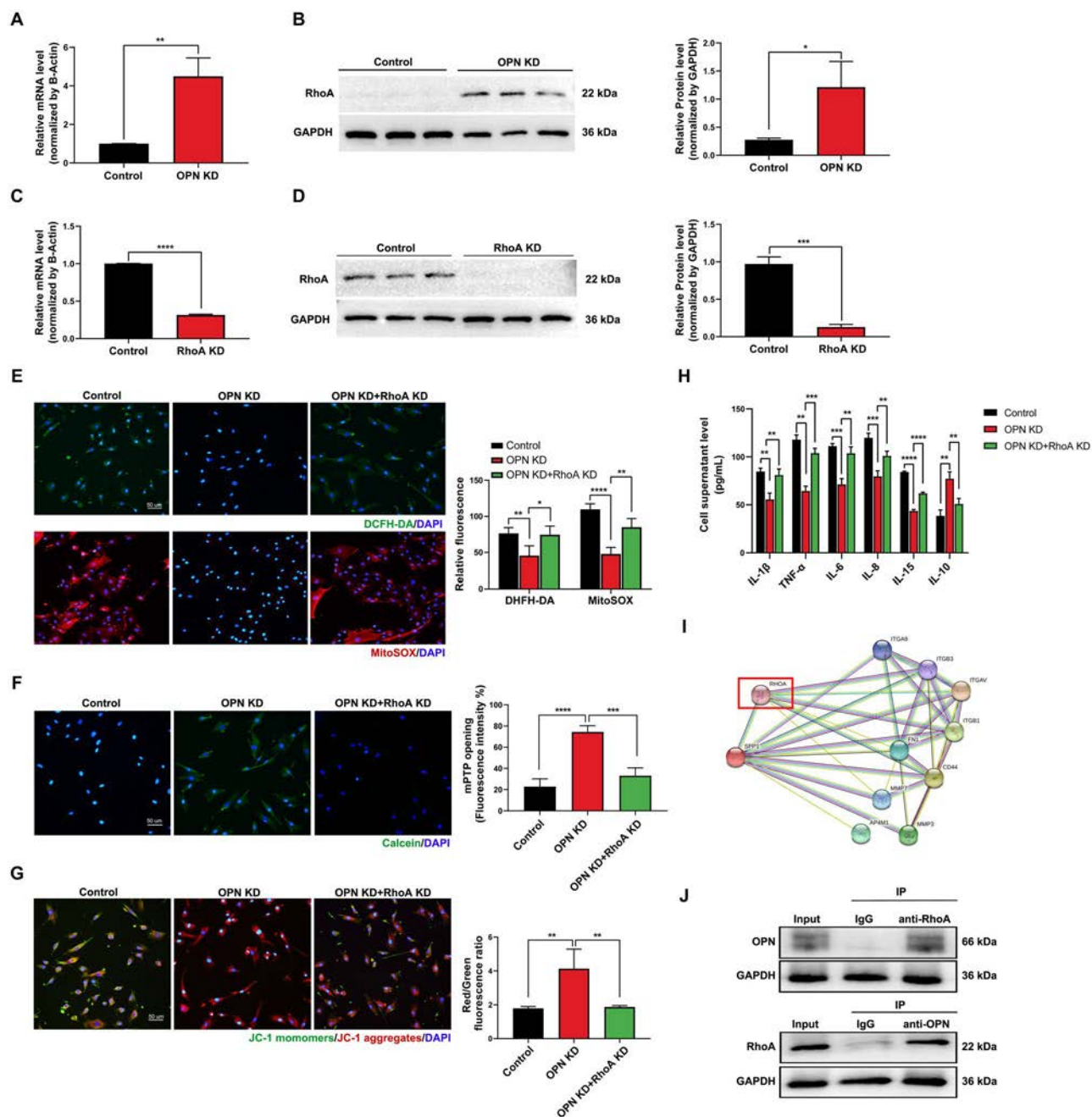


Figure 5. OPN knockdown inhibits ROS and release of inflammatory factors by promoting RhoA. (A and B) qRT-PCR and WB assays confirmed RhoA mRNA and protein levels in EESCs transfected with OPN knockdown (red). (C and D) qRT-PCR and WB assays demonstrated the knockdown efficiency of RhoA in OPN knockdown treatment-EESCs (red). (E) MitoSOX and DCFH-DA staining images in EESCs treated with OPN knockdown or RhoA knockdown (green, DCHA-DA; red, MitoSOX; blue, DAPI; original magnification 400×), and fluorescence intensity was quantified (black, control; red, EESCs treated with OPN knockdown; green, EESCs treated with OPN knockdown and RhoA knockdown). (F) mPTP opening in EESCs treated with OPN knockdown or RhoA knockdown measured by Calcein loading/CoCl₂ quenching (green, Calcein; original magnification 400×), and fluorescence intensity was quantified. (G) $\Delta\Psi_m$ assessment using JC-1 staining in EESCs treated with OPN knockdown or RhoA knockdown (green, JC-1 momomers; red, JC-1 aggregates; original magnification 400×) and fluorescence intensity was quantified. (H) ELISA assays measured inflammatory factor levels in EESCs treated with (red) or without OPN knockdown (black) and with both OPN and RhoA knockdown (green). (I) Prediction of endogenous interaction between OPN and other proteins using the STRING website (<https://string-db.org>). (J) Co-IP assays confirmed the binding relationship between OPN and RhoA. All data are shown as the mean $\pm$ SD of three independent experiments (* P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001). Co-IP: co-immunoprecipitation; EESCs: endometriosis ectopic stromal cells; mPTP: mitochondrial permeability transition pore; OPN knockdown: OPN knockdown; OPN: oseopontin; qRT-PCR: quantitative real-time PCR; ROS: reactive oxygen species; WB: Western blot; $\Delta\Psi_m$: mitochondrial membrane potential.

ROS promoting the necroptosis

To investigate the relationship between ROS and necroptosis, qRT-PCR and WB assays showed that DPI, as a ROS inhibitor, suppressed necroptosis (Fig. 6A and B). OPN knockdown inhibited the expression of necroptosis marker proteins, while RhoA knockdown

reversed the outcomes. Consistently, DPI reversed the trends promoted by RhoA knockdown, similar to OPN knockdown (Fig. 6C). LDH assays also demonstrated the effect of OPN, RhoA knockdown, and DPI on necroptosis (Fig. 6D), suggesting that OPN promoted necroptosis via the RhoA/ROS pathway. Membrane MLKL

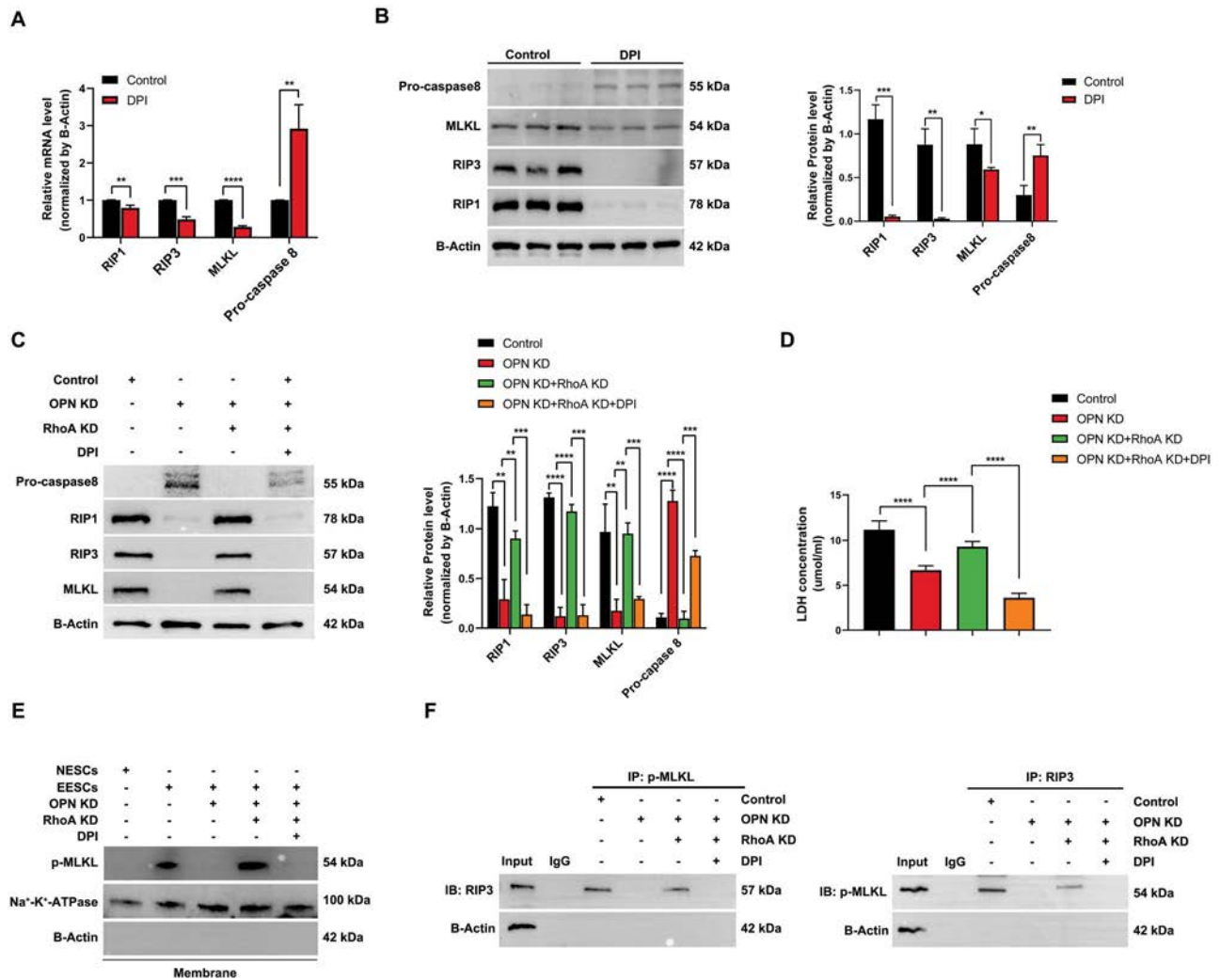


Figure 6. ROS promote necroptosis. (A and B) qRT-PCR and WB assays of RIP1, RIP3, MLKL, and Pro-caspase8 mRNA and protein levels in control EESCs (black) and EESCs treated with DPI (red). (C) WB assays of RIP1, RIP3, MLKL, and Pro-caspase8 protein levels in EESCs treated with OPN knockdown (red), combined OPN and RhoA knockdown (green), or combined OPN and RhoA knockdown with DPI (orange). (D) LDH production in EESCs treated with OPN knockdown (red), combined OPN and RhoA knockdown (green), or OPN and RhoA knockdown with DPI (orange). (E) WB assays showed p-MLKL protein levels in cell membranes of NESCs and EESCs transfected with OPN knockdown, RhoA knockdown, or DPI. (F) Co-IP experiments confirmed the binding relationship between RIP3 and p-MLKL. All data are shown as the mean $\pm$ SD of three independent experiments (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). Co-IP: co-immunoprecipitation; DPI: ROS inhibitor; EESCs: endometriosis ectopic stromal cells; LDH: lactate dehydrogenase; MLKL: mixed lineage kinase domain-like protein; NESCs: normal endometrial stromal cells; OPN: osteopontin; p-MLKL: phosphorylated mixed lineage kinase domain-like protein; qRT-PCR: quantitative real-time PCR; RIP1: receptor-interacting protein kinase 1; RIP3: receptor-interacting protein kinase 3; ROS: reactive oxygen species; WB: Western blot.

phosphorylation (p-MLKL) is required for necroptosis (Wen et al., 2023). To quantify p-MLKL levels, WB assays were performed using isolated cell membranes. The results demonstrated significant suppression of OPN knockdown upon p-MLKL, whereas RhoA knockdown reversed the effect. Remarkably, DPI counteracted the effect of RhoA knockdown equally to OPN knockdown (Fig. 6E). Binding of RIP3 to p-MLKL is a key step in necroptosis (Liu et al., 2021). Co-IP assays showed that OPN knockdown inhibited the binding of RIP3 and p-MLKL, while RhoA knockdown reversed the effect of OPN knockdown (Fig. 6F). These findings indicated that ROS promotes necroptosis, and OPN knockdown inhibits necroptosis via the RhoA/ROS pathway.

OPN regulating endometriosis progression, necroptosis, ROS, and inflammatory factors release in vitro and in vivo

To investigate whether OPN regulates the proliferation and metastasis of EESCs through the RhoA/ROS axis in vitro, we used

CCK8 and Transwell assays. These showed that OPN knockdown inhibited EESCs proliferation and metastasis via the RhoA/ROS axis (Fig. 7A and B). To further investigate the therapeutic potential of OPN in vivo, we intravenously administered AAV9 vectors containing OPN knockdown sequences or empty vectors to mice with established endometriosis models. The administration of AAV9 vectors did not result in any obvious toxicity. After 14 days, OPN knockdown suppressed the volume and weight of endometriosis lesions in vivo (Fig. 7C–E). IHC assays revealed that OPN knockdown decreased the protein levels of OPN, RIP1, RIP3, and MLKL, while increasing RhoA and Pro-caspase8 levels, indicating that OPN knockdown attenuated necroptosis in vivo (Fig. 7F). qRT-PCR and WB assays also confirmed the effect of OPN knockdown on RIP1, RIP3, MLKL, RhoA, and Pro-caspase8 (Fig. 7G and H). ELISA assays showed that OPN knockdown inhibited ROS release and affected the secretion of inflammatory factors (Fig. 7I and J). Collectively, these data suggested that OPN knockdown inhibits

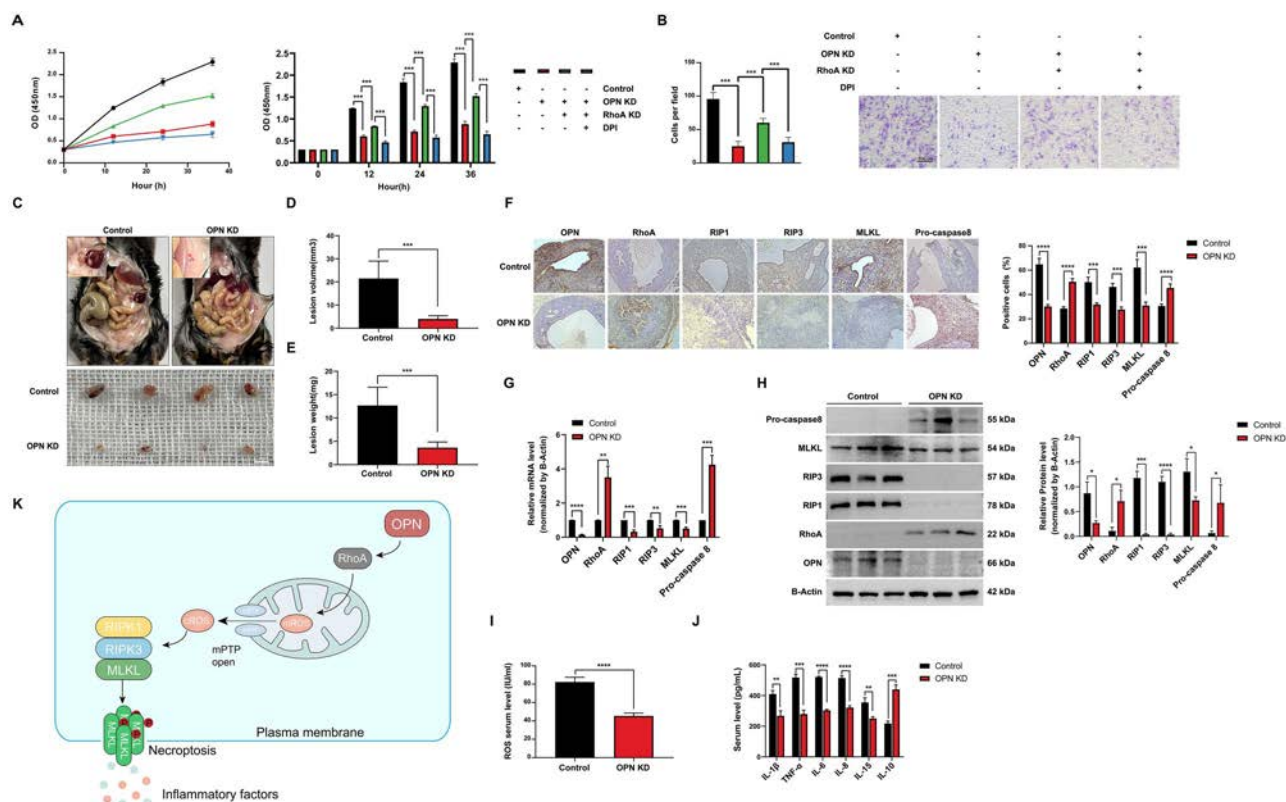


Figure 7. OPN regulating endometriosis progression, necroptosis, ROS, and inflammatory factors release in vitro and in vivo. (A) CCK8 assays measured the proliferation of EESCs treated with OPN knockdown (red), combined OPN and RhoA knockdown (green), or combined OPN and RhoA knockdown with DPI (blue). (B) Transwell assays evaluated the migratory capability of EESCs treated with OPN knockdown (red), combined OPN and RhoA knockdown (green), or combined OPN and RhoA knockdown with DPI (blue) (original magnification 200×). (C) Development of ectopic lesions (white arrows) on Day 14 in mice injected with or without AAV9 (OPN knockdown). (D and E) Measurement of ectopic lesion volume and weight in control (black) (n = 4) and OPN knockdown groups (red) (n = 4). (F) IHC assays showed the expression of OPN, RhoA, RIP1, RIP3, MLKL, and Pro-caspase8 in control (black) and OPN knockdown lesions (red) (original magnification 200×). (G and H) qRT-PCR and WB assays confirmed OPN, RhoA, RIP1, RIP3, MLKL, and Pro-caspase8 levels in control (black) and OPN knockdown groups (red). (I) Measurement of ROS levels in the blood plasma of control (black) and OPN knockdown groups (red). (J) ELISA assays measured the levels of inflammatory factors in the blood plasma of control (black) and OPN knockdown groups (red). (K) Cartoon illustration of the role of OPN in necroptosis and inflammatory factor release through the RhoA/ROS pathway. All data are shown as the mean ± SD of three independent experiments (**P < 0.01, ***P < 0.001, ****P < 0.0001). CCK8: cell counting kit-8; DPI: ROS inhibitor; EESCs: endometriosis ectopic stromal cells; IHC: immunohistochemical staining; MLKL: mixed lineage kinase domain-like protein; OPN: oseopontin; qRT-PCR: quantitative real-time PCR; RIP1: receptor-interacting protein kinase 1; RIP3: receptor-interacting protein kinase 3; ROS: reactive oxygen species; WB: Western blotting.

endometriosis progression, necroptosis, ROS, and inflammatory factors release in vivo.

OPN and inflammatory factors are significantly upregulated in E-MESCs

To investigate the utility of OPN as a potential molecular marker for endometriosis and overcome the limitation of peripheral circulating blood not accurately reflecting the inflammatory status of endometriosis, we adopted a non-invasive approach to collect menstrual blood samples. By utilizing samples from this specific source, we aimed to enhance the diagnostic potential for endometriosis. MESCs were extracted from menstrual blood and the expression level of OPN was detected. RNA-seq assays revealed the presence of OPN among the differentially expressed genes in the top 100. Moreover, the expression level of OPN was notably higher in E-MESCs compared to N-MESCs (Fig. 8A and B). Significantly higher expression of OPN in E-MESCs was also observed by qRT-PCR and WB assays (Fig. 8C and D). ELISA assays showed significant changes in the expression of inflammatory factors, with higher levels observed in E-MESCs compared to N-MESCs (Fig. 8E). Together, these findings suggested that OPN and

inflammatory factors are significantly upregulated in E-MESCs, highlighting their potential as molecular markers for endometriosis and the possibility of non-invasive diagnosis through menstrual blood testing.

Discussion

Endometriosis is well-known as a common benign gynecologic inflammation, and understanding the mechanisms of related inflammation is crucial for its treatment and management. In this study, we elucidated the intricate communication pathways involving OPN, mitochondria, and the plasma membrane, which play pivotal roles in driving inflammation in endometriosis (Fig. 7K). Furthermore, we observed that OPN was highly expressed in MESCs, suggesting that non-invasive detection of OPN expression in the endometrium can be achieved in vitro. In conclusion, targeting OPN enables to provides a promising approach to improving the inflammatory state and enhancing the therapeutic efficacy of endometriosis. Additionally, non-invasive detection of OPN in vitro through menstrual blood further establishes OPN as a reliable molecular marker of endometriosis.

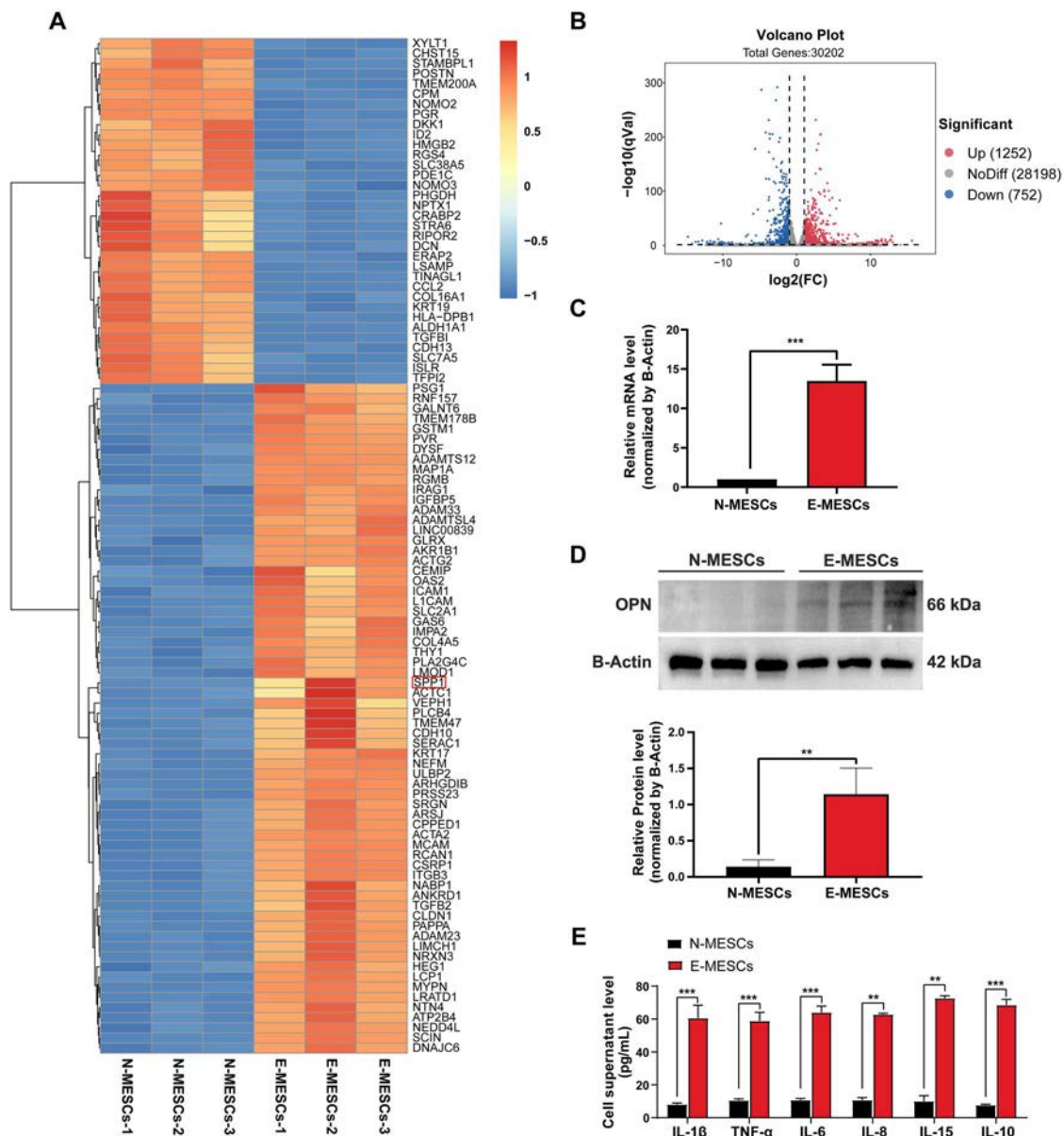


Figure 8. OPN and inflammatory factors were significantly upregulated in E-MESCs. (A) Heatmap showing the top 100 DEGs. The color coding in the legend represents the range of different values, with darker colors indicating higher values. Red, upregulated; blue, downregulated (N-MESCs vs E-MESCs, $q < 0.01$). (B) Volcano plot of DEGs. The colors in the diagram represent different types of volcanic material. Red, upregulated; gray, no different; blue, downregulated (N-MESCs vs E-MESCs, $q < 0.05$). (C and D) qRT-PCR and WB assays confirmed OPN mRNA and protein levels in N-MESCs (black) and E-MESCs (red). (E) ELISA assays measured the levels of inflammatory factors in N-MESCs (black) and E-MESCs (red). All data are shown as the mean $\pm$ SD of three independent experiments (** $P < 0.01$, *** $P < 0.001$). DEGs, differentially expressed genes; E-MESCs, endometrial cells derived from menstrual blood of endometriosis patients; N-MESCs, normal menstrual blood endometrial cells; OPN: osteopontin; qRT-PCR: quantitative real-time PCR; WB: Western blotting.

Although the critical role of necroptosis in initiating inflammation has been documented, its regulatory mechanisms remain largely unknown (Pasparakis and Vandenabeele, 2015). In this study, we observed the occurrence of necroptosis in endometriosis, and its alteration was associated with significant changes in the levels of inflammatory factors. Specifically, inhibiting necroptosis led to a decrease in the release of pro-inflammatory factors and an increase in anti-inflammatory factors. Notably, OPN knockdown inhibited necroptosis in endometriosis, thereby counteracting its effect on inflammatory factors. The reduction of necroptosis theoretically corresponds to a decrease in cell death. Interestingly, the knockdown of OPN did not improve cell death, which aligns with the therapeutic effect of OPN. We hypothesized that OPN knockdown may activate alternative

pathways of cell death, a hypothesis that warrants further investigation in future experiments.

Targeting ROS for therapeutic purposes poses a challenge due to its dual nature. Despite the various deleterious effects associated with ROS, recent studies have highlighted their role in maintaining cellular homeostasis and prolonging lifespan (Lagouge and Larsson et al., 2013; Zorov et al., 2014). Our findings suggested that OPN knockdown may offer a potential solution to this challenge. Endometriosis exhibited mitochondrial respiratory dysfunction and elevated levels of both mROS and cROS. However, OPN knockdown significantly reduced mROS and cROS levels in EESCs. Furthermore, we found that OPN fine-tunes ROS levels during excessive mPTP opening. Importantly, the opening of mPTP and ROS levels are closely linked to the release of

inflammatory factors. Inhibitors of mPTP opening and ROS production, similar to OPN knockdown, were able to reduce the release of pro-inflammatory cytokines. Our study provided conceptual evidence that OPN knockdown can mitigate mitochondrial stress, attenuate ROS release, and improve the inflammatory state in endometriosis.

RhoA, a small G protein and intracellular signal transducer, plays a crucial role in various cellular processes, including cell proliferation and survival (Zaoui et al., 2019; Fernández-Simón et al., 2022). Previous studies have demonstrated the protective role of RhoA signaling against oxidative stress-induced cardiac damage by inhibiting mitochondrial death pathways (Ren et al., 2020). However, its role in regulating mitochondrial quality control in response to mitochondrial stress remains unknown. Our findings suggest that RhoA is involved in OPN-mediated regulation of the mitochondrial stress pathway. Blocking downstream RhoA expression resulted in mPTP closure, inhibiting ROS release. Furthermore, OPN and RhoA exhibited endogenous interactions. Manipulating RhoA levels modulated the occurrence of necroptosis and the levels of inflammatory factors, emphasizing the pivotal role of RhoA in the OPN-mediated regulation of necroptosis. *In vivo* experiments confirmed the role of OPN in the treatment of endometriosis. OPN knockdown significantly inhibited lesion growth, necroptosis, ROS secretion, and the release of pro-inflammatory factors. Collectively, our results indicated that targeting OPN expression holds promise as a clinical treatment strategy for endometriosis. Finally, to overcome the limitations of diagnostic testing for endometriosis, we developed an assay route using menstrual blood. We successfully detected significantly high expression of OPN and abnormal secretion of inflammatory factors in E-MESCs.

In summary, our study unveils the critical role of OPN in regulating the RhoA/ROS signaling pathway, thereby controlling necroptosis and the release of inflammatory factors. OPN knockdown effectively impeded the development of endometriosis, and the non-invasive and specific detection of endometriosis can be achieved through menstrual blood, providing a promising treatment approach by targeting inflammation through OPN regulation.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

All data generated or analyzed during this study are included in this published article.

Acknowledgments

We thank the members of the Department of Pathology at Wenzhou Medical University for kindly providing technical support.

Authors' roles

All the authors reviewed and approved the final manuscript. P.D. supervised the project and designed the experiments for this study. H.W. conducted the experiments and data analysis. H.W. and B.W. prepared all the figures and tables and the manuscript. H.W. and M.W. organized the public data. H.W., B.W., and J.L. conducted the experimental execution. H.W., M.W., and J.L. were

responsible for the identification of patients. All authors read and approved the final manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (Grant No. 82071626), the Zhejiang Province Public Welfare Technology Application Research Project (Grant No. LGF21H040010), and the Clinical Research project of the Second Affiliated Hospital of Wenzhou Medical University (Grant No. 1010293).

Conflict of interest

The authors declare that they have no competing interests.

References

- Andrieux P, Chevillard C, Cunha-Neto E, Nunes JPS. Mitochondria as a cellular hub in infection and inflammation. *Int J Mol Sci* 2021; **22**:11338.
- Bischoff FZ, Simpson JL. Heritability and molecular genetic studies of endometriosis. *Hum Reprod Update* 2000; **6**:37–44.
- Cheng J, Duan X, Fu X, Jiang Y, Yang P, Cao C, Li Q, Zhang J, Hu X, Zhang X et al. RIP1 perturbation induces chondrocyte necroptosis and promotes osteoarthritis pathogenesis via targeting BMP7. *Front Cell Dev Biol* 2021; **9**:638382.
- de Almagro MC, Goncharov T, Izrael-Tomasevic A, Duttler S, Kist M, Varfolomeev E, Wu X, Lee WP, Murray J, Webster JD et al. Coordinated ubiquitination and phosphorylation of RIP1 regulates necroptotic cell death. *Cell Death Differ* 2017; **24**:26–37.
- Dey S, DeMazumder D, Sidor A, Foster DB, O'Rourke B. Mitochondrial ROS drive sudden cardiac death and chronic proteome remodeling in heart failure. *Circ Res* 2018; **123**:356–371.
- Dondelinger Y, Hulpiau P, Saeys Y, Bertrand MJM, Vandenabeele P. An evolutionary perspective on the necroptotic pathway. *Trends Cell Biol* 2016; **26**:721–732.
- Donnez J, Binda MM, Donnez O, Dolmans MM. Oxidative stress in the pelvic cavity and its role in the pathogenesis of endometriosis. *Fertil Steril* 2016; **106**:1011–1017.
- Fernández-Simón E, Suárez-Calvet X, Carrasco-Rozas A, Piñol-Jurado P, López-Fernández S, Pons G, Bech Serra JJ, de la Torre C, de Luna N, Gallardo E et al. RhoA/ROCK2 signalling is enhanced by PDGF-AA in fibro-adipogenic progenitor cells: implications for Duchenne muscular dystrophy. *J Cachexia Sarcopenia Muscle* 2022; **13**:1373–1384.
- Gargett CE, Schwab KE, Deane JA. Endometrial stem/progenitor cells: the first 10 years. *Hum Reprod Update* 2016; **22**:137–163.
- Huang HS, Chen J, Chen CF, Ma MC. Vitamin E attenuates crystal formation in rat kidneys: roles of renal tubular cell death and crystallization inhibitors. *Kidney Int* 2006; **70**:699–710.
- Khan KN, Masuzaki H, Fujishita A, Kitajima M, Kohno T, Sekine I, Matsuyama T, Ishimaru T. Regulation of hepatocyte growth factor by basal and stimulated macrophages in women with endometriosis. *Hum Reprod* 2005; **20**:49–60.
- Kwong JQ, Molkentin JD. Physiological and pathological roles of the mitochondrial permeability transition pore in the heart. *Cell Metab* 2015; **21**:206–214.
- Lagouge M, Larsson NG. The role of mitochondrial DNA mutations and free radicals in disease and ageing. *J Intern Med* 2013; **273**:529–543.
- Liu C, Chen Y, Cui W, Cao Y, Zhao L, Wang H, Liu X, Fan S, Huang K, Tong A et al. Inhibition of neuronal necroptosis mediated by RIP1/

- RIP3/MLKL provides neuroprotective effects on kaolin-induced hydrocephalus in mice. *Cell Prolif* 2021;**54**:e13108.
- Lok ZSY, Lyle AN. Osteopontin in vascular disease. *Arterioscler Thromb Vasc Biol* 2019;**39**:613–622.
- Mizumura K, Cloonan SM, Nakahira K, Bhashyam AR, Cervo M, Kitada T, Glass K, Owen CA, Mahmood A, Washko GR et al. Mitophagy-dependent necroptosis contributes to the pathogenesis of COPD. *J Clin Invest* 2014;**124**:3987–4003.
- Newton K, Manning G. Necroptosis and Inflammation. *Annu Rev Biochem* 2016;**85**:743–763.
- Oikonomou N, Schuijs MJ, Chatzigiagkos A, Androulidaki A, Aidinis V, Hammad H, Lambrecht BN, Pasparakis M. Airway epithelial cell necroptosis contributes to asthma exacerbation in a mouse model of house dust mite-induced allergic inflammation. *Mucosal Immunol* 2021;**14**:1160–1171.
- Pasparakis M, Vandenabeele P. Necroptosis and its role in inflammation. *Nature* 2015;**517**:311–320.
- Picard M, Shirihaï OS. Mitochondrial signal transduction. *Cell Metab* 2022;**34**:1620–1653.
- Qin X, Yan M, Wang X, Xu Q, Wang X, Zhu X, Shi J, Li Z, Zhang J, Chen W. Cancer-associated fibroblast-derived IL-6 promotes head and neck cancer progression via the osteopontin-NF-kappa B signaling pathway. *Theranostics* 2018;**8**:921–940.
- Ren R, Fang Y, Sherchan P, Lu Q, Lenahan C, Zhang JH, Zhang J, Tang J. Kynurenine/aryl hydrocarbon receptor modulates mitochondria-mediated oxidative stress and neuronal apoptosis in experimental intracerebral hemorrhage. *Antioxid Redox Signal* 2022;**37**:1111–1129.
- Ren YS, Li HH, Yao JC, Tan YJ, Pan LH, Peng T, Zhao LL, Zhang GM, Yue J, Hu XM et al. Application quantitative proteomics approach to identify differentially expressed proteins associated with cardiac protection mediated by cycloastragenol in acute myocardial infarction rats. *J Proteomics* 2020;**222**:103691.
- Shirakawa K, Endo J, Kataoka M, Katsumata Y, Yoshida N, Yamamoto T, Isobe S, Moriyama H, Goto S, Kitakata H et al. IL (interleukin)-10-STAT3-galectin-3 axis is essential for osteopontin-producing reparative macrophage polarization after myocardial infarction. *Circulation* 2018;**138**:2021–2035.
- Song Z, Chen W, Athavale D, Ge X, Desert R, Das S, Han H, Nieto N. Osteopontin takes center stage in chronic liver disease. *Hepatology* 2021;**73**:1594–1608.
- Wang H, Chen C, Wang D, Zhu Y, Chen P. Correlation of clinicopathological and prognostic characteristics between endometriosis-associated and primary ovarian cancer. *BMC Cancer* 2023;**23**:1210.
- Wang H, Liang Z, Gou Y, Li Z, Cao Y, Jiao N, Tan J, Yu Y, Zhang Z. FTO-dependent N(6)-methyladenosine regulates the progression of endometriosis via the ATG5/PKM2 axis. *Cell Signal* 2022;**98**:110406.
- Wang XQ, Zhou WJ, Luo XZ, Tao Y, Li DJ. Synergistic effect of regulatory T cells and proinflammatory cytokines in angiogenesis in the endometriotic milieu. *Hum Reprod* 2017;**32**:1304–1317.
- Weindel CG, Martinez EL, Zhao X, Mabry CJ, Bell SL, Vail KJ, Coleman AK, VanPortfliet JJ, Zhao B, Wagner AR et al. Mitochondrial ROS promotes susceptibility to infection via gasdermin D-mediated necroptosis. *Cell* 2022;**185**:3214–3231.e23.
- Wen Q, Zhang J, Zhang Z, Chen L, Liu H, Han Z, Chen Y, Wang K, Liu J, Sai N et al. Cisatracurium besylate rescues *Mycobacterium tuberculosis*-infected macrophages from necroptosis and enhances the bactericidal effect of isoniazid. *Int Immunopharmacol* 2023;**120**:110291.
- Wen Y, Jeong S, Xia Q, Kong X. Role of osteopontin in liver diseases. *Int J Biol Sci* 2016;**12**:1121–1128.
- Zaoui K, Boudhraa Z, Khalifé P, Carmona E, Provencher D, Mes-Masson AM. Ran promotes membrane targeting and stabilization of RhoA to orchestrate ovarian cancer cell invasion. *Nat Commun* 2019;**10**:2666.
- Zhang T, Yin C, Boyd DF, Quarato G, Ingram JP, Shubina M, Ragan KB, Ishizuka T, Crawford JC, Tummers B et al. Influenza virus Z-RNAs induce ZBP1-mediated necroptosis. *Cell* 2020;**180**:1115–1129.e13.
- Zhang T, Zhou J, Man GCW, Leung KT, Liang B, Xiao B, Ma X, Huang S, Huang H, Hegde VL et al. MDSCs drive the process of endometriosis by enhancing angiogenesis and are a new potential therapeutic target. *Eur J Immunol* 2018;**48**:1059–1073.
- Zhao M, Wang Y, Li L, Liu S, Wang C, Yuan Y, Yang G, Chen Y, Cheng J, Lu Y et al. Mitochondrial ROS promote mitochondrial dysfunction and inflammation in ischemic acute kidney injury by disrupting TFAM-mediated mtDNA maintenance. *Theranostics* 2021;**11**:1845–1863.
- Zhao Q, Liu J, Deng H, Ma R, Liao JY, Liang H, Hu J, Li J, Guo Z, Cai J et al. Targeting mitochondria-located circRNA SCAR alleviates NASH via reducing mROS output. *Cell* 2020;**183**:76–93.e22.
- Zhou B, Wang DD, Qiu Y, Airhart S, Liu Y, Stempien-Otero A, O'Brien KD, Tian R. Boosting NAD level suppresses inflammatory activation of PBMCs in heart failure. *J Clin Invest* 2020;**130**:6054–6063.
- Zhu Y, Yang J, Xu D, Gao XM, Zhang Z, Hsu JL, Li CW, Lim SO, Sheng YY, Zhang Y et al. Disruption of tumour-associated macrophage trafficking by the osteopontin-induced colony-stimulating factor-1 signalling sensitises hepatocellular carcinoma to anti-PD-L1 blockade. *Gut* 2019;**68**:1653–1666.
- Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med* 2020;**382**:1244–1256.
- Zorov DB, Juhaszova M, Sollott SJ. Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release. *Physiol Rev* 2014;**94**:909–950.

Reproductive endocrinology

Adolescents diagnosed with polycystic ovary syndrome under the Rotterdam criteria but not meeting the diagnosis under the updated guideline

Jun Ju Kim^{1,2}, Kyu Ri Hwang ^{2,3,*}, Dayong Lee ^{2,3}, Sunmie Kim^{1,2}, and Young Min Choi^{4,5}

¹Department of Obstetrics and Gynecology, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, South Korea

²Department of Obstetrics and Gynecology, Seoul National University College of Medicine, Seoul, South Korea

³Department of Obstetrics and Gynecology, SMG-SNU Boramae Medical Center, Seoul, South Korea

⁴Department of Obstetrics and Gynecology, Grace Hospital, Goyang, South Korea

⁵The Institute of Reproductive Medicine and Population, Medical Research Center, Seoul National University College of Medicine, Seoul, South Korea

*Correspondence address. Department of Obstetrics and Gynecology, SMG-SNU Boramae Medical Center, 20, Boramaero 5 Gil, Dongjak-gu, Seoul, South Korea.
Tel: +82-2-870-2344; E-mail: orangemd@snu.ac.kr  <https://orcid.org/0000-0001-6845-1260>

ABSTRACT

STUDY QUESTION: What are the characteristics of adolescents diagnosed with polycystic ovary syndrome (PCOS) based on the 2003 Rotterdam criteria, but who do not meet the diagnosis according to the international evidence-based guideline?

SUMMARY ANSWER: Adolescents who had features of PCOS but did not meet the evidence-based guideline adolescent criteria exhibited unfavorable metabolic profiles compared to controls and shared considerable metabolic and hormonal features with adolescents who did meet the adolescent criteria.

WHAT IS KNOWN ALREADY: The international evidence-based PCOS guideline recommended that ultrasound should not be used for the diagnosis of PCOS in girls with a gynecological age of <8 years. Thus far, few studies have evaluated the clinical characteristics of the girls diagnosed with PCOS based on the Rotterdam criteria but who do not meet the diagnosis according to the updated guideline.

STUDY DESIGN, SIZE, DURATION: This is a retrospective study, and subjects attended for care from 2004 to 2022.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Adolescent girls with PCOS diagnosed according to the 2003 Rotterdam criteria and healthy controls. All participants were between 2 and 8 years since menarche.

MAIN RESULTS AND THE ROLE OF CHANCE: Of the 315 girls diagnosed with PCOS according to the Rotterdam criteria, those with irregular menstruation (IM)/hyperandrogenism (HA)/polycystic ovary (PCO), IM/HA, HA/PCO, and IM/PCO phenotypes accounted for 206 (65.4%), 30 (9.5%), 12 (3.8%), and 67 (21.3%) participants, respectively. According to the evidence-based guideline, 79 girls (25.1%) with the HA/PCO or IM/PCO phenotypes were not diagnosed with PCOS, and aligned to the international guideline; they were designated as the ‘at-risk’ group. As expected, the girls meeting the evidence-based guideline adolescent criteria showed the worst metabolic profiles (degree of generalized or central obesity, frequency of insulin resistance, prediabetes or diabetes, and metabolic syndrome) and higher hirsutism scores than the at-risk group or controls. Approximately 90% of the at-risk group were not overweight or obese, which was similar to the controls. However, they showed worse metabolic profiles, with higher blood pressure, triglyceride, and insulin resistance parameters than controls; furthermore, these profiles were similar to those of the girls meeting the adolescent criteria. The at-risk group showed similarly elevated serum LH levels and LH/FSH ratio with the girls meeting adolescent criteria.

LIMITATIONS, REASONS FOR CAUTION: We could not evaluate hormonal or ultrasound parameters in controls.

WIDER IMPLICATIONS OF THE FINDINGS: Compared to the conventional Rotterdam criteria, the recent international evidence-based guideline—avoiding ultrasound in PCOS diagnosis in adolescents—still gives the opportunity to identify young girls at risk, aligned to the findings in this study. A practical approach to this adolescent population would involve establishing IM or HA (with ultrasound not indicated) and designating ‘at-risk’ PCOS status with regular check-ups for newly developed or worsening PCOS-related symptoms or metabolic abnormalities, with subsequent reassessment including ultrasound or anti-Müllerian hormone, once 8 years post-menarche.

STUDY FUNDING/COMPETING INTEREST(S): No funding was received in support of this study. The authors have no conflicts of interest to disclose.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: adolescent / anti-Müllerian hormone / diagnosis / hyperandrogenism / menstrual cycle irregularities / polycystic ovary syndrome

Received: August 11, 2023. Revised: January 26, 2024. Editorial decision: February 15, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

Introduction

Polycystic ovary syndrome (PCOS) is common in women of reproductive age, and the 2003 Rotterdam consensus has been widely used for its diagnosis ([The Rotterdam ESHRE/ASRM-sponsored PCOS consensus workshop group, 2004](#)) ([Table 1](#)). Women with PCOS usually develop the associated symptoms during adolescence. However, the diagnosis of PCOS in adolescents is challenging due to the overlap between physiological pubertal development and the typical manifestations of PCOS. Unfortunately, there is no specific diagnostic guideline for adolescents in the Rotterdam criteria.

Since the Rotterdam consensus, many researchers have tried to define adolescent-specific diagnostic recommendations, and most have reported that polycystic ovary (PCO) is not necessary for the diagnosis of PCOS in adolescents ([Fauser et al., 2012](#); [Ibáñez et al., 2017](#); [Teede et al., 2018, 2023](#)). In 2018, an international evidence-based PCOS guideline development group published recommendations, now updated in 2023, spanning all topics for the assessment and management of PCOS ([Teede et al., 2018, 2023](#)). The guideline was a collaboration across 39 professional and consumer organizations from six continents, and the 2023 version provides 77 evidence-based and 54 consensus recommendations, with 123 practice points. Each individual diagnostic component of the Rotterdam criteria was upgraded on the best available evidence ([Table 1](#)). In adolescents, the evidence-based guideline recommended that both hyperandrogenism (HA) and ovulatory dysfunction are required and ultrasound is not recommended for the diagnosis of PCOS in girls with a gynecological age of <8 years (defined as <8 years since menarche). Thus, the irregular menstruation (IM)/PCO or HA/PCO phenotype would not be diagnosed as PCOS if a girl had not yet past 8 years since menarche.

Thus far, few studies have evaluated the clinical characteristics of the girls diagnosed with PCOS based on the Rotterdam criteria but who do not meet the diagnosis according to the recent evidence-based guideline. The present study therefore aimed to investigate the impact of the diagnostic criteria change among adolescent girls with PCOS and to compare the clinical characteristics of the girls not meeting the PCOS diagnosis any longer with the girls meeting the adolescent criteria or healthy controls.

Materials and methods

Subjects

This is a retrospective study, and subjects attended for care from 2004 to 2022. According to the evidence-based guideline, ultrasound is not recommended for the diagnosis of PCOS in girls with

a gynecological age of <8 years. Therefore, we included subjects who were between 2 and 8 years since menarche, all of whom were ethnically Korean. The inclusion and exclusion criteria are presented as a flowchart in [Fig. 1](#). The study starts from the girls who were diagnosed as PCOS based on the Rotterdam criteria, with the following definitions of the three components. Oligomenorrhea was defined as fewer than eight periods per year, or cycles longer than 35 days, whereas amenorrhea was defined as the absence of menstruation for >3 months, provided it persisted for >2 years after menarche. Hirsutism was assessed using a modified Ferriman and Gallwey system (mF-G score), on which clinical HA was defined as a score of ≥ 6 , as reported in our previous study ([Kim et al., 2011](#)). Conversely, biochemical HA was defined as the presence of one of the following three criteria: total testosterone (T) >0.68 ng/ml, free T >1.72 pg/ml, and free androgen index >5.36 ([Chae et al., 2008](#)). PCO was defined as either 12 or more follicles measuring 2–9 mm in diameter or an increased ovarian volume (>10 cm³). All ultrasound evaluations were performed via transrectal route with informed consent. Ultrasound examinations were carried out with wide band frequency (5–9 MHz) transducers with automatic optimization, and the center frequency was 8 MHz (Voluson E8, GE Healthcare, Milwaukee, WI, USA). Patients with suspected hypothalamic amenorrhea were excluded and none of the adolescent girls with PCOS had taken insulin sensitizers or lipid-lowering agents during the initial evaluation. For patients using oral contraceptives at their first visit, the medication was discontinued for a minimum of 3 months prior to evaluation.

Next, in this PCOS group, we examined whether each subject also met the evidence-based guideline adolescent criteria. Thus, menstrual cycle irregularity was redefined by time post-menarche ([Table 1](#)). Definition of clinical and biochemical HA was applied the same way as described above. The presence of PCO was not reevaluated since it is not part of the adolescent criteria in the evidence-based guideline. If a girl had both IM and HA, she was classified as ‘meeting the adolescent criteria’. If a girl had either IM or HA, she was defined as being ‘at risk’.

A total of 428 girls with a gynecological age of 2–8 years who visited the Seoul National University Hospital Healthcare System as part of a regular health and gynecological development check-up were enrolled as controls. In these girls, information about menstrual cycle and the presence of hirsutism was evaluated before enrollment, and all controls had absolutely regular (21–35 days) menstrual cycles and an mF-G score below 6. Girls taking hormonal medication were excluded from the control group.

This study was approved by the Institutional Review Board of Seoul National University Hospital (approval no. 1407-129-597), and written informed consent was obtained from all patients.

Table 1. Diagnostic criteria for polycystic ovary syndrome in adolescence.

Rotterdam criteria (2 out of 3)	International evidence-based guideline (both 1 and 2)
1. Oligo- and/or anovulation*	1. Oligo- and/or anovulation* Irregular menstrual cycles are defined as follows • Normal in the first year post-menarche as part of the pubertal transition. • 1 to <3 years post-menarche: <21 or >45 days. • 3 years post-menarche to perimenopause: <21 or >35 days or <8 cycles per year. • 1 year post-menarche >90 days for any one cycle. • Primary amenorrhea by age 15 or >3 years post thelarche
2. Clinical and/or biochemical signs of hyperandrogenism*	2. Clinical and/or biochemical signs of hyperandrogenism*
3. Polycystic ovaries	Ultrasound should not be used for the diagnosis of PCOS in those with a gynecological age of <8 years (<8 years after menarche).

* Provided other causes are excluded.

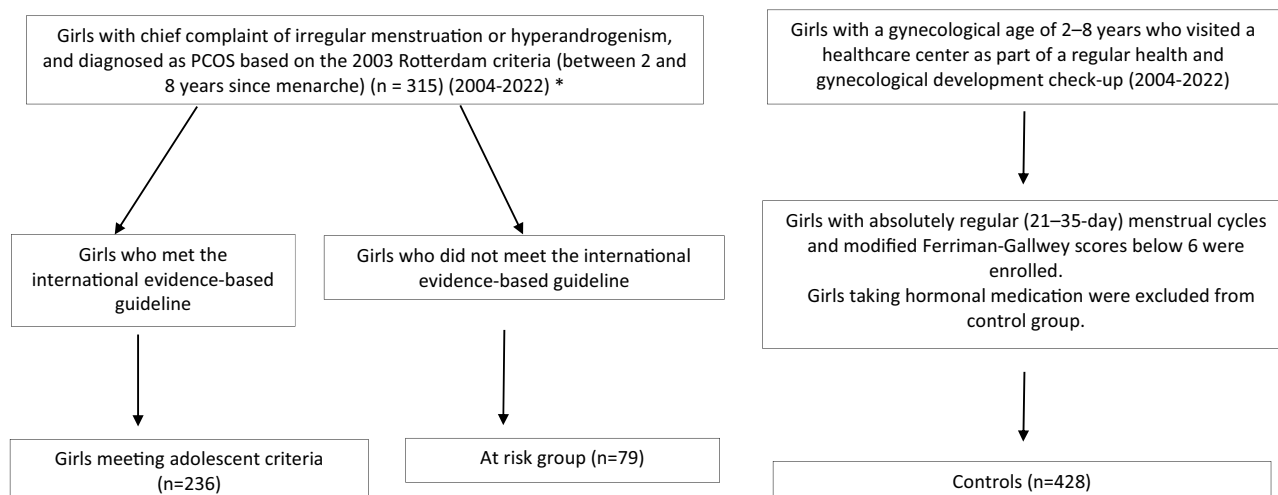


Figure 1. Flowchart showing subjects' inclusion and exclusion. *Patients with suspected hypothalamic amenorrhea were excluded. None of the adolescent girls with PCOS had taken insulin sensitizers or lipid-lowering agents during the initial evaluation. For patients using oral contraceptives at their first visit, medication was discontinued for a minimum of 3 months prior to evaluation.

Clinical and biochemical measurements

Clinical variables including body weight, height, waist circumference (WC), and blood pressure (BP) were assessed. BMI was calculated as the weight (kg) divided by the height in meters squared (m^2). The degree of obesity was classified by BMI for age and sex percentiles, in accordance with the 2017 Korean National Growth Charts (Kim et al., 2018). Fasting glucose, insulin, hemoglobin A_{1c} (A1C), triglycerides (TG), and high-density lipoprotein (HDL) cholesterol were measured in all participants. Further, the homeostatic model for insulin resistance (HOMA-IR) was calculated as $\text{glucose (mg/dl)} \times \text{insulin (mU/ml)} / 405$. IR was defined as a $\text{HOMA-IR} \geq 2.64$, as reported in our previous study (Kim et al., 2019). In girls with PCOS, serum levels of total T, free T, sex hormone-binding globulin, and dehydroepiandrosterone sulfate were measured using radioimmunoassay (Siemens, Los Angeles, CA, USA). In girls with PCOS, the serum LH, FSH, estradiol, and anti-Müllerian hormone (AMH) levels were measured, and 2-h glucose and insulin levels were evaluated using a 75-g oral glucose tolerance test (OGTT). Metabolic syndrome was diagnosed if ≥ 3 of the following five components were met: (i) $\text{WC} > 88$ cm, (ii) $\text{TG} \geq 150$ mg/dl, (iii) systolic BP ≥ 130 mmHg and/or diastolic BP ≥ 85 mmHg, (iv) fasting glucose 110 – 126 mg/dl and/or 2-h glucose from OGTT 140 – 199 mg/dl, and (v) HDL cholesterol < 50 mg/dl based on the 2003 Rotterdam consensus. Diabetes was defined as a fasting glucose level ≥ 126 mg/dl, 2-h glucose in the 75-g OGTT ≥ 200 mg/dl, or A1C $\geq 6.5\%$ (American Diabetes Association, 2019). Prediabetes was defined as fasting glucose level of 100 – 125 mg/dl and/or 2-h glucose level in the 75-g OGTT of 140 – 199 mg/dl, and/or A1C level of 5.7 – 6.4% .

Statistical analysis

Data are shown as the mean $\pm$ SD, geometric mean and 95% confidence intervals, or median value with first and third quartile. Data normality was tested using the Kolmogorov–Smirnov test. For parameters that did not follow a Gaussian distribution, normality was achieved using a natural logarithmic or square root transformation. Subsequently, continuous parameters were compared using the Student's t-test or ANOVA and categorical variables were compared using the chi-square test. All data analyses were performed using the Statistical Package for the Social Sciences software (version 27.0, IBM SPSS, NY, USA), and statistical significance was set at a two-sided P value of < 0.05 .

Results

Diagnostic status according to the international evidence-based guideline

The average age of the participants was 18 years, and the average number of years since menarche was 5 (Table 2). Among 315 girls diagnosed with PCOS by the original Rotterdam criteria, 206 (65.4%), 30 (9.5%), 12 (3.8%), and 67 (21.3%) were classified as having the IM/HA/PCO, IM/HA, HA/PCO, and IM/PCO phenotypes, respectively. Menstrual cycle irregularity was redefined by time post-menarche (Table 1) and all subjects had no changes in IM status when we applied the adolescent criteria. Thus, 236 girls with both IM and HA ($\pm$ PCO) were still diagnosed as PCOS by the new criteria, and they formed the girls meeting adolescent criteria group. Seventy-nine girls (25.1%) with HA/PCO or IM/PCO phenotypes did not meet the diagnosis of PCOS and they formed the at-risk group.

Comparison of clinical features among the girls meeting adolescent criteria, at-risk group, and controls

We investigated whether the at-risk group showed differences in clinical characteristics compared with the girls meeting adolescent criteria and controls (Table 2). As expected, the girls meeting adolescent criteria showed the worst metabolic profiles, including the highest degree of generalized and central obesity, frequency of IR, prediabetes or diabetes and metabolic syndrome, and a higher degree of androgenicity (hirsutism score) than the other two groups. The frequencies of IR were highest in the girls meeting adolescent criteria, followed by the at-risk group and controls, in that order.

The at-risk group was similar to controls in terms of central or generalized obesity parameters, and most (about 90%) were not overweight or obese. However, interestingly, they showed worse metabolic profiles in terms of mean BP, TG, fasting insulin, and HOMA-IR levels than controls, with elevated profiles similar to the girls meeting adolescent criteria. The at-risk group showed elevated mean serum LH level and LH/FSH ratio similar to the girls meeting adolescent criteria. The mean serum AMH level was lower in the at-risk group than the girls meeting adolescent criteria. Regarding the 75-g OGTT 2-h glucose/insulin levels, no differences were observed between the girls meeting adolescent criteria and the at-risk group. Finally, the mean serum HDL-

Table 2. Comparison of clinical features of the girls meeting adolescent criteria, at-risk group, and controls.

	Girls meeting adolescent criteria (n = 236)	At-risk group (n = 79)	Controls (n = 428)	P value
Anthropometric parameters				
Age (years) ^a	18.2 ± 1.9	18.1 ± 2.0	17.9 ± 1.7	0.165
Years since menarche ^a	5.4 ± 2.0	5.1 ± 2.1	5.4 ± 1.9	0.534
BMI (kg/m ²) ^a	22.8 ± 4.6 ^d	20.6 ± 2.7 ^e	20.6 ± 2.9 ^e	<0.001
Obesity classification*				
Underweight (<5th %)	4.3% (8/184)	6.3% (4/64)	12.5% (42/336)	<0.001
Overweight (≤85th and <95th %)	10.3% (19/184)	6.3% (4/64)	4.2% (14/336)	
Obesity (≥95th %)	25.0% (46/184)	4.7% (3/64)	7.4% (25/336)	
Waist circumference (cm) ^a	73.3 ± 11.1 ^d	67.7 ± 6.8 ^e	69.4 ± 8.4 ^e	<0.001
Hirsutism score ^b	7 (5, 9) ^d	3 (2, 5) ^e	1 (0, 2) ^f	<0.001
Systolic blood pressure (mmHg) ^a	112.0 ± 14.4 ^d	109.3 ± 9.9 ^d	104.5 ± 11.0 ^e	<0.001
Diastolic blood pressure (mmHg) ^a	72.6 ± 10.5 ^d	71.0 ± 9.7 ^d	64.0 ± 8.0 ^e	<0.001
Hormonal parameters				
Total testosterone (ng/ml) ^c	0.51 (0.46, 0.56)	0.31 (0.29, 0.337)		<0.001
Free testosterone (pg/ml) ^c	1.67 (1.48, 1.86)	0.82 (0.75, 0.89)		<0.001
Free androgen index ^c	5.63 (4.03, 7.23)	2.18 (2.01, 2.35)		<0.001
Sex hormone binding globulin ^c	39.8 (35.5, 44.1)	51.9 (48.8, 55.0)		<0.001
Dehydroepiandrosterone sulfate ^b	2171 (2007, 2335)	1609 (1495, 1723)		<0.001
LH ^c	9.2 (7.9, 10.5)	9.4 (8.2, 10.6)		0.872
FSH ^c	4.9 (4.6, 5.2)	5.4 (5.2, 5.6)		0.185
LH/FSH ^b	1.7 (1.2, 2.6)	1.8 (1.0, 2.4)		0.555
Anti-Müllerian hormone (ng/ml) ^c	11.7 (10.7, 12.7)	8.8 (7.9, 9.7)		0.004
Metabolic parameters				
Triglyceride (mg/dl) ^b	86 (64, 117) ^d	80 (61, 104) ^d	65 (50, 84) ^e	<0.001
HDL cholesterol (mg/dl) ^a	58.1 ± 16.2	60.4 ± 13.7	57.6 ± 11.2	0.263
Fasting glucose (mg/dl) ^a	89.3 ± 11.0	88.4 ± 7.4	88.1 ± 6.5	0.233
Fasting insulin (μU/ml) ^c	13.9 (11.6, 16.2) ^d	11.7 (11.0, 12.4) ^d	5.7 (5.3, 6.1) (n = 78) ^e	<0.001
HOMA-IR ^c	3.06 (2.53, 3.59) ^d	2.54 (2.37, 2.71) ^d	1.19 (1.11, 1.27) (n = 78) ^e	<0.001
HOMA-IR ≥ 2.64	50.5% (104/206) ^d	31.3% (21/67) ^e	5.1% (4/78) ^f	<0.001
Hemoglobin A1c (%) ^a	5.33 ± 0.55	5.29 ± 0.31	5.34 ± 0.28	0.628
75 g OGTT 2 h glucose (mg/dl) ^a	107.8 ± 27.9	99.5 ± 16.7		0.006
75 g OGTT 2 h insulin (μU/ml) ^c	66.0 (52.3, 79.7)	55.3 (46.7, 63.9)		0.172
Presence of prediabetes or diabetes ⁺	24.2% (53/219) ^d	17.8% (13/73) ^e	13.7% (41/299) ^e	0.009
Presence of metabolic syndrome	11.1% (21/189) ^d	0% (0/68) ^e	1.0% (4/406) ^e	<0.001

Values with different superscripts are significantly different.

P-values are indicated for the differences between groups, as analyzed using Student's t-test, the chi-square test, or ANOVA. A post hoc test was performed using Scheffe's method.

^a Data are shown as the mean ± SD.

^b Median (interquartile range).

^c Geometric mean and 95% confidence interval.

* The degree of obesity was classified by BMI for age and sex percentiles, in accordance with the 2017 Korean National Growth Charts (Kim et al., 2018).

⁺ Only in the girls meeting adolescent criteria, subjects who were diagnosed as having diabetes (n = 4) existed. In controls, fasting glucose and hemoglobin A1c data were available.

HDL, high-density lipoprotein; HOMA-IR, homeostasis model assessment of insulin resistance; OGTT, oral glucose tolerance test.

cholesterol, fasting glucose, and A1C levels were similar among the three groups.

Discussion

This study examined the characteristics of adolescents diagnosed with PCOS based on the original Rotterdam criteria but not meeting the evidence-based guideline adolescent criteria. Approximately one-quarter (25.1%) did not meet the PCOS diagnosis based on the updated guideline. Although most (90%) of them were not overweight or obese, their mean BP, TG, fasting insulin, and HOMA-IR levels were significantly higher than those of the controls, and similar to those of the girls who still remained with PCOS. The serum LH levels and LH/FSH ratio of these at-risk girls were also elevated to levels similar to the girls meeting adolescent criteria. However, the degree of obesity, diabetic status, and frequencies of IR or metabolic syndrome were significantly different between the girls meeting adolescent criteria and the other two groups. Overall, our data show that the application of

the evidence-based guideline substantially reduced the number of adolescent girls diagnosed with PCOS, predominantly identifying subjects with the greatest risk of metabolic abnormalities, but also suggest that the at-risk group may exist in a 'gray zone' between PCOS and normal controls. Continuous follow-up for newly developed or worsening PCOS-related symptoms or metabolic abnormalities is required in these girls. According to the evidence-based guideline, for adolescents who have features of PCOS but do not meet diagnostic criteria, an increased risk could be considered, and symptomatic treatment and reassessment at or before 8 years post-menarche is recommended. The results of our study are in line with this recommendation.

The at-risk group showed an elevated mean serum LH level and LH/FSH ratio similar to the girls meeting adolescent criteria. Compared with women with regular menstruation, women with PCOS generally exhibit increased serum LH with low/normal FSH levels. IR and increased gonadotropin-releasing hormone pulsatile secretion frequency can increase the serum LH concentrations or LH/FSH ratios, which is a core pathophysiology of PCOS.

Therefore, the at-risk girls with symptoms of either IM or HA may be pathophysiologically aligned with PCOS, and the degree of obesity may divide the symptom presentation. However, we did not evaluate LH and FSH levels in the controls; therefore, we were unable to judge whether an increase in LH levels in the remained girls with PCOS and in the at-risk group was associated with PCOS pathophysiology, or merely reflected the physiological stage of adolescent development.

In our study, the at-risk group showed lower serum AMH levels (8.8 ng/ml, 95% CI 7.9–9.7) than the girls meeting adolescent criteria (11.7 ng/ml, 95% CI 10.7–12.7). In the updated 2023 evidence-based guideline, AMH can be used for defining PCO in adults, but it is not recommended in adolescents. However, several studies have found increased AMH levels in adolescent PCOS and reported that levels above 3.4 ng/ml in non-obese adolescent girls in the USA (Sopher et al., 2014), 4.16 ng/ml in Koreans (Song et al., 2021), 6.32 ng/ml in Chinese (Li et al., 2022), 6.26 ng/ml in US obese adolescent girls (Kim et al., 2017), or 7.25 ng/ml in Turkey (Savas-Erdeve et al., 2016) best distinguish adolescents with PCOS from controls. Recently, a systematic review and meta-analysis reported that the cutoff values of AMH for the diagnosis of adolescent PCOS were estimated to be 6–7 ng/ml (Tsukui et al., 2022). Although ethnic differences, definition of PCOS, and different AMH assays impact on the interpretation of AMH levels, the at-risk group in our study showed higher AMH levels than the previously reported cutoffs. A long-term follow-up study reported that AMH either alone or adjuvant to adolescent oligomenorrhea does not contribute as a prognostic marker for PCOS in adulthood (Caanen et al., 2021). However, our results and the above-mentioned AMH studies suggest that at-risk girls need to be monitored for the possible development of PCOS in later life.

In the present study, we observed differences in obesity classification and median BMI percentile between the girls meeting adolescent criteria and the at-risk group. The risk of developing PCOS-like features increases with increasing obesity in vulnerable subjects, as obesity aggravates the degree of IR. Further investigations are therefore needed to determine whether weight gain in the at-risk girls results in a diagnosis of PCOS due to overt symptom presentation, or a deterioration in diabetic or metabolic syndrome classification. A recent study showed that adolescents who met the updated 2018 PCOS criteria had a higher BMI than those without PCOS from the prepubertal period to adulthood (Tay et al., 2020). However, adolescents diagnosed with PCOS under the original Rotterdam criteria but excluded under the 2018 updated guideline showed comparable BMI changes over time with controls. Thus, the authors reported that the use of the 2018 updated adolescent criteria can limit the over-diagnosis of PCOS, to identify girls at the greatest risk of long-term weight gain as a core driver of PCOS severity.

There are some limitations in this study. First, serum androgen measurements were not done via liquid chromatography-tandem mass spectrometry, which is considered to be highly accurate. Second, in the international evidence-based guideline, ultrasound is not used in adolescent PCOS diagnosis in the consensus recommendations (2018) or practice points (2023) levels, similar recommendation levels with that of the Rotterdam criteria. We did not immediately change the approach to adolescent PCOS diagnosis after the 2018 guideline was released and still evaluated the three components of the Rotterdam criteria if a girl visited a gynecology clinic due to menstrual disturbance or hyperandrogenic symptoms. And ultrasound was performed via transrectal route. Transrectal scanning might be an alternative technique when transvaginal scanning is not feasible and the

images obtained were reported to be comparable with those obtained by transvaginal scanning (Timor-Tritsch et al., 2003; Güdücü et al., 2013; Lee et al., 2015). One thing to note is that when we determined the PCO morphology, the follicle number cutoff was 12. Since the Rotterdam criteria, thresholds for follicle excess have been updated, such as 20 (Teede et al., 2018), 22 (Carmina et al., 2016), or 25 (Dewailly et al., 2014), with the increasing sensitivity of ultrasound, but we neither reflected the updated follicle counts nor performed a time course analysis to evaluate PCO changes over time. Thus, substantial numbers of the subjects might not have been categorized as having PCO according to the updated cutoffs, and identification of the at-risk group in our study might actually have been based on either IM or HA. In this respect, we can assume that the evaluation of PCO in adolescents offered no clinical advantage in assessing the 'gray zone' or the at-risk group. Third, we did not measure hormonal levels in controls, and the lack of data is an obstacle in determining the hormonal position of the at-risk group. Finally, the current study was performed from a homogenous population, Koreans, and care must be taken in applying our findings to other ethnic groups.

The at-risk group can be divided into girls with HA only ($n=12$) and girls with IM only ($n=67$). Additional analysis of these two sub-groups did not show any significant differences since girls with IM only made up bulk (85%) of the at-risk group (data not shown). Studies are needed with a larger number of subjects to determine whether there are metabolic abnormality differences among girls with PCOS, with HA only, with IM only, and controls.

Overall, approximately one-quarter of adolescent girls diagnosed with PCOS based on the Rotterdam criteria did not meet the international evidence-based guideline adolescent criteria. These girls nevertheless visited a hospital due to menstrual abnormalities or hyperandrogenic symptoms. Further, they showed worse metabolic profiles than controls and shared considerable metabolic and hormonal features with adolescents who still remained with PCOS. Compared to the conventional Rotterdam criteria, the recent international evidence-based guideline can limit the over-diagnosis of PCOS in adolescents but also provides the opportunity to identify young girls who are at risk. The results of our study are in line with the recommendation by the evidence-based guideline, and a practical approach to this population would involve symptom control and regular check-ups for newly developed or worsening PCOS-related symptoms or metabolic abnormalities.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' roles

J.J.K. developed the study design and drafted the manuscript. K. R.H. developed the study design and revised the manuscript for important intellectual content. D.L. revised the manuscript for important intellectual content. S.K. interpreted the data. Y.M.C. offered material support. All authors read and approved the final manuscript.

Funding

No funding was received in support of this study.

Conflict of interest

The authors have no conflicts of interest to disclose.

References

- American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes—2019. *Diabetes Care* 2019;**42**:S13–S28.
- Caanen MR, Peters HE, van de Ven PM, Jüttner AMFM, Laven JSE, van Hooff MHA, Lambalk CB. Anti-Müllerian hormone levels in adolescence in relation to long-term follow-up for presence of polycystic ovary syndrome. *J Clin Endocrinol Metab* 2021;**106**:e1084–e1095.
- Carmina E, Campagna AM, Fruzzetti F, Lobo RA. AMH measurement versus ovarian ultrasound in the diagnosis of polycystic ovary syndrome in different phenotypes. *Endocr Pract* 2016;**22**:287–293.
- Chae SJ, Kim JJ, Choi YM, Hwang KR, Jee BC, Ku SY, Suh CS, Kim SH, Kim JG, Moon SY. Clinical and biochemical characteristics of polycystic ovary syndrome in Korean women. *Hum Reprod* 2008;**23**:1924–1931.
- Dewailly D, Lujan M, Carmina E, Cedars M, Laven J, Norman R, Escobar-Morreale H. Definition and significance of polycystic ovarian morphology: a task force report from the Androgen Excess and Polycystic Ovary Syndrome Society. *Hum Reprod Update* 2014;**20**:334–352.
- Fauser BCJM, Tarlatzis BC, Rebar RW, Legro RS, Balen AH, Lobo R, Carmina E, Chang J, Yildiz BO, Laven JS et al. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertil Steril* 2012;**97**:28–38.e25.
- Güdücü N, Sidar G, İşçi H, Yiğiter AB, Dündar İ. The utility of transrectal ultrasound in adolescents when transabdominal or transvaginal ultrasound is not feasible. *J Pediatr Adolesc Gynecol* 2013;**26**:265–268.
- Ibáñez L, Oberfield SE, Witchel S, Auchus RJ, Chang RJ, Codner E, Dabaghao P, Darendeliler F, Elbarbary NS, Gambineri A et al. An international consortium update: pathophysiology, diagnosis, and treatment of polycystic ovarian syndrome in adolescence. *Horm Res Paediatr* 2017;**88**:371–395.
- Kim JJ, Chae SJ, Choi YM, Hwang SS, Hwang KR, Kim SM, Yoon SH, Moon SY. Assessment of hirsutism among Korean women: results of a randomly selected sample of women seeking pre-employment physical check-up. *Hum Reprod* 2011;**26**:214–220.
- Kim JH, Yun S, Hwang SS, Shim JO, Chae HW, Lee YJ, Lee JH, Kim SC, Lim D, Yang SW et al.; Division of Health and Nutrition Survey, Korea Centers for Disease Control and Prevention. The 2017 Korean National Growth Charts for children and adolescents: development, improvement, and prospects. *Korean J Pediatr* 2018;**61**:135–149.
- Kim JJ, Hwang KR, Oh SH, Chae SJ, Yoon SH, Choi YM. Prevalence of insulin resistance in Korean women with polycystic ovary syndrome according to various homeostasis model assessment for insulin resistance cutoff values. *Fertil Steril* 2019;**112**:959–966.e1.
- Kim JY, Tfayli H, Michaliszyn SF, Lee SJ, Nasr A, Arslanian S. Anti-Müllerian hormone in obese adolescent girls with polycystic ovary syndrome. *J Adolesc Health* 2017;**60**:333–339.
- Lee DE, Park SY, Lee SR, Jeong K, Chung HW. Diagnostic usefulness of transrectal ultrasound compared with transvaginal ultrasound assessment in young Korean women with polycystic ovary syndrome. *J Menopausal Med* 2015;**21**:149–154.
- Li M, Ruan X, Ju R, Min M, Xu Z, Luo S, Wang H, Mueck AO. Is anti-Müllerian hormone a useful biomarker in the diagnosis of polycystic ovary syndrome in Chinese adolescents? *Gynecol Endocrinol* 2022;**38**:148–152.
- Savas-Erdeve S, Keskin M, Sagsak E, Cenesiz F, Cetinkaya S, Aycan Z. Do the anti-Müllerian hormone levels of adolescents with polycystic ovary syndrome, those who are at risk for developing polycystic ovary syndrome, and those who exhibit isolated oligomenorrhea differ from those of adolescents with normal menstrual cycles? *Horm Res Paediatr* 2016;**85**:406–411.
- Song J, Park Y, Cho HW, Lee SG, Kim S, Lim JB. Age-group-specific reference intervals for anti-Müllerian hormone and its diagnostic performance for polycystic ovary syndrome in a Korean population. *J Clin Lab Anal* 2021;**35**:e23861.
- Sopher AB, Grigoriev G, Laura D, Cameo T, Lerner JP, Chang RJ, McMahon DJ, Oberfield SE. Anti-Müllerian hormone may be a useful adjunct in the diagnosis of polycystic ovary syndrome in nonobese adolescents. *J Pediatr Endocrinol Metab* 2014;**27**:1175–1179.
- Tay CT, Hart RJ, Hickey M, Moran LJ, Earnest A, Doherty DA, Teede HJ, Joham AE. Updated adolescent diagnostic criteria for polycystic ovary syndrome: impact on prevalence and longitudinal body mass index trajectories from birth to adulthood. *BMC Med* 2020;**18**:389.
- Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod* 2018;**33**:1602–1618.
- Teede HJ, Tay CT, Laven JJE, Dokras A, Moran LJ, Piltonen TT, Costello MF, Boivin J, Redman LM, Boyle JA et al. Recommendations from the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *J Clin Endocrinol Metab* 2023;**108**:2447–2469.
- The Rotterdam ESHRE/ASRM-sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod* 2004;**19**:41–47.
- Timor-Tritsch IE, Monteagudo A, Rebarber A, Goldstein SR, Tsymbal T. Transrectal scanning: an alternative when transvaginal scanning is not feasible. *Ultrasound Obstet Gynecol* 2003;**21**:473–479.
- Tsukui Y, Kitahara Y, Hasegawa Y, Kobayashi M, Osuka S, Iwase A. Anti-Müllerian hormone levels in the diagnosis of adolescent polycystic ovarian syndrome: a systematic review and meta-analysis. *Endocr J* 2022;**69**:897–906.

Reproductive endocrinology

Resumption of ovulation in anovulatory women with PCOS and obesity is associated with reduction of 11 β -hydroxyandrostenedione concentrations

Z. Wang¹, M. Van Faassen², H. Groen³, A.E.P. Cantineau¹, A. Van Oers¹, A. Van der Veen², J.M. Hawley⁴, B.G. Keevil⁴, I.P. Kema², and A. Hoek^{1,*}

¹Department of Obstetrics and Gynecology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

²Department of Laboratory Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

³Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

⁴Department of Clinical Biochemistry, Wythenshawe Hospital, Manchester NHS Foundation Trust, Manchester, UK

*Correspondence address. Department of Obstetrics and Gynecology, University of Groningen, University Medical Centre Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands. E-mail: a.hoek@umcg.nl <https://orcid.org/0000-0003-4441-7142>

ABSTRACT

STUDY QUESTION: Is resumption of ovulation after a 6-month lifestyle intervention in women with PCOS and obesity associated with differential changes in endocrine and metabolic parameters (weight, insulin resistance, anti-Müllerian hormone (AMH), and androgens) compared to women with PCOS who remained anovulatory?

SUMMARY ANSWER: Resumption of ovulation after a 6-month lifestyle intervention in women with PCOS and obesity is associated with changes in serum 11 β -hydroxyandrostenedione (11OHA4) concentrations.

WHAT IS KNOWN ALREADY: Lifestyle interventions have been shown to reduce clinical and biochemical hyperandrogenism in women with PCOS. Weight loss of 5–10% may reverse anovulatory status, thereby increasing natural conception rates. However, the mechanisms underlying why some women with PCOS remain anovulatory and others resume ovulation after weight loss are unclear. Reproductive characteristics at baseline and a greater degree of change in endocrine and metabolic features with lifestyle intervention may be crucial for ovulatory response.

STUDY DESIGN, SIZE, DURATION: We used data and samples originating from an earlier randomized controlled trial (RCT), which examined the efficacy of a 6-month lifestyle intervention prior to infertility treatment compared to prompt infertility treatment on live birth rate in women with obesity. A total of 577 women with obesity (BMI > 29 kg/m²) were randomized between 2009 and 2012. Anovulatory women with PCOS who were allocated to the intervention arm of the original RCT (n = 95) were included in the current analysis.

PARTICIPANTS/MATERIALS, SETTING, METHODS: We defined women as having resumed ovulation (RO+) based on the following criteria: spontaneous pregnancy; or assignment to expectant management; or IUI in natural cycles as the treatment strategy after lifestyle intervention. Steroid hormones were measured using liquid chromatography tandem mass spectrometry. Generalized estimating equations with adjustment for baseline measures and interaction between group and time was used to examine differences in changes of endocrine and metabolic parameters between RO+ (n = 34) and persistently anovulatory women (RO–, n = 61) at 3 and 6 months after intervention.

MAIN RESULTS AND THE ROLE OF CHANCE: At baseline, the mean $\pm$ SD age was 27.5 $\pm$ 3.6 years in the RO+ group and 27.9 $\pm$ 4.1 years in the RO– group (P = 0.65), and the mean $\pm$ SD weights were 101.2 $\pm$ 9.5 kg and 105.0 $\pm$ 14.6 kg, respectively (P = 0.13). Baseline AMH concentrations showed significant differences between RO+ and RO– women (median and interquartile range [IQR] 4.7 [3.2; 8.3] versus 7.2 [5.3; 10.8] ng/ml, respectively). Baseline androgen concentrations did not differ between the two groups. During and after lifestyle intervention, both groups showed weight loss; changes in 11OHA4 were significantly different between the RO+ and RO– groups (P-value for interaction = 0.03). There was a similar trend for SHBG (interaction P-value = 0.07), and DHEA-S (interaction P-value = 0.06), with the most pronounced differences observed in the first 3 months. Other parameters, such as AMH and FAI, decreased over time but with no difference between the groups.

LIMITATIONS, REASONS FOR CAUTION: No high-resolution transvaginal ultrasonography was used to confirm ovulatory status at the end of the lifestyle program. The small sample size may limit the robustness of the results.

WIDER IMPLICATIONS OF THE FINDINGS: Reduction of androgen concentrations during and after lifestyle intervention is associated with recovery of ovulatory cycles. If our results are confirmed in other studies, androgen concentrations could be monitored during lifestyle intervention to provide individualized recommendations on the timing of resumption of ovulation in anovulatory women with PCOS and obesity.

STUDY FUNDING/COMPETING INTEREST(S): The study was supported by a grant from ZonMw, the Dutch Organization for Health Research and Development (50-50110-96-518). The Department of Obstetrics and Gynecology of the UMCG received an unrestricted

Received: March 30, 2023. Revised: February 02, 2024. Editorial decision: February 19, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

educational grant from Ferring Pharmaceuticals BV, The Netherlands. A.H. reports consultancy for the development and implementation of a lifestyle App MyFertiCoach developed by Ferring Pharmaceutical Company. All other authors have no conflicts to declare.

TRIAL REGISTRATION NUMBER: The LIFeStyle RCT was registered at the Dutch trial registry (NTR 1530).

Keywords: PCOS / lifestyle intervention / resumption of ovulation / anti-Müllerian hormone / androgens / obesity / 11 β -hydroxyandrostenedione

Introduction

Polycystic ovary syndrome (PCOS) is the leading cause of anovulatory infertility, characterized by chronic anovulation, hyperandrogenemia, and polycystic ovaries. The available evidence suggests that more than one pathway is involved in the pathophysiological mechanisms leading to anovulation in PCOS (Franks and Hardy, 2020). Among these pathways, androgens play a key role in linking hypothalamic or pituitary dysfunction to ovarian dysfunction (Franks and Hardy, 2020). The production of excess androgens in PCOS is complex and has a heterogeneous pathophysiological background. Excess androgen levels increase the number of small antral follicles by stimulating the formation of primordial, primary, secondary and early preantral follicles but without progression to the dominant follicle as a result of compromised granulosa cell function (Lebbe and Woodruff, 2013).

Anti-Müllerian hormone (AMH), a glycoprotein hormone, is expressed by granulosa cells of growing follicles from the primary up to the small antral stage, and AMH expression in granulosa cells disappears after FSH-dependent selection (Broer et al., 2014). Serum AMH has been recognized as a marker of ovarian function, particularly in assessing ovarian reserve (Moolhuijsen and Visser, 2020). It has been suggested that elevated AMH concentrations may result in acceleration of maturation and premature luteinization of antral follicles, which contributes to problems with ovulation and fertility in PCOS (Dumont et al., 2015). Fig. 1 shows AMH actions in normal and polycystic ovaries.

Up to 60% of women with PCOS are overweight or obese (Lim et al., 2012). Excess body weight and visceral fat promote the development of insulin resistance and compensatory hyperinsulinemia, which stimulates the synthesis of androgens in the ovaries (Diamanti-Kandarakis and Dunaif, 2012). Elevated androgen concentrations in turn disrupt glucose and insulin regulation (Baptiste et al., 2010). In addition, the production of sex hormone-binding globulin (SHBG) is reduced, which results in an increase

in 'bioavailable' androgens (Zhu et al., 2019). Women with PCOS and obesity are at high risk of chronic diseases, such as metabolic syndrome and cardiovascular diseases (Anagnostis et al., 2018). Moreover, adipose tissue is an important site for steroid storage and metabolism (Li et al., 2015). An important link between androgen excess and obesity in fat cells is the conversion of androstenedione (A4) to testosterone (T) and 11-ketoandrostenedione (11KA4) to 11-ketotestosterone (11KT) through the activity of 17 β -hydroxysteroid dehydrogenase type 5 (HSD17B5) (Fig. 2) (Torchen et al., 2020).

Recent international evidence-based guidelines on PCOS highlight the important role of diet and exercise in the management of women with PCOS, particularly in those who are overweight or obese (Teede et al., 2018). Lifestyle interventions have been shown to reduce clinical and biochemical hyperandrogenism in women with PCOS (Lim et al., 2019). Even a 5–10% loss of weight may reverse anovulatory status, thereby increasing natural conception rates (van Oers et al., 2016). However, the underlying mechanisms of why some women with PCOS remain anovulatory and others resume ovulation after weight loss are unclear. Reproductive characteristics at baseline and a greater degree of change in endocrine and metabolic features with lifestyle intervention in some women may be crucial for ovulatory response (Jarrett and Lujan, 2016). To illustrate, women with lower AMH and higher SHBG at baseline more often experience sporadic ovulations compared to women who remained anovulatory after weight loss, as shown in previous small studies (van Dam et al., 2004; Moran et al., 2007; Thomson et al., 2009). Furthermore, a reduction in BMI, hyperandrogenism and insulin resistance might mediate the occurrence of ovulation resumption in women with PCOS and obesity undergoing weight loss (Pasquali et al., 2006).

In a previously published pilot study of our lifestyle intervention (Kuchenbecker et al., 2011), we investigated body fat distribution, particularly intra-abdominal and abdominal subcutaneous fat, in a group of anovulatory women with PCOS and obesity. We

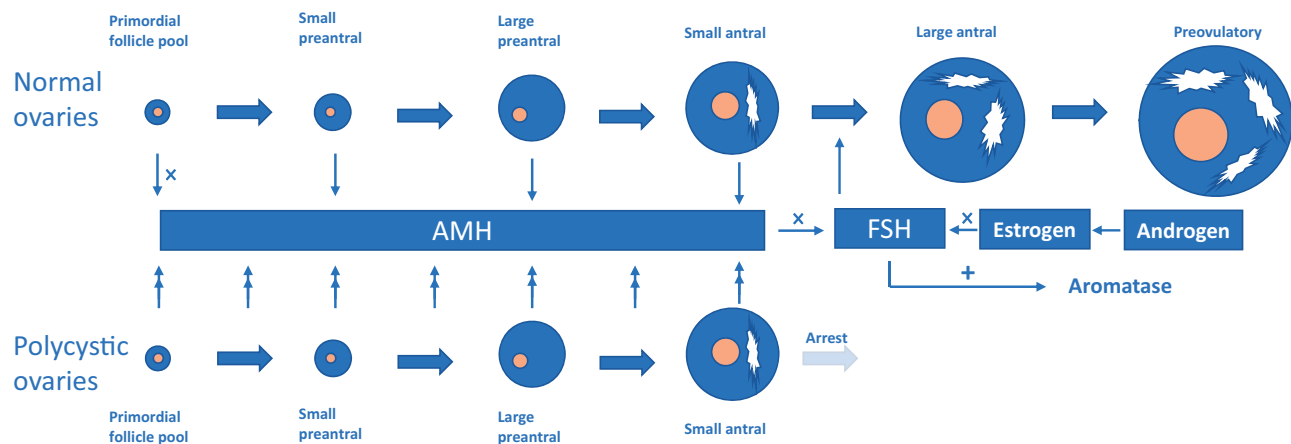


Figure 1. Schematic of anti-Müllerian hormone actions in normal and polycystic ovaries. Upper panel: normal ovaries, lower panel: polycystic ovaries. The yellow center represents the oocyte, the blue area represents the granulosa cell layer and the white area represents follicle fluid in the antrum. High levels of anti-Müllerian hormone (AMH) in polycystic ovary syndrome (PCOS) cause increased suppression of FSH signaling and result in arrest in follicular growth.

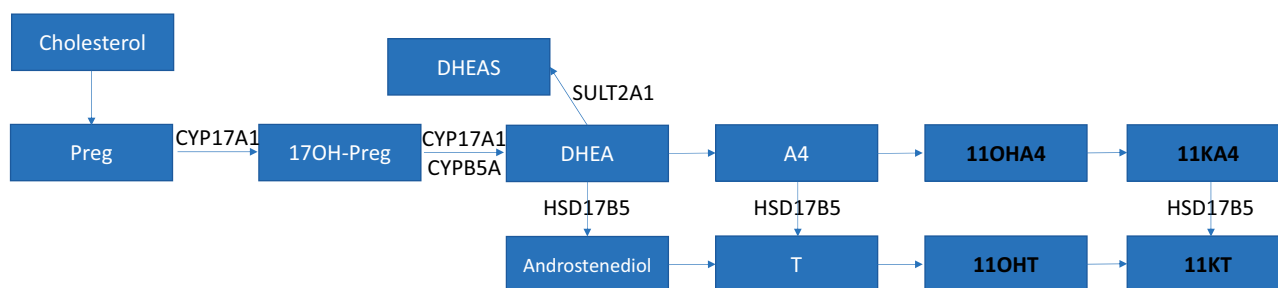


Figure 2. Schematic of 11-oxygenated androgen synthesis in adipose tissue. 11-oxygenated androgen synthesis is indicated by bold text. Preg: pregnenolone; 17OH-Preg: 17-hydroxypregnenolone; DHEA(S): dehydroepiandrosterone (sulfate); A4: androstenedione; T: testosterone; 11OHA4, 11β-hydroxyandrostenedione, 11KA4: 11-ketoandrostenedione; 11OHT: 11β-hydroxytestosterone; 11KT: 11-ketotestosterone; CYP17A1: 17α-hydroxylase/17,20-lyase; CYPB5A: cytochrome b5 type A; SULT2A1: sulfotransferase 2A1; HSD17B5: 17β-hydroxysteroid dehydrogenase type 5.

found that women who regained ovulation lost more weight and intra-abdominal fat than those who remained anovulatory after lifestyle intervention (Kuchenbecker et al., 2011). In the current *post hoc* study, we aimed to investigate if resumption of ovulation after a 6-month lifestyle intervention in women with PCOS and obesity was associated with changes in endocrine and metabolic parameters (weight, insulin resistance, AMH, and androgens). It is hypothesized that achieving a certain level of weight loss can lead to a clinically relevant decrease in AMH levels and an improvement in hyperandrogenism, which are essential for resumption of ovulation.

Materials and methods

Study design and population

We used samples originating from a multi-center randomized controlled trial (RCT). The RCT was approved by the medical ethics committee of the University Medical Center Groningen (METc code: 2008/284), in combination with approval of the board of directors from the other 22 participating hospitals. All women included in the study provided written informed consent. The study was registered in the Netherlands Trial Registry (NTR 1530). Briefly, a total of 577 infertile women with a BMI > 29 kg/m² were randomly assigned to either a 6-month lifestyle intervention followed by 18 months of fertility treatment or 24 months of fertility treatment between 2009 and 2012, to investigate the effect of lifestyle intervention preceding fertility treatment on live birth (Mutsaerts et al., 2010, 2016). The lifestyle intervention consisted of an energy-restricted diet (a reduction of 600 kcal/day but at least 1200 kcal/day), an increase in physical activity (10 000 steps daily and two to three moderate-to-vigorous sessions of exercise a week) and motivational counseling, with a goal of weight reduction of at least 5% of the original body weight or a reduction in BMI to below 29 kg/m² within the intervention period. All women underwent fertility investigation before randomization and were categorized as unexplained infertility, tubal factor, male factor, anovulation (World Health Organization (WHO) I; WHO II: PCOS; WHO II: non-PCOS), endometriosis, and cervix factor. Women were diagnosed as PCOS if the Rotterdam criteria were met: oligo-ovulation or anovulation, clinical manifestations of hyperandrogenism and/or hyperandrogenemia, or ovarian polycystic changes (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004).

Anovulatory women with PCOS who were allocated to the intervention arm of the original RCT were included in the current analysis. They were divided into two groups based on their ovulation status: resumed ovulation (RO+) or remaining anovulatory (RO-) at the end of intervention (6 months after randomization).

Women who conceived did not continue in the lifestyle program. Women who stopped the lifestyle program before the sixth month were considered as dropouts. Numbers of available serum hormone measurements decreased over the course of the study period owing to the nature of the RCT with dropouts, pregnancies, failure to visit the hospital, or blood samples being exhausted for other measurements.

Resumption of ovulation

After completion of the intervention (6 months or in case their weight reduction goal of decrease in body weight was met), women were evaluated to determine any further infertility treatment strategy. During the lifestyle intervention women would monitor their menstrual cycle pattern. Based on information available at the end of the intervention period, we defined women as having resumed ovulation using the following criteria: if they became pregnant spontaneously before the end of the intervention period; if the infertility treatment strategy was expectant management, which is common practice in the Netherlands in ovulatory women with good partner semen quality and no tubal infertility; or, if IUI treatment in natural cycles was started. Women who did not fulfill these criteria were classified as 'remaining anovulatory' after the lifestyle intervention.

Clinical and laboratory measurements

During the hospital visits at randomization, and at 3 and 6 months after randomization, body weight in kg, height in cm, and waist and hip circumference in cm were measured by research nurses not involved in the lifestyle intervention coaching. Fasting blood samples were collected by venipuncture into a collection tube. Cardiometabolic parameters, including fasting glucose and insulin, and serum hormone measurements were performed in the central laboratory of the University Medical Centre Groningen (UMCG, Groningen, The Netherlands) after the trial had been finished. Homeostasis model assessment of insulin resistance (HOMA-IR) was calculated as fasting insulin concentration (μU/ml) multiplied by fasting glucose concentration (mmol/l) divided by 22.5 (Matthews et al., 1985). Detailed information, including measurement method, as well as intra- and inter-assay variations of these outcomes, can be found in our previous study investigating effect of a lifestyle intervention on cardiometabolic health in women with infertility and obesity (van Dammen et al., 2018). The steroid hormones A4, T, dihydrotestosterone (DHT), dehydroepiandrosterone (DHEA) and dehydroepiandrosterone sulfate (DHEA-S) were measured using liquid chromatography tandem mass spectrometry (LC-MS/MS) (van der Veen et al., 2019). SHBG was measured with the Architect manufactured by Abbott Diagnostics (Abbott park, IL, USA), using a chemiluminescent micro particle immunoassay. The free

androgen index (FAI) was calculated by dividing the total serum T level by the SHBG level and multiplying this by 100. 11 β -Hydroxyandrostenedione (11OHA4) and 11KT were analyzed using liquid-liquid extraction followed by LC-MS/MS (Hawley *et al.*, 2020) in Wytheshawe hospital (Manchester, UK). AMH was measured with a chemiluminescent microparticle immunoassay on the Roche Cobas 6000 system (Roche, Mannheim, Germany). Imprecision for AMH was 2.0%, 2.6%, and 4.9% at 1.66, 4.87, and 8.4 ng/ml, respectively, while the lower limit of quantification was 0.03 ng/ml.

Statistical analysis

No sample size calculation could be performed owing to the post hoc nature of this analysis. Our analyses should therefore be regarded as explorative. Baseline characteristics were compared between RO+ and RO- groups. Data were expressed as mean $\pm$ SD or median (interquartile range: IQR) for continuous variables, and proportions (percentage, %) for categorical variables. Normality testing was performed using histograms and normal probability plots (Q-Q plots) combined with the Kolmogorov-Smirnov (K-S) test. The differences between the two groups were compared with Student's t-test or Mann-Whitney U-test where relevant for continuous variables and the Chi-square test for categorical variables. Anthropometric measurements and endocrine and metabolic parameters at 3, 3 and 6 months for the two groups were described separately.

It is important to note that while all these women had baseline anthropometric measures (weight, BMI), not all of them had complete data for anthropometric measures or serum hormone measures at 3 or 6 month time points, for various reasons (Fig. 3). We used the generalized estimating equations (GEE) method with exchangeable correlation structure to analyze longitudinal data, which is preferable over complete-case analysis, as the latter does not reflect a real-life situation and poses the risk of selection bias. To answer our research questions, we performed GEE with weight, AMH, HOMA-IR, and androgen concentrations as the dependent variables and with 'time' (baseline, 3 months, and 6 months), and group (RO+/RO-) as the independent variables. Subsequently, we added an interaction term between time and group to the models to evaluate differences between the RO+ and RO- groups in the course over time for each of the dependent variables separately. All analyses were corrected for baseline for the respective dependent variable. Moreover, estimated marginal means and SEs by time and group were plotted to visualize patterns of change by group. To identify confounders, we included potential confounders (age, smoking status, education, ethnicity) in the respective GEE models, one at a time. If addition of a confounder changed the effect estimate in the models by more than 10%, the variable was regarded as a relevant confounder. Lastly, we compared the differences in endocrine and metabolic parameters between 3 months and baseline and between 6 months and baseline for the RO+ group and the RO- group separately, to quantify the changes over time within both groups. Only 'time' (baseline, 3 months, and 6 months) was included in these models as a categorical independent variable with baseline as the reference. Estimated marginal means at baseline, 3 and 6 months were calculated in the RO+ and RO- groups, respectively.

The Statistical Package for Social Science (IBM SPSS, Armonk, NY, USA, version 27.0) was used to perform statistical analyses and GraphPad Prism (San Diego, CA, USA, version 8.0) was used for data visualization. A *P*-value <0.05 was regarded as statistically significant.

Results

In total, 577 women were included in the original RCT and 289 women were allocated to the lifestyle intervention arm. Ninety-seven anovulatory women with PCOS who were allocated to the lifestyle intervention arm were selected for the current analysis. After the 6-month lifestyle intervention, two women were excluded owing to unknown ovulation status (lost to follow-up), leaving 61 women in RO+ group and 34 women in RO- group. The flow chart of the current study is shown in Fig. 3.

Table 1 shows baseline measurements of the women included in the current study according to whether or not they resumed ovulation at 6 months. The mean age was 27.5 ± 3.6 years in the RO+ group and 27.9 ± 4.1 years in the RO- group (*P* = 0.65). The mean weight was 101.2 ± 9.5 and 105.0 ± 14.6 kg, respectively (*P* = 0.13). There were no statistically significant differences in other anthropometric measurements and baseline characteristics including ethnicity, education, and smoking status. As for the measured serum hormones, only AMH concentrations at baseline showed statistically significant differences between the RO+ women and the RO- women (median and IQR 4.7 [3.2; 8.3] ng/ml in the RO+ group and 7.2 [5.3; 10.8] ng/ml in the RO- group, *P* = 0.03). Other serum hormone measures, including insulin, HOMA-IR, T, and FAI, did not differ significantly between the two groups. Moreover, none of the variables turned out to be relevant confounders, so we did not further adjust for confounders in the GEE model. Supplementary Table S1 shows endocrine and metabolic parameters at 3 and 6 months in the two groups with decreasing numbers of observations in both groups owing to loss to follow-up over time.

Results of interaction terms between time and group from GEE analyses of weight, AMH, HOMA-IR, and serum hormones at baseline, 3 months and 6 months in two groups are shown in Table 2. The only parameter showing a statistically significant difference over time between the RO+ and RO- groups was 11OHA4 (*P*-value for interaction term between time and RO group: 0.03). A similar trend towards differences over time between the RO+ and RO- groups was observed for SHBG (*P*-value for interaction: 0.07), and DHEA-S (*P*-value for interaction: 0.06). The most pronounced differences were observed in the first 3 months. We did not observe differences between the RO+ and RO- groups for other parameters such as BMI (*P* = 0.85), HOMA-IR (*P* = 0.58) or AMH (*P* = 0.24). For some parameters estimated marginal means and SEs by time and group (Table 2) were also plotted (Fig. 4) to visualize the patterns of change by group.

In Table 3 we show the estimated marginal means of weight, AMH, HOMA-IR, and serum hormone concentrations at 3 and 6 months for RO+ and RO- groups separately, to quantify the changes over time within both groups. Weight and BMI significantly decreased during the course of the lifestyle intervention for both groups (*P* < 0.001). AMH concentrations were lower at 3 and 6 months compared to baseline in the RO+ group (baseline estimated marginal means and SE 6.7 ng/ml [1.0] versus 3 months: 4.7 [0.9], *P* = 0.01, and 6.7 [1.0] versus 6 months: 3.0 [1.2], *P* = 0.02) but not in the RO- group (baseline: 8.8 [0.9], 3 months: 8.3 [0.8], and 6 months: 7.9 [0.8]). FAI was significantly lower at 6 months compared to baseline in the RO+ group (baseline 5.5 [0.5] versus 6 months: 3.5 [0.5], *P* = 0.003) and in the RO- group (baseline: 7.2 [0.6], 3 months: 6.0 [0.4], and 6 months: 5.9 [0.5]; *P* = 0.007 for the difference between 3 months and baseline and *P* = 0.049 for the difference between 6 months and baseline). There were no statistically significant differences between 3 months and baseline or between 6 months and baseline for

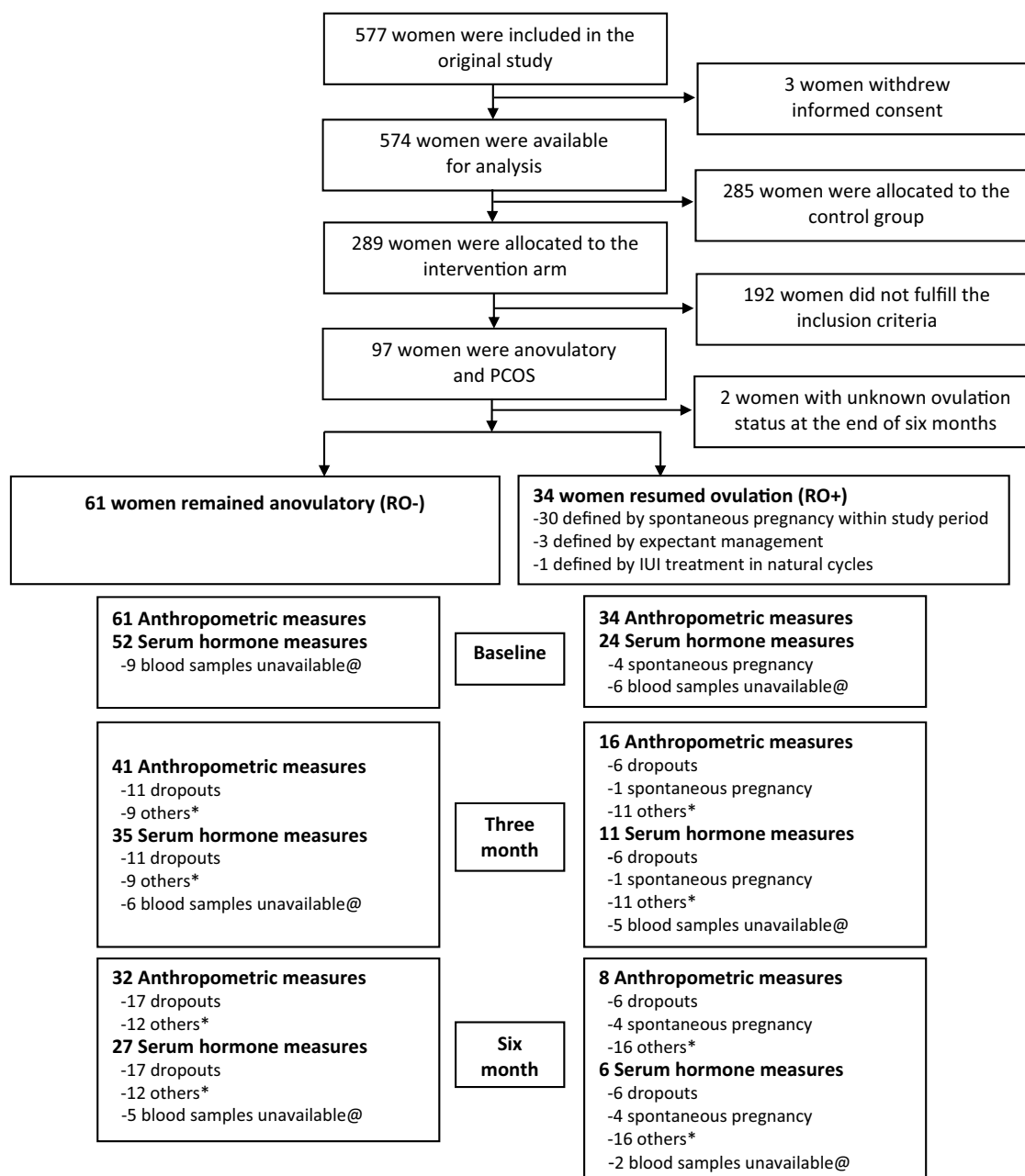


Figure 3. Flow chart of participants for the current analysis of resumption of ovulation in anovulatory women with PCOS and obesity. *Failure to attend the hospital visit or other personal reasons. @blood samples were used in other investigations or missing.

either SHBG or DHEA-S concentrations for both RO+ and RO- groups.

Discussion

In a *post hoc* analysis of our LIFEstyle RCT, we investigated if resumption of ovulation after a 6-month lifestyle intervention in women with PCOS and obesity was associated with changes in endocrine and metabolic parameters (weight, insulin resistance, AMH, and androgens). At baseline, RO+ women had lower serum AMH concentrations than RO- women. During and after lifestyle intervention, only 11OHA4 showed a statistically significant difference between the RO+ and RO- groups over the course of the intervention, while SHBG and DHEA-S showed a similar trend. Moreover, RO+ women showed a significant reduction of serum

AMH concentrations at 3 and 6 months compared to baseline measures while RO- women did not.

The association between greater degree of weight loss and restored ovarian function in anovulatory women with PCOS has been supported by evidence from a small study involving bariatric surgery in 12 anovulatory women with PCOS, in which all women regained normal ovulation and menstrual cycles after a weight loss of 41 kg on average (Escobar-Morreale et al., 2005). Previous studies of lifestyle intervention in women with PCOS and obesity demonstrated that, compared to non-responders, those women who resumed ovulation after weight loss had experienced greater reductions in weight/BMI (Kuchenbecker et al., 2011; Jarrett and Lujan, 2016). Although the optimal degree of weight loss to restore ovulatory function is not clear, at least 5–10% weight loss is generally recommended (Norman et al., 2004; Teede et al., 2018). In the current study, there was no significant

Table 1. Baseline characteristics of anovulatory women with PCOS and obesity with *post hoc* allocation into RO+ and RO– groups at the end of 6 months.

	RO+ (n = 34)	RO– (n = 61)	P-value
Anthropometric measurements			
Weight (kg)	101.2 ± 9.5	105.0 ± 14.6	0.13
BMI (kg/m ²)	35.5 ± 2.9	36.0 ± 3.6	0.44
Waist circumference (cm)	108.7 ± 7.8	108.0 ± 10.1	0.75
Hip circumference (cm)	122.1 ± 8.2	124.2 ± 8.6	0.26
Waist-hip circumference ratio	0.89 ± 0.06	0.87 ± 0.08	0.28
Baseline characteristics			
Age (years)	27.5 ± 3.6	27.9 ± 4.1	0.65
Western European ethnicity	32 (94.1%)	55 (90.2%)	0.51
Current smoker	8 (24.2%)	20 (33.3%)	0.36
Education (number)			0.53
Primary school	1 (2.9%)	3 (4.9%)	
Secondary education	7 (20.6%)	14 (23.0%)	
Intermediate vocational education	20 (58.8%)	26 (42.6%)	
Advanced vocational education or university	5 (14.7%)	12 (19.7%)	
Unknown	1 (2.9%)	6 (9.8%)	
Primary infertility	24 (70.6%)	45 (73.8%)	0.74
Serum measurements			
Insulin (pmol/l)	90.2 (70.1; 131.9)	95.5 (73.8; 142.5)	0.48
HOMA-IR	3.1 (2.3; 4.3)	3.3 (2.5; 4.9)	0.54
AMH (ng/ml)	4.7 (3.2; 8.3)	7.2 (5.3; 10.8)	0.03
SHBG (nmol/l)	28.7 (21.3; 46.3)	27.2 (18.3; 38.6)	0.46
LH (U/l)	10.6 ± 4.9	11.8 ± 5.1	0.37
FSH (U/l)	5.0 ± 2.0	5.3 ± 1.6	0.48
A4 (nmol/l)	6.8 ± 2.6	6.8 ± 2.2	0.99
T (nmol/l)	1.6 ± 0.7	1.8 ± 0.8	0.34
DHT (nmol/l)	0.4 ± 0.3	0.3 ± 0.2	0.10
DHEA (nmol/l)	23.3 ± 10.7	19.6 ± 11.0	0.17
DHEA-S (nmol/l)	6.0 ± 2.8	5.2 ± 2.4	0.24
11KT (nmol/l)	1.3 ± 0.5	1.3 ± 0.6	0.69
11OHA4 (nmol/l)	5.1 ± 2.5	4.6 ± 2.4	0.51
FAI	5.5 ± 2.5	7.2 ± 4.1	0.07

RO+: resumed ovulation at the end of 6 months; RO–: remained anovulatory at the end of 6 months; HOMA-IR: homeostatic model assessment for insulin resistance; AMH: anti-Müllerian hormone; SHBG: sex hormone-binding globulin; A4: androstenedione; T: testosterone; DHT: dihydrotestosterone; DHEA: dehydroepiandrosterone; DHEA-S: dehydroepiandrosterone sulfate; 11KT: 11-ketotestosterone; 11OHA4: 11β-hydroxyandrostenedione; FAI: free androgen index. Data are presented as mean ± SD or median with interquartile (Q25; Q75) or proportion (percentage). The differences between two groups were compared using Fisher's exact test or chi-square test for categorical variables, and Student's t-test or Mann-Whitney U test for continuous variables.

Table 2. Estimated marginal means for anthropometric measurements and endocrine and metabolic parameters from generalized estimating equations analyses.

	Baseline		3 months		6 months		Interaction between time and RO group Overall P-value
	RO+ (n = 34)	RO– (n = 61)	RO+ (n = 16)	RO– (n = 41)	RO+ (n = 8)	RO– (n = 32)	
Weight (kg)	103.4 (0.05)	103.4 (0.04)	99.2 (1.1)	100.3 (0.69)	98.5 (1.0)	99.1 (0.87)	0.72
BMI (kg/m ²)	35.7 (0.01)	35.7 (0.01)	34.3 (0.3)	34.5 (0.3)	34.0 (0.4)	34.3 (0.3)	0.85
Waist circumference (cm)	108.2 (0.16)	108.1 (0.15)	103.2 (2.2)	105.5 (0.8)	102.8 (2.2)	103.5 (1.1)	0.60
Hip circumference (cm)	123.3 (0.04)	123.3 (0.03)	121.3 (1.7)	121.4 (0.8)	120.0 (1.8)	119.3 (0.9)	0.89
Waist-hip circumference ratio	0.88 (0.002)	0.88 (0.001)	0.85 (0.01)	0.87 (0.01)	0.87 (0.01)	0.87 (0.01)	0.29
Insulin (pmol/l)	109.9 (2.6)	113.8 (2.9)	95.7 (9.7)	101.8 (9.0)	117.4 (24.0)	100.0 (11.0)	0.57
HOMA-IR	3.8 (0.1)	3.9 (0.1)	3.3 (0.3)	3.6 (0.4)	4.2 (1.0)	3.4 (0.4)	0.58
AMH (ng/ml)	8.3 (0.1)	8.6 (0.1)	6.4 (0.6)	8.0 (0.3)	4.9 (1.5)	7.8 (0.4)	0.24
SHBG (nmol/l)	31.1 (0.5)	31.3 (0.2)	60.6 (25)	31.6 (1.2)	41.9 (4.6)	30.4 (2.5)	0.07
LH (U/l)	11.3 (0.3)	11.6 (0.2)	9.1 (1.5)	10.6 (0.8)	8.9 (2.9)	11.2 (1.0)	0.67
FSH (U/l)	5.1 (0.1)	5.1 (0.06)	4.0 (0.6)	4.8 (0.3)	3.7 (1.0)	5.1 (0.3)	0.27
A4 (nmol/l)	6.9 (0.1)	6.9 (0.1)	6.4 (0.5)	6.8 (0.3)	5.7 (0.9)	7.1 (0.5)	0.36
T (nmol/l)	1.8 (0.03)	1.8 (0.03)	1.7 (0.2)	1.6 (0.07)	1.4 (0.2)	1.7 (0.1)	0.50
DHT (nmol/l)	0.35 (0.01)	0.33 (0.004)	0.34 (0.06)	0.31 (0.02)	0.27 (0.07)	0.37 (0.03)	0.33
DHEA (nmol/l)	21.1 (0.3)	20.6 (0.2)	18.9 (3.1)	25.3 (2.3)	26.9 (6.2)	20.7 (2.0)	0.13
DHEA-S (nmol/l)	5.4 (0.02)	5.4 (0.01)	5.1 (0.3)	5.8 (0.2)	5.9 (0.6)	5.9 (0.3)	0.06
11KT (nmol/l)	1.3 (0.02)	1.3 (0.02)	1.1 (0.2)	1.3 (0.1)	1.3 (0.2)	1.7 (0.4)	0.65
11OHA4 (nmol/l)	4.7 (0.03)	4.7 (0.01)	3.7 (0.5)	5.5 (0.4)	4.9 (0.5)	5.3 (0.4)	0.03
FAI	6.3 (0.1)	6.6 (0.1)	5.3 (0.7)	5.8 (0.3)	3.9 (0.8)	5.8 (0.6)	0.33

GEE: generalized estimating equations; HOMA-IR: homeostatic model assessment for insulin resistance; AMH: anti-Müllerian hormone; SHBG: sex hormone-binding globulin; A4: androstenedione; T: testosterone; DHT: dihydrotestosterone; DHEA: dehydroepiandrosterone; DHEA-S: dehydroepiandrosterone sulfate; 11KT: 11-ketotestosterone; 11OHA4: 11β-hydroxyandrostenedione; FAI: free androgen index. Data are presented as estimated marginal means (SE) from GEE analyses. The analyses are corrected for baseline, with time and group as the independent variables and an interaction term for time × group.

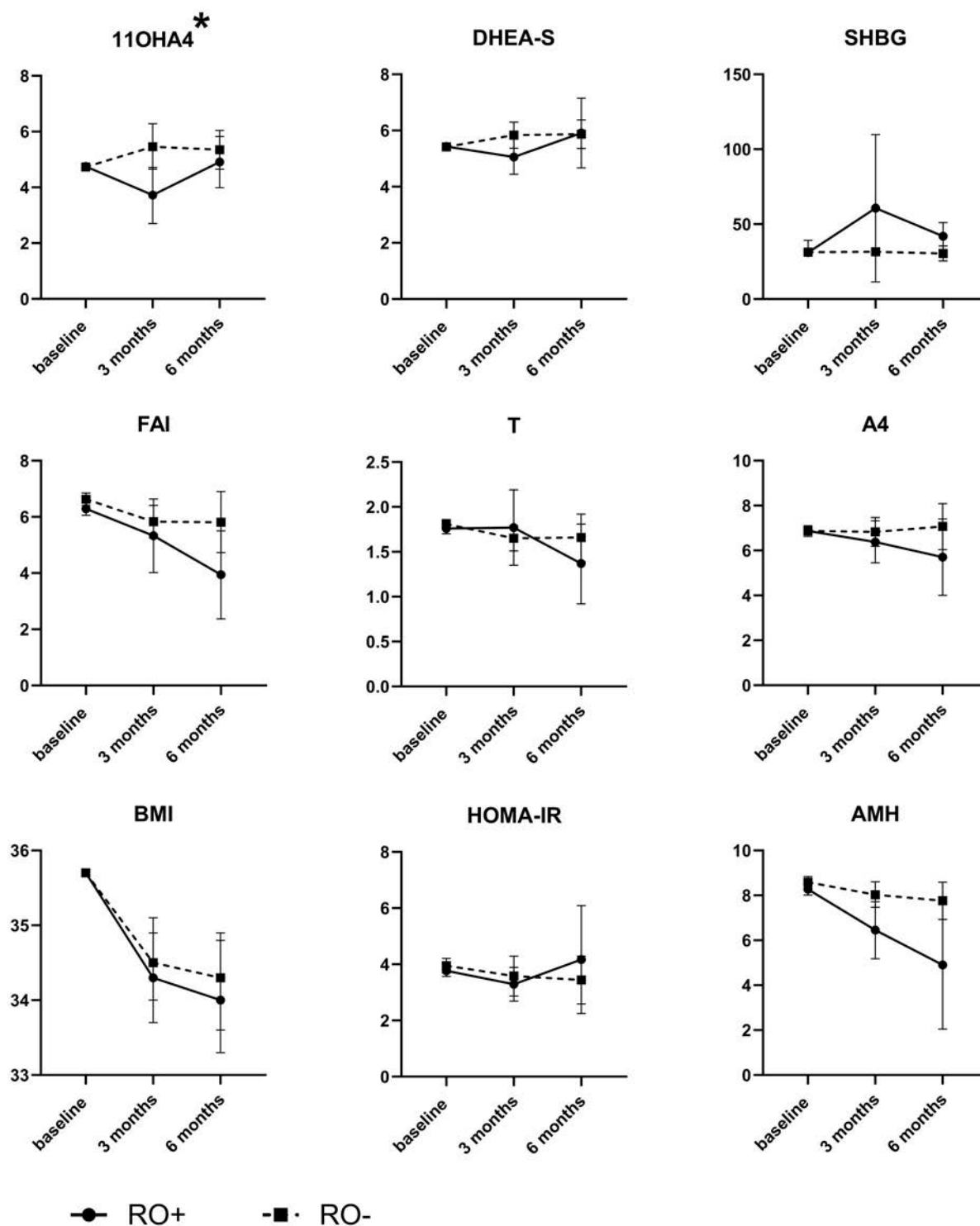


Figure 4. Baseline-corrected generalized estimating equations analysis of parameters in women who had or had not resumed ovulation after 3 and 6 months of lifestyle intervention. Data are from women who had (RO+) or had not (RO-) resumed ovulation. GEE: generalized estimating equations, AMH: anti-Müllerian hormone; HOMA-IR: homeostatic model assessment for insulin resistance; SHBG: sex hormone-binding globulin; T: testosterone; A4: androstenedione; FAI: free androgen index; 11OHA4: 11 β -hydroxyandrostenedione. Y-axes show GEE estimates. Error bars indicate standard errors of GEE estimates. As GEE was performed with correction for baseline, the values at baseline differ from those shown in Table 1. * $P < 0.05$ for interaction term group and time.

difference between the RO+ and RO- groups over the course of the intervention for body weight or BMI. This might be explained by the small sample size because of attrition of participants in our study. Additionally, both insulin concentrations and insulin resistance represented by HOMA-IR did not show significant differences in RO+ and RO- women. Possible explanations might

include the limited statistical power of the current study, the high average BMI within our population ($\text{BMI} > 35 \text{ kg/m}^2$), or the limited degree of weight reduction. It is important to note that moderate weight loss in women with a high BMI does not consistently lead to an improvement in insulin resistance (Tang et al., 2006).

Table 3. Estimated marginal means for anthropometric measurements and endocrine and metabolic parameters at baseline, 3 months, and 6 months after starting lifestyle intervention, stratified by group.

	RO+					RO–				
	Baseline	3 months	P-value	6 months	P-value	Baseline	3 months	P-value	6 months	P-value
Weight (kg)	101.2 (1.6)	97.0 (1.9)	<0.001	96.1 (1.9)	<0.001	105.0 (1.9)	101.7 (1.9)	<0.001	100.4	<0.001
BMI (kg/m ²)	35.5 (0.5)	34.1 (0.6)	<0.001	33.7 (0.6)	<0.001	36.0 (0.5)	34.8 (0.5)	<0.001	34.5 (0.5)	<0.001
Waist circumference (cm)	108.7 (1.3)	104.0 (2.0)	0.03	104.1 (1.9)	0.03	108.2 (1.3)	105.4 (1.3)	0.001	103.8 (1.4)	<0.001
Hip circumference (cm)	122.1 (1.4)	119.8 (2.1)	0.18	118.6 (2.0)	0.04	124.3 (1.1)	122.3 (1.3)	0.02	120.2 (1.2)	<0.001
Waist-hip circumference ratio	0.89 (0.01)	0.86 (0.02)	0.04	0.89 (0.01)	0.77	0.87 (0.01)	0.86 (0.01)	0.27	0.86 (0.01)	0.44
Insulin (pmol/l)	103.1 (8.5)	85.2 (11.8)	0.09	106.3 (22.8)	0.90	116.7 (9.5)	103.0 (10.8)	0.16	103.1 (9.3)	0.23
HOMA-IR	3.5 (0.3)	2.8 (0.4)	0.045	3.7 (0.9)	0.85	4.0 (0.4)	3.6 (0.4)	0.26	3.5 (0.3)	0.24
AMH (ng/ml)	6.7 (1.0)	4.7 (0.9)	0.01	3.0 (1.2)	0.02	8.8 (0.9)	8.3 (0.8)	0.12	7.9 (0.8)	0.08
SHBG (nmol/l)	33.4 (3.9)	62.2 (27.2)	0.28	45.7 (9.2)	0.14	30.5 (2.4)	30.5 (2.1)	0.98	31.0 (2.4)	0.82
LH (U/l)	10.6 (1.0)	9.1 (1.8)	0.37	8.9 (2.5)	0.53	11.7 (0.7)	10.7 (0.8)	0.20	11.1 (0.9)	0.55
FSH (U/l)	5.0 (0.4)	3.7 (0.7)	0.08	3.7 (0.8)	0.19	5.3 (0.2)	4.9 (0.3)	0.19	5.1 (0.3)	0.53
A4 (nmol/l)	6.8 (0.5)	6.4 (0.4)	0.43	5.8 (0.8)	0.27	6.8 (0.3)	6.7 (0.4)	0.84	7.0 (0.5)	0.63
T (nmol/l)	1.6 (0.1)	1.6 (0.2)	0.99	1.3 (0.2)	0.26	1.8 (0.1)	1.6 (0.1)	0.01	1.7 (0.1)	0.26
DHT (nmol/l)	0.4 (0.05)	0.4 (0.06)	0.73	0.3 (0.05)	0.08	0.3 (0.02)	0.3 (0.03)	0.33	0.4 (0.03)	0.17
DHEA (nmol/l)	23.3 (2.1)	20.9 (3.2)	0.48	27.5 (7.0)	0.48	20.0 (1.6)	25.1 (2.8)	0.04	20.6 (2.2)	0.75
DHEA-S (nmol/l)	6.0 (0.6)	5.5 (0.6)	0.17	6.3 (0.8)	0.71	5.2 (0.3)	5.7 (0.4)	0.06	5.7 (0.4)	0.08
11KT (nmol/l)	1.3 (0.1)	1.1 (0.2)	0.26	1.3 (0.2)	0.93	1.3 (0.08)	1.3 (0.1)	0.90	1.6 (0.4)	0.48
11OHA4 (nmol/l)	5.1 (0.5)	4.3 (0.6)	0.14	5.3 (0.8)	0.74	4.7 (0.3)	5.6 (0.5)	0.05	5.2 (0.4)	0.24
FAI	5.5 (0.5)	4.8 (0.7)	0.3	3.5 (0.5)	0.003	7.2 (0.6)	6.0 (0.4)	0.007	5.9 (0.5)	0.049

RO+: resumed ovulation at the end of 6 months; RO: remained anovulatory at the end of 6 months; HOMA-IR: homeostatic model assessment for insulin resistance; AMH: anti-Müllerian hormone; SHBG: sex hormone-binding globulin; A4: androstenedione; T: testosterone; DHT: dihydrotestosterone; DHEA: dehydroepiandrosterone; DHEA-S: dehydroepiandrosterone sulfate; 11KT: 11-ketotestosterone; 11OHA4: 11 β -hydroxyandrostenedione; FAI: free androgen index. Data are presented as estimated marginal means (SE) from generalized estimating equations analyses. P-values refer to change from baseline at 3 months and at 6 months in RO+ and RO– women, respectively.

Hyperandrogenism is one of the most dominant features of PCOS (Azziz *et al.*, 2006). However, biochemical hyperandrogenism remains a diagnostic challenge and recent research advocates the quantification of several steroid hormones simultaneously using LC-MS/MS (Keevil, 2014, 2019). For instance, it has been demonstrated that 11-oxygenated androgens represent the majority of circulating androgens in women with PCOS. Consequently, focus has recently shifted towards including these 11-oxygenated C19 steroids in biochemical hyperandrogenism analysis, although definitive conclusions have not been reached (O'Reilly *et al.*, 2017; Hawley *et al.*, 2020). Therefore, there is room for discussion as to which androgen or combination of androgens is a better marker associated with restored ovulation in anovulatory women with PCOS and obesity after weight loss. Weight loss results in decreased androgen production in the ovarian theca cells owing to a reduction in insulin as co-factor to LH. The combination of reduced insulin and androgens concentrations consequently leads to more SHBG production in the liver (Poretsky, 1991). In our study, 11OHA4 was the only parameter that showed a statistically significant difference between the RO+ and RO– groups over the course of the intervention. Regarding 11OHA4, as one of the most abundant unconjugated androgens produced by the human adrenals, our results imply a contribution of adrenal hormones in this group of women with obesity and PCOS. We also found a trend towards significant differences in SHBG and DHEA-S between two groups. This is somewhat in line with a previous study in women with PCOS undergoing a weight loss program (Guzick *et al.*, 1994). Moreover, FAI in both groups decreased over the course of the intervention period although no significant differences between groups were found. All these results suggest that there is an important role of androgen concentrations, especially 11OHA4, in restoring ovulation in women with PCOS but this needs to be explored in larger studies in the future.

Furthermore, we did not observe a significant interaction between time and group for serum AMH. Serum AMH is considered a stable marker of the oocyte pool with limited variation throughout the menstrual cycle (Broer *et al.*, 2014). Data regarding the association between AMH changes and resumption of ovulation

after weight loss intervention in women with PCOS and obesity are limited. Thomson *et al.* (2009) has demonstrated that in a 20-week weight loss program (a hypocaloric dietary with 1430kcal per day) in 52 women, changes in AMH concentrations did not significantly differ for women who responded with improved reproductive function (identified as having either improvement in ovulatory function and/or improvements in menstrual status) and those who did not. There are some differences between our studies: First, inconsistencies exist in the measurement of ovulation across studies. In our study, we solely used resumption of ovulation as the indicator, whereas in their study, either improvement in ovulatory function and/or improvements in menstrual status were considered. Second, is heterogeneity in lifestyle interventions. Indeed, there is evidence from one study suggesting that the response of serum AMH concentrations to interventions, such as diet, exercise, and a combination of diet and exercise, may vary in women with PCOS and obesity. Specifically, serum AMH concentrations were significantly decreased only in the diet group, but not in the exercise group or the combined intervention group (Nybacka *et al.*, 2013). Third, there is heterogeneity in the diagnosis of PCOS. Although both studies used the Rotterdam criteria to diagnose PCOS and baseline body weight and age were similar, potential differences in other characteristics cannot be ruled out as PCOS is such a multi-phenotype syndrome. Fourth, there were differences in methods of data analysis. Thomson *et al.* (2009) calculated the changes of AMH concentrations before and after the intervention and compared changes in concentrations directly between two groups, while we compared the concentrations during and after the intervention with the correction for baseline measurements. Notably, we revealed a significant reduction in serum AMH concentrations among RO+ women at 3 and 6 months compared to baseline measures while this was not observed in RO– women. It is reasonable to assume that no significant differences in AMH concentrations over time were detected between the two groups, possibly because of the small sample size. We clearly need more studies to investigate the changes in AMH concentrations and their putative association with the resumption of ovulatory

cycles following weight loss in a range of PCOS phenotypes (Azziz et al., 2006). However, we can get a glimpse of possible results by looking at the existing research on bariatric surgery, which yields greater weight loss than lifestyle intervention. Bariatric surgery has been shown to reduce AMH concentrations in 14 women with PCOS and is associated with weight loss (Chiofalo et al., 2017). Importantly, both the Thomson et al. (2009) study and our study showed lower AMH concentrations at baseline in women with improved ovulation status than non-responders. This might suggest, hypothetically, that anovulatory women with higher baseline AMH concentrations would require greater weight loss in order to decrease AMH concentrations and facilitate the restoration of ovulation.

The mechanism of resumption of ovulation in women with obesity and PCOS after weight loss is far from clear. Our results suggest an important role for reduction of androgens concentrations, especially 11-oxygenated androgens, in the resumption of ovulation. With the conduct of larger-scale studies, this can presumably provide significant clinical guidance for anovulatory women with PCOS, i.e., monitoring androgens concentrations to provide individualized recommendations for lifestyle interventions (amount of weight loss) before ovulatory cycles resume.

The strength of this study is the use of well-characterized participants of an RCT, close follow-up of participants during and after the lifestyle program, and the measurement of a wide range of steroid hormones, including adrenal oxygenated steroids using LC-MS/MS methods. However, several limitations in this study need to be mentioned. First, no direct observation of follicular growth and rupture by high-resolution transvaginal ultrasonography was performed to confirm ovulations for all women with PCOS at the end of lifestyle program. We did not evaluate ovulation status in the third month after randomization thus we were unable to investigate the association over time between resumption of ovulation and weight loss or insulin resistance improvement at 3 months. Second, the number of available serum hormone measurements decreased over the course of the intervention period owing to the nature of the RCT with dropouts during the intervention, pregnancies, missed hospital visits, or blood samples being exhausted for other measurements. The remaining relatively small sample size might produce a false negative result. Third, we could not perform subgroup analyses to explore the optimal degree of weight loss required to restore ovulatory function in PCOS owing to the small sample size. The nature of the study design also did not allow us to explore the causal relation between lower androgen concentrations and ovulation resumption. Last, we cannot divide the subtypes of PCOS (Azziz et al., 2006) in this analysis and our results cannot be generalized to other populations of anovulatory PCOS women without obesity.

In conclusion, in this *post hoc* exploratory analysis, we showed that 11OHA4 was the only parameter showing a statistically significant difference between the RO+ and RO- groups over the course of the intervention, although a similar trend was observed for DHEA-S and SHBG. We have also observed that RO+ women had lower concentrations of AMH than RO- women at baseline. A significant reduction in serum AMH concentrations was shown among RO+ women at 3 and 6 months compared to baseline measures while this was not observed in RO- women. In the future, if our results can be confirmed in other studies, monitoring of androgen concentrations during lifestyle intervention may

help to provide individualized recommendations on resumption of ovulatory cycles in anovulatory women with PCOS and obesity.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Authors' roles

Z.W., M.V.F., and A.V.d.V. drafted the manuscript. J.M.H. and B.G. K. performed the measurement of androgens. Z.W. and H.G. carried out the statistical analyses. A.H., I.P.K., and A.E.P.C. designed the study. All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors revised and approved the final manuscript. All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

This work was supported by ZonMw, the Dutch Organization for Health Research and Development (50-50110-96-518). ZonMw had no role in data collection, analysis, interpretation of data or writing the report. The Department of Obstetrics and Gynecology of the University Medical Centre Groningen received an unrestricted educational grant from Ferring Pharmaceuticals BV, The Netherlands.

Conflict of interest

A.H. reports consultancy for development and implementation of a lifestyle App MyFertiCoach developed by Ferring Pharmaceutical Company.

References

- Anagnostis P, Tarlatzis BC, Kauffman RP. Polycystic ovarian syndrome (PCOS): long-term metabolic consequences. *Metabolism* 2018;**86**:33–43.
- Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE et al. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab* 2006;**91**:4237–4245.
- Baptiste CG, Battista MC, Trottier A, Baillargeon JP. Insulin and hyperandrogenism in women with polycystic ovary syndrome. *J Steroid Biochem Mol Biol* 2010;**122**:42–52.
- Broer SL, Broekmans FJ, Laven JS, Fauser BC. Anti-Müllerian hormone: ovarian reserve testing and its potential clinical implications. *Hum Reprod Update* 2014;**20**:688–701.
- Chiofalo F, Ciuoli C, Formichi C, Selmi F, Forleo R, Neri O, Vuolo G, Paffetti P, Pacini F. Bariatric surgery reduces serum anti-

- Müllerian hormone levels in obese women with and without polycystic ovarian syndrome. *Obes Surg* 2017;**27**:1750–1754.
- Diamanti-Kandarakis E, Dunaif A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocr Rev* 2012;**33**:981–1030.
- Dumont A, Robin G, Catteau-Jonard S, Dewailly D. Role of anti-Müllerian hormone in pathophysiology, diagnosis and treatment of polycystic ovary syndrome: a review. *Reprod Biol Endocrinol* 2015;**13**:137.
- Escobar-Morreale HF, Botella-Carretero JJ, Alvarez-Blasco F, Sancho J, San Millán JL. The polycystic ovary syndrome associated with morbid obesity may resolve after weight loss induced by bariatric surgery. *J Clin Endocrinol Metab* 2005;**90**:6364–6369.
- Franks S, Hardy K. What causes anovulation in polycystic ovary syndrome? *Curr Opin Endocr Metab Res* 2020;**12**:59–65.
- Guzick DS, Wing R, Smith D, Berga SL, Winters SJ. Endocrine consequences of weight loss in obese, hyperandrogenic, anovulatory women. *Fertil Steril* 1994;**61**:598–604.
- Hawley JM, Adaway JE, Owen LJ, Keevil BG. Development of a total serum testosterone, androstenedione, 17-hydroxyprogesterone, 11 β -hydroxyandrostenedione and 11-ketotestosterone LC-MS/MS assay and its application to evaluate pre-analytical sample stability. *Clin Chem Lab Med* 2020;**58**:741–752.
- Jarrett BY, Lujan ME. Impact of hypocaloric dietary intervention on ovulation in obese women with PCOS. *Reproduction* 2016;**153**:R15–R27.
- Keevil B. Steroid mass spectrometry for the diagnosis of PCOS. *Med Sci (Basel)* 2019;**7**:78.
- Keevil BG. How do we measure hyperandrogenemia in patients with PCOS? *J Clin Endocrinol Metab* 2014;**99**:777–779.
- Kuchenbecker WKH, Groen H, van Asselt SJ, Bolster JHT, Zwerver J, Slart RHJ, Vd Jagt EJ, Muller Kobold AC, Wolffenbuttel BHR, Land JA et al. In women with polycystic ovary syndrome and obesity, loss of intra-abdominal fat is associated with resumption of ovulation. *Hum Reprod* 2011;**26**:2505–2512.
- Lebbe M, Woodruff TK. Involvement of androgens in ovarian health and disease. *Mol Hum Reprod* 2013;**19**:828–837.
- Li J, Papadopoulos V, Vihma V. Steroid biosynthesis in adipose tissue. *Steroids* 2015;**103**:89–104.
- Lim SS, Davies MJ, Norman RJ, Moran LJ. Overweight, obesity and central obesity in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update* 2012;**18**:618–637.
- Lim SS, Hutchison SK, Van Ryswyk E, Norman RJ, Teede HJ, Moran LJ; Cochrane Gynaecology and Fertility Group. Lifestyle changes in women with polycystic ovary syndrome. *Cochrane Database Syst Rev* 2019;**3**:CD007506.
- Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985;**28**:412–419.
- Moolhuijsen LME, Visser JA. Anti-Müllerian hormone and ovarian reserve: update on assessing ovarian function. *J Clin Endocrinol Metab* 2020;**105**:3361–3373.
- Moran LJ, Noakes M, Clifton PM, Norman RJ. The use of anti-müllerian hormone in predicting menstrual response after weight loss in overweight women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2007;**92**:3796–3802.
- Mutsaerts MA, van Oers AM, Groen H, Burggraaf JM, Kuchenbecker WK, Perquin DA, Koks CA, van Golde R, Kaaijk EM, Schierbeek JM et al. Randomized trial of a lifestyle program in obese infertile women. *N Engl J Med* 2016;**374**:1942–1953.
- Mutsaerts MAQ, Groen H, ter Bogt NCW, Bolster JHT, Land JA, Bemelmans WJE, Kuchenbecker WKH, Hompes PGA, Macklon NS, Stolk RP et al. The LIFESTYLE study: costs and effects of a structured lifestyle program in overweight and obese subfertile women to reduce the need for fertility treatment and improve reproductive outcome. A randomised controlled trial. *BMC Womens Health* 2010;**10**:22–22.
- Norman RJ, Noakes M, Wu R, Davies MJ, Moran L, Wang JX. Improving reproductive performance in overweight/obese women with effective weight management. *Hum Reprod Update* 2004;**10**:267–280.
- Nybacka Å, Carlström K, Fabri F, Hellström PM, Hirschberg AL. Serum antimüllerian hormone in response to dietary management and/or physical exercise in overweight/obese women with polycystic ovary syndrome: secondary analysis of a randomized controlled trial. *Fertil Steril* 2013;**100**:1096–1102.
- O'Reilly MW, Kempegowda P, Jenkinson C, Taylor AE, Quanson JL, Storbeck KH, Arlt W. 11-Oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. *J Clin Endocrinol Metab* 2017;**102**:840–848.
- Pasquali R, Gambineri A, Pagotto U. The impact of obesity on reproduction in women with polycystic ovary syndrome. *BJOG* 2006;**113**:1148–1159.
- Poretsky L. On the paradox of insulin-induced hyperandrogenism in insulin-resistant states. *Endocr Rev* 1991;**12**:3–13.
- Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod* 2004;**19**:41–47.
- Tang T, Glanville J, Hayden CJ, White D, Barth JH, Balen AH. Combined lifestyle modification and metformin in obese patients with polycystic ovary syndrome. A randomized, placebo-controlled, double-blind multicentre study. *Hum Reprod* 2006;**21**:80–89.
- Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ; International PCOS Network International PN. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertil Steril* 2018;**110**:364–379.
- Thomson RL, Buckley JD, Moran LJ, Noakes M, Clifton PM, Norman RJ, Brinkworth GD. The effect of weight loss on anti-Müllerian hormone levels in overweight and obese women with polycystic ovary syndrome and reproductive impairment. *Hum Reprod* 2009;**24**:1976–1981.
- Torchen LC, Sisk R, Legro RS, Turcu AF, Auchus RJ, Dunaif A. 11-Oxygenated C19 steroids do not distinguish the hyperandrogenic phenotype of PCOS daughters from girls with obesity. *J Clin Endocrinol Metab* 2020;**105**:e3903–e3909.
- van Dam EW, Roelfsema F, Veldhuis JD, Hogendoorn S, Westenberg J, Helmerhorst FM, Frölich M, Krans HM, Meinders AE, Pijl H. Retention of estradiol negative feedback relationship to LH predicts ovulation in response to caloric restriction and weight loss in obese patients with polycystic ovary syndrome. *Am J Physiol Endocrinol Metab* 2004;**286**:E615–E620.
- van Dammen L, Wekker V, van Oers AM, Mutsaerts MAQ, Painter RC, Zwinderman AH, Groen H, van de Beek C, Muller Kobold AC, Kuchenbecker WKH et al.; LIFESTYLE study group. Effect of a lifestyle intervention in obese infertile women on cardiometabolic health and quality of life: a randomized controlled trial. *PLoS One* 2018;**13**:e0190662.
- van der Veen A, van Faassen M, de Jong WHA, van Beek AP, Dijk-Brouwer DAJ, Kema IP. Development and validation of a LC-MS/MS method for the establishment of reference intervals and

- biological variation for five plasma steroid hormones. *Clin Biochem* 2019;**68**:15–23.
- van Oers AM, Groen H, Mutsaerts MA, Burggraaff JM, Kuchenbecker WK, Perquin DA, Koks CA, van Golde R, Kaaijk EM, Schierbeek JM et al.; LIFEstyle study group. Effectiveness of lifestyle intervention in subgroups of obese infertile women: a subgroup analysis of a RCT. *Hum Reprod* 2016;**31**:2704–2713.
- Zhu JL, Chen Z, Feng WJ, Long SL, Mo ZC. Sex hormone-binding globulin and polycystic ovary syndrome. *Clin Chim Acta* 2019;**499**:142–148.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Human Reproduction, 2024, 39, 1078–1088
<https://doi.org/10.1093/humrep/deae058>
Original Article

Reproductive endocrinology

Natural proliferative phase frozen embryo transfer—a new approach which may facilitate scheduling without hindering pregnancy outcomes

Catarina Mendes Godinho ^{1,*}, Sérgio Reis Soares ¹, Sofia Gouveia Nunes¹, Juan M Mascarós Martínez², and Samuel Santos-Ribeiro ^{1,3}

¹Reproductive Medicine, IVI RMA Lisbon, Lisbon, Portugal

²IVI Foundation, Valencia, Spain

³Department of Obstetrics and Gynaecology, Faculty of Medicine, University of Lisbon, Lisbon, Portugal

*Correspondence address. Clínica IVI Lisboa, Avenida Infante D. Henrique, 333H esc 1-8, 1800-282 Lisboa, Portugal. E-mail: catarina.godinho@ivirma.com

 <https://orcid.org/0000-0001-7045-9237>

ABSTRACT

STUDY QUESTION: How does a natural proliferative phase (NPP) strategy for frozen embryo transfer (FET) compare with the conventional artificial (AC) and natural (NC) endometrial preparation protocols in terms of live birth rates (LBR)?

SUMMARY ANSWER: This study supports the hypothesis that, just as for NC, NPP-FET may be a superior alternative to AC in terms of LBR.

WHAT IS KNOWN ALREADY: Although FETs are increasing worldwide, the optimal FET protocol is still largely controversial. Despite recent evidence supporting a possibly higher efficacy and safety of NC FETs, their widespread use is limited by the difficulties encountered during cycle monitoring and scheduling.

STUDY DESIGN, SIZE, DURATION: In this single center retrospective cohort study, we describe the NPP-FET protocol, in which vaginal progesterone is initiated during the proliferative phase as soon as an endometrium with a thickness of at least 7 mm is identified and ovulation is ruled out, regardless of mean diameter of the dominant follicle.

PARTICIPANTS/MATERIALS, SETTING, METHODS: For comparison, we considered all blastocyst stage FET cycles preformed at a private infertility center between January 2010 and June 2022, subdivided according to the following subgroups of endometrial preparation: AC, NPP, and NC. We performed multivariable generalized estimating equations regression analysis to account for the following potential confounding variables: oocyte age at retrieval, oocyte source (autologous without preimplantation genetic testing for aneuploidies (PGT-A) versus autologous with PGT-A versus donated), number of oocytes retrieved/donated, embryo developmental stage (Day 5 versus Day 6), number of embryos transferred, quality of the best embryo transferred, and year of treatment. The main outcome measure was LBR. The secondary outcomes included hCG positive, clinical pregnancy and miscarriage rates, and the following perinatal outcomes: first trimester bleeding, second/third trimester bleeding, preterm rupture of membranes, gestational diabetes, gestational hypertensive disorders (GHD), and gestational age at delivery.

MAIN RESULTS AND THE ROLE OF CHANCE: A total of 5791 FET cycles were included in this analysis (2226 AC, 349 NPP, and 3216 NC). The LBR for FET was lower in the AC subgroup when compared to the NPP and NC (38.4%, 49.1%, and 45.2%, respectively; $P < 0.01$ AC versus NPP and AC versus NC). The rates of miscarriage were also lower in the NPP and NC subgroups when compared to AC (19.7%, 25.0%, and 34.9%, respectively; $P < 0.01$ NPP versus AC and NC versus AC). Considering perinatal outcomes, NPP-FET and NC were associated with a significantly lower first trimester bleeding compared to AC (17.3%, 14.7%, and 37.6%, respectively; $P < 0.01$ NPP versus AC and NC versus AC). Additionally, NC was associated with a lower rate of GHD when compared with AC (8.6% versus 14.5%, $P < 0.01$), while the rate following NPP-FET was 9.4%.

LIMITATIONS, REASONS FOR CAUTION: This study is limited by its retrospective design. Moreover, there was also a low number of patients in the NPP subgroup, which may have led the study to be underpowered to detect clinically relevant differences between the subgroups.

WIDER IMPLICATIONS OF THE FINDINGS: Our study posits that the NPP-FET protocol may be an effective and safe alternative to both NC and AC, while still allowing for enhanced practicality in patient follow-up and FET scheduling. Further investigation on NPP-FET is warranted, with prospective studies including a larger and more homogeneous subsets of patients.

STUDY FUNDING/COMPETING INTEREST(S): This research was supported by the IVI-RMA-Lisbon (2008-LIS-053-CG). The authors did not receive any funding for this study. The authors have no competing interests.

TRIAL REGISTRATION NUMBER: Not applicable.

Keywords: natural cycle / artificial cycle / frozen embryo transfer / gestational hypertensive disorders / live birth rate / miscarriage

Introduction

Frozen embryo transfer (FET) is increasing worldwide (De Geyter et al., 2018), not only owing to its ever-mounting list of indications, but also to the continuous optimization of the cryopreservation techniques (Loutradi et al., 2008) and increased reassurance regarding both neonatal safety (Zhang et al., 2018) and subsequent risk for congenital abnormality (Belva et al., 2016). Nonetheless, further knowledge regarding the different FET protocols available is still of paramount importance, not only regarding the practicality of their follow-up, but also in terms of their efficacy and safety for both the mother and newborn (Jing et al., 2019).

Thus far, the FET endometrial preparation protocols used most often have been the artificial and the natural cycle (AC and NC, respectively). ACs entail the use of estrogen and progesterone (P4) supplementation for endometrial preparation, thus replacing endogenous hormone production. The AC advantages—and the reasons for its widespread use—are the requirement of minimal monitoring and the flexibility in terms of FET scheduling. Although this approach has been proven on multiple occasions to be effective (Younis et al., 1996; Groenewoud et al., 2016; Mumusoglu et al., 2021), there are some drawbacks with such a strategy. Specifically, the medication used entails a higher cost and, also, the potential long-term estrogen supplementation is not free of potential risks regarding the health of the patient (Maas et al., 2021; Dalsgaard et al., 2022). Moreover, previous studies have shown that AC-FET may be associated with lower implantation (Altmäe et al., 2016; Mumusoglu et al., 2021) and higher miscarriage rates (Veleva et al., 2008; Tomás et al., 2012; Mumusoglu et al., 2021), although others have not confirmed these findings (Groenewoud et al., 2013, 2016). Equally important, there is also recent evidence suggesting that the AC may circumvent the physiological vascular adaptation occurring during pregnancy and ultimately increase the risk of gestational hypertensive disorders (GHD) (Ginström et al., 2019; Chih et al., 2021; Hu et al., 2021) and caesarean section delivery (Jing et al., 2019; Hu et al., 2021).

In NCs, the timing of FET is determined by the timing of ovulation. In true NCs, the endometrium proliferates along with follicular growth until natural ovulation occurs with minimal medical intervention. Transvaginal ultrasonographic monitoring is started on Day 8–10 of the cycle and serial endocrine profiling, using serum estradiol (E2), LH and P4 measurements, is frequently performed, at least when the leading follicle reaches a mean diameter of approximately 16–17 mm. Ovulation in our setting is usually considered to occur a day after LH ≥ 18 IU/l, especially when accompanied with a concomitant rise in serum P4 (>1.5 ng/ml). This phenomenon is closely monitored by often exhaustive consecutive transvaginal ultrasounds and hormone assays (Casper and Yanushpolsky, 2016). Meanwhile, in the case of the modified NC, once adequate follicular and endometrial growth are documented, hCG is administered to promote ovulation and subsequent luteinization. When compared to ACs, both these approaches have the disadvantage of potentially requiring more visits to the clinic to determine the ideal timing for FET, and a decreased flexibility for scheduling of FET (Mumusoglu et al., 2021). Moreover, the proper timing of ovulation may prove to be difficult to detect especially owing to the inconsistency of definitions of the LH surge currently in use (Testart et al., 1981; Groenewoud et al., 2017). This fact, associated with the risk of a non-ovulatory cycle, leads to an increased risk of cycle cancellation when the NC protocol is used (Sathanandan et al., 1991; Groenewoud et al., 2017). Although the administration of hCG in the modified NC generally overcomes the difficulties related to

identification of the LH surge, it implies an extra cost and has been associated with conflicting data regarding clinical success (Fatemi et al., 2010; Chang et al., 2011; Tomás et al., 2012; Montagut et al., 2016; Huberlant et al., 2018), possibly because of an embryo-endometrium asynchrony generated when hCG is administered concomitantly with an unidentified spontaneous LH surge (Fatemi et al., 2010). Therefore, there is still a lively debate on whether FETs occurring under the more physiologically prepared endometrium in the NC might result in higher pregnancy rates (Shapiro et al., 2011; Roque et al., 2013; Mumusoglu et al., 2021) and decreased maternal and neonatal morbidity (Evans et al., 2014; Ishihara et al., 2014; Asserhøj et al., 2021; Hu et al., 2021).

Our center has developed an alternative approach for FET endometrial preparation, which we coined the natural proliferative phase FET (NPP-FET). In this protocol, vaginal progesterone is initiated during a natural menstrual cycle once a proliferative endometrium with a thickness of at least 7 mm is identified and a dominant follicle is observed, regardless of its mean diameter. The FET may be scheduled according to patient and clinic preference, so long as ovulation is excluded and hCG triggering is not performed. A similar protocol has been previously described in a proof of concept study in which clinical pregnancy rates did not seem to be compromised (42.9%), providing flexibility regarding the day of the transfer (Weiss et al., 2021). This strategy, indicated for patients with regular ovulatory cycles, has the potential to combine the advantages of both the AC (regarding simplicity and flexibility during monitoring and scheduling) and NC (namely the purported optimization regarding implantation and feto-maternal safety ascribed to a more physiological endometrial proliferation). The aim of this study was to describe this new FET strategy further and to assess its efficacy and safety in comparison to both ACs and NCs.

Materials and methods

Population

This was a retrospective single center cohort study of blastocyst stage FET cycle attempts performed at IVI Lisboa between January 2010 and June 2022, subdivided into the following three subgroups: conventional AC preparation, those performed using NPP-FET, and those following a NC (Fig. 1). Subjects included underwent at least one FET at blastocyst stage with a maximum of two embryos transferred. Those who used exogenous ovarian stimulation, that had an endometrium thickness of less than 7 mm, or with known untreated uterine abnormalities (e.g. polyps, submucous myomas, and severe adenomyosis) at the time of transfer were excluded (Fig. 2). This study was approved by the ethics committee of IVI Lisboa prior to commencing to analysis (study code 2008-LIS-053-CG). All subjects signed a legally required informed consent form for cryopreserved embryo transfer that includes the possibility to use, in scientific works, data and results from their Assisted Reproductive Technology treatments, in a completely anonymous regimen.

IVF

All embryos were obtained after IVF using either autologous or donated oocytes. Female patients or donors underwent exogenous ovarian stimulation followed by transvaginal ultrasound-guided oocyte retrieval under superficial general anesthesia (anesthetic conscious sedation), as previously described (Cozzolino et al., 2021). Fertilization was performed *in vitro* by either conventional IVF or ICSI, according to sperm quality, followed by fertilization and embryo development assessment

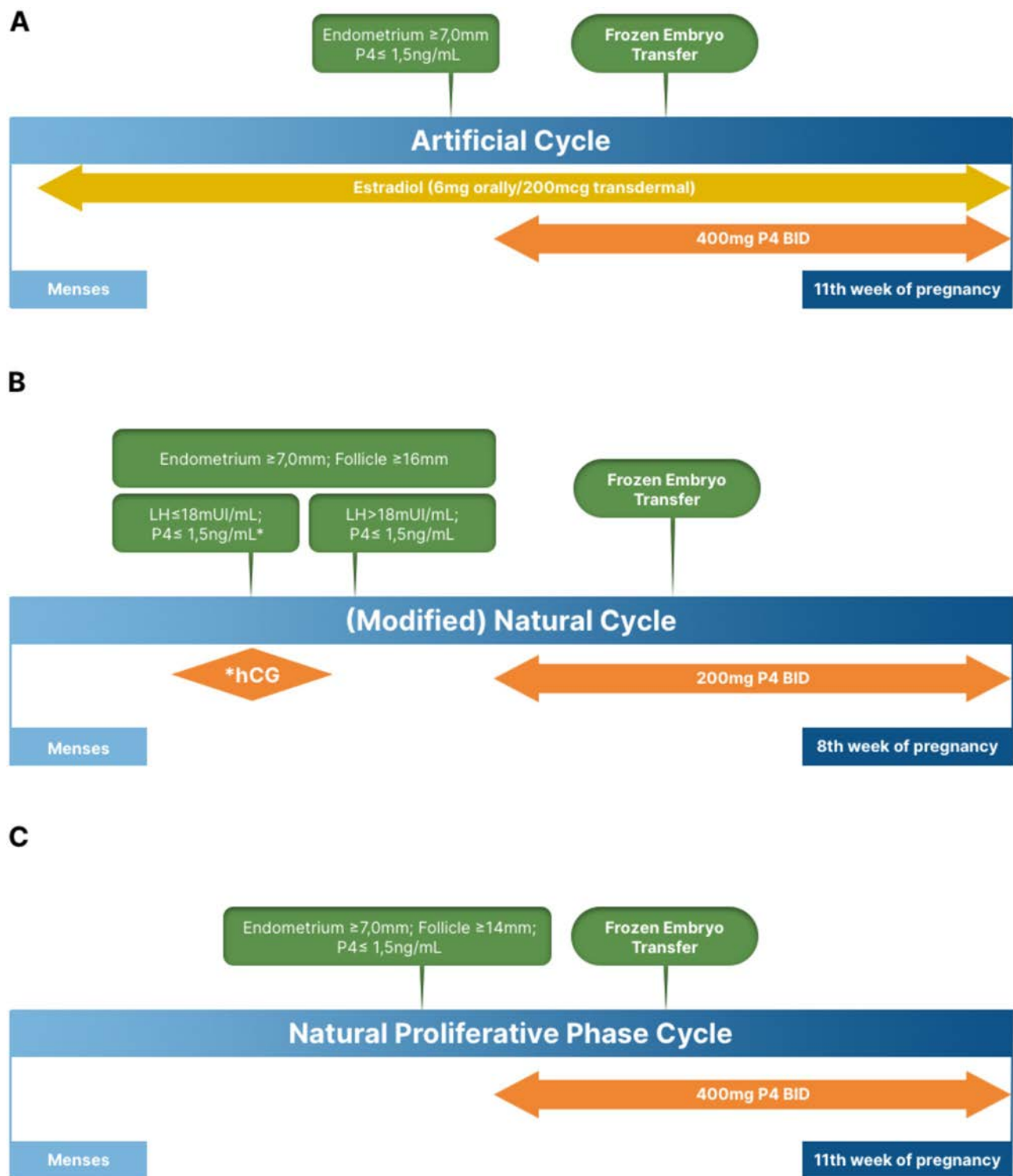


Figure 1. Summary of the different endometrial preparations for frozen embryo transfer included in the study. (A) Artificial cycle, (B) natural cycle, and (C) natural proliferative phase cycle. P4, progesterone; BID, twice per day.

according to our center's routine practice. All good quality embryos were selected for vitrification according to the Spanish Association for the Study of Reproductive Biology (ASEBIR) scoring classification system (ASEBIR, 2008). The embryo vitrification conditions have been described elsewhere (Martínez-Burgos *et al.*, 2011). In our center, until 2014, embryos were mostly transferred and cryopreserved at cleavage stage, unless preimplantation genetic testing for aneuploidies (PGT-A) was performed, in which case embryos were cryopreserved at blastocyst stage. Between

2014 and 2016, if there were more than three embryos available at Day 3, all remaining embryos following transfer were cryopreserved at the blastocyst stage. From 2016 onwards, all transfers were generally performed at the blastocyst stage and couples were routinely advised to perform single embryo transfer.

Endometrial preparation protocols

The different endometrial preparations for FET are summarized in Fig. 1. ACs were initiated with either transdermal E2 (200 μg

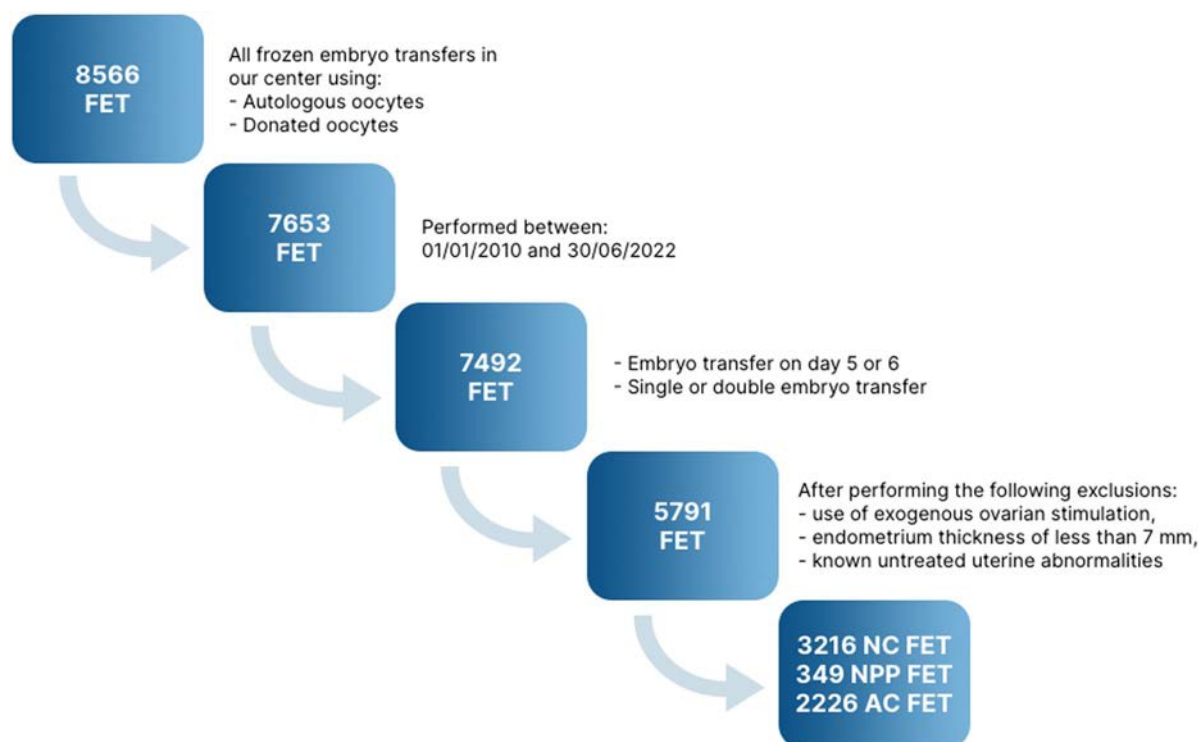


Figure 2. Flowchart of the frozen embryo transfers in the study. NC, natural cycle; NPP, natural proliferative cycle; AC, artificial cycle FET: frozen embryo transfer.

every other day) or oral E2 (2 mg every 8 h) from the second day of the menstrual cycle onwards. Endometrial thickness and echogenicity were evaluated by vaginal ultrasound between the 10th and the 14th day of preparation. Whenever the endometrium measured above 7 mm and serum P4 was < 1.5 ng/ml, vaginal micronized P4 was initiated using a dose of 400 mg every 12 h, maintaining E2 supplementation.

For NC FET preparation, endometrial and dominant follicle growth were tracked generally from the eighth day of the cycle onwards by serial transvaginal ultrasound. When the mean dominant follicle diameter reached at least 17 mm and the endometrial thickness was above 7 mm, serum E2, P4, and LH were evaluated. Ovulation was considered imminent whenever a LH peak of >18 mIU/ml was detected, as described previously (Frydman et al, 1982), with vaginal micronized P4 being started the day after at night, at a dose of 200 mg every 12 h. Conversely, whenever a LH peak was not detected, recombinant hCG 250 μ g was administered s.c followed 48 h later by daily administration of 200 mg micronized vaginal P4 every 12 h.

Finally, in the NPP cycle, endometrial thickness was generally monitored from the eighth day of the cycle onwards by serial vaginal ultrasound. Whenever the endometrial thickness reached at least 7 mm, vaginal micronized P4 was initiated at 400 mg every 12 h if serum P4 levels were <1.5 ng/ml, regardless of the dominant follicle diameter and circulating LH serum levels.

In all FET protocols, the embryos were transferred on the fifth day of P4. In AC and NPP cycles, the FET was scheduled on the sixth day of exogenous P4 exposure. In NCs, exogenous P4 was started from the night of ovulation day and transfer was scheduled with similar P4 exposure as in non-NC cycles. On the day of the transfer, all patients had serum P4 measured by chemiluminescence and additional exogenous P4 (i.m. 100 mg P4 every 2 days) was administered whenever the result was one standard

deviation below our center's mean (i.e. below 8.5 ng/ml). Luteal support was interrupted at either the eighth (NC) or 11th week of gestation (NPP and AC).

Outcome measures

The primary outcome of this study was live birth rate (LBR) per FET, defined as the delivery of one or more neonates after 24 weeks of gestation. The secondary outcomes included a positive hCG, clinical pregnancy and miscarriage rates, and the following perinatal outcomes: first trimester bleeding, second/third trimester bleeding, preterm rupture of membranes, gestational diabetes, GHD, and gestational age at delivery. Clinical pregnancy was defined by the presence of intrauterine gestational sac on the fifth week of gestation, while clinical miscarriage was defined as spontaneous loss of a previously diagnosed pre-clinical/clinical pregnancy before 24 completed weeks of gestational age.

Statistical analysis

Categorical data were presented using absolute and relative frequencies, while continuous values were summarized with means and SDs. Baseline between-group comparisons were performed using univariable logistic or linear generalized estimating equations (GEE) regression, as appropriate, to account for the clustering derived from FET cycles using embryos derived from the same fresh treatment cycle. Confounder adjustment was performed using multivariable GEE regression analysis, adjusting for the potentially relevant confounding variables: female age at oocyte retrieval, oocyte source (autologous without PGT-A versus autologous with PGT-A versus donated), number of oocytes retrieved/donated, embryo developmental stage (Day 5 versus Day 6), number of embryos transferred, quality of the best embryo transferred (A versus B versus C) and year of treatment. An additional sensitivity analysis was performed dichotomizing the

sample in the following subgroups: oocyte age (<40 years old, ≥40 years old), female BMI (<25 kg/m², ≥25 kg/m²), source of oocytes (autologous without PGT-A versus autologous with PGT-A or donated), FET cycle rank (first versus others) and endometrial thickness (<9 mm, ≥9 mm).

Stata Software version 16.0 (StataCorp, College Station, TX, USA) was used for statistical analysis, with a *P*-value < 0.05 being considered as statistically significant (followed by Bonferroni-adjusted pairwise comparisons whenever significant).

Results

A total of 5791 FET cycles (performed in 3625 patients and 3993 IVF cycles) were included in this analysis; 3216 in NC, 349 in NPP, and 2226 in AC cycles (Fig. 2). In the NC group, there were 44% modified NCs, where exogenous hCG was administered to induce ovulation. The distribution of FET over time studied is shown on Fig. 3. The baseline female patient characteristics, as well as the IVF cycle and embryo transfer features, are summarized in Table 1. The mean female age at oocyte retrieval was higher for the NC subgroup compared to the AC group (33.3±6.0 versus 30.5±6.4 years, respectively, *P*<0.01). P4 level on the day of the transfer was significantly higher in NPP and NC (*P*<0.01) and 26.1% of the AC cycles were rescued with i.m. P4, which was also significantly higher compared with the other protocols (*P*<0.01). Finally, single embryo transfers were more frequent in the NPP subgroup, when compared to the AC and NC subgroups (*P*<0.01).

Pregnancy outcomes

Pregnancy outcomes are shown in Table 2. The LBR was higher in the NPP and NC group compared to the AC (49.1%, 45.2% versus 38.4%, respectively, *P*<0.01 versus NPP and AC versus NC). The rate of miscarriage was higher in the AC subgroup when compared to the NPP and NC subgroups (34.9%, 19.7%, and 25.0%, respectively, *P*<0.01 NPP versus AC and NC versus AC). These results remained mostly unaltered following confounder adjustment and in the five sensitivity analyses performed (Supplementary Fig. S1).

Perinatal outcomes

With regards to perinatal outcomes after FET (Table 3), AC had significantly higher first trimester bleeding when compared to NPP and NC (37.6%, 17.4%, and 14.7%, respectively, *P*<0.01 NPP versus AC and NC versus AC). The incidence per pregnancy of GHD was 14.5%, 9.4%, and 8.6% for the AC, NPP, and NC subgroups, respectively. The incidence of GHD was statistically significant between the AC and NC subgroups (*P*<0.01) following pairwise comparison.

Discussion

There is still controversy regarding the best FET protocol. This study presents an alternative protocol to the two most currently used (AC and NC), which combines the advantages of transfer scheduling of the AC with the potential clinical benefits associated with NC FET. A similar protocol was previously described by Weiss *et al.* (2021), albeit in a small group of 42 patients with no control group. To that extent, our study, seems to confirm the hypothesis that NPP may also provide high LBR that do not vary significantly from those observed after a NC FET.

Our retrospective analysis showed lower miscarriage rates for the NPP and NC compared with AC. This reiterates the conclusions found in previous literature describing potentially worse clinical outcomes (namely, higher miscarriage and lower live-birth rates) following AC-FET (Tomás C *et al.*, 2012; Mumusoglu *et al.*, 2021). The lack of endogenous progesterone production and absence of luteinization with exogenous progesterone administration have been described as possible explanations for such outcomes (Roelens *et al.*, 2020). Specifically, multiple thresholds of luteal serum progesterone have been proposed as the minimal clinical requirement to ensure an optimal endocrine milieu during embryo implantation and early pregnancy after FET treatment (Labarta *et al.*, 2017; Melo *et al.*, 2021). In our study, serum progesterone levels were also routinely assessed on the day of transfer and, if below 8.5 ng/ml, rescue intramuscular progesterone was initiated. This strategy of individualized luteal phase support has been shown to enhance LBR in AC FET (Álvarez *et al.*, 2021; Yarali *et al.*, 2021; Labarta *et al.*, 2022). However, our results

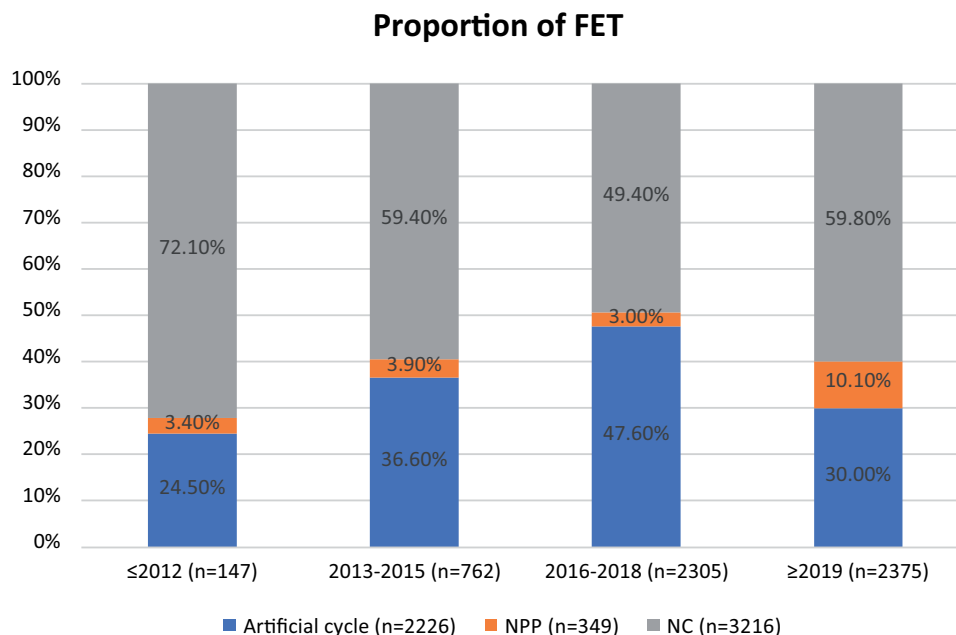


Figure 3. Proportion of each frozen embryo transfer cycle type over time. NPP, natural proliferative phase; NC, natural cycle.

Table 1. Comparison of baseline clinical characteristics based on subgroups of frozen embryo transfer.

	AC (n = 2527)	NPP (n = 369)	NC (n = 3771)	P-value
Primary infertility	1669/2105 (77.3%)	259/317 (81.7%)	2357/2965 (79.6%)	0.61
Duration of infertility (in years)	3.4 (2.5)	3.2 (2.5)	3.1 (2.3)	<0.01 [†]
Female BMI (kg/m²)	23.8 (4.0)	23.0 (3.7)	23.2 (3.7)	<0.01 ^{*,†}
IVF cycle characteristics				
Source of oocytes				<0.01 ^{*,†}
Autologous oocytes	880 (39.5%)	197 (56.5%)	1928 (60.0%)	
Autologous oocytes with PGT-A	127 (5.7%)	38 (10.9%)	363 (11.3%)	
Donated oocytes	1219 (54.8%)	114 (32.7%)	925 (28.8%)	
Female age at oocyte retrieval	30.5 (6.4)	32.7 (6.2)	33.3 (6.0)	<0.01 ^{*,†}
Oocytes retrieved/donated	14.9 (6.0)	14.2 (6.5)	14.3 (6.3)	<0.01 [†]
Number of embryos for transfer/cryopreservation	5.3 (2.9)	4.5 (2.8)	4.7 (2.8)	<0.01 ^{*,†}
Frozen embryo transfer characteristics				
Female patient age	39.3 (5.4)	38.5 (4.3)	38.1 (4.3)	<0.01 ^{*,†}
Cycle duration until transfer (in days)	20.1 (6.5)	18.1 (2.9)	18.9 (4.0)	<0.01 ^{*,†,‡}
Number of visits to clinic until planning	1.3 (0.5)	1.8 (0.9)	2.1 (1.0)	<0.01 ^{*,†,‡}
Endometrial thickness	9.1 (1.5)	9.0 (1.2)	9.1 (1.4)	0.34
Dominant follicle mean diameter (mm)	—	16.3 (2.7)	17.3 (2.2)	<0.01
E2 on day of schedule	402.4 (361.4)	254.9 (119.8)	248.7 (130.2)	<0.01 ^{*,†}
P4 on day of schedule	0.25 (0.24)	0.36 (0.36)	0.54 (0.49)	<0.01 ^{*,†,‡}
LH on the day of schedule	11.7 (17.8)	7.9 (4.5)	21.9 (25.1)	<0.01 ^{†,‡}
Serum Progesterone on day of FET	10.4 (6.2)	20.6 (12.3)	20.5 (6.7)	<0.01 ^{*,†}
Rescue progesterone administered	540/2071 (26.1%)	10/339 (2.9%)	39/3090 (1.3%)	<0.01 ^{*,†}
Embryo development stage				<0.01 [†]
Blastocyst Day 5	1772 (79.6%)	266 (76.2%)	2387 (74.2%)	
Blastocyst Day 6	454 (20.4%)	83 (23.8%)	829 (25.8%)	
Single embryo transfer	1919 (86.2%)	327 (93.7%)	2779 (86.4%)	<0.01 ^{*,‡}
Rank of FET				0.80
First	1422 (63.9%)	226 (64.8%)	2082 (64.7%)	
Second	535 (24.0%)	80 (22.9%)	777 (24.2%)	
Third and beyond	269 (12.1%)	43 (12.3%)	357 (11.1%)	

Pairwise $P < 0.05$ for the following comparisons:

* AC versus NPP.

† AC versus NC.

‡ NPP versus NC.

AC, artificial cycle; NPP, natural proliferate phase; NC, natural cycle; FET, frozen embryo transfer.

Table 2. Comparison of pregnancy outcomes after frozen embryo transfer considering each subgroup with confounder adjustment.

	AC (n = 2226)	NPP (n = 349)	NC (n = 3216)	P-value
hCG-positive pregnancy	1343 (60.4%) 0.96 (0.93–0.99)	219 (62.8%) 1.01 (0.96–1.07)	1978 (61.6%) Reference	0.56
Clinical pregnancy	1122 (50.5%) 0.94 (0.91–0.96)	199 (57.0%) 1.01 (0.95–1.07)	1748 (54.4%) Reference	<0.01 [†]
Miscarriage	466/1335 (34.9%) 1.11 (1.07–1.15)	43/218 (19.7%) 0.98 (0.92–1.04)	491/1962 (25.0%) Reference	<0.01 ^{*,†}
Live birth	848/2206 (38.4%) 0.91 (0.88–0.94)	168/342 (49.1%) 1.02 (0.96–1.09)	1437/3182 (45.2%) Reference	<0.01 ^{*,†}

Pairwise $P < 0.05$ for the following comparisons:

* AC versus NPP.

† AC versus NC.

PGT-A, preimplantation genetic testing for aneuploidies; AC, artificial cycle; NPP, natural proliferative phase; NC, natural cycle. Confounder-adjustment was performed using multivariable GEE regression analysis, adjusting for the potentially relevant confounding variables: female age at oocyte retrieval, oocyte source (autologous without PGT-A versus autologous with PGT-A versus donated), number of oocytes retrieved/donated, embryo developmental stage (Day 5 versus Day 6), number of embryos transferred, quality of best embryo transferred (A versus B versus C) and year of treatment. Confounder-adjusted data are presented with odds-ratios (95% CIs), in which results with $P < 0.05$ from the reference group are highlighted in bold.

Table 3. Perinatal outcome after frozen embryo transfer from different subgroups.

	AC (n = 804)	NPP (n = 160)	NC (n = 1348)	P-value
First trimester bleeding	246/655 (37.6%)	25/144 (17.4%)	163/1110 (14.7%)	<0.01 ^{*,†}
Second/third trimester bleeding	39/653 (3.7%)	8/144 (5.6%)	41/1110 (3.7%)	0.08
Preterm rupture of membranes	41/655 (6.3%)	7/144 (4.9%)	44/1101 (4.0%)	0.10
Gestational diabetes	113/653 (17.3%)	22/144 (15.3%)	159/1113 (14.3%)	0.24
Gestational hypertensive disorders	92/636 (14.5%)	13/138 (9.4%)	95/1099 (8.6%)	<0.01 [†]
Gestational age at delivery (weeks)	38.5 (2.8)	39.1 (4.7)	38.8 (2.6)	0.02 [†]

Pairwise $P < 0.05$ for the following comparisons:

* AC versus NPP.

† AC versus NC.

FET, frozen embryo transfer; NPP, natural proliferative phase; NC, natural cycle; AC, artificial cycle.

raise the hypothesis that the use of luteal phase progesterone monitoring and subsequent rescue strategies may still be insufficient to prevent the increased risk of miscarriage in AC cycles when compared to the NC. This finding may be of major clinical significance and requires confirmation in future studies.

The incidence of GHD per pregnancy, in general population, is difficult to access and the information is not always available. A recent study demonstrated an incidence of 7.3% per pregnancy (Garovic et al., 2020). The description of an increased risk of GHD in FET in comparison to fresh ET has been recently narrowed specifically to FET carried out in ACs (Ginström et al., 2019; Hu et al., 2021; Rafael et al., 2022). The corpus luteum (CL) is responsible for producing progesterone, estradiol, relaxin, as well as other vasoactive and angiogenic substances which play important roles in regulating and influencing physiological processes that take place early in pregnancy, such as decidualization, implantation, angiogenesis, and maternal hemodynamic adaptation (Pereira et al., 2021). Typically, no follicular growth, ovulation and CL formation are observed in ACs. Therefore, substances that are relevant for maternal cardiovascular and renal adaptation to early pregnancy may be under produced. This could, at least partially, contribute to the increased frequency of GHD observed in such cases (Conrad, 2020), a finding that was also observed in our study. In NPP group cycles, progesterone exposure begins before CL formation, and although a CL was frequently observed during the early gestation scan carried out in patients belonging to this group, such finding was not systematically assessed. We believe that the formation of a CL may be induced by either the beginning of higher doses of exogenous progesterone (Dozortsev et al., 2020) or by the non-opposition of the natural evolution of follicular growth and luteinization. Additionally, the fact that the incidence of GHD in NPP cycles seemed comparable to NCs gives support to the hypothesis that ovulation does occur with such protocol and reinforces the interest in the appraisal of this issue (Dozortsev et al., 2020). In an attempt to conclusively reply to these scientific hypotheses, a prospective study to evaluate the NPP endocrine profile following exogenous progesterone exposure and to assess whether patients undergo ovulation during NPP-FET in currently underway.

Beyond the absence of CL, other potential reasons for an increase of adverse outcomes in AC FET could be the effect continued estradiol has also previously been associated with an increased risk of miscarriage rates in an, albeit small, RCT (Ghaffari et al., 2022). To this extent, some aspects related to the practicality of different FET protocols should be further scrutinized as they are frequent justification for the overuse of AC-FET.

The reasons to avoid the NC in ovulatory patients are mainly: the need for a higher number of visits to the clinic to document follicular and endometrial growth and spontaneous LH surge, higher cancellation rates and the lack of flexibility regarding FET scheduling. However, considering the increasing evidence regarding the potential clinical disadvantages of AC, it may be interesting to assess further alternative strategies such as NPP, in which the initiation of exogenous progesterone is determined more loosely by the documentation of a triple layered above 7 mm endometrium and the presence of a dominant follicle. Hence, using this approach, FET scheduling gains flexibility comparing to the NC, as progesterone can be initiated in a wider range of dates during the last days of the proliferative phase of the menstrual cycle. Moreover, if during monitoring a spontaneous LH surge is detected on, for example a Sunday, that would determine FET taking place during the following weekend, while

in the NPP protocol progesterone might be initiated that same night and the FET would be performed the following Friday.

Some clear limitations are present in the study and must be acknowledged. Owing to its retrospective nature, the potential for selection bias and unmeasured confounding may still exist, despite our attempt to minimize its effect by performing confounder adjustment using multivariable GEE regression. Another limitation worth mentioning is the low number of patients in the NPP subgroup, which may have led the study to be underpowered for the detection of statistical significance of clinically relevant differences.

In conclusion, we posit that the NPP-FET protocol may be an effective and safe alternative to NC and AC FETs, while still allowing for enhanced practicality in patient follow-up and FET scheduling. However, further investigation is warranted, with prospective studies using larger and more homogeneous subsets of patients.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article will be shared upon reasonable request to IVI-RMA Lisbon and the corresponding author.

Acknowledgements

The authors thank all couples who trusted our clinic.

Authors' roles

C.M.G.: conception of study design, screening, data collection and analysis, manuscript draft and revision. S.R.S.: conception of study design, screening, and manuscript revision. S.G.N. and J.M.M.: data collection, statistical analysis, and manuscript revision. S.S.-R.: conception of study design, screening, data collection, statistical analysis, manuscript revision. All authors approved the final version of the manuscript to be published. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article and take responsibility for the integrity of the work.

Funding

This research was supported by the IVI-RMA-Lisbon (2008-LIS-053-CG). The authors did not receive any funding for this study.

Conflict of interest

The authors declare that there are no conflicts of interest in connection with this article.

References

- Altmäe S, Tamm-Rosenstein K, Esteban FJ, Simm J, Kolberg L, Peterson H, Metsis M, Haldre K, Horcajadas JA, Salumets A et al. Endometrial transcriptome analysis indicates superiority of natural over artificial cycles in recurrent implantation failure patients undergoing frozen embryo transfer. *Reprod Biomed Online* 2016;**32**:597–613.

- Álvarez M, Gaggiotti-Marre S, Martínez F, Coll L, García S, González-Foruria I, Rodríguez I, Parriego M, Polyzos NP, Coroleu B et al. Individualized luteal phase support in artificially prepared frozen embryo transfer cycles based on serum progesterone levels: a Prospective Cohort Study. *Hum Reprod* 2021;**36**:1552–1560.
- ASEBIR. *Cuadernos De Embriología Clínica II. Criterios ASEBIR De Valoración Morfológica De Oocitos, Embriones Tempranos y Blastocistos Humanos*, 2nd edn. Madrid: Gobalo, 2008.
- Asserhøj LL, Spangmose AL, Aaris Henningsen A-K, Clausen TD, Ziebe S, Jensen RB, Pinborg A. Adverse obstetric and perinatal outcomes in 1,136 singleton pregnancies conceived after programmed frozen embryo transfer (FET) compared with natural cycle FET. *Fertil Steril* 2021;**115**:947–956.
- Belva F, Bonduelle M, Roelants M, Verheyen G, van Landuyt L. Neonatal health including congenital malformation risk of 1072 children born after vitrified embryo transfer. *Hum Reprod* 2016;**31**:1610–1620.
- Casper RF, Yanushpolsky EH. Optimal endometrial preparation for frozen embryo transfer cycles: window of implantation and progesterone support. *Fertil Steril* 2016;**105**:867–872.
- Chang EM, Han JE, Kim YS, Lyu SW, Lee WS, Yoon TK. Use of the natural cycle and vitrification thawed blastocyst transfer results in better in vitro fertilization outcomes: cycle regimens of vitrification thawed blastocyst transfer. *J Assist Reprod Genet* 2011;**28**:369–374.
- Chih HJ, Elias FTS, Gaudet L, Velez MP. Assisted reproductive technology and hypertensive disorders of pregnancy: systematic review and meta-analyses. *BMC Pregnancy Childbirth* 2021;**21**:449.
- Conrad K. Evidence for corpus luteal and endometrial origins of adverse pregnancy outcomes in women conceiving with or without assisted reproduction. *Obstet Gynecol Clin North Am* 2020;**47**:163–181.
- Cozzolino M, Cecchino GN, Bosch E, Garcia-Velasco JA, Garrido N. Minimal ovarian stimulation is an alternative to conventional protocols for older women according to Poseidon's stratification: a retrospective multicenter cohort study. *J Assist Reprod Genet* 2021;**38**:1799–1807.
- Dalsgaard TH, Hvas AM, Kirkegaard KS, Jensen MV, Knudsen UB. Impact of frozen thawed embryo transfer in hormone substituted cycles on thrombotic risk markers. *Thromb Res* 2022;**209**:23–32.
- De Geyter C, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, Scaravelli G, Smeenk J, Vidakovic S, Goossens V; European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE)ART in Europe, 2014: results generated from European registries by ESHRE: the European IVF-monitoring Consortium (EIM) for the European society of human reproduction and embryology (ESHRE). *Hum Reprod* 2018;**33**:1586–1601.
- Dozortsev D, Pellicer A, Diamond MP. Progesterone is a physiological trigger of ovulatory gonadotropins. *Fertil Steril* 2020;**113**:923–924.
- Evans J, Hannan NJ, Edgell TA, Vollenhoven BJ, Lutjen PJ, Osianlis T, Salomonsen LA, Rombauts LJF. Fresh versus frozen embryo transfer: backing clinical decisions with scientific and clinical evidence. *Hum Reprod Update* 2014;**20**:808–821.
- Fatemi HM, Kyrrou D, Bourgain C, van den Abbeel E, Griesinger G, Devroey P. Cryopreserved-thawed human embryo transfer: spontaneous natural cycle is superior to human chorionic gonadotropin-induced natural cycle. *Fertil Steril* 2010;**94**:2054–2058.
- Frydman R, Testart J, Fernandez H, Arvis P, Belaisch JC. La prédiction de l'ovulation [Prediction of ovulation]. *J Gynecol Obstet Biol Reprod* 1982;**11**:793–799.
- Garovic VD, White W, MD, Vaughan L, Saiki M, Parashuram S, Garcia-Valencia O, Weissgerber TL, Milic N, Weaver A, Mielke MM. Incidence and long-term outcomes of hypertensive disorders of pregnancy. *J Am Coll Cardiol* 2020;**75**:2323–2334.
- Ghaffari F, Chekini Z, Vesali S. Duration of estradiol supplementation in luteal phase support for frozen embryo transfer in hormone replacement treatment cycles: a randomized, controlled phase III trial. *Arch Gynecol Obstet* 2022;**305**:767–775.
- Ginström E, Wennerholm UB, Khatibi A, Petzold M, Bergh C. Neonatal and maternal outcome after frozen embryo transfer: increased risks in programmed cycles. *Am J Obstetrics Gynecol* 2019;**221**:126 e1–126.e18.
- Groenewoud ER, Cantineau AEP, Kollen BJ, Macklon NS, Cohlen BJ. What is the optimal means of preparing the endometrium in frozen-thawed embryo transfer cycles? A systematic review and meta-analysis. *Hum Reprod Update* 2013;**19**:458–470.
- Groenewoud ER, Cohlen BJ, Al-Oraiby A, Brinkhuis EA, Broekmans FJM, de Bruin JP, van den Dool G, Fleisher K, Friederich J, Goddijn M et al. A randomized controlled, non-inferiority trial of modified natural versus artificial cycle for cryo-thawed embryo transfer. *Hum Reprod* 2016;**31**:1483–1492.
- Groenewoud ER, Macklon NS, Cohlen BJ; ANTARCTICA Study Group. The effect of elevated progesterone levels before hCG triggering in modified natural cycle frozen-thawed embryo transfer cycles. *Reprod Biomed Online* 2017;**34**:546–554.
- Hu KL, Zhang D, Li R. Endometrium preparation and perinatal outcomes in women undergoing single-blastocyst transfer in frozen cycles. *Fertil Steril* 2021;**115**:1487–1494.
- Huberlant S, Vaast M, Anahory T, Tailland ML, Rougier N, Ranisavljevic N, Hamamah S. Natural cycle for frozen-thawed embryo transfer: spontaneous ovulation or triggering by hCG. *Gynecol Obstet Fertil Senol* 2018;**46**:466–473.
- Ishihara O, Araki R, Kuwahara A, Itakura A, Saito H, Adamson GD. Impact of frozen-thawed single-blastocyst transfer on maternal and neonatal outcome: an analysis of 277,042 single-embryo transfer cycles from 2008 to 2010 in Japan. *Fertil Steril* 2014;**101**:128–133.
- Jing S, Li XF, Zhang S, Gong F, Lu G, Lin G. Increased pregnancy complications following frozen-thawed embryo transfer during an artificial cycle. *J Assist Reprod Genet* 2019;**36**:925–933.
- Labarta E, Mariani G, Holtmann N, Celada P, Remohi J, Bosch E. Low serum progesterone on the day of embryo transfer is associated with a diminished ongoing pregnancy rate in oocyte donation cycles after artificial endometrial preparation: a prospective study. *Hum Reprod* 2017;**32**:2437–2442.
- Labarta E, Mariani G, Rodríguez-Varela C, Bosch E. Individualized luteal phase support normalizes live birth rate in women with low progesterone levels on the day of embryo transfer in artificial endometrial preparation cycles. *Fertil Steril* 2022;**117**:96–103.
- Loutradi KE, Kolibianakis EM, Venetis CA, Papanikolaou EG, Pados G, Bontis I, Tarlatzis BC. Cryopreservation of human embryos by vitrification or slow freezing: a systematic review and meta-analysis. *Fertil Steril* 2008;**90**:186–193.
- Maas AHEM, Rosano G, Cifkova R, Chieffo A, van Dijken D, Hamoda H, Kunadian V, Laan E, Lambrinoudaki I, Maclaran K et al. Cardiovascular health after menopause transition, pregnancy disorders, and other gynaecologic conditions: a consensus document from European cardiologists, gynaecologists, and endocrinologists. *Eur Heart J* 2021;**42**:967–984.
- Martínez-Burgos M, Herrero L, Megías D, Salvanes R, Montoya MC, Cobo AC, Garcia-Velasco JA. Vitrification versus slow freezing of oocytes: effects on morphologic appearance, meiotic spindle configuration, and DNA damage. *Fertil Steril* 2011;**95**:374–377.

- Melo P, Chung Y, Pickering O, Price MJ, Fishel S, Khairy M, Kingsland C, Lowe P, Petsas G, Rajkhowa M *et al*. Serum luteal phase progesterone in women undergoing frozen embryo transfer in assisted conception: a systematic review and meta-analysis. *Fertil Steril* 2021;**116**:1534–1556.
- Montagut M, Santos-Ribeiro S, De Vos M, Polyzos NP, Drakopoulos P, Mackens S, van de Vijver A, van Landuyt L, Verheyen G, Tournaye H *et al*. Frozen-thawed embryo transfers in natural cycles with spontaneous or induced ovulation: the search for the best protocol continues. *Hum Reprod* 2016;**31**:2803–2810.
- Mumusoglu S, Polat M, Ozbek I, Bozdog G, Papanikolaou E, Esteves S, Humaidan P, Yarali H. Preparation of the endometrium for frozen embryo transfer: a systematic review. *Front Endocrinol (Lausanne)* 2021;**12**:688237.
- Pereira M, Mainigi M, Strauss J. Secretory products of the corpus luteum and preeclampsia. *Hum Reprod Update* 2021;**27**:651–672.
- Rafael F, Robles G, Navarro A, Garrido N, Garcia-Velasco J, Bosch E, Nunes S, Soares S, Santos-Ribeiro S. Perinatal outcomes in children born after fresh or frozen embryo transfer using donated oocytes. *Hum Reprod* 2022;**37**:1642–1651.
- Roelens C, Santos-Ribeiro S, Becu L, Mackens S, Van Landuyt L, Racca A, De Vos M, van de Vijver A, Tournaye H, Blockeel C *et al*. Frozen-warmed blastocyst transfer after 6 or 7 days of progesterone administration: impact on live birth rate in hormone replacement therapy cycles. *Fertil Steril* 2020;**114**:125–132.
- Roque M, Lattes K, Serra S, Solà I, Geber S, Carreras R, Checa MA. Fresh embryo transfer versus frozen embryo transfer in in vitro fertilization cycles: a systematic review and meta-analysis. *Fertil Steril* 2013;**99**:156–162.
- Sathanandan M, Macnamee MC, Rainsbury P, Wick K, Brinsden P, Edwards RG. Replacement of frozen-thawed embryos in artificial and natural cycles: a prospective semi-randomized study. *Hum Reprod* 1991;**6**:685–687.
- Shapiro BS, Daneshmand ST, Garner FC, Aguirre M, Hudson C, Thomas S. Evidence of impaired endometrial receptivity after ovarian stimulation for in vitro fertilization: a prospective randomized trial comparing fresh and frozen-thawed embryo transfer in normal responders. *Fertil Steril* 2011;**96**:344–348.
- Testart J, Frydman R, Feinstein MC, Thebault A, Roger M, Scholler R. Interpretation of plasma luteinizing hormone assay for the collection of mature oocytes from women: definition of a luteinizing hormone surge-initiation rise. *Fertil Steril* 1981;**36**:50–54.
- Tomás C, Alsbjerg B, Martikainen H, Humaidan P. Pregnancy loss after frozen-embryo transfer—a comparison of three protocols. *Fertil Steril* 2012;**98**:1165–1169.
- Weiss A, Baram S, Geslevich Y, Goldman S, Nothman S, Beck-Fruchter R. Should the modified natural cycle protocol for frozen embryo transfer be modified? A prospective case series proof of concept study. *Eur J Obstet Gynecol Reprod Biol* 2021;**258**:179–183.
- Veleva Z, Tiitinen A, Vilska S, Hydén-Granskog C, Tomás C, Martikainen H, Tapanainen JS. High and low BMI increase the risk of miscarriage after IVF/ICSI and FET. *Hum Reprod* 2008;**23**:878–884.
- Yarali H, Polat M, Mumusoglu S, Ozbek IY, Erden M, Bozdog G, Humaidan P. Subcutaneous luteal phase progesterone rescue rectifies ongoing pregnancy rates in hormone replacement therapy vitrified-warmed blastocyst transfer cycles. *Reprod Biomed Online* 2021;**43**:45–51.
- Younis JS, Simon A, Laufer N. Endometrial preparation: lessons from oocyte donation. *Fertil Steril* 1996;**66**:873–884.
- Zhang J, Du M, Li Z, Wang L, Hu J, Zhao B, Feng Y, Chen X, Sun L. Fresh versus frozen embryo transfer for full term singleton birth: a retrospective cohort study. *J Ovarian Res* 2018;**11**:59.

Reproductive endocrinology

Ovarian response and embryo ploidy following oral micronized progesterone-primed ovarian stimulation versus GnRH antagonist protocol. A prospective study with repeated ovarian stimulation cycles

M. del Mar Vidal, Francisca Martínez , Ignacio Rodríguez, and Nikolaos P. Polyzos *

Dexeus Fertility, Hospital Universitari Dexeus, Barcelona 08014, Spain

*Correspondence address. Dexeus Fertility, Hospital Universitari Dexeus, Gran Via Carles III 71-75, Barcelona, 08014, Spain. E-mail: nikpol@dexeus.com  <https://orcid.org/0000-0001-6161-0011>

ABSTRACT

STUDY QUESTION: Is there any difference in ovarian response and embryo ploidy following progesterone-primed ovarian stimulation (PPOS) using micronized progesterone or GnRH antagonist protocol?

SUMMARY ANSWER: Pituitary downregulation with micronized progesterone as PPOS results in higher number of oocytes retrieved and a comparable number of euploid blastocysts to a GnRH antagonist protocol.

WHAT IS KNOWN ALREADY: Although the GnRH antagonist is considered by most the gold standard protocol for controlling the LH surge during ovarian stimulation (OS) for IVF/ICSI, PPOS protocols are being increasingly used in freeze-all protocols. Still, despite the promising results of PPOS protocols, an early randomized trial reported potentially lower live births in recipients of oocytes resulting following downregulation with medroxyprogesterone acetate as compared with a GnRH antagonist protocol. The scope of the current prospective study was to investigate whether PPOS with micronized progesterone results in an equivalent yield of euploid blastocysts to a GnRH antagonist protocol.

STUDY DESIGN, SIZE, DURATION: In this prospective study, performed between September 2019 to January 2022, 44 women underwent two consecutive OS protocols within a period of 6 months in a GnRH antagonist protocol or in a PPOS protocol with oral micronized progesterone.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Overall, 44 women underwent two OS cycles with an identical fixed dose of rFSH (225 or 300 IU) in both cycles. Downregulation in the first cycles was performed with the use of a flexible GnRH antagonist protocol (0.25 mg per day as soon as one follicle of 14 mm) and consecutively, after a washout period of 1 month, control of LH surge was performed with 200 mg of oral micronized progesterone from stimulation Day 1. After the completion of both cycles, all generated blastocysts underwent genetic analysis for aneuploidy screening (preimplantation genetic testing for aneuploidy, PGT-A).

MAIN RESULTS AND THE ROLE OF CHANCE: Comparisons between protocols did not reveal differences between the duration of OS. The hormonal profile on the day of trigger revealed statistically significant differences between protocols in all the tested hormones except for FSH: with significantly higher serum E2 levels, more elevated LH levels and higher progesterone levels in PPOS cycles as compared with antagonist cycles, respectively. Compared with the GnRH antagonist protocol, the PPOS protocol resulted in a significantly higher number of oocytes (12.7 ± 8.09 versus 10.3 ± 5.84 ; difference between means [DBM] -2.4 [95% CI -4.1 to -0.73]), metaphase II (9.1 ± 6.12 versus 7.3 ± 4.15 ; DBM -1.8 [95% CI -3.1 to -0.43]), and 2 pronuclei (7.1 ± 4.99 versus 5.7 ± 3.35 ; DBM -1.5 [95% CI -2.61 to -0.32]), respectively. Nevertheless, no differences were observed regarding the mean number of blastocysts between the PPOS and GnRH antagonist protocols (2.9 ± 2.11 versus 2.8 ± 2.12 ; DBM -0.07 [95% CI -0.67 to 0.53]) and the mean number of biopsied blastocysts (2.9 ± 2.16 versus 2.9 ± 2.15 ; DBM -0.07 [95% CI -0.70 to 0.56]), respectively. Concerning the euploidy rates per biopsied embryo, a 29% [95% CI 21.8–38.1%] and a 35% [95% CI 26.6–43.9%] were noticed in the PPOS and antagonist groups, respectively. Finally, no difference was observed for the primary outcome, with a mean number of euploid embryos of 0.86 ± 0.90 versus 1.00 ± 1.12 for the comparison of PPOS versus GnRH antagonist.

LIMITATIONS, REASONS FOR CAUTION: The study was powered to detect differences in the mean number of euploid embryos and not in terms of pregnancy outcomes. Additionally, per protocol, there was no randomization, the first cycle was always a GnRH antagonist cycle and the second a PPOS with 1 month of washout period in between.

WIDER IMPLICATIONS OF THE FINDINGS: In case of a freeze-all protocol, clinicians may safely consider oral micronized progesterone to control the LH surge and patients could benefit from the advantages of a medication of oral administration, with a potentially higher number of oocytes retrieved at a lower cost, without any compromise in embryo ploidy rates.

STUDY FUNDING/COMPETING INTEREST(S): This research was supported by an unrestricted grant from Theramex. N.P.P. has received Research grants from Merck Serono, Organon, Ferring Pharmaceutical, Roche, Theramex, IBSA, Gedeon Richter, and Besins

Received: August 12, 2023. Revised: January 21, 2024. Editorial decision: February 19, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

Healthcare; honoraria for lectures from: Merck Serono, Organon, Ferring Pharmaceuticals, Besins International, Roche Diagnostics, IBSA, Theramex, and Gedeon Richter; consulting fees from Merck Serono, Organon, Besins Healthcare, and IBSA. M.d.M.V., F.M., and I.R. declared no conflicts of interest.

TRIAL REGISTRATION NUMBER: The study was registered at Clinical Trials Gov. (NCT04108039).

Keywords: blastocyst euploidy / non-conventional ovarian stimulation / progestin-primed ovarian stimulation / micronized progesterone / preimplantation genetic testing for aneuploidy

Introduction

Ovarian stimulation (OS) has led to a significant increase in live birth rates in IVF/ICSI cycles, mainly due to the induction of multifollicular development that allowed extended embryo culture and embryo selection. However, one of the main disadvantages of OS is the premature LH surge, which is caused by increasing plasma estradiol produced by multiple growing follicles (Ata *et al.*, 2021). The use of GnRH analogues has undeniably been a landmark in the history of OS, since through the suppression of endogenous LH secretion, premature LH surges and cycle cancellation are prevented in more than 20% of OS cycles (The ESHRE Guideline Group on Ovarian Stimulation *et al.*, 2020). With that purpose, GnRH agonists were initially used but, nowadays, the antagonist protocol has become the standard of care in OS treatments, mainly owing to its patients friendly profile, short administration and increased patient convenience (Verberg *et al.*, 2008), and the increased safety profile via a drastic reduction or even elimination of ovarian hyperstimulation syndrome through GnRH agonist triggering and freeze-all policy (Devroey *et al.*, 2011). Nevertheless, despite the numerous advantages associated with the GnRH antagonist protocol, antagonists are still an injectable medication, with significant costs requiring adherence to therapy.

Recently, owing to the significant increase of freeze-all protocols, accumulating evidence have been published with novel regimens to control the LH surge in freeze-all cycles. The use of oral progestins have been proposed, in the so-called progesterone-primed ovarian stimulation (PPOS) protocols (The ESHRE Guideline Group on Ovarian Stimulation *et al.*, 2020), taking into consideration basic physiology evidence suggesting that progesterone can block the pre-ovulatory LH surge, by a direct inhibitory action on estradiol-responsive cells within the hypothalamus, and its disruptive effect in response to estradiol is rapid but transient (Skinner *et al.*, 1998). PPOS are claimed to be safe, patient-friendly, cost-effective (La Marca and Capuzzo, 2019; Ata *et al.*, 2021), and they emerge as an effective alternative to antagonist cycles for those treatments in which there is no need for a receptive endometrium (such as deferred embryo transfer, fertility preservation, oocyte donation, and preimplantation genetic screening) (Massin, 2017; Martínez *et al.*, 2021).

Over the last years, different progestins such as medroxyprogesterone, micronized progesterone, dydrogesterone, or desogestrel have been evaluated in freeze-all IVF cycles, as a regimen to control the LH surge. Although the number of oocytes retrieved and the clinical outcomes appeared to be similar as compared with GnRH antagonist protocols (Kuang *et al.*, 2015; Zhu *et al.*, 2015; Martínez *et al.*, 2020; Ata *et al.*, 2021), most of the evidence is based on observational studies. Nevertheless, as the ploidy status is an unquestionably relevant marker of embryo quality and directly related to the live birth rate, it seems fair enough to evaluate the efficacy of the PPOS protocols compared to the more universal use of GnRH antagonist protocol in those terms.

Taking into account the conflicting evidence, the scope of the current study was to evaluate for the first time in a prospective

study whether the use of a PPOS protocol with micronized progesterone in women planned to undergo OS for PGT-A may result in comparable number of oocytes and number of euploid embryos as compared to a GnRH antagonist protocol, in order to justify its use in all freeze-all PGT-A cycles in the future as a more patient-friendly, equally, efficient, and cheaper protocol.

Material and methods

This prospective study was performed at the Reproductive Medicine Department of Dexeus University Hospital, from September 2019 to January 2022.

The study was approved by the Ethical Committee of University Hospital Dexeus and was prospectively registered at Clinical Trials Gov (NCT04108039).

Study population

Infertile patients planning to undergo two OS cycles, in order to accumulate embryos to increase the chance of obtaining euploid embryos, and fulfilling the inclusion criteria were recruited after signing the proper informed consent (IC).

Inclusion criteria were age 36–40 years, BMI 18–30 kg/m², undergoing preimplantation genetic screening, and planning at least two cycles. The use of partner's or donor's sperm was allowed, excluding cases with severe male factor requiring testicular sperm extraction. Candidates with low ovarian reserve (defined as anti-Müllerian hormone (AMH) < 1.2 ng/ml) or who were using any other drug, potentially interfering with the treatment, were excluded.

Treatment

Overall, 44 women underwent two OS cycles with identical fixed dose of rFSH (Ovaleap®, Theramex, Madrid, Spain) 225 or 300 IU in both cycles. A washout period of one menstrual cycle was always maintained between both cycles in order to eliminate the risk of a potential carry-over effect from the first cycle.

The stimulation was started on the second day of the menstrual cycle or after a pill-free interval of 5 days if they were previously taking oral contraceptives. The treatment allocation to the intervention was fixed to first OS with an antagonist protocol and second with the oral progesterone.

In the antagonist cycle, endogenous LH suppression was accomplished by s.c., injections of 0.25 mg of Ganirelix (Orgalutran®, Organon, Barcelona, Spain) started in the presence of follicles >14 mm or E2 levels >400 pg/ml and continued until ovulation triggering.

In the progesterone cycle, the LH surge was controlled by oral administration of 200 mg of micronized progesterone (Utrogestan®, SeidLab, Barcelona, Spain) once a day, from stimulation Day 1 (SD1) until the day of the trigger.

Final oocyte maturation was induced with a GnRH agonist bolus (triptorelin 0.2 mg; Decapeptyl®, Ipsen Pharma, Spain) as soon as three follicles ≥18 mm were observed by transvaginal ultrasound in both cycles, followed by oocyte retrieval 34–36 h thereafter.

Blood samples and transvaginal ultrasound were performed before administration of any medication (SD1) and on stimulation days 6, 8, day of triggering, and Day +1 after trigger.

Fertilized oocytes were cultured up to the blastocyst stage and PGT-A was performed by trophectoderm biopsy. Zona pellucida opening was performed on Day 3 embryos presenting ≥ 5 cells and $<30\%$ fragmentation at 66 ± 2 h post-ICSI, by laser thermolysis (Boada et al., 1998). Finally, blastocysts were biopsied by laser thermolysis as previously described (Veiga et al., 1997). Three to eight cells were biopsied for each embryo. Biopsied blastocysts were subsequently vitrified afterwards according to Kitazato methodology and kits (Kitazato Medical Group, Japan) (Parriego et al., 2007; Coll et al., 2018). Chromosome analysis was performed by next-generation sequencing of whole genome amplification products using commercially available kits and following the protocols provided by the manufacturer (VeriSeq. PGS Kit, Illumina, UK). Cytogenetic assessment was performed using BlueFuse Multi v3 software (Illumina, UK). Euploid blastocysts were warmed for transfer in a subsequent cycle.

Outcomes measures

The primary endpoint was to evaluate the mean number of euploid embryos generated from antagonist or progesterone down-regulated cycles, in the same patient. Secondary endpoints were the endocrine profile at specific intervals, namely on Day 1, Day 6, Day 8, day of triggering and the day after triggering, the length of stimulation, and total dose of gonadotropins required, number of oocytes and metaphase II (MIIs) retrieved, fertilization rate, and blastocyst formation rate.

Statistical analysis

Sample size calculation

The study was designed as an equivalence trial to show that the two interventions (antagonist OS or PPOS) do not differ in either direction by more than a pre-specified unimportant or insignificant amount (i.e. a two-sided test). In a two one-sided tests analysis for additive equivalence of paired means with bounds -0.5 and 0.5 for the mean difference and a significance level of 0.05 , assuming a mean difference of 0 , a common SD of 2.02 and correlation 0.85 , a sample size of 44 patients (pairs of stimulation cycles) is required to obtain a power of at least 0.80 , assuming a normal distribution, and a correlation 0.85 (between the two stimulation cycles) with equivalence limits $(-0.5, 0.5)$. The data to obtain the sample size were calculated from a retrospective analysis of a database: 37 pairs of consecutive PGT-A cycles in young patients with normal AMH and antral follicle count (AFC) were identified. Analysis of the results in these patients showed a mean difference of 0.5 with a SD of 2.02 in the number of euploid embryos between the two consecutive stimulation cycles.

Statistical analysis

Descriptive summary measures expressed as mean (SD) were used for continuous variables and number (%) for categorical variables, in order to provide a summary estimate of patient demographics and baseline characteristics.

For primary outcome, we compared the mean number of euploid embryos between antagonist and progesterone OS: (i) All pairwise differences within one patient were calculated from difference in the number of euploid embryos between antagonist and progesterone OS cycles, (ii) results of the percentage pairwise differences are shown with boxplots and summary statistics were calculated (mean, median, SD, percentiles, and range), (iii) P-values from paired t-tests were calculated for each comparison.

All tests were two-sided with a level of significance $P < 0.05$. All analyses were performed in R software.

Ethical issues

Patients signed an IC, accepting to undergo two stimulation cycles with an interval of at least 1 month between treatments. However, they might have dropped the study before the two cycles were completed. Total dose of rFSH (Ovaleap®, Theramex, Madrid, Spain) was provided without cost and a reduced price in the embryo biopsy and the PGT-A was applied.

Results

In total, 44 women underwent two OS cycles and 255 blastocysts were biopsied: 129 in the PPOS group and 126 in the antagonist group. Although patients did undergo first stimulation cycle with antagonist and after the washout period received the PPOS protocol, for easiness of reading, results are presented comparing PPOS versus antagonist cycles.

The main demographic characteristics of population were a mean age of 39.59 years, BMI 21.0 kg/m^2 , AFC 12 , and AMH 1.81 ng/ml (Table 1).

OS outcomes

Comparative analysis of the two cycles did not reveal differences between PPOS and antagonist in the duration of OS (difference between means (DBM) 0.43 [95% CI -0.08 to 0.94]).

Hormonal profile on the day of trigger revealed statistically significant differences between both cycles in all the tested hormones except for FSH. Significantly higher serum estradiol levels, more elevated LH levels and higher progesterone levels were noticed in PPOS cycles as compared with antagonist cycles, respectively (Table 2; Fig. 1).

Compared with the GnRH antagonist protocol, PPOS protocol resulted in a significantly higher number of oocytes (12.68 ± 8.09 versus 10.27 ± 5.84 ; DBM -2.4 [95% CI -4.1 to -0.73]), MIIs (9.09 ± 6.12 versus 7.34 ± 4.15 ; DBM -1.8 [95% CI -3.1 to -0.43]), and 2 pronuclei (2PNs; 7.14 ± 4.99 versus 5.66 ± 3.35 ; DBM -1.5 [95% CI $-2.6.1$ to -0.32]), respectively (Table 2; Fig. 2).

Embryological data

Embryological outcomes were comparable between groups with a mean number of blastocysts (2.91 ± 2.11 versus 2.84 ± 2.12 ; DBM -0.07 [95% CI -0.67 to 0.53]) and a mean number of biopsied blastocysts (2.93 ± 2.16 versus 2.86 ± 2.15 ; DBM -0.07 [95% CI -0.70 to 0.56]) in the PPOS versus antagonist group, respectively. Concerning the euploidy rates per biopsied embryo, a 29.45% [95% CI $21.76\text{--}38.12\%$] and a 34.92% [95% CI $26.64\text{--}43.92\%$] were noticed in the PPOS and antagonist groups, respectively (Fig. 1).

Finally, no difference was observed for the primary outcome, with a mean number of euploid embryos of 0.86 ± 0.90 versus 1.00 ± 1.12 in the comparison of PPOS versus GnRh antagonist protocol (Table 3).

Table 1. Demographic characteristics of population (n = 44).

Age (years) ¹	39.59 (38.73, 40.34)
Antral follicle count (AFC) ¹	12 (9, 14)
Anti-Müllerian hormone (AMH) ¹	1.81 (1.47, 2.42)
BMI (kg/m ²) ¹	21.0 (19.6, 24.8)
Smoker ²	14 (32%)

¹ Mean (quartile 1, quartile 3).

² n (%).

Table 2. Characteristics and outcomes of ovarian stimulation.

	PPOS ¹	Antagonist ¹	Difference ²	95% CI ^{2,3}	P-value ²
Duration of stimulation (days)	9.36 (1.51)	9.80 (1.47)	0.43	−0.08 to 0.94	0.095
Total FSH consumption (IU)	2684.66 (537.33)	2800.57 (557.96)	116	−48 to 280	0.2
Estradiol on triggering day (pg/ml)	2743.40 (1402.45)	2224.05 (875.07)	−519	−1039	0.021
LH on triggering day (mIU/ml)	2.96 (2.58)	1.80 (1.79)	−1.2	−2.2 to −0.14	0.027
Progesterone on triggering day (ng/ml)	7.11 (4.39)	1.47 (0.84)	−5.6	−7.3 to −4.0	<0.001
Number of COCs	12.68 (8.09)	10.27 (5.84)	−2.4	−4.1 to −0.73	0.006
Number of MII	9.09 (6.12)	7.34 (4.15)	−1.8	−3.1 to −0.43	0.011
Number of 2PN	7.14 (4.99)	5.66 (3.35)	−1.5	−2.6 to −0.32	0.014

¹ Mean (SD).

² Paired t-test.

³ CI.

2PN: 2 pronuclei; COC: cumulus oocyte complex; MII: metaphase II; PPOS: progesterone-primed ovarian stimulation.

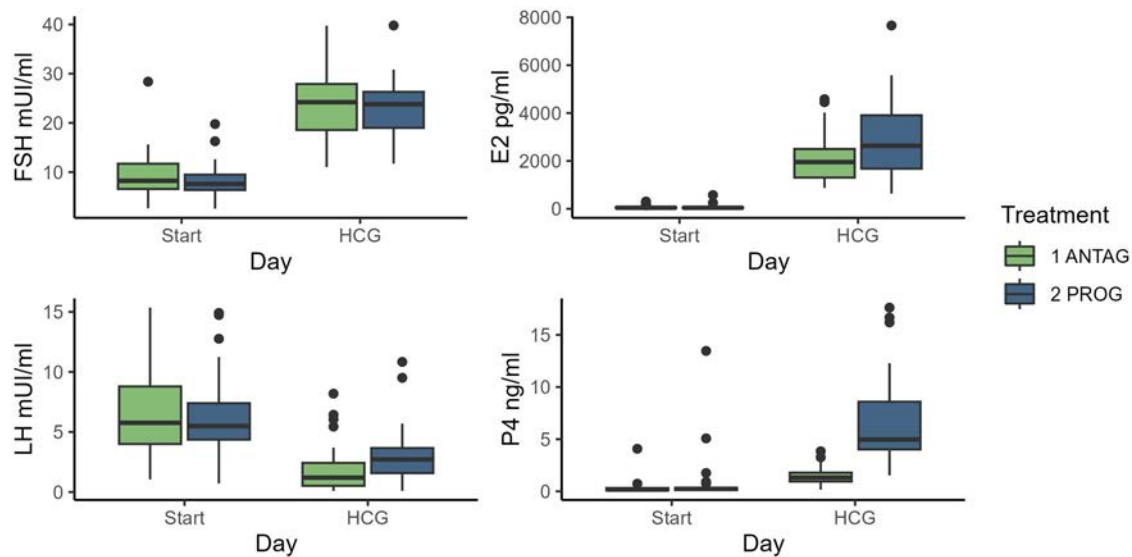


Figure 1. Hormonal profile of patients at ovarian stimulation Day 1 (SD1) and trigger day in antagonist protocol (ANT) and progesterone-primed ovarian stimulation protocol (PROG).

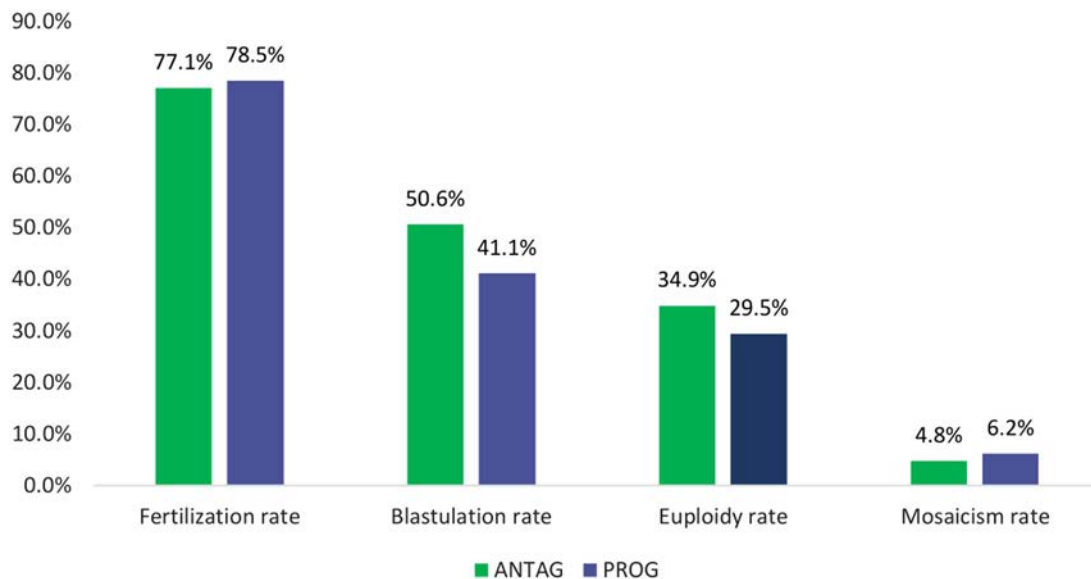


Figure 2. Outcomes according to antagonist protocol (ANT) or progesterone-primed ovarian stimulation protocol (PROG) for fertilization, blastocyst formation, euploidy, and mosaicism rates.

Discussion

To our knowledge, this is the first prospective study comparing the euploidy among embryos generated from different OS protocols (GnRH antagonist versus oral micronized progesterone PPOS) in the same patient. According to our findings, GnRH antagonist and oral micronized progesterone result in equivalent number of euploid blastocysts suggesting that PPOS can be a safe and friendly alternative, not only for oocyte donors or for patients undergoing fertility preservation (Massin, 2017; Martinez et al., 2021) but also for patients undergoing IVF in case of a freeze-all protocol (PGT-A or 'non-conventional' protocols) (La Marca et al., 2020; Ata et al., 2021).

Most of the available studies did not demonstrate any difference in terms of retrieved oocytes, number of embryos, and pregnancy rates between PPOS and GnRH analogues protocols (Kuang et al., 2015; Zhu et al., 2015; Ata et al., 2021).

Nevertheless, up to date, evidence from randomized clinical trials (RCTs) concerning clinical outcomes in PPOS as compared with antagonist cycles remain controversial. Although some authors demonstrated that PPOS protocols with medroxyprogesterone acetate (MPA) may results in inferior ongoing pregnancy rates in recipients (Beguiria et al., 2019) or women with PCOS (Eftekhar et al., 2019), studies carried out by other groups with desogestrel (Martínez et al., 2020; Devesa et al., 2022) and other RCTs with MPA (Giles et al., 2021) failed to find any difference in clinical outcomes. Besides, Yildiz et al. (2019) also reported similar clinical pregnancy rates between PPOS and antagonist cycles for oocyte recipients derived from the same donors.

Our study further reinforces these observations with the use of micronized progesterone by showing that PPOS with micronized progesterone results in comparable number of euploid blastocysts.

A recent observational study has been published during the elaboration of this manuscript by Giles et al. (2023), comparing oral MPA versus GnRH antagonist in fertility preservation and PGT-A treatments. Although their findings confirm our results in terms of retrieved oocytes, euploidy rate, and clinical outcomes, results should be interpreted with caution for several reasons. First of all, this was a retrospective study of PGT-A cycles and several of the baseline characteristics (such as AFC 14.3 ± 8.2 in the MPA group 13.5 ± 8.0 in the antagonist group, $P = 0.052$) may have differed between groups. Second, the starting stimulation dose was not standardized and patients in the MPA group appeared to have received higher total dose of stimulation (2861.8 ± 1020.6 , 2387.1 ± 1064.1 , $P < 0.0001$) despite the comparable duration of stimulation (10.5 ± 1.5 versus 10.8 ± 9.2 , $P = 0.537$). This in addition to the lack of multivariable regression analysis to control for such confounders cannot exclude the risk of residual bias. On the contrary, our prospective design with the use of identical doses and type of gonadotropins and the use of a very standardized protocol in the same patients allows us to support our conclusions with higher certainty.

Regarding euploid blastocyst rate, other authors have reported no differences between PPOS (either medroxyprogesterone 10 mg/day or dydrogesterone 20 mg/day) and GnRH antagonist protocols cycles in retrospective studies (Yang et al., 2022; Giles et al., 2023; Pai et al., 2023; Wang et al., 2023) or in a prospective age-matched case-control study (La Marca et al., 2020). Interestingly, Pai et al. reported a reduced euploidy rate per injected oocyte in 19 women ≥ 38 years treated with PPOS compared to 47 women treated with antagonist protocol (8.3 ± 6.1 versus 24.0 ± 4.2 , $P = 0.003$), while no difference was found in women < 38 years ($n = 15$ PPOS versus $n = 47$ antagonist). In our study, regardless all patients were ≥ 38 years, such difference was not noticed.

An interesting observation of our study is the difference observed between the two cycles in terms of endocrine profile. Although the hormonal profile of the participants was comparable at the initiation of OS in both treatment cycles, on the day of trigger, statistically significant differences between cycles were observed in all the tested hormones, except for FSH. As expected, the LH values decreased during the OS, and the LH levels in both groups were significantly lower the day of trigger compared to SD1, in agreement with others (Kuang et al., 2015). However, we found more elevated LH levels and higher progesterone levels in PPOS cycles as compared with antagonist cycles, in contrast with previous studies also using micronized progesterone to prevent LH surge (Zhu et al., 2015). Moreover, estradiol also was higher in the PPOS group. A potential explanation could be that the suppression effect of oral micronized progesterone on the pituitary may be much weaker than the GnRH antagonist, despite being efficient enough to effectively control LH pulses. This potentially lower pituitary inhibition potency of micronized progesterone (Schindler et al., 2003) may actually explain the higher serum LH levels observed in the oral micronized progesterone downregulated cycles.

In addition, given the fact that all cycles were triggered by a GnRH agonist, and taking into account that serum LH levels on the day of trigger are strictly associated with suboptimal response (Meyer et al., 2015) this may explain the higher number of retrieved oocytes, MII, and 2PN observed in our study. Nevertheless, a recent meta-analysis of prospective, retrospective, and randomized studies found no differences in the oocyte yield between PPOS and GnRH antagonists (Ata et al., 2021).

An important strength of our study is its prospective design aimed to minimize possible confounding factors. First of all, both OS cycles were performed in the same patient, with identical clinical protocols and with a fixed dose of rFSH that could not be adjusted with the only difference being the medication used to prevent the LH surge. Second, each patient underwent a washout period of at least one cycle. This was in order to minimize the risk for potential bias associated with a possible carry-over effect from the first stimulation cycle. As described in previous reports in the context of Duo_stim cycles, the number of oocytes

Table 3. Embryological and ploidy data.

	PPOS ¹	Antagonist ¹	Difference ²	95% CI ^{2,3}	P-value ²
Blastocysts	2.91 (2.11)	2.84 (2.12)	-0.07	-0.67 to 0.53	0.8
Biopsied blastocysts	2.93 (2.16)	2.86 (2.15)	-0.07	-0.70 to 0.56	0.8
Euploid embryos	0.86 (0.90)	1.00 (1.12)	0.14	-0.18 to 0.45	0.4
Aneuploid embryos	1.77 (1.57)	1.68 (1.43)	-0.09	-0.62 to 0.44	0.7
Mosaic embryos	0.18 (0.39)	0.14 (0.35)	-0.05	-0.19 to 0.10	0.5

¹ Mean (SD).

² Paired t-test.

³ CI.

retrieved may be higher in the second cycle (Alsbjerg et al., 2019). For this reason, we considered appropriate to adopt a 1-month washout interval. Finally, a severe male factor was excluded to avoid possible differences regarding that component.

Still, we also acknowledge some limitations. The study was powered to detect differences in the mean number of euploid embryos and not in terms of pregnancy outcomes. Additionally, per protocol, there was no randomization, the first cycle was always a GnRH antagonist cycle, and the second a PPOS with at least 1 month of washout period in between. Another limitation of our study is a lack of data regarding reproductive outcomes following the transfer of all the embryos.

In terms of clinical practice and future research, the approach with progestins allows an efficient control of pituitary LH secretion at lower costs of medication and a more-friendly access and administration for women. Our study for the first time further supports that the use of 200 mg of oral micronized progesterone may indeed result in higher number of oocytes, MIIs, and 2PNs as compared with a GnRH antagonist protocol, probably do the weaker, still efficient ovulation inhibition relative potency. In this context, it could be an efficient option in cycles in which freeze-all strategy is planned. This finding, in addition with the evidence provided that micronized progesterone PPOS results in equivalent euploid blastocyst as compared with the antagonist protocol, may imply that in case of a freeze-all protocol, clinicians may safely consider 200 mg of micronized progesterone in a PPOS protocol to control the LH surge and patients could benefit from the advantages.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Acknowledgements

This work was carried out under the auspices of the Càtedra d' Investigació en Obstetrícia i Ginecologia of the Department of Obstetrics and Gynecology, Hospital Universitari Quirón Dexeus, Universitat Autònoma de Barcelona.

Author's roles

M.d.M.V., F.M., and N.P.P. conceived and designed the study. All the authors analyzed and interpreted the data. M.d.M.V. contributed to the data collection and I.R. performed the statistical analysis. M.d.M.V., N.P.P., and F.M. wrote the article. All the authors approved the final version of the manuscript.

Funding

This research was supported by an unrestricted grant from Theramex.

Conflict of interest

N.P.P. has received Research grants from Merck Serono, Organon, Ferring Pharmaceutical, Roche, Theramex, IBSA, Gedeon Richter, and Besins Healthcare; honoraria for lectures from Merck Serono, Organon, Ferring Pharmaceuticals, Besins International, Roche Diagnostics, IBSA, Theramex, and Gedeon Richter; consulting fees from Merck Serono, Organon, Besins Healthcare, and IBSA. M.d.M.V., F.M., and I.R. declared no conflicts of interest.

References

- Alsbjerg B, Haahr T, Elbaek HO, Laursen R, Povlsen BB, Humaidan P. Dual stimulation using corifollitropin alfa in 54 Bologna criteria poor ovarian responders—a case series. *Reprod Biomed Online* 2019;**38**:677–682.
- Ata B, Capuzzo M, Turkogeldi E, Yildiz S, La Marca A. Progestins for pituitary suppression during ovarian stimulation for ART: a comprehensive and systematic review including meta-analyses. *Hum Reprod Update* 2021;**27**:48–66.
- Beguiria R, García D, Vassena R, Rodríguez A. Medroxyprogesterone acetate versus ganirelix in oocyte donation: a randomized controlled trial. *Hum Reprod* 2019;**34**:872–880.
- Boada M, Carrera M, De La Iglesia C, Sandalinas M, Barri PN, Veiga A. Successful use of a laser for human embryo biopsy in preimplantation genetic diagnosis: report of two cases. *J Assist Reprod Genet* 1998;**15**:302–307.
- Coll L, Parriego M, Boada M, Devesa M, Arroyo G, Rodríguez I, Coroleu B, Vidal F, Veiga A. Transition from blastomere to trophectoderm biopsy: comparing two preimplantation genetic testing for aneuploidies strategies. *Zygote* 2018;**26**:191–198.
- Devesa M, Racca A, Clúa E, Casato C, García S, Polyzos NP, Martínez F. Progesterone-primed ovarian stimulation in oocyte donation: a model for elective fertility preservation? *Reprod Biomed Online* 2022;**44**:1015–1022.
- Devroey P, Polyzos NP, Blockeel C. An OHSS-free clinic by segmentation of IVF treatment. *Hum Reprod* 2011;**26**:2593–2597.
- Eftekhari M, Hoseini M, Saeed L. Progesterone-primed ovarian stimulation in polycystic ovarian syndrome: an RCT. *Int J Reprod Biomed* 2019;**19**:579.
- Giles J, Alama P, Gamiz P, Vidal C, Badia P, Pellicer A, Bosch E. Medroxyprogesterone acetate is a useful alternative to a gonadotropin-releasing hormone antagonist in oocyte donation: a randomized, controlled trial. *Fertil Steril* 2021;**116**:404–412.
- Giles J, Cruz M, Cobo A, Vidal C, Requena A, Remohi J, Bosch E. Medroxyprogesterone acetate: an alternative to GnRH-antagonist in oocyte vitrification for social fertility preservation and preimplantation genetic testing for aneuploidy. *Reprod Biomed Online* 2023;**48**:103841.
- Kuang Y, Chen Q, Fu Y, Wang Y, Hong Q, Lyu Q, Ai A, Shoham Z. Medroxyprogesterone acetate is an effective oral alternative for preventing premature luteinizing hormone surges in women undergoing controlled ovarian hyperstimulation for in vitro fertilization. *Fertil Steril* 2015;**104**:62–70.e3.
- La Marca A, Capuzzo M. Use of progestins to inhibit spontaneous ovulation during ovarian stimulation: the beginning of a new era? *Reprod Biomed Online* 2019;**39**:321–331.
- La Marca A, Capuzzo M, Sacchi S, Imbrogno MG, Spinella F, Varricchio MT, Minasi MG, Greco P, Fiorentino F, Greco E. Comparison of euploidy rates of blastocysts in women treated with progestins or GnRH antagonist to prevent the luteinizing hormone surge during ovarian stimulation. *Hum Reprod* 2020;**35**:1325–1331.
- Martínez F, Clúa E, García S, Coroleu B, Polyzos NP, Barri PN. Does LH suppression by progesterone-primed ovarian stimulation compared with GnRH antagonist affect live birth rate among oocyte recipients? *Reprod Biomed Online* 2020;**40**:661–667.
- Martínez F, Racca A, Rodríguez I, Polyzos NP. Ovarian stimulation for oocyte donation: a systematic review and meta-analysis. *Hum Reprod Update* 2021;**27**:673–696.
- Massin N. New stimulation regimens: endogenous and exogenous progesterone use to block the LH surge during ovarian stimulation for IVF. *Hum Reprod Update* 2017;**23**:211–220.

- Meyer L, Murphy LA, Gumer A, Reichman DE, Rosenwaks Z, Cholist IN. Risk factors for a suboptimal response to gonadotropin-releasing hormone agonist trigger during in vitro fertilization cycles. *Fertil Steril* 2015;**104**:637–642.
- Pai AH-Y, Sung YJ, Li C-J, Lin C-Y, Chang CL. Progesterin primed ovarian stimulation (PPOS) protocol yields lower euploidy rate in older patients undergoing IVF. *Reprod Biol Endocrinol* 2023;**21**:72.
- Parriego M, Solé M, Aurell R, Barri PN, Veiga A. Birth after transfer of frozen-thawed vitrified biopsied blastocysts. *J Assist Reprod Genet* 2007;**24**:147–149.
- Schindler AE, Campagnoli C, Druckmann R, Huber J, Pasqualini JR, Schweppe KW, Thijssen JHH. Classification and pharmacology of progestins. *Maturitas* 2003;**46 Suppl 1**:S7–S16.
- Skinner DC, Evans NP, Delaleu B, Goodman RL, Bouchard P, Caraty A. The negative feedback actions of progesterone on gonadotropin-releasing hormone secretion are transduced by the classical progesterone receptor. *Proc Natl Acad Sci U S A* 1998;**95**:10978–10983.
- The ESHRE Guideline Group on Ovarian Stimulation; Bosch E, Broer S, Griesinger G, Grynberg M, Humaidan P, Kolibianakis E, Kunicki M, La Marca A, Lainas G, et al. ESHRE guideline: ovarian stimulation for IVF/ICSI†. *Hum Reprod Open* 2020;**2020**:hoaa009.
- Veiga A, Sandalinas M, Benkhalifa M, Boada M, Carrera M, Santaló J, Barri PN, Ménéz Y. Laser blastocyst biopsy for preimplantation diagnosis in the human. *Zygote* 1997;**5**:351–354.
- Verberg MFG, Eijkemans MJC, Heijnen EMEW, Broekmans FJ, de Klerk C, Fauser BCJM, Macklon NS. Why do couples drop-out from IVF treatment? A prospective cohort study. *Hum Reprod* 2008;**23**:2050–2055.
- Wang L, Wang J, Zhang Y, Qian C, Wang X, Bai J, Li F, Chen Z, Ai A. Analysis of euploidy rates in preimplantation genetic testing for aneuploidy cycles with progestin-primed versus GnRH agonist/antagonist protocol. *Eur J Med Res* 2023;**28**:28.
- Yang L, Luo K, Lu G, Lin G, Gong F. Euploidy rates among preimplantation genetic testing for aneuploidy cycles with oral dydrogesterone primed ovarian stimulation or GnRH antagonist protocol. *Reprod Biomed Online* 2022;**45**:721–726.
- Yildiz S, Turkgeldi E, Angun B, Eraslan A, Urman B, Ata B. Comparison of a novel flexible progestin primed ovarian stimulation protocol and the flexible gonadotropin-releasing hormone antagonist protocol for assisted reproductive technology. *Fertil Steril* 2019;**112**:677–683.
- Zhu X, Zhang X, Fu Y. Utrogestan as an effective oral alternative for preventing premature luteinizing hormone surges in women undergoing controlled ovarian hyperstimulation for in vitro fertilization. *Medicine (Baltimore)* 2015;**94**:e909.

Reproductive epidemiology

A history of recurrent pregnancy loss is associated with increased perinatal complications, but not necessarily a longer birth interval: a population study spanning 18 years

Kristy Cho ^{1,2}, Bahi Faye ^{1,2,3}, Yang (Doris) Liu ^{1,2}, Arianne Albert⁴, E. Wiesenthal^{1,2}, Sabina Dobrer⁴, Faten F. AbdelHafez ^{1,5}, Sarka Lisonkova^{2,4,6}, and Mohamed A. Bedaiwy ^{1,2,*}

¹Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

²Department of Obstetrics and Gynecology, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

³Department of Obstetrics and Gynecology, Faculty of Medicine, Menoufia University, Menoufia, Egypt

⁴Women's Health Research Institute, Vancouver, BC, Canada

⁵Department of Obstetrics and Gynecology, Faculty of Medicine, Assiut University, Assiut, Egypt

⁶School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

*Correspondence address. Department of Obstetrics and Gynaecology, BC Women's Hospital and Health Centre, Faculty of Medicine, The University of British Columbia, D415-4500 Oak Street, Vancouver, BC V6H 3N1, Canada. E-mail: mohamed.bedaiwy@cw.bc.ca  <https://orcid.org/0000-0002-3454-8555>

ABSTRACT

STUDY QUESTION: Is there a difference in the time interval between the first and second live births among individuals with and without recurrent pregnancy loss (RPL)?

SUMMARY ANSWER: Primary RPL (two or more pregnancy losses before the first live birth) is associated with a shorter time interval between the first and second live births compared with individuals without RPL, but this association is reversed in patients with secondary RPL (RPL patients with no or one pregnancy loss before the first live birth).

WHAT IS KNOWN ALREADY: There is limited information regarding the ability to have more than one child for patients with RPL. Previous studies have investigated the time to live birth and the live birth rate from the initial presentation to clinical providers. Most of the previous studies have included only patients treated at specialized RPL clinics and thus may be limited by selection bias, including patients with a more severe condition.

STUDY DESIGN, SIZE, DURATION: We conducted a population-based retrospective cohort study of 184 241 participants who delivered in British Columbia, Canada, and had at least two recorded live births between 2000 and 2018. The aim was to study the differences in the time interval between the first and second live births and the prevalence of pregnancy complications in patients with and without RPL. Additionally, 198 319 individuals with their first live birth between 2000 and 2010 were studied to evaluate cumulative second live birth rates.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Among individuals with at least two recorded live births between 2000 and 2018, 12 321 patients with RPL and 171 920 participants without RPL were included. RPL was defined as at least two pregnancy losses before 20 weeks gestation. Patients with primary RPL had at least two pregnancy losses occurring before the first live birth, while patients with secondary RPL had no or one pregnancy loss before the first live birth. We compared the time interval from the first to second live birth in patients with primary RPL, those with secondary RPL, and participants without RPL using generalized additive models to allow for a non-linear relationship between maternal age and time interval between first and second live births. We also compared prevalence of pregnancy complications at the first and second live births between the groups using non-parametric Kruskal–Wallis *H* test and Fisher's exact test for continuous and categorical variables, respectively. We assessed the cumulative second live birth rates in patients with primary RPL and those without RPL, among participants who had their first live birth between 2000 and 2010. Cox proportional hazards model was used to estimate and compare hazard ratios between the two groups using a stratified modelling approach.

MAIN RESULTS AND THE ROLE OF CHANCE: The adjusted time interval between the first and second live births was the longest in patients with secondary RPL, followed by individuals without RPL, and the shortest time interval was observed in patients with primary RPL: 4.34 years (95% CI: 4.09–4.58), 3.20 years (95% CI: 3.00–3.40), and 3.05 years (95% CI: 2.79–3.32). A higher frequency of pregnancy losses was associated with an increased time interval between the first and second live births. The prevalence of pregnancy complications at the first and second live births, including gestational diabetes, hypertensive disorder of pregnancy, preterm birth, and multiple gestations was significantly higher in patients with primary RPL compared with those without RPL. The cumulative second live birth rate was significantly lower in patients with primary RPL compared with individuals without RPL.

LIMITATIONS, REASONS FOR CAUTION: This study may be limited by its retrospective nature. Although we adjusted for multiple potential confounders, there may be residual confounding due to a lack of information about pregnancy intentions and other factors, including unreported pregnancy losses.

Received: November 10, 2021. Revised: January 24, 2024. Editorial decision: January 30, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved. For permissions, please email: journals.permissions@oup.com

WIDER IMPLICATIONS OF THE FINDINGS: The results of this study provide information that will help clinicians in the counselling of RPL patients who desire a second child.

STUDY FUNDING/COMPETING INTEREST(S): This study was supported in part by a grant from the Canadian Institutes of Health Research (CIHR): Reference Number W11-179912. M.A.B. reports research grants from CIHR and Ferring Pharmaceutical. He is also on the advisory board for AbbVie, Pfizer, and Baxter. The other authors report no conflict of interest.

TRIAL REGISTRATION NUMBER: NCT04360564.

Keywords: recurrent pregnancy loss / time to live birth / pregnancy complications / cumulative live birth rate / population study

Introduction

Recurrent pregnancy loss (RPL) is defined by the European Society of Human Reproduction and Embryology (ESHRE) as the loss of two or more pregnancies, including all pregnancy losses from conception until 24 weeks of gestation (Bender Atik et al., 2023). RPL is a distressing clinical condition, which affects around 2.5% of couples attempting to achieve pregnancy (Dimitriadis et al., 2020). Despite the rigorous research in this field, 50% of RPL patients have no explanation for the cause of their condition (Practice Committee of the American Society for Reproductive Medicine, 2012). Pregnancy loss can be classified as either visualized (clinical) or non-visualized (pre-clinical), depending on ultrasonographic and histopathological findings. The term 'non-visualized pregnancy loss' is employed when there is no clinical evidence of pregnancy observed through ultrasound evaluation or products of conception identified in histopathological examination (Kolte et al., 2015). There is limited information regarding the overall prognosis and ability to have more than one child for patients with RPL. Previous studies have investigated the time to live birth and live birth rate. However, it is worth noting that the majority of these studies have primarily examined the time interval between the initial healthcare provider visit and the subsequent pregnancy (Brigham et al., 1999; Lund et al., 2012; Shapira et al., 2012; Kaandorp et al., 2014). Moreover, the patient population in some prior studies has been limited to those who were referred to specialized RPL clinics or tertiary centres (Brigham et al., 1999; Kaandorp et al., 2010; Lund et al., 2012; Shapira et al., 2012). Thus, the reported time to subsequent live birth may be falsely prolonged in older patients who reside in urban settings or those with more severe medical conditions. The 2016 Statistics Canada population census reported the average fertility rate as 1.40 to 2.99 children per woman depending on the Canadian province (Provencher et al., 2018). In British Columbia (BC), the mean fertility rate was 1.45 children per woman during the period from 2006 to 2016 (Provencher et al., 2018) so we can extrapolate that around 50% of individuals with one child will attempt to conceive again.

Our study aimed to (i) compare the time interval between the first and second live births among patients with primary RPL (two or more pregnancy losses before the first live birth), those with secondary RPL (RPL patients with no or one pregnancy loss before the first live birth), and individuals without RPL, (ii) to compare the prevalence of pregnancy complications at the first and second live births among these three groups, and (iii) to investigate the cumulative rates of a second live birth among individuals with primary RPL compared with individuals without RPL in those who had their first live birth between 2000 and 2010. We also evaluated the impact of any adverse reproductive outcomes during the first live birth on the odds of having a second live birth.

Materials and methods

BC Perinatal Data Registry

This is a retrospective population-based observational study. We utilized data from the BC Perinatal Data Registry (BCPDR), which

is a provincially inclusive database that contains data abstracted from obstetrical and neonatal medical records on >99% of births in the province of BC, including births in over 60 hospitals and at-home births attended by BC registered midwives (Perinatal Services BC, 2014). There have been over 700 000 births included in this registry since data collection started in April 2000 (Perinatal Services BC, 2014). The study was approved by the University of British Columbia Ethics Committee (H20-01181).

Time interval between the first and second live births and prevalence of pregnancy complications

To compare the average time interval from the first to second live birth and the prevalence of pregnancy complications among patients with primary and secondary RPL and individuals without RPL, all participants with at least two live births recorded between 2000 and 2018 were included (Fig. 1A). A live birth was defined as the birth of a viable foetus at ≥ 20 weeks gestational age. Individuals with incomplete records and those who were multiparous at the first birth recorded in the BCPDR after 2000 were excluded in order to limit our cohort to a group with detailed information about both the first and second live births. This has also made the comparison groups more similar with respect to fecundability and family planning desires. Data on any pregnancy beyond the second live birth were not utilized. The study population was divided into three groups: individuals without RPL, patients with primary RPL, and those with secondary RPL. Primary RPL was defined as at least two pregnancy losses before the first live birth. Secondary RPL was defined as at least two pregnancy losses before the second live birth and no or one pregnancy loss prior to the first live birth.

Cumulative second live birth rate

To investigate the cumulative rates of a second live birth among primary RPL patients compared with individuals without RPL, individuals aged 15–49 years who had their first live birth between 2000 and 2010 were included in this analysis (Fig. 1B). Individuals included in this analysis were followed up until the 31st of March 2018 or 49 years of age, whichever came first, regardless whether they had a second live birth or not. This analysis provided more detailed results with respect to probability of having a second live birth, because it included not only participants who achieved two live births but also individuals who had no subsequent births at all after the first recorded birth between 2000 and 2010. The study population was stratified into two cohorts, with each cohort divided into RPL and non-RPL groups. Cohort I consisted of patients whose obstetric history included pregnancy losses without other adverse reproductive outcomes. Cohort II included patients with pregnancy losses and other adverse reproductive outcomes including ectopic and molar pregnancies, pregnancy termination, and stillbirths. In Cohort I, our aim was to quantify the influence of pregnancy losses on the cumulative second live birth rates. In Cohort II, we explored whether the presence of other adverse reproductive events,

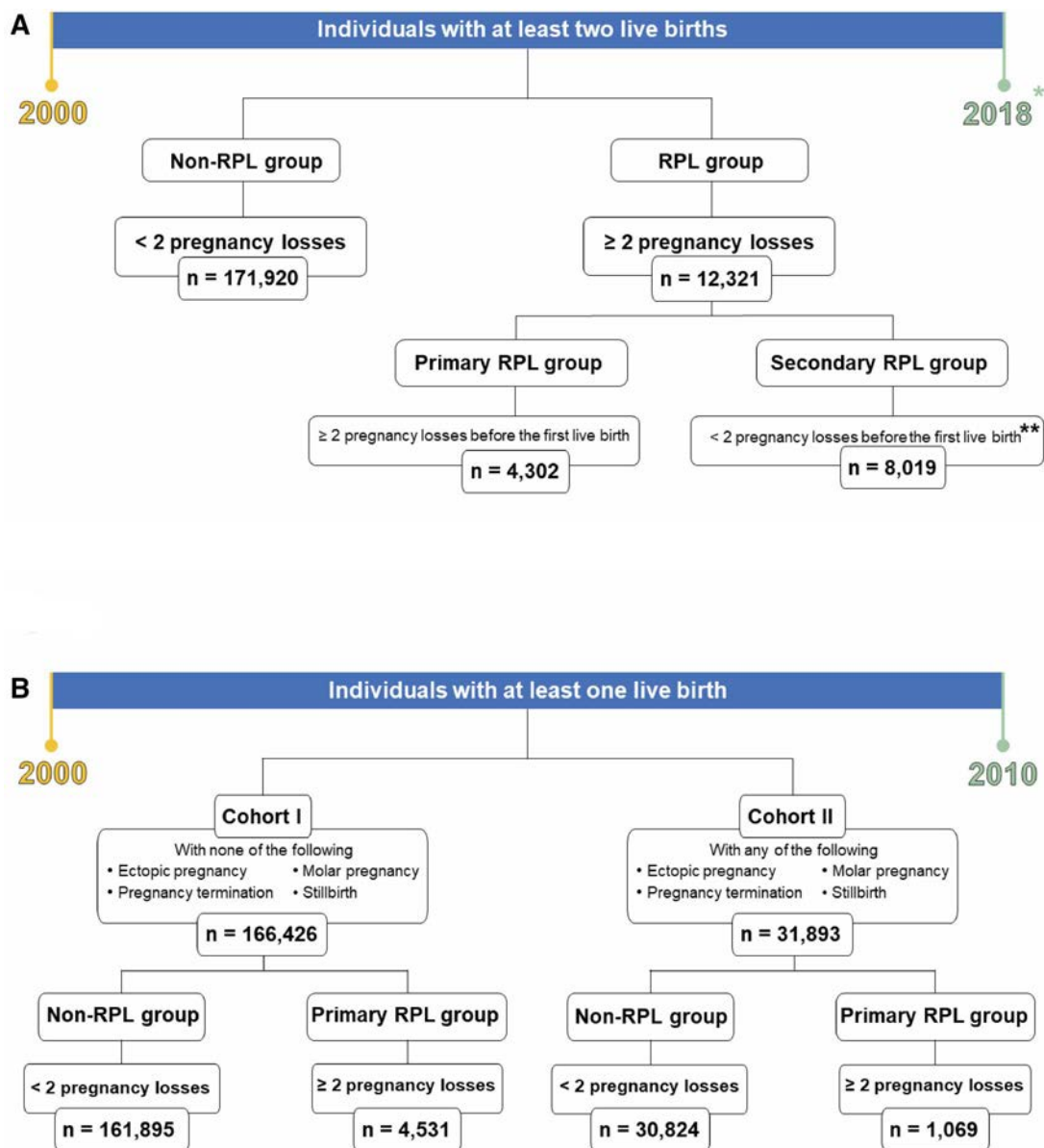


Figure 1. Flow chart of study population. (A) Flow chart of eligible patients for studying time interval between the first and second live births and prevalence of pregnancy complications. Patients were followed up until the 31st of March 2018 or to 49 years old, whichever came first. Secondary RPL patients: RPL patients with no or one pregnancy loss before first live birth. (B) Flow chart of eligible patients for studying the cumulative second live birth rate. RPL: recurrent pregnancy loss.

including stillbirth, ectopic pregnancy, molar pregnancy, and pregnancy termination, had any additional impact on the cumulative rates of second live birth. These comparative analyses allowed us to gain valuable insights into the complex dynamics affecting the cumulative rates of a second live birth following pregnancy losses and other adverse reproductive events.

Statistical analysis

Time interval between the first and second live births

Descriptive analyses were performed to compare the time interval between the first and second live births among patients with primary RPL, those with secondary RPL, and individuals without RPL. In multivariable analyses, we constructed a series of generalized additive models to allow for non-linear relationship between age at the first live birth and time interval between the first and second live births (Supplementary Table S1). Factors associated with the main exposure (RPL) and the main outcome of

interest (time interval between the first and second live births) were considered potential confounders. Factors associated with the exposure were evaluated using the non-parametric Kruskal-Wallis *H* test and Fisher's exact test for continuous and categorical data, respectively. Factors associated with the main outcome were evaluated using unadjusted generalized linear models. Additionally, in selecting the covariates for our statistical models, we carefully considered prior literature to identify variables that have been consistently associated with the time interval between the first and second live births. These covariates were chosen based on their established relevance to reproductive outcomes, as reported in previous studies. The potential confounders included maternal age (Zhang et al., 2019), geographical region (Zhang et al., 2019), pre-pregnancy BMI and gravidity (number of prior pregnancies) at the first live birth, as well as hypertensive disorders of pregnancy (HDP) (Roberts et al., 2005), gestational diabetes mellitus (GDM) (Venkatesh et al., 2022), preterm

birth (PTB) (Yee et al., 2016), multiple gestations (Fuchs and Senat, 2016), at both the first and second live births, and history of stillbirths prior to the first and second live births (İmamoğlu et al., 2022). All of the above-mentioned potential confounding factors were significantly different between the groups in unadjusted analyses ($P < 0.05$). The final model included all potential confounders. We examined age at first live birth as a potential effect modifier among individuals with two live births. We also assessed the effect of the number of pregnancy losses on the time interval between the first and second live births in the primary and secondary RPL groups.

Prevalence of pregnancy complications

The prevalence of pregnancy complications among patients with primary RPL, those with secondary RPL, and individuals without RPL was compared using the non-parametric Kruskal-Wallis H test and Fisher's exact test for continuous and categorical data, respectively.

Cumulative second live birth rate

The second set of analyses aimed to assess the cumulative second live birth rate, examining whether and when the second live birth occurred following the first live birth. A stratified analytical approach was used. Kaplan-Meier cumulative incidence curves with Gray's tests were used to calculate and compare the unadjusted cumulative second live birth rates between RPL and non-RPL groups in both Cohort I and Cohort II. Adjusted Cox proportional hazards model controlling for confounders (maternal age, geographical region, pre-pregnancy BMI, HDP, GDM, PTB, and multiple gestations at the first live birth) was used to compare cumulative second live birth rates in each cohort. Mann-Whitney U test and Fisher's exact test were used to compare the characteristics between RPL and non-RPL groups (a reference category) in each cohort for continuous and categorical variables, respectively.

Proportional hazard assumptions were tested using log (–log (survival)) vs log of survival time for categorical variables, and Schoenfeld residuals were examined for continuous variables, indicating no violations of the proportional hazard assumptions.

Data analysis

Data analysis was performed using the statistical software R (version 4.0.3; 2020, R core team) and SAS software version 9.4 (SAS Institute, Inc., Cary, NC, USA). Complete-case analyses were performed.

Results

Time interval between the first and second live births

A total of 184 241 individuals with two or more live births recorded between 2000 and 2018 were included in the analysis. Among these, 12 321 (6.7%) participants had an RPL diagnosis.

In the unadjusted analysis, the time interval between the first and second live births was significantly shorter for the primary RPL group (2.69 years, 95% CI: 2.62–2.75) compared with the non-RPL group (3.05 years, 95% CI: 3.04–3.06) and longer for the secondary RPL group (4.15 years, 95% CI: 4.10–4.20).

In the generalized additive model adjusted for maternal age, geographical region, pre-pregnancy BMI, and gravidity at the first live birth as well as HDP, GDM, PTB, multiple gestations at both the first and second live births, and history of stillbirths prior to the first and second live births, the time interval between the first and second live births in the primary RPL group was only

marginally shorter compared with the non-RPL group, 3.05 years (95% CI 2.79–3.32) vs 3.20 years (95% CI 3.00–3.40), respectively. It was relatively longer, 4.34 years (95% CI 4.09–4.58), in the secondary RPL group (Table 1).

Across all the three groups, older maternal age was associated with a shorter time interval between the first and second live births (Fig. 2). The time interval between the first and second live births changed from 3.17 years (95% CI 2.98–3.36) at the age of 25–2.33 years (95% CI 2.14–2.54) at the age of 40 among individuals in the non-RPL group, from 3.02 years (95% CI 2.82–3.23) to 2.19 years (95% CI 1.98–2.40) in the primary RPL group, and from 4.31 years (95% CI 4.11–4.50) to 3.47 years (95% CI 3.27–3.68) in the secondary RPL group at the specified maternal ages, respectively.

Prevalence of pregnancy complications

The same study population of 184 241 individuals with two or more live births recorded between 2000 and 2018 was included in the analysis. Patients with primary RPL were older (30.8 years [± 5.2] vs 28.2 years [± 5.1], $P < 0.001$) and had a higher gravidity at their first live birth compared with those without RPL ($P < 0.001$). There was also a slightly higher pre-pregnancy BMI in both RPL groups (Table 2). The prevalence of HDP, GDM, PTB and multiple gestations was significantly higher in primary RPL patients compared to those without RPL (Table 2).

Impact of the number of pregnancy losses on time interval between the first and second live births

We examined the effect of the number of pregnancy losses on the time interval between the first and second live births in the RPL groups (primary and secondary). The majority of individuals had three or fewer pregnancy losses prior to the second live birth, 3584/4302 (83.31%) in the primary RPL group and 7497/8019 (93.49%) in the secondary RPL group. There were 3470 (43.27%) patients in the secondary RPL group (of 8.019 patients) with one pregnancy loss prior to the first birth. There were 3302 (76.75%) patients in the primary RPL group did not have additional pregnancy losses between the two births (Table 2). The time interval between the first and second live births for primary RPL patients with no additional losses after the first live birth was 2.51 years (95% CI: 2.45–2.58), in contrast to 3.57 years (95% CI: 3.26–3.89) in patients with six or more pregnancy losses. Among secondary RPL patients with two pregnancy losses prior to the second live birth, the time interval between the first and second live births was 3.96 years (95% CI: 3.83–4.10), in contrast to 5.02 years (95% CI: 4.64–5.41) in patients with six or more pregnancy losses.

A larger number of pregnancy losses was associated with a longer time interval between the first and second live births.

Cumulative live birth rates

There were 198 319 participants included in cumulative live birth rates analysis that were stratified into two cohorts. Cohort I included 166 426 individuals, while Cohort II included 31 893 participants. Each cohort was divided into primary RPL and non-RPL groups (Fig. 1B).

Cohort I (patients with pregnancy losses without other adverse reproductive outcomes)

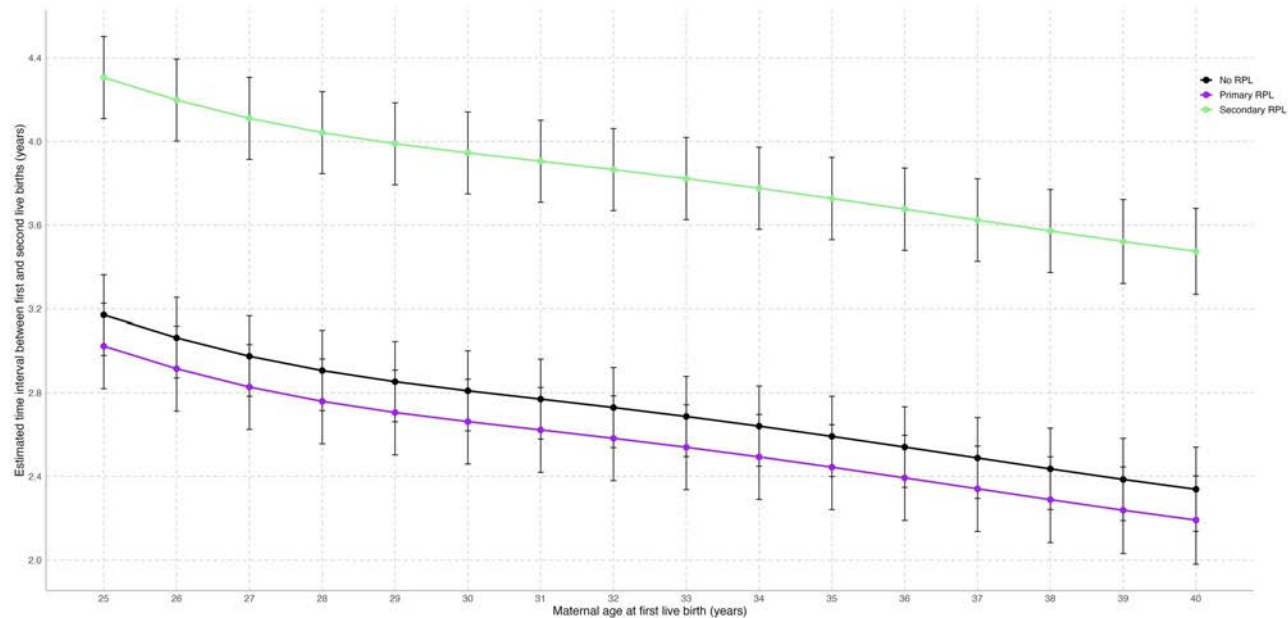
Cohort I included 4531 (2.7%) patients in the primary RPL group and 161 895 (97.3%) individuals in the non-RPL group. The median follow-up time was 5.50 years (interquartile range [IQR]: 7.97) in primary RPL patients and 3.75 years (IQR: 6.67) in individuals without RPL. Patients in the primary RPL group were older at

Table 1. Adjusted estimates of the time interval between the first and the second live births (years).

Predictors	Time between first and second live births (years)		
	Estimates	95% CI	P value*
Non-RPL group	3.20	3.00–3.40	
Primary RPL group	3.05	2.79–3.32	<0.001
Secondary RPL group	4.34	4.09–4.58	<0.001

RPL: recurrent pregnancy loss.

* Adjusted for maternal age, geographical health region, pre-pregnancy BMI, and gravidity at the first live birth as well as hypertensive disorders of pregnancy (HDP), gestational diabetes mellitus (GDM), preterm birth (PTB), multiple gestations, and previous stillbirths at both the first and second live births.

**Figure 2.** Relationship between the adjusted time interval between the first and second live births and maternal age at the time of the first live birth in the non-RPL group, primary RPL group, and secondary RPL group. RPL: recurrent pregnancy loss.

the first live birth compared with those in the non-RPL group, (32.2 ± 5.9 vs 28.5 ± 5.6 years; $P < 0.001$), respectively. Primary RPL patients also had a higher pre-pregnancy BMI at the first birth compared with non-RPL individuals (24.5 ± 5.3 vs 23.6 ± 4.9 kg/m², $P < 0.001$, respectively). RPL patients were more likely to have diagnosis of GDM or HDP and had higher proportions of a PTB or multiple-fetal pregnancy at the first live birth ($P < 0.001$ for all comparisons; Table 3).

The overall cumulative rate of the second live birth at the end of the study period in this cohort was 68.3%, with 52.3% in the primary RPL group and 68.7% in the non-RPL group (Fig. 3A).

The unadjusted hazard ratio of the second live birth in the RPL group was 0.70 (95% CI: 0.67–0.73) compared with the non-RPL group, and it was 0.86 (95% CI: 0.83–0.90) when adjusted for the covariates described above (Supplementary Table S2).

Cohort II (patients with pregnancy losses and other adverse reproductive outcomes)

Cohort II included 1069 (3.4%) patients in the primary RPL group and 30 824 (96.6%) participants in the non-RPL group at the first live birth. The median follow-up time was 7.33 years (IQR: 7.75) for patients with primary RPL and 5.33 years (IQR: 8.08) for individuals in the non-RPL group. The difference in the clinical characteristics between the primary RPL group and non-RPL group was similar to that observed in Cohort I (Table 4). The overall cumulative rate of the second live birth at the end of the study

period in this cohort was 58.3%, with 44.2% in the primary RPL group and 58.7% in the non-RPL group (Fig. 3B).

The unadjusted hazard ratio of the second live birth in the RPL compared with the non-RPL group was 0.73 (95% CI: 0.66–0.80) and it changed slightly to 0.88 (95% CI: 0.80–0.97) in the adjusted analysis (Supplementary Table S3). Clinical and socio-demographic characteristics of Cohort I and Cohort II are presented in Supplementary Table S4.

Overall, individuals in Cohort II had lower cumulative rates of the second live birth compared with those in Cohort I (Fig. 3).

Discussion

Summary of the main findings

To our knowledge, this study is the first to report specifically on the time interval between the first and second live births among patients with RPL as compared to those without RPL. The population-based birth data were collected over nearly two decades and showed that primary RPL was associated with a shorter time interval to second live birth compared with individuals without RPL; however, this association was reversed for patients with secondary RPL. Perinatal complications were elevated in patients with RPL. Patients with primary RPL without other adverse reproductive outcome had a lower cumulative second live birth rate than individuals without RPL or other adverse outcomes. Patients who experienced adverse reproductive outcomes in addition to pregnancy loss had a further decreased odds of

Table 2. Clinical and sociodemographic characteristics of individuals with at least two live births between 2000 and 2018.

	Total 184 241	Non-RPL group 171 920 (93.3%)	Primary RPL group 4302 (2.3%)	Secondary RPL group 8019 (4.4%)	P value*
Geographical region					
Fraser	71 350 (38.73%)	66 771 (38.84%)	1614 (37.52%)	2965 (36.97%)	<0.001
Interior	25 746 (13.97%)	24 064 (14.00%)	581 (13.51%)	1101 (13.73%)	
Island	26 064 (14.15%)	24 126 (14.03%)	729 (16.95%)	1209 (15.08%)	
Northern	14 395 (7.81%)	13 380 (7.78%)	321 (7.46%)	694 (8.65%)	
Vancouver Coastal	46 121 (25.03%)	43 044 (25.04%)	1047 (24.34%)	2030 (25.31%)	
Missing	565 (0.31%)	535 (0.31%)	10 (0.23%)	20 (0.25%)	
Age at first live birth (years)					
Mean ($\pm$ SD)	28.3 ($\pm$ 5.1)	28.2 ($\pm$ 5.1)	30.8 ($\pm$ 5.2)	29.2 ($\pm$ 5.6)	<0.001
Pre-pregnancy BMI at first live birth					
Mean ($\pm$ SD)	23.55 ($\pm$ 4.75)	23.52 ($\pm$ 4.73)	24.41 ($\pm$ 5.18)	23.87 ($\pm$ 4.85)	<0.001
Missing	45 492 (24.69%)	42 360 (24.64%)	1119 (26.01%)	2013 (25.10%)	
Gravida at first live birth					
Mean ($\pm$ SD)	1.34 ($\pm$ 0.83)	1.26 ($\pm$ 0.71)	3.61 ($\pm$ 1.09)	1.79 ($\pm$ 1.15)	<0.001
Used IVF for first live birth					
Yes	2160 (1.17%)	1870 (1.09%)	160 (3.72%)	130 (1.62%)	<0.001
No	76 338 (41.43%)	71 278 (41.46%)	1927 (44.79%)	3133 (39.07%)	
Missing	105 743 (57.39%)	98 772 (57.45%)	2215 (51.49%)	4756 (59.31%)	
Hypertensive disorder of pregnancy at first live birth					
Yes	12 010 (6.52%)	11 133 (6.48%)	330 (7.67%)	547 (6.81%)	0.004
No	172 231 (93.48%)	160 787 (93.52%)	3972 (92.33%)	7472 (93.18%)	
Hypertensive disorder of pregnancy at second live birth					
Yes	5773 (3.13%)	5287 (3.08%)	175 (4.07%)	311 (3.88%)	<0.001
No	178 468 (96.87%)	166 633 (96.92%)	4127 (95.93%)	7708 (96.12%)	
Gestational diabetes mellitus at first live birth					
Yes	11 561 (6.27%)	10 578 (6.15%)	421 (9.79%)	562 (7.01%)	<0.001
No	172 680 (93.73%)	161 342 (93.85%)	3881 (90.21%)	7457 (92.99%)	
Gestational diabetes mellitus at second live birth					
Yes	16 853 (9.15%)	15 434 (8.98%)	517 (12.02%)	902 (11.25%)	<0.001
No	167 388 (90.85%)	156 486 (91.02%)	3785 (87.98%)	7117 (88.75%)	
Preterm at first live birth					
Extreme preterm (<32 weeks)	2156 (1.17%)	1913 (1.11%)	87 (2.02%)	156 (1.95%)	<0.001
Preterm (32–36 + 6 weeks)	12 488 (6.78%)	11 501 (6.69%)	353 (8.21%)	634 (7.91%)	
Term	169 384 (91.94%)	158 314 (92.09%)	3854 (89.59%)	7216 (89.99%)	
Missing	213 (0.12%)	192 (0.11%)	8 (0.19%)	13 (0.16%)	
Preterm at second live birth					
Extreme preterm (<32 weeks)	1760 (0.96%)	1561 (0.91%)	57 (1.32%)	142 (1.77%)	<0.001
Preterm (32–36 + 6 weeks)	12 945 (7.03%)	11 741 (6.83%)	442 (10.27%)	762 (9.50%)	
Term	169 382 (91.94%)	158 472 (92.18%)	3800 (88.33%)	7110 (88.66%)	
Missing	154 (0.08%)	146 (0.08%)	3 (0.07%)	5 (0.06%)	
Multiple gestations at first live birth					
Singleton	182 998 (99.33%)	170 784 (99.34%)	4257 (98.95%)	7957 (99.23%)	0.007
Multiple gestations	1243 (0.67%)	1136 (0.66%)	45 (1.05%)	62 (0.77%)	
Multiple gestations at second live birth					
Singleton	181 693 (98.62%)	169 624 (98.66%)	4195 (97.51%)	7874 (98.19%)	<0.001
Multiple gestations	2548 (1.38%)	2296 (1.34%)	107 (2.49%)	145 (1.81%)	
Number of patients with stillbirths prior to their first live birth					
Yes	53 (0.03%)	34 (0.02%)	13 (0.30%)	6 (0.07%)	<0.001
No	184 188 (99.97%)	171886 (99.98%)	4289 (99.70%)	8013 (99.93%)	
Number of patients with stillbirths prior to their first and second live births					
Yes	1259 (0.68%)	1104 (0.64%)	62 (1.44%)	93 (1.16%)	<0.001
No	182 982 (99.32%)	170 816 (99.36%)	4240 (98.56%)	7926 (98.84%)	
Number of patients with stillbirths between their first and second live births					
Yes	1206 (0.65%)	1070 (0.62%)	49 (1.14%)	87 (1.08%)	<0.001
No	183 035 (99.35%)	170 850 (99.38%)	4253 (98.86%)	7932 (98.92%)	
Number of patients with pregnancy losses before the first live birth within the RPL groups					
Patients without pregnancy loss			0 (0%)	4549 (56.73%)	
Patients with 1 pregnancy loss			0 (0%)	3470 (43.27%)	
Patients with 2 pregnancy losses			3197 (74.31%)	0 (0%)	
Patients with 3 pregnancy losses			784 (18.22%)	0 (0%)	
Patients with 4–5 pregnancy losses			267 (6.21%)	0 (0%)	
Patients with $\geq$ 6 pregnancy losses			54 (1.26%)	0 (0%)	
Number of patients with cumulative pregnancy losses before the first and second live births within the RPL groups					
Patients with 2 pregnancy losses			2505 (58.23%)	6136 (76.52%)	
Patients with 3 pregnancy losses			1079 (25.08%)	1361 (16.97%)	
Patients with 4–5 pregnancy losses			572 (13.30%)	439 (5.47%)	
Patients with $\geq$ 6 pregnancy losses			146 (3.39%)	83 (1.04%)	
Number of patients with pregnancy losses that occurred between the first and second live births within the RPL groups					
Patients without pregnancy loss			3302 (76.75%)	0 (0%)	
Patients with 1 pregnancy loss			690 (16.04%)	2695 (33.61%)	
Patients with 2 pregnancy losses			203 (4.72%)	4002 (49.91%)	

(continued)

Table 2. Continued.

	Total 184 241	Non-RPL group 171 920 (93.3%)	Primary RPL group 4302 (2.3%)	Secondary RPL group 8019 (4.4%)	P value*
Patients with 3 pregnancy losses			59 (1.37%)	944 (11.77%)	
Patients with 4–5 pregnancy losses			36 (0.84%)	313 (3.90%)	
Patients with ≥6 pregnancy losses			12 (0.28%)	65 (0.81%)	

RPL: recurrent pregnancy loss.

* Kruskal–Wallis *H* test for continuous variables; Fisher's exact test for categorical variables.

Table 3. Clinical and sociodemographic characteristics of Cohort I (patients without pregnancy complications).

	Total 166 426	Non-RPL group 161 895 (97.3%)	Primary RPL group 4531 (2.7%)	P value*
Geographical region				
Fraser	63 968 (38.4%)	62 343 (38.5%)	1625 (35.9%)	<0.001
Interior	23 141 (13.9%)	22 525 (13.9%)	616 (13.6%)	
Island	22 990 (13.8%)	22 277 (13.8%)	713 (15.7%)	
Northern	12 640 (7.6%)	12 306 (7.6%)	334 (7.4%)	
Vancouver Coastal	42 869 (25.8%)	41 650 (25.7%)	1219 (26.9%)	
Unknown	818 (0.5%)	794 (0.5%)	24 (0.5%)	
Age at first live birth (years)				
Mean (±SD)	28.6 (±5.7)	28.5 (±5.6)	32.2 (±5.9)	<0.001
<30	96 710 (58.1%)	95 186 (58.8%)	1524 (33.6%)	
30–35	46 986 (28.2%)	45 518 (28.1%)	1468 (32.4%)	
35–40	19 270 (11.6%)	18 163 (11.2%)	1107 (24.4%)	
40–45	3460 (2.1%)	3028 (1.9%)	432 (9.5%)	
Pre-pregnancy BMI at first live birth				
Mean (±SD)	23.6 (±4.9)	23.6 (±4.9)	24.5 (±5.3)	<0.001
Missing	44 915 (27.0%)	43 625 (27.0%)	1290 (28.5%)	
Hypertensive disorder of pregnancy in first live birth				
Yes	12 019 (7.2%)	11 602 (7.2%)	417 (9.2%)	<0.001
Gestational diabetes mellitus in first birth				
Yes	2515 (1.5%)	2359 (1.5%)	156 (3.4%)	<0.001
Preterm at first live birth				
Extreme preterm (<32 weeks)	2448 (1.5%)	2322 (1.4%)	126 (2.8%)	<0.001
Preterm (32–36 + 6 weeks)	12 392 (7.4%)	11 937 (7.4%)	455 (10.0%)	
Term	151 586 (91.1%)	147 636 (91.2%)	3950 (87.2%)	
Multiple gestations at first live birth				
Singleton	163 682 (98.4%)	159 299 (98.4%)	4383 (96.7%)	<0.001
Multiple gestations	2744 (1.6%)	2596 (1.6%)	148 (3.3%)	

RPL: recurrent pregnancy loss.

* Mann–Whitney *U* test for continuous variables; Fisher's exact test for categorical variables.

having a second live birth compared to those who experienced pregnancy loss without adverse reproductive outcomes. The study results provide information that will help clinicians in the counselling of RPL patients who desire a second child.

Discussion of findings and comparison with literature

Time interval between the first and second live births

Our study showed an adjusted time interval to second live birth of 3.05 years (95% CI: 2.79–3.32) for primary RPL patients and 4.34 years (95% CI: 1.91–4.58) for those with secondary RPL. A previous study reported that the time from RPL diagnosis to the subsequent live birth was 102 weeks or 1.96 years (IQR 82–115 weeks), and this interval was considered to be comparable to that reported for individuals without RPL (Kaandorp *et al.*, 2014). The relatively longer time interval between the first and second live births in our study is likely due to the fact that we report the interval from the first to second live birth rather than the time to live birth from the RPL diagnosis or enrolment in an RPL clinic. The time interval between the first and second live births is an important outcome measure, which provides information on the longer-term childbearing trajectory of RPL patients. Moreover, the population included in our study comprised all patients with

RPL in the province. Thus, our study provides unbiased and likely a more accurate prognosis for RPL patients than prior studies which were limited by a sample of patients with access to an RPL clinic or with more severe conditions (e.g., greater number of pregnancy losses and/or older age).

A previous study by Kaandorp *et al.* suggested that there was no statistically significant difference in the time to live birth associated with secondary RPL compared with primary RPL (hazard ratio of 1.14, $P = 0.46$) (Kaandorp *et al.*, 2014). Similarly, other observational studies found no significant difference in the likelihood of subsequent live birth in patients with secondary RPL compared with those with primary RPL (Brigham *et al.*, 1999; Shapira *et al.*, 2012; Lee *et al.*, 2016). In our study, primary RPL was associated with the shorter interval to the second live birth, 3.05 years (95% CI: 2.62–2.75), compared with individuals without RPL, 3.20 years (95% CI: 3.00–3.40), which can be explained by the fact that primary RPL patients may intend to conceive sooner after the first live birth compared with those without RPL. However, a longer interval to the second child was observed in our cohort of secondary RPL patients, 4.34 years (95% CI 1.09–1.18), which may reflect the definition of secondary RPL that includes additional pregnancy losses after the first live birth. Our study showed that an increased number of pregnancy losses led

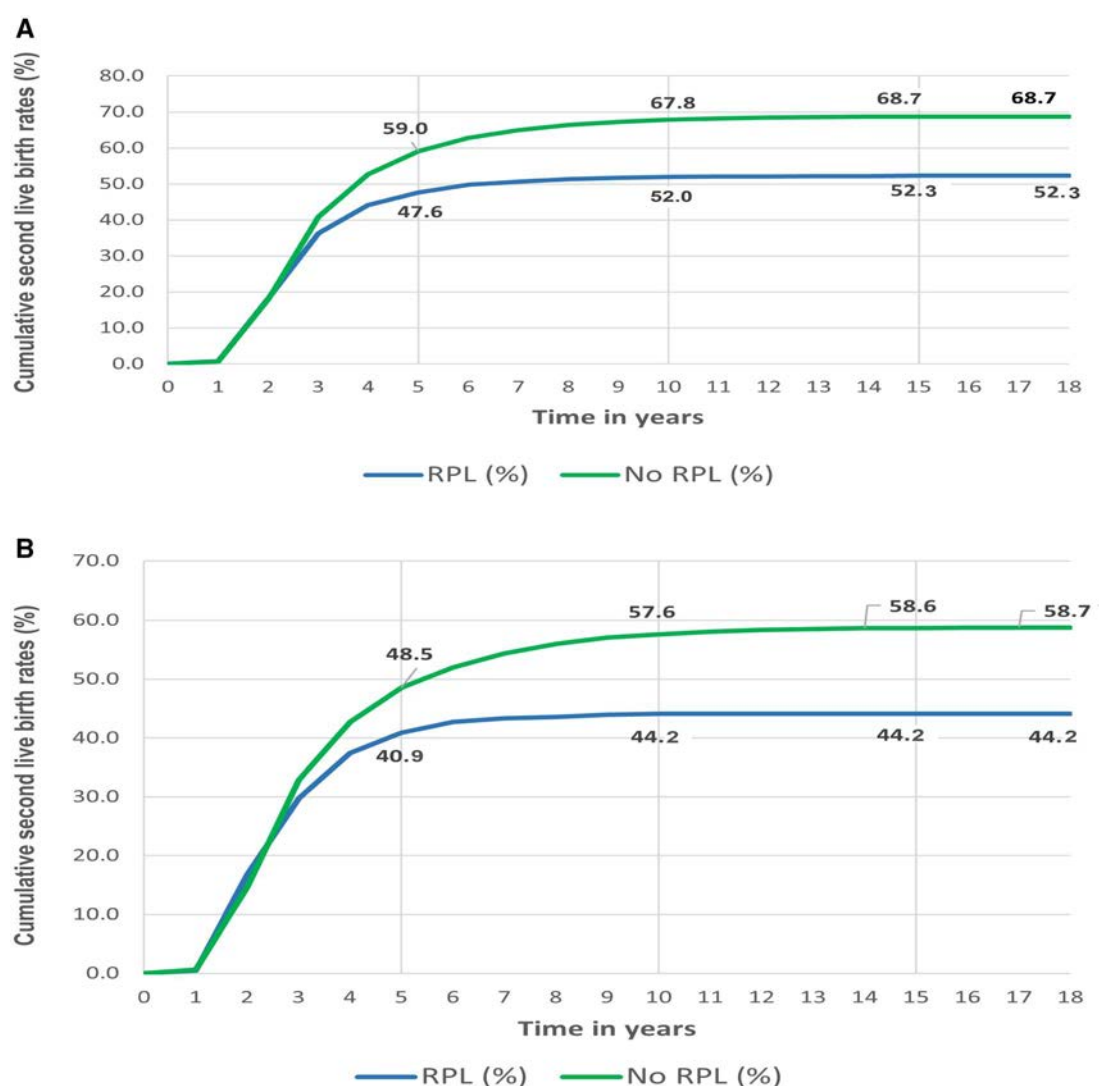


Figure 3. Cumulative second live birth rates. (A) Cumulative second live birth rates in Cohort I (those without other adverse reproductive outcomes). (B) Cumulative second live birth rates in Cohort II (those with other adverse reproductive outcomes). RPL: recurrent pregnancy loss.

to a longer time interval between the first and second live births. Likewise, in prior studies, the number of prior miscarriages was the sole predictive factor related to the time to achieve a pregnancy ending in a live birth in patients with unexplained recurrent miscarriages (Kaandorp et al., 2014). A greater number of pregnancy losses was also found to be associated with a longer time to conception in RPL patients in other publications (Lund et al., 2012; Kaandorp et al., 2014).

It is possible that family planning intentions differed among the groups. However, since the population database does not contain pregnancy intentions or other qualitative data regarding patients' attitudes, this study was unable to provide any insight regarding the influence of this factor. Our study contributes to the literature by showing that from a prognostic perspective, primary RPL is not associated with a delayed time interval between the first and second live births.

Traumatic childbirth experiences can profoundly affect a woman's decision to have a second child. Such experiences may lead to emotional distress, anxiety, and post-traumatic stress disorder, making the prospect of another pregnancy and childbirth overwhelming (Ertan et al., 2021; Parry-Smith et al., 2021; Shorey et al., 2023). Differences in the attitudes towards and desire for a

second pregnancy between the RPL and non-RPL, if any, may have resulted in residual confounding.

Perinatal complications

In our study, patients with primary RPL had a significantly higher prevalence of HDP, GDM, PTB and multiple gestations at the first and second live births compared with those without RPL, which is consistent with the existing literature (van Oppenraaij et al., 2009; Nielsen et al., 2010; Shapira et al., 2012; Field and Murphy, 2015; Rasmak Roepke et al., 2021). Shapira et al. reported that the rates of PTB, GDM, and fetal intrauterine growth restriction were higher in patients with primary RPL, compared with secondary RPL (Shapira et al., 2012). In a Danish population study, secondary RPL was associated with higher incidence of stillbirth, pre-eclampsia, placental abruption, PTB, and birth weight below 2500g, compared with the control group (Nielsen et al., 2010). While a larger epidemiological cohort study of over 30 000 participants with singleton pregnancies showed increased rates of PTB and perinatal death in those with a history of RPL, they did not find such differences between those with primary vs secondary RPL (Field and Murphy, 2015). Our study results are similar to a recent Swedish population-based cohort study, which showed

Table 4. Clinical and sociodemographic characteristics of Cohort II (patients with pregnancy complications).

	Total 31 893	Non-RPL group 30 824 (96.6%)	Primary RPL group 1069 (3.4%)	P value*
Geographical region				
Fraser	10 065 (31.6%)	9701 (31.5%)	364 (34.1%)	<0.001
Interior	3975 (12.5%)	3852 (12.5%)	123 (11.5%)	
Island	5729 (18.0%)	5518 (17.9%)	211 (19.7%)	
Northern	2049 (6.4%)	1973 (6.4%)	76 (7.1%)	
Vancouver Coastal	9922 (31.1%)	9632 (31.1%)	290 (27.1%)	
Unknown	153 (0.5%)	148 (0.5%)	5 (0.5%)	
Age at first live birth (years)				
Mean ($\pm$ SD)	29.6 ($\pm$ 5.8)	29.4 ($\pm$ 5.7)	33.0 ($\pm$ 6.1)	<0.001
<30	16 484 (51.7%)	16 133 (52.3%)	351 (32.8%)	
30–35	9331 (29.3%)	9041 (29.3%)	290 (27.1%)	
35–40	4981 (15.6%)	4701 (15.3%)	280 (26.2%)	
40–45	1097 (3.4%)	949 (3.1%)	148 (13.8%)	
Pre-pregnancy BMI at first live birth				
Mean ($\pm$ SD)	23.5 ($\pm$ 4.7)	23.5 ($\pm$ 4.7)	24.4 ($\pm$ 4.8)	<0.001
Missing	7814 (24.5%)	7527 (24.4%)	287 (26.8%)	
Hypertensive disorder of pregnancy in first live birth				
Yes	1992 (6.2%)	1909 (6.2%)	83 (7.8%)	0.037
Gestational diabetes mellitus in first birth				
Yes	536 (1.7%)	505 (1.6%)	31 (2.9%)	0.002
Preterm at first live birth				
Extreme preterm (<32wks)	515 (1.6%)	485 (1.6%)	30 (2.8%)	<0.001
Preterm (32–36 + 6wks)	2516 (7.9%)	2396 (7.8%)	120 (11.2%)	
Term	28 862 (90.5%)	27 943 (90.7%)	919 (86%)	
Multiple gestations at first live birth				
Singleton	31 350 (98.3%)	30 311 (98.3%)	1039 (97.2%)	0.005
Multiple gestations	543 (1.7%)	513 (1.7%)	30 (2.8%)	

RPL: recurrent pregnancy loss.

* Mann-Whitney U test for continuous variables; Fisher's exact test for categorical variables.

that primary and secondary RPL was associated with higher risk of PTB, placental abruption and preeclampsia, compared with individuals without RPL (Rasmak Roepke *et al.*, 2021). We were unable to assess the contribution of confounding factors, such as pre-existing maternal medical conditions and types of treatment used, as these were not collected in this database. However, pregnancy loss itself can be a risk factor for developing pregnancy complications, and their incidence increases with increasing number of pregnancy losses. A recent retrospective cohort study concluded that a history of pregnancy losses was related to a higher risk of GDM in future pregnancies (Zhao *et al.*, 2022). In a population-based retrospective analysis comparing the risk of diabetes mellitus and hypertension between individuals who had experienced miscarriage and those who had not, those who had miscarriages were more likely to develop HDP or GDM (Okoth *et al.*, 2022). Advanced maternal age is also a contributing factor for perinatal complications in the RPL groups. Many studies found that pregnant individuals, who developed gestational complications such as HDP and GDM, were of a significantly higher maternal age (Li *et al.*, 2020; McLaughlin *et al.*, 2021). However, in our study, the difference in age, although statistically significant, was not clinically meaningful because the patients in all three groups were of a similar age.

A large retrospective cohort study showed that a history of RPL was associated with obesity and the use of assisted reproductive technologies (Field and Murphy, 2015). We also observed a comparatively higher prevalence of IVF use and higher BMI in patients with RPL and non-missing data.

Cumulative second live birth rates

Among patients with a history of RPL, the cumulative second birth rate was 47.6% by the fifth year and 52.0% by the tenth year in Cohort I, and 40.9% by the fifth year and 44.2% by the tenth

year in Cohort II. The cumulative birth rates reported in this study are overall comparable to prior studies, including those by Brigham *et al.* reporting a 51% live birth rate over 10 years, Lund *et al.* showing a rate of 66.7% over 5 years, and Green *et al.* reporting a 60% cumulative rate over 5 years (Brigham *et al.*, 1999; Lund *et al.*, 2012; Kling *et al.*, 2018; Green and O'Donoghue, 2019; Bashiri *et al.*, 2022).

Our results indicate that the overall prognosis for individuals with RPL is favourable even though the cumulative live birth rate is lower than those without RPL. There was no difference in the cumulative second live birth rate between patients with RPL and those without until the third year of follow-up, with a slight change in the probability of conceiving between the third and the fifth year in favour of non-RPL participants. The gap between cumulative incidences after year 5 remained steady. The results were consistent for both Cohort I and Cohort II, with slightly lower rates over time in Cohort II. This is consistent with the findings from Wildenschild *et al.*, who reported a similar probability of conception within the 12 cycles of follow-up of RPL patients compared with individuals without RPL, 85% (95% CI: 73–92%) vs 88% (95% CI: 87–89%), respectively (Wildenschild *et al.*, 2019).

Our findings are consistent with a large nationwide registry-based cohort study which found that pregnancy trajectory affected the estimated likelihood of a subsequent live birth (Kolte *et al.*, 2021). However, the authors concluded that the number of pregnancy losses by itself was not sufficient to predict the live birth outcome, as obstetric morbidities, stillbirth, and ectopic pregnancies also had negative impacts on achieving live birth in the subsequent pregnancy (Kolte *et al.*, 2021).

Strengths and limitations

To our knowledge, this is the first study to date that examines the time interval between the first and second live births in

individuals with RPL on a population level. Most of the previous studies reported the time interval to live birth from the time of RPL clinic visit, which may be biased by inclusion of patients with more severe conditions. Our study is adequately large to provide robust results. In addition, we adjusted for pregnancy complications at the first and second live births, which may have impacted the time interval between the first and second live births. We also adjusted for geographic regions to control for potential variations in clinical management of RPL and/or the post-miscarriage counselling by different healthcare providers in the different geographic regions of the province, which might have differentially affected the time interval between the first and second live births among the studied groups. Furthermore, our study included both pre-clinical pregnancy losses as well as clinically recognized miscarriages. Pre-clinical pregnancy losses were reported to affect reproductive outcomes in the same way as clinical miscarriages (Kolte et al., 2014; Bedaiwy et al., 2023). We excluded multiparous women at the first delivery reported occurring after the year 2000 in order to create a more homogeneous cohort in terms of fecundability and, possibly, similar family planning desires. However, we acknowledge that fecundability is a complex concept influenced by a diverse range of factors and it is not exclusively determined by an individual's age, reproductive history or desire for more children. Lastly, we provided prognostic information with respect to the cumulative second live birth rates for individuals with pregnancy loss who did not experience any other adverse reproductive outcomes as their prognosis differ from those who experienced, in addition to pregnancy loss, adverse reproductive outcomes, such as ectopic or molar pregnancy, pregnancy termination or stillbirth.

Our study has also several limitations related to the retrospective cohort design of this observational study. First, the incidence of RPL was slightly higher in our study population (6.6%) than that reported in previous studies (Larsen et al., 2013; Sugiura-Ogasawara et al., 2013). This difference could be due to the self-reported pregnancy losses, which is subject to recall. The observed discrepancy could stem from various factors, including variations in the definitions of RPL adopted across different studies. In our investigation, we adhered to the recently published guidelines by the ESHRE (Bender Atik et al., 2023) for defining our RPL cohort. We have included a broad spectrum of pregnancy losses, encompassing both non-visualized (pre-clinical) and visualized (clinical) losses, regardless of whether they are consecutive or not. Conversely, RPL defined as three consecutive pregnancy losses before 20 weeks of gestation from the last menstrual period, results in a lower prevalence of RPL (1–2% of women) (Ford and Schust, 2009). In a different study conducted by Larsen and colleagues in 2013, the prevalence of RPL ranged from 0.8% to 1.4% when examining clinical miscarriages confirmed by ultrasound or histology. The inclusion of biochemical losses increased the prevalence to 2–3% (Larsen et al., 2013). Differences in reported data across studies can also be attributed to the utilization of different denominators. Some studies employed women of reproductive age as their denominator, while others focused on populations identified to have higher risks of RPL or on women or couples actively engaged in conception attempts. In a retrospective analysis of a registry involving individuals with RPL, two RPL incidences were derived utilizing different groups as the denominator for calculation. The initial incidence was 0.053% when computed from women aged 18–42 years, while the other incidence was 0.65% among pregnant women (Rasmak Roepke et al., 2017). These differing denominators highlight the importance of considering specific study populations when interpreting

and comparing findings related to RPL. Furthermore, we utilized data from BCPDR, a database covering over 99% of births in BC, for a more comprehensive dataset. This broader scope likely captured a greater range of pregnancy losses, as it includes data from the vast majority of women giving birth in BC during the study period compared to other studies restricted to patients who present to health care providers. Second, we did not have information about family planning intentions, which affect decisions about a second pregnancy and the time interval between the first and second live births; thus, we were unable to control for these factors regarding the time interval between the first and second live births. While the literature showed that various demographic or behavioural factors and adverse or traumatic experience during childbirth affect individuals' desire for more children (Shreffler et al., 2015; Mumford et al., 2016; Ertan et al., 2021; Parry-Smith et al., 2021; Shorey et al., 2023), we did not attempt to speculate about such desires as the objective of our study was not to ascertain causal associations. Instead, our aim was to provide prognostic information with respect to reproductive prospects of women with RPL, drawing on an unselected large population data. Lastly, our study was not powered to examine a possible effect modification by multiple gestations on the relationship between RPL and time interval between the first and second live births. Although we adjusted for multiple gestations at the first and second live births in the model, the number of individuals who had multiple gestations at the first live birth was insufficient to perform stratified analyses. Nevertheless, from a prognostic perspective, our study provides information on the average time interval between the first and second live births among patients with RPL who succeeded in having a second child.

Implications for practice and future research

Our study results provide additional information that might help clinicians in the counselling of RPL patients, specifically those who desire to have a second child. The results indicate that the time interval between the first and second live births is not prolonged for patients with primary RPL compared with those without RPL; thus, our results are specifically applicable to long-term family planning counselling for RPL patients who desire more than one child. Future research should include longitudinal studies of RPL patients with a follow-up over an extended period. Such studies could also explore the impact of various interventions on subsequent pregnancy outcomes. Furthermore, effects of behavioural factors, family pregnancy intentions, and experiences of traumatic or adverse events during childbirth, that are not typically available in population databases, should be addressed in the future studies. In addition, future research should focus specifically on patients with unexplained RPL to uncover specific risk factors and pathophysiology that contribute to these pregnancy losses and on developing predictive models for future reproductive outcomes.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data utilized in this study were sourced from BCPDR under data use agreements and confidentiality commitments, ensuring that the data would remain private and not be shared or disseminated.

Acknowledgements

We thank Emma Branch and Nicole Prestley for their assistance with accessing the BCPDR. We thank Perinatal Services BC and the Women's Health Research Institute for maintaining the BCPDR. We thank Jeyah Cruz for editing and improving the language of the manuscript.

Authors' roles

All authors met the criteria for authorship. M.A.B. granted external funding for the study. K.C., A.A., and M.A.B. designed the study. K.C., A.A., and M.A.B. executed the study. K.C., A.A., and M.A.B. collected the data. K.C., A.A., and S.D. analysed the data. K.C., B.F., Y.D.L., S.D., F.F.A., and M.A.B. interpreted the results. K.C., A.A., E.W., and M.A.B. prepared the first draft of the manuscript. B.F., Y.D.L., and F.F.A. edited the manuscript. B.F., Y.D.L., S.D., F.F.A., S.L., and M.A.B. critically revised the article. K.C., B.F., S.D., S.L., and M.A.B. prepared the responses to the reviewers' comments and finalized the manuscript. All authors approved the final version of the article and were accountable for all aspects of the work.

Funding

This study was supported in part by a grant from the Canadian Institutes of Health Research (CIHR): Reference Number W11-179912.

Conflict of interest

M.A.B. is on the advisory board for AbbVie, Pfizer, and Baxter. M.A.B. also reports research grants from the Canadian Institutes of Health Research (CIHR) and Ferring Pharmaceutical. The other authors report no conflict of interest.

References

- Bashiri A, Giliutin M, Ziedenberg H, Plakht Y, Baumfeld Y. A proposed prognostic prediction tool for a live birth among women with recurrent pregnancy loss. *J Matern Fetal Neonatal Med* 2022; **35**:3736–3742.
- Bedaiwy MA, Fayek B, Yang EC, Iews MS, Elgendi M, Abdelkareem AO, Abdallah AH, Liu YD, Jones A, Mayer U et al. Prevalence, causes, and impact of non-visualized pregnancy losses in a recurrent pregnancy loss population. *Hum Reprod* 2023; **38**:830–839.
- Brigham SA, Conlon C, Farquharson RG. A longitudinal study of pregnancy outcome following idiopathic recurrent miscarriage. *Hum Reprod* 1999; **14**:2868–2871.
- Dimitriadis E, Menkhorst E, Saito S, Kuttah WH, Brosens JJ. Recurrent pregnancy loss. *Nat Rev Dis Primers* 2020; **6**:98.
- Ertan D, Hingray C, Burlacu E, Sterlé A, El-Hage W. Post-traumatic stress disorder following childbirth. *BMC Psychiatry* 2021; **21**:155.
- Bender Atik R, Christiansen OB, Elson J, Kolte AM, Lewis S, Middeldorp S, Mcheik S, Peramo B, Quenby S, Nielsen HS et al.; ESHRE Guideline Group on RPL. ESHRE guideline: recurrent pregnancy loss: an update in 2022. *Hum Reprod Open* 2023; **2023**:hoad002.
- Field K, Murphy DJ. Perinatal outcomes in a subsequent pregnancy among women who have experienced recurrent miscarriage: a retrospective cohort study. *Hum Reprod* 2015; **30**:1239–1245.
- Ford HB, Schust DJ. Recurrent pregnancy loss: etiology, diagnosis, and therapy. *Rev Obstet Gynecol* 2009; **2**:76–83.
- Fuchs F, Senat MV. Multiple gestations and preterm birth. *Semin Fetal Neonatal Med* 2016; **21**:113–120.
- Green DM, O'Donoghue K. A review of reproductive outcomes of women with two consecutive miscarriages and no living child. *J Obstet Gynaecol* 2019; **39**:816–821.
- İmamoğlu M, Şimşek D, Dincgez B, Ünal S, Demirci A, İlhan O, İmamoğlu AG, Özçeltik G. Short interdelivery interval in modern obstetrics: maternal and neonatal outcomes. *Turk J Obstet Gynecol* 2022; **19**:295–301.
- Kaandorp SP, van Mens TE, Middeldorp S, Hutten BA, Hof MH, van der Post JA, van der Veen F, Goddijn M. Time to conception and time to live birth in women with unexplained recurrent miscarriage. *Hum Reprod* 2014; **29**:1146–1152.
- Kaandorp SP, Goddijn M, van der Post JA, Hutten BA, Verhoeve HR, Hamulyák K, Mol BW, Folkeringa N, Nahuis M, Papatsonis DN et al. Aspirin plus heparin or aspirin alone in women with recurrent miscarriage. *N Engl J Med* 2010; **362**:1586–1596.
- Kling C, Hedderich J, Kabelitz D. Fertility after recurrent miscarriages: results of an observational cohort study. *Arch Gynecol Obstet* 2018; **297**:205–219.
- Kolte AM, van Oppenraaij RH, Quenby S, Farquharson RG, Stephenson M, Goddijn M, Christiansen OB; ESHRE Special Interest Group Early Pregnancy. Non-visualized pregnancy losses are prognostically important for unexplained recurrent miscarriage. *Hum Reprod* 2014; **29**:931–937.
- Kolte AM, Bernardi LA, Christiansen OB, Quenby S, Farquharson RG, Goddijn M, Stephenson MD; ESHRE Special Interest Group, Early Pregnancy. Terminology for pregnancy loss prior to viability: a consensus statement from the ESHRE early pregnancy special interest group. *Hum Reprod* 2015; **30**:495–498.
- Kolte AM, Westergaard D, Lidegaard Ø, Brunak S, Nielsen HS. Chance of live birth: a nationwide, registry-based cohort study. *Hum Reprod* 2021; **36**:1065–1073.
- Larsen EC, Christiansen OB, Kolte AM, Macklon N. New insights into mechanisms behind miscarriage. *BMC Med* 2013; **11**:154.
- Lee GS, Park JC, Rhee JH, Kim JI. Etiologic characteristics and index pregnancy outcomes of recurrent pregnancy losses in Korean women. *Obstet Gynecol Sci* 2016; **59**:379–387.
- Li G, Wei T, Ni W, Zhang A, Zhang J, Xing Y, Xing Q. Incidence and risk factors of gestational diabetes mellitus: a prospective cohort study in Qingdao, China. *Front Endocrinol (Lausanne)* 2020; **11**:636.
- Lund M, Kamper-Jørgensen M, Nielsen HS, Lidegaard Ø, Andersen AM, Christiansen OB. Prognosis for live birth in women with recurrent miscarriage: what is the best measure of success? *Obstet Gynecol* 2012; **119**:37–43.
- McLaughlin K, Audette M, Carmona S, Kingdom J. Clinical phenotype & outcomes of women who develop severe hypertension in pregnancy 2020. *Canadian Journal of Cardiology* 2021; **37**:e7.
- Mumford SL, Sapra KJ, King RB, Louis JF, Buck Louis GM. Pregnancy intentions-a complex construct and call for new measures. *Fertil Steril* 2016; **106**:1453–1462.
- Nielsen HS, Steffensen R, Lund M, Egestad L, Mortensen LH, Andersen AM, Lidegaard Ø, Christiansen OB. Frequency and impact of obstetric complications prior and subsequent to unexplained secondary recurrent miscarriage. *Hum Reprod* 2010; **25**:1543–1552.
- Okoth K, Subramanian A, Chandan JS, Adderley NJ, Thomas GN, Nirantharakumar K, Antza C. Long term miscarriage-related hypertension and diabetes mellitus. Evidence from a United Kingdom population-based cohort study. *PLoS One* 2022; **17**:e0261769.
- Parry-Smith W, Okoth K, Subramanian A, Gokhale KM, Chandan JS, Humpston C, Coomarasamy A, Nirantharakumar K, Šumilo D. Postpartum haemorrhage and risk of mental ill health: a

- population-based longitudinal study using linked primary and secondary care databases. *J Psychiatr Res* 2021;**137**:419–425.
- Perinatal Services BC. Perinatal Data Registry, 2014. <http://www.perinatal-servicesbc.ca/health-professionals/data-surveillance/perinatal-data-registry>.
- Practice Committee of the American Society for Reproductive Medicine. Evaluation and treatment of recurrent pregnancy loss: a committee opinion. *Fertil Steril* 2012;**98**:1103–1111.
- Provencher C, Milan A, Hallman S, Aoust CD. Fertility: overview, 2012 to 2016. *Stat Canada Cat* 2018;**91**:1–7.
- Rasmak Roepke E, Christiansen OB, Källén K, Hansson SR. Women with a history of recurrent pregnancy loss are a high-risk population for adverse obstetrical outcome: a retrospective cohort study. *J Clin Med* 2021;**10**:179.
- Rasmak Roepke E, Matthiesen L, Rylance R, Christiansen OB. Is the incidence of recurrent pregnancy loss increasing? A retrospective register-based study in Sweden. *Acta Obstet Gynecol Scand* 2017;**96**:1365–1372.
- Roberts CL, Algert CS, Morris JM, Ford JB, Henderson-Smart DJ. Hypertensive disorders in pregnancy: a population-based study. *Med J Aust* 2005;**182**:332–335.
- Shapira E, Ratzon R, Shoham-Vardi I, Serjienko R, Mazor M, Bashiri A. Primary vs. secondary recurrent pregnancy loss—epidemiological characteristics, etiology, and next pregnancy outcome. *J Perinat Med* 2012;**40**:389–396.
- Shorey S, Downe S, Chua JYX, Byrne SO, Fobelets M, Lalor JG. Effectiveness of psychological interventions to improve the mental well-being of parents who have experienced traumatic childbirth: a systematic review and meta-analysis. *Trauma Violence Abuse* 2023;**24**:1238–1253.
- Shreffler KM, Greil AL, Mitchell KS, McQuillan J. Variation in pregnancy intendedness across U.S. women's pregnancies. *Matern Child Health J* 2015;**19**:932–938.
- Sugiura-Ogasawara M, Suzuki S, Ozaki Y, Katano K, Suzumori N, Kitaori T. Frequency of recurrent spontaneous abortion and its influence on further marital relationship and illness: the Okazaki Cohort Study in Japan. *J Obstet Gynaecol Res* 2013;**39**:126–131.
- van Oppenraaij RHF, Jauniaux E, Christiansen OB, Horcajadas JA, Farquharson RG, Exalto N; ESHRE Special Interest Group for Early Pregnancy (SIGEP). Predicting adverse obstetric outcome after early pregnancy events and complications: a review. *Hum Reprod Update* 2009;**15**:409–421.
- Venkatesh KK, Lynch CD, Powe CE, Costantine MM, Thung SF, Gabbe SG, Grobman WA, Landon MB. Risk of adverse pregnancy outcomes among pregnant individuals with gestational diabetes by race and ethnicity in the United States, 2014–2020. *JAMA* 2022;**327**:1356–1367.
- Wildenschild C, Riis AH, Ehrenstein V, Hatch EE, Wise LA, Rothman KJ, Sørensen HT, Mikkelsen EM. Fecundability among Danish women with a history of miscarriage: a prospective cohort study. *BMJ Open* 2019;**9**:e023996.
- Yee LM, Truong YN, Caughey AB, Cheng YW. The association between interdelivery interval and adverse perinatal outcomes in a diverse US population. *J Perinatol* 2016;**36**:593–597.
- Zhang Y, Quist A, Enquobahrie D. Short birth-to-pregnancy intervals among African-born black women in Washington State. *J Matern Fetal Neonatal Med* 2019;**32**:947–953.
- Zhao Y, Zhao Y, Fan K, Jin L. Association of history of spontaneous or induced abortion with subsequent risk of gestational diabetes. *JAMA Netw Open* 2022;**5**:e220944.

Reproductive genetics

Ready for polygenic risk scores? An analysis of regulation of preimplantation genetic testing in European countries

M. Siemann ^{1,2,*}, V. van der Schoot³, E.M. Bunnik ⁴, and P. Borry ¹

¹Centre of Biomedical Ethics and Law, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

²Department of Physiology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

³Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands

⁴Department of Medical Ethics, Philosophy and History of Medicine, Erasmus MC, Rotterdam, The Netherlands

*Correspondence address. Centre for Biomedical Ethics and Law, Department of Public Health and Primary Care, KU Leuven, Kapucijnenvoer 7, Box 7001, 3000 Leuven, Belgium. E-mail: maria.siemann@kuleuven.be  <https://orcid.org/0000-0002-1289-1512>

ABSTRACT

STUDY QUESTION: Would the different regulatory approaches for preimplantation genetic testing (PGT) in Europe permit the implementation of preimplantation genetic testing using polygenic risk scores (PGT-P)?

SUMMARY ANSWER: While the regulatory approaches for PGT differ between countries, the space provided for potential implementation of PGT-P seems limited in all three regulatory models.

WHAT IS KNOWN ALREADY: PGT is a reproductive genetic technology that allows the testing for hereditary genetic disorders and chromosome abnormalities in embryos before implantation. Throughout its history, PGT has largely been regarded as an ethically sensitive technology. For example, ethical questions have been raised regarding the use of PGT for adult-onset conditions, non-medical sex selection, and human leukocyte antigen typing for the benefit of existing siblings. Countries in which PGT is offered each have their own approach of regulating the clinical application of PGT, and a clear overview of legal and practical regulation of PGT in Europe is lacking. An emerging development within the field of PGT, namely PGT-P, is currently bringing new ethical tensions to the forefront. It is unclear whether PGT-P may be applied within the current regulatory frameworks in Europe. Therefore, it is important to investigate current regulatory frameworks in Europe and determine whether PGT-P fits within these frameworks.

STUDY DESIGN, SIZE, DURATION: The aim of this study was to provide an overview of the legal and practical regulation of the use of PGT in seven selected European countries (Belgium, France, Germany, Italy, the Netherlands, Spain, and the UK) and critically analyse the different approaches with regards to regulatory possibilities for PGT-P. Between July and September 2023, we performed a thorough and extensive search of websites of governments and governmental agencies, websites of scientific and professional organizations, and academic articles in which laws and regulations are described.

PARTICIPANTS/MATERIALS, SETTING, METHODS: We investigated the legal and regulatory aspects of PGT by analysing legal documents, regulatory frameworks, scientific articles, and guidelines from scientific organizations and regulatory bodies to gather relevant information about each included country. The main sources of information were national laws relating to PGT.

MAIN RESULTS AND THE ROLE OF CHANCE: We divided the PGT regulation approaches into three models. The regulation of PGT differs per country, with some countries requiring central approval of PGT for each new indication (the medical indication model: the UK, the Netherlands), other countries evaluating each individual PGT request at the local level (the individual requests model: France, Germany), and countries largely leaving decision-making about clinical application of PGT to healthcare professionals (the clinical assessment model: Belgium, Italy, Spain). In the countries surveyed that use the medical indication model and the individual requests model, current legal frameworks and PGT criteria seem to exclude PGT-P. In countries using the clinical assessment model, the fact that healthcare professionals and scientific organizations in Europe are generally negative about implementation of PGT-P due to scientific and socio-ethical concerns, implies that, even if it were legally possible, the chance that PGT-P would be offered in the near future might be low.

LIMITATIONS, REASONS FOR CAUTION: The results are based on our interpretation of publicly available written information and documents, therefore not all potential discrepancies between law and practice might have been identified. In addition, our analysis focuses on seven—and not all—European countries. However, since these countries are relevant players within PGT in Europe and since they have distinct PGT regulations, the insights gathered give relevant insights into diverse ways of PGT regulation.

WIDER IMPLICATIONS OF THE FINDINGS: To the best of our knowledge, this is the first paper that provides a thorough overview of the legal and practical regulation of PGT in Europe. Our analysis of how PGT-P fits within current regulation models provides guidance for healthcare professionals and policymakers in navigating the possible future implementation of PGT-P within Europe.

Received: November 29, 2023. Revised: February 19, 2024. Editorial decision: February 23, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

All rights reserved. For permissions, please email: journals.permissions@oup.com

STUDY FUNDING/COMPETING INTEREST(s): This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 813707. The authors declare no conflict of interest.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: preimplantation genetic testing / PGT-P / polygenic risk scores / regulation / legislation / Europe / ethics

Introduction

Preimplantation genetic testing (PGT) is a reproductive genetic technology that allows the testing or screening of embryos for hereditary genetic disorders and chromosome abnormalities. PGT is mostly used by prospective parents who want to avoid passing on a genetic condition to their offspring (De Rycke et al., 2020). Scientific advancements and the expanding scope of PGT raise ethical and regulatory questions that are important to address.

For PGT, embryos are created with IVF or ICSI before being biopsied and genetically screened or tested (De Rycke et al., 2020). Various types of PGT exist. PGT-M tests embryos for the presence of a monogenic condition and can be used if prospective parents know they are at high risk of having a child with a monogenic condition. Monogenic conditions are caused by pathogenic variants in one gene with an autosomal recessive, autosomal dominant, or X-linked inheritance. In case of autosomal recessive monogenic conditions, one is affected if one inherits the gene with a mutation from both parents. Those with a mutation on one gene ('carriers') have the potential to pass the condition to their offspring if their partner carries a mutation in the same gene, but they are generally not affected themselves. With autosomal dominant monogenic conditions, one is at risk of developing the condition if the mutation is inherited from one of the parents. X-linked conditions are caused by mutations in genes on the X-chromosome. The risk of developing X-linked conditions partially depends on one's combination of sex chromosomes. Generally, men have an increased risk of being affected by an X-linked recessive condition (Chial, 2008). PGT-SR can be used for prospective parents who are at risk of structural chromosomal rearrangements in order to avoid an unbalanced chromosome abnormality, mostly leading to miscarriages, congenital anomalies, and developmental problems (De Rycke et al., 2020). PGT-A aims to select euploid (i.e. chromosomally normal) embryos in order to increase IVF success rate per embryo transfer, though its clinical benefit remains a topic of discussion (Simpson et al., 2019; De Rycke et al., 2020). PGT can also be used for selection based on sex chromosomes. This selection could either be performed for medical purposes, for example for X-linked conditions or for conditions that largely affect one sex, or more controversially, in a few countries for non-medical purposes, for example social sex selection including family balancing (Lemke and Rüppel, 2018; De Rycke et al., 2020). Human leukocyte antigen (HLA) typing could be used, often in combination with PGT-M, to select embryos that are HLA-compatible with an affected sibling in need of a hematopoietic stem cell transplantation (Simpson et al., 2019; De Rycke et al., 2020; Ginoza and Isasi, 2020).

PGT is regarded as an ethically sensitive technology and throughout its 30-year history, various socio-ethical concerns have been raised regarding its applications and expanding scope. For example, the use of PGT for conditions that will not develop until adulthood has led to concerns around its benefits, where to draw the line, and the person's 'right to an open future'. Furthermore, the introduction of PGT for indications without full penetrance, such as hereditary breast and ovarian cancer (BRCA1/2 mutations), have led to ethical questions not only about adult-onset versus childhood-onset conditions, but also

about risk versus certainty. The use of PGT for non-medical sex selection has also been met with ethical debates and is associated with concerns about sex discrimination (Lemke and Rüppel, 2018; Ginoza and Isasi, 2020; Siermann et al., 2022a). PGT-HLA raised concerns about creating so-called 'saviour siblings' and the instrumentalization of one person for the health of another person (Lemke and Rüppel, 2018). Considering these ethical concerns, many countries restrict the use of PGT and have legal frameworks in place that limit the scope of PGT, for example prohibition of non-medical sex selection and requirements related to 'seriousness' or 'severity' of a condition (Lemke and Rüppel, 2018; Ginoza and Isasi, 2020; Siermann et al., 2022b).

An emerging development within the field of PGT is currently bringing new ethical tensions to the forefront. Recently, the use of polygenic risk scoring in embryo selection, in which embryos are screened for their genetic susceptibility to various multifactorial conditions or potentially even non-medical traits, has been introduced in a commercial setting in the USA. This is referred to as PGT-P or polygenic embryo screening (Treff et al., 2019a,b). The validity and utility of PGT-P have been quantified in several scientific publications (Treff et al., 2019a,b, 2020; Lello et al., 2020, 2023; Lello and Telier, 2021; Lencz et al., 2021; Widen et al., 2022). This research, for example based on simulated data derived from biobanks, has shown that relative risk reduction via embryo selection based on polygenic risk scores (PRS) is possible for various polygenic indications, such as type 1 diabetes and hypothyroidism (Treff et al., 2019a,b, 2020; Widen et al., 2022). The expected risk reduction depends on the selection strategy used, with the achieved risk reduction being higher when selecting the embryo with the lowest PRS than when excluding the embryo with very high PRS (Lencz et al., 2021).

However, some see the potential of PGT-P to reduce risk as rather limited, and are concerned about its validity, utility, and probabilistic nature. Concerns have furthermore been raised regarding psychological impact on prospective parents and socio-ethical implications related to eugenics, commercialization, and discrimination (Karavani et al., 2019; Lázaro-Muñoz et al., 2021; Siermann et al., 2023, 2024). While many healthcare professionals in Europe and North-America have concerns about the clinical introduction of PGT-P (Barlevy et al., 2024; Siermann et al., 2023, 2024), the general population and some IVF users in the United States are seemingly more positive about PGT-P (Eccles et al., 2021; Zhang et al., 2021; Barlevy et al., 2024; Furrer et al., 2023; Meyer et al., 2023; Neuhausser et al., 2023).

In light of this new and controversial development within PGT, it is important to investigate whether PGT-P fits within current regulatory and legal frameworks in Europe. Because of the many discussions surrounding PGT-P's validity, utility, and ethics, it is imperative to see whether implementation of PGT-P would currently be possible within European countries at all. This is especially relevant because the characteristics of PGT-P distinctly differ from current PGT offerings. By investigating how the current regulatory frameworks are applicable to PGT-P, the future of PGT-P and potential need for more guidance can be navigated. Little information has yet been published on how exactly the (clinical) application of PGT is legally and practically

arranged in different European countries, and how and by whom decisions are made regarding approval for PGT requests or indications. Bayefsky (2016) previously compared PGT policies of a few countries, but did not focus specifically on how the approval process is organized practically for each included country. For this reason, we present an overview of the current regulatory frameworks for PGT in seven European countries that have diverse legislation and history regarding PGT, focussed particularly on how PGT approval functions in practice. By analysing how the demarcation of the scope of PGT is currently arranged in Belgium, France, Germany, Italy, the Netherlands, Spain, and the UK, we aim to provide a baseline of possible models for PGT regulation. The aim is not to give an overview of PGT and its regulation in Europe at large, but rather to give in-depth insights into the regulatory processes of a few selected countries. This overview of current PGT regulation and models in the selected countries provides a basis to examine what these regulatory models would mean for the opportunities to clinically introduce PGT-P in these countries.

Materials and methods

We have reviewed legal documents, regulatory frameworks, scientific articles, guidelines from scientific organizations and regulatory bodies, and relevant online material to gather relevant information about each country included in the analysis. Since much of the information relevant for our research question is published in grey literature, we did not focus on academic databases, but performed a thorough and extensive search of websites of governments and governmental agencies, websites of scientific and professional organizations, and academic articles in which laws and regulations are described. We used snowballing (i.e. reviewing the references of relevant documents) to find

additional sources relevant for this analysis. The search took place between July and September 2023. The main sources of information were national laws relating to PGT. Our analysis is based on publicly available written information and documents, with efforts to map the clinical practice that follow from laws and regulation. Not all potential discrepancies between law and practice might have been identified. More insight and openness about the way PGT evaluation and approval are organized in practice in each country would be valuable. Importantly, this analysis is based on our interpretations of the law and other relevant documents. It is not meant to provide a legal standpoint or recommendation regarding the use of PGT-P and its regulatory possibilities in Europe.

We have summarized all relevant aspects of legal and regulatory aspects of PGT for each included country, in particular related to how approval decisions are made for PGT requests and indications, after which we investigated the main characteristics and themes of the country's regulations in order to find similarities and differences. Based on these similarities and differences in regulatory characteristics, we divided the approaches into three models.

Results

The mechanisms and criteria of each country for decision-making regarding acceptability of PGT can be found in Table 1. In the seven countries analysed, we identified three models for the regulation of PGT: the medical indication model, the individual requests model, and the clinical assessment model. An overview of these models and included countries can be found in Table 2. An overview of which types of PGT are permitted in each analysed country can be found in Table 3.

Table 1. Decision-making for PGT requests for each country.

	National laws relating to PGT	PGT-M/SR ^a approval regulated by ...	Regulatory mechanism for PGT indications	Criteria for PGT-M/SR ^a
Belgium	– ‘Wet betreffende de medisch begeleide voortplanting en de bestemming van de overtallige embryo's en de gameten’/‘Law on Medically Assisted Reproduction and the Disposition of Supernumerary Embryos and Gametes’ (6 July 2007)	PGT centres	Internal decision-making	• Not for ‘eugenic’ selection or non-medical sex selection, otherwise no specifications (Belgisch Staatsblad 2007)
France	– ‘LOI no. 2004-800 du 6 août 2004 relative à la bioéthique’/‘LAW no. 2004-800 of August 6, 2004 relating to bioethics’ (6 August 2004) – ‘LOI no. 2011-814 du 7 juillet 2011 relative à la bioéthique’/‘LAW no. 2011-814 of July 7, 2011 relating to bioethics’ (7 July 2011) – ‘LOI no. 2021-1017 du 2 août 2021 relative à la bioéthique’/‘LAW no. 2021-1017 of August 2, 2021 relating to bioethics’ (2 August 2021) – ‘Code de la santé Publique, Article L2131-4’/‘Public Health Code, Article L2131-4’ (4 August 2021)	Centres Pluridisciplinaire de Diagnostic Prénatal	Approval for each PGT request	• High probability of having a child with a severe and incurable disease (LOI no. 2021-1017 relative à la bioéthique 2021; Article L2131-4 2021)

(continued)

Table 1. (continued)

	National laws relating to PGT	PGT-M/SR ^a approval regulated by ...	Regulatory mechanism for PGT indications	Criteria for PGT-M/SR ^a
Germany	– ‘Code de la santé Publique, Article L2131-4-1’/‘Public Health Code, Article L2131-4-1’ (4 August 2021) – ‘Gesetz zum Schutz von Embryonen (Embryonenschutzgesetz—EschG)’/‘Act for the Protection of Embryos (The Embryo Protection Act)’ (19 December 1990, amended on 21 November 2011)	PGT centres together with interdisciplinary independent PGT ethics commissions	Approval for each PGT request	PGT is not allowed, except if: • It is highly likely for the potential parent’s offspring to have a serious genetic illness • To identify an abnormality that would be highly likely to lead to still-birth or miscarriage (Embryo Protection Act (amendment) 2011)
Italy	– ‘Legge 194/1978: Norme per la tutela sociale della maternità e sull’interruzione volontaria della gravidanza’/‘Law 194/1978: Rules for the social protection of maternity and on the voluntary termination of pregnancy’ (22 May 1978) – ‘Legge 40/2004: Norme in materia di procreazione medicalmente assistita’/‘Law 40/2004: Regulations on medically assisted procreation’ (24 February 2004) – ‘Sentenza 96/2015: procreazione medicalmente assistita e diagnosi genetica preimpianto’/‘Sentence 96/2015: medically assisted procreation and preimplantation genetic diagnosis’ (5 June 2015) – ‘Sentenza 229/2015: procreazione medicalmente assistita e diagnosi genetica preimpianto’/‘Sentence 96/2015: medically assisted procreation and preimplantation genetic diagnosis’ (11 November 2015)	PGT centres	Internal decision-making	• For genetic diseases meeting severity criteria relating to voluntary termination of pregnancy, after the first ninety days (significant anomalies or malformations of the unborn child that could cause serious danger to the mental or physical health of the woman) (Legge no. 194 1978; La Corte Costituzionale 2015b)
The Netherlands	– ‘Regeling pre-implantatie genetische diagnostiek’/‘Regulation preimplantation genetic diagnosis’ (16 February 2009, 23 March 2013) – ‘Embryo law’/‘Embryowet’ (20 June 2002, 1 July 2021)	Landelijke Indicatiecommissie PGT	Approval for new PGT indications	Four criteria have to be considered in the decision framework: • Type and seriousness of the condition (e.g. severity, age of onset, risk) • Treatment options • Other medical criteria (e.g. age, egg cell quality) • Psychological and ethical factors (e.g. psychological burden experienced by the gene carrier and previous experience of parents related to, e.g. pregnancy termination) (Regeling preimplantatie genetische diagnostiek 2009)
Spain	– ‘Ley 35/1988, de 22 de noviembre, sobre Técnicas de Reproducción Asistida’/‘Law 35/1988, of November 22, on Assisted Reproduction Techniques’ (24 November 1988) – ‘Ley 14/2006, de 26 de mayo, sobre técnicas de reproducción humana asistida’/‘Law 14/2006, of May 26, on assisted human reproduction techniques’ (28 May 2006)	PGT centres and Comisión Nacional de Reproducción Humana Asistida for exceptions	Internal decision-making, approval from CNRHA when uncertainty about appropriateness of PGT	PGT is authorized for: • A condition that is ‘serious, early-onset and for which no treatment exist’ • Or ‘to detect the alterations that may affect negatively the viability of the embryos’ (Ley 14 sobre técnicas de reproducción humana asistida 2006)

(continued)

Table 1. (continued)

	National laws relating to PGT	PGT-M/SR ^a approval regulated by ...	Regulatory mechanism for PGT indications	Criteria for PGT-M/SR ^a
The UK	– 1990 Human Fertilisation and Embryology Act (1 November 1990) – 2008 Human Fertilisation and Embryology Act (13 November 2008)	Human Fertilisation and Embryology Authority	Approval for new PGT indication	'Preimplantation genetic diagnosis (PGD) can be carried out for a heritable condition only in two circumstances: • Where there is a particular risk that the embryo to be tested may have a genetic, mitochondrial, or chromosomal abnormality, and the Authority is satisfied that a person with the abnormality will have or develop a serious disability, illness, or medical condition, or • Where there is a particular risk that any resulting child will have or develop a gender-related serious disability, illness, or medical condition. A condition is gender related if the Authority is satisfied that it affects only one sex, or affects one sex significantly more than the other. In the first situation, PGD may be carried out to establish whether the embryo has the suspected abnormality; in the second, PGD may be carried out to establish the sex of the embryo.' (Human Fertilisation & Embryology Authority 2021)

^a Preimplantation genetic testing for monogenic conditions (M) or structural chromosome defects (SR).

The medical indication model

The first model for PGT regulation requires authorization for new PGT indications, based on laws that specify and limit the medical indications for which PGT can be used. Both the UK and the Netherlands have laws in place that dictate that PGT can only be used for specific indications and that a national regulatory body has to authorize the use of PGT for each newly requested medical indication.

The UK

In the UK, ARTs including PGT are regulated by the Human Fertilisation and Embryology Authority (HFEA) ([Corveleyn et al., 2007](#)). The HFEA is a statutory body and independent regulator of fertility treatments and research that uses human embryos in the UK. Clinics carrying out PGT need to have a license from the HFEA and are obliged to follow their regulations ([Soini et al., 2006](#); [Nakasato et al., 2022](#)). HFEA regulation is based on the 1990 and

2008 Human Fertilisation and Embryology Acts and is summarized in the HFEA Code of Practice (1990, 2008; [Human Fertilisation & Embryology Authority, 2021](#)). PGT is allowed 'where there is a particular risk that the embryo to be tested may have a genetic, mitochondrial, or chromosomal abnormality, and the Authority is satisfied that a person with the abnormality will have or develop a serious disability, illness, or medical condition' or for sex selection of embryos for conditions that affect only or mostly one sex and that meet the same requirements (1990, 2008; [Human Fertilisation & Embryology Authority, 2021](#)). Non-medical sex selection is forbidden by law (1990, 2008; [Human Fertilisation & Embryology Authority, 2021](#)). PGT-A is not explicitly included in the 1990 and 2008 Human Fertilisation and Embryology Act, but according to the HFEA's interpretations, use of PGT-A to test embryos for a particular chromosomal abnormality is only allowed if '(a) that abnormality may affect its capacity to result in a live birth, or (b) there is a particular risk that it has that abnormality, and where the Authority is satisfied that there is a significant risk that a person with that abnormality will have or develop a serious medical condition' ([Human Fertilisation & Embryology Authority, 2021](#)). PGT-M/SR with HLA matching is permitted (1990, 2008). The HFEA requires that the situation of the affected child, the child who will possibly be born, and the family situation are considered ([Human Fertilisation & Embryology Authority, 2021](#)). HFEA's interpretations of the 1990 and 2008 Human Fertilisation and Embryology Act further state that selecting an embryo that will have significant risk of a medical condition is prohibited, except if there is no other embryo available for transfer. In this case, approval from a clinical ethics

Table 2. Regulatory models for preimplantation genetic testing.

Model type	Countries included
The medical indication model	• The UK • The Netherlands
The individual requests model	• France • Germany
The clinical assessment model	• Belgium • Italy • Spain

Table 3. Types of PGT that are permitted—with restrictions—according to national laws.

	Target or purpose of PGT			
	Monogenic conditions (PGT-M) or structural chromosome defects (PGT-SR)	Aneuploidy (PGT-A)	Human leukocyte antigen (PGT-HLA)	Non-medical sex selection
Belgium	Permitted	Permitted	Permitted	Not permitted
France	Permitted	Not permitted	Permitted	Not permitted
Germany	Permitted*	Not permitted	Not permitted	Not permitted
Italy	Permitted	Permitted**	Permitted**	Not permitted
The Netherlands	Permitted	Not permitted	Permitted	Not permitted
Spain	Permitted	Permitted	Permitted	Not permitted
The UK	Permitted	Permitted	Permitted	Not permitted

* PGT is prohibited in Germany with narrow exceptions, therefore for these narrow exceptions we see PGT as permitted.

** Not included in legislation, but recommendations of Società Italiana di Genetica Umana also suggest PGT-A and PGT-HLA are possible in specific circumstances.

committee is needed and the welfare of the future child needs to be considered (Human Fertilisation & Embryology Authority, 2021). According to HFEA's Code of Practice, if patients are already using PGT-M/SR for an approved indication, centres can offer PGT-M or PGT-SR for additional condition(s) that do not meet risk requirements but are seen to be of significant risk, if patients provide consent for this (Human Fertilisation & Embryology Authority, 2021). While PGT-P is not yet included in HFEA's Code of Practice, the HFEA has already issued a statement saying that 'The use of polygenic risk scores in PGT is currently illegal in the UK. Embryo selection is only legal in the UK to avoid serious inherited illnesses, usually by means of pre-implantation genetic diagnosis authorized by the HFEA, known as PGT-M or PGT-SR' (Blakely, 2022).

The HFEA creates a list of conditions for which PGT is permitted, which currently includes over 600 conditions. The list of conditions that have been approved for PGT or are under consideration by the HFEA for PGT is publicly available (Human Fertilisation & Embryology Authority, 2023a). The HFEA and its Statutory Approvals Committee (SAC) decide on approval of PGT indications that are not already on the approved list (Human Fertilisation & Embryology Authority, 2023b). When PGT is requested for an indication that is not on the list, fertility clinics can submit an application to the HFEA on behalf of their patients (Corveleyn et al., 2007; Bayefsky, 2016; Nakasato et al., 2022; Human Fertilisation & Embryology Authority, 2023a). The HFEA considers various elements when deciding whether to approve PGT for an indication, including the likelihood of inheritance, the seriousness of the condition, and experiences of people affected by the condition (Human Fertilisation & Embryology Authority, 2023a). It is stated that 'When deciding if it is appropriate to provide PGT-M or PGT-SR in particular cases, the centre should consider the circumstances of those seeking treatment rather than the particular heritable condition' (Human Fertilisation & Embryology Authority, 2021). This suggests that not only the pre-established criteria for what is considered a 'serious' condition, but also the specific situation and experiences of those requesting PGT are considered relevant to take into account. Approval of PGT-HLA is given on a case-by-case basis for the specific situation of those requesting it, while approval for PGT-M is given for the indication (Human Fertilisation & Embryology Authority, 2023a). Decisions about approval of PGT for a condition are usually made within two to three months (Human Fertilisation & Embryology Authority, 2023a). According to the HFEA's ruling (Blakely, 2022), PGT-P is currently not allowed in the UK.

The Netherlands

PGT in the Netherlands is regulated by the Ministry of Health, Welfare and Sports (Corveleyn et al., 2007). In the Netherlands, the Special Medical Procedures Act ('Wet op Bijzondere Medische Verrichtingen') states that certain medical acts are not allowed unless the Minister of Health, Welfare and Sports provides a license for these acts (Nederlandse Overheid, 2021). The regulation for PGT ('Regeling preïmplantatie genetische diagnostiek') is part of this Act and was created in 2009 and revised in 2013 (2009). The 2009 regulation indicates that an institution needs a license in order to perform PGT and states that University Hospital Maastricht is designated as the only licensed Dutch PGT centre (though 'transport PGT' is allowed with other Dutch clinics having a specific agreement) (2009). The regulation furthermore states PGT is permitted only if there is a high risk of a serious genetic condition in the offspring (2009). According to Dutch regulation, the following decision framework needs to be used when eligibility for PGT is evaluated: the type and seriousness of the condition (e.g. severity, age of onset, risk), any existing treatment options, other medical criteria (e.g. age of the requesting woman, egg cell quality), and psychological and ethical factors (e.g. psychological burden experienced by the gene carrier and previous experience of parents related to, e.g. pregnancy termination) (2009). Sex selection for conditions with unequal sex incidence is permitted since 2020 (2002 (revised in 2021)). Non-medical sex selection is not allowed according to Article 26 of the Dutch Embryo Law (2002 (revised in 2021)). Other types of non-medical selection via PGT are also not permitted in The Netherlands (2009). PGT-A is not performed in the Netherlands, as it does not fall under the scope of PGT in the Netherlands (Geraedts et al., 2016; Calhaz-Jorge et al., 2020; PGT Nederland, 2020; Paternotte and Hermans, 2023). HLA typing in PGT is only allowed to also prevent a serious genetic condition in the future child itself, not solely to provide treatment possibilities via stem cells for a sibling (2009). Selection for non-carriers, that will most likely not be affected but could pass the condition to their offspring, is not a permitted indication for PGT. When carriers of a condition are identified in a PGT-M cycle that was initiated to prevent a recessive condition, prioritization of non-carriers for transfer is permitted (2009). In 2023, a bill was proposed to allow PGT specifically to select for carrier status, for which no final decision has been made (Paternotte and Hermans, 2023).

The Netherlands has a national indication committee ('Landelijke Indicatiecommissie PGT' or 'LIC') that decides whether PGT can be used for indications that have not previously been approved for PGT. This committee consists of clinical

geneticists, ethicists, gynaecologists, and a policy officer on behalf of patient organizations (PGT Nederland, 2023). When PGT is requested for a new indication, the case is first discussed with a multidisciplinary team (including at least one ethicist and if relevant an oncologist) at University Hospital Maastricht (2009). If they conclude that the medical indication is indeed new and the couple's request for PGT is eligible, they ask the LIC to discuss the indication. The LIC looks mainly at two criteria of the aforementioned decision framework of the regulation of the Ministry of Health, Welfare and Sports (severity and nature of the disease, and treatment options) (2009; PGT Nederland, 2023). The advice the LIC gives is conditional: the LIC can conclude with 'yes, provided that ...' or 'no, except ...' for an indication. Taking the advice of the LIC and the specific context of the case into consideration, the University Hospital Maastricht decides on PGT approval for the specific case. In case of a 'no, except ...' conclusion from the LIC, Maastricht can decide in individual cases to allow PGT (PGT Nederland, 2023). If the LIC allows PGT for a certain indication, there is permission to allow PGT for the same genetic predisposition in the future, and the LIC does not need to be asked for permission again (PGT Nederland, 2023). As PGT-P does not meet the Dutch regulatory requirements pertaining to high risk of a serious genetic condition, it seems unlikely that PGT-P would be approved by the LIC.

The individual requests model

The individual requests model stipulates that every single request for PGT must be evaluated by a regulatory body, according to whether the request meets the country's legal criteria for PGT. France and Germany have laws in place that restrict the use of PGT and require approval on a case-by-case basis by a regional regulatory body.

France

PGT in France is regulated by the Law on Bioethics (enacted in 1994, revised in 2004, 2011, and 2021) (Duguet and Boyer-Beviere, 2017; Creux et al., 2020; 2021). Centres offering PGT in France have to be licensed and monitored (2021; 2021). The 2021 Law on Bioethics and Article L.2131-4 of the French Public Health Code state that PGT is only allowed for couples and women having a high probability of having a child with a severe and incurable disease (2021; 2021). In France, there is no list available of conditions for which PGT is seen as appropriate, as there are concerns that this could lead to stigmatization of people with these conditions (Duguet and Boyer-Beviere, 2017). PGT-A is not included in the PGT regulation and its application does not meet the established PGT criteria, meaning PGT-A is not authorized in France (2011; Calhaz-Jorge et al., 2020; Creux et al., 2020; 2021). France has ratified the Oviedo Convention, which states that 'techniques of medically assisted procreation shall not be allowed for the purpose of choosing a future child's sex, except where serious hereditary sex-related disease is to be avoided' (Oviedo Convention, 1997). This would however exclude PGT for non-medical indications. The Agence de la Biomédecine, created by the Law on Bioethics (No. 2004–800), oversees assisted reproductive technologies and PGT and can extend the uses of PGT that are specified in the bioethics law of France (2004; Bayefsky, 2016). The Agence de la Biomédecine extended the use of PGT in 2013 to allow PGT for HLA tissue matching to select a potential tissue donor for the existing child on a case-by-case basis (Article L.2131-4-1) (L'Agence de la biomédecine, 2013, 2022a).

In France, every individual PGT request is reviewed by a regional Centre Pluridisciplinaire de Diagnostic Prénatal (CPDPN) on a case-by-case basis (Bayefsky, 2016; Duguet and Boyer-

Beviere, 2017). A CPDPN consist of practitioners from different fields, such as genetics, gynaecology, paediatrics, biology, and psychiatry (L'Agence de la Biomédecine, 2022b). First, the physician in one of the five PGT centres has to confirm that the person or couple is likely to have a child that has a genetic condition that is severe or severely disabling and is incurable (Duguet and Boyer-Beviere, 2017; L'Agence de la Biomédecine, 2022b). The genetic condition must have been identified in at least one of the parents or in one of their direct offspring (Corveleyn et al., 2007; Council of Europe, 2015; Duguet and Boyer-Beviere, 2017; 2021). Then the CPDPN evaluates and decides whether or not to allow PGT, based on the severity and incurability of the condition and the potential suffering of the prospective parents (Bayefsky, 2016; Duguet and Boyer-Beviere, 2017). For PGT with HLA matching, a request has to be authorized by the Agence de la Biomédecine. The agency considers the suffering of the affected sibling and the benefit of PGT-HLA for the affected sibling as well as the integrity of the future child, for example preventing that the child is solely used as an 'instrument' for the existing sibling (Duguet and Boyer-Beviere, 2017). Seeing as every PGT request has to be approved on a case-by-case basis according to criteria related to probability, severity, and incurability, it would be unlikely that PGT-P would be accepted in France.

Germany

The legislation for PGT in Germany has historically been restrictive, due to interpretations of the Embryo Protection Act/ Embryonenschutzgesetz/EschG (1990), which grants a high level of protection to the human embryo (Krones and Richter, 2004; Fahrenkrog, 2006; Nippert, 2016). While no explicit regulation for PGT existed, discarding embryos was not permitted under the 1990 Embryo Protection Act, thereby disallowing selective transfer of embryos (1990). While direct testing of embryos (including cleavage stage embryos and blastocysts) was not allowed, an exception was made for the analysis of polar bodies, which is conducted in the oocytes and is thus not considered testing of embryos (Soini et al., 2006; Corveleyn et al., 2007, 2008). As polar body biopsy only gives information about the egg cell, PGT via polar body testing is not possible for conditions inherited via the father (Corveleyn et al., 2007; Bock von Wulffingen, 2016). These restrictions led German patients to go abroad for PGT, though formal referral by German physicians of patients to other countries in order to access PGT was prohibited (Corveleyn et al., 2007). In 2010, a court case centred around a German gynaecologist who, in 2005 and 2006, had previously performed PGT on blastocyst embryos for three couples that carried a predisposition for a severe genetic condition. While his acts seemed in violation of interpretations of the German law of that time, the German Federal Court of Justice ruled in favour of the gynaecologist, and stated that his application of PGT was in accordance with the law and was not in violation of the Embryo Protection Act, as it did not breach the law's purpose of protecting embryos from abuse (Bundesgerichtshof, 2010; Tuffs, 2010; Bock von Wulffingen, 2016; Nippert, 2016). Following this court case, the German parliament then voted in favour of creating regulation for PGT in 2011 and an amendment was made to the Embryo Protection Act (Federal Law Gazette I, p. 2228, section 3a). This amendment states that PGT is generally prohibited but allowed for limited exceptions (2011).

The regulation detailing the necessary procedures and organization for PGT came into force in 2014 (Council of Europe, 2015; Nippert, 2016). PGT can only be performed by qualified physicians in centres that have been approved for PGT (2011; Nippert, 2016). Every state has to establish licensed PGT centres and an

interdisciplinary independent PGT ethics commission, which has to approve the PGT request on a case-by-case basis (2011; Nippert, 2016). All requested cases of PGT—including those not approved by an ethics committee—need to be reported in anonymized form to a national registry (2011; Nippert, 2016). At the moment, the limited exceptions for which PGT is allowed in Germany are: if one of the parents has a severe hereditary genetic condition that can be passed on to the offspring, or if there is a high risk of stillbirth or miscarriage (2011; Koehler et al., 2013; Kim, 2015). Non-medical sex selection is not allowed (2011). PGT-A and PGT-HLA are not included in the exceptional uses of PGT that the German law allows. Because PGT-P does not fall under the limited exceptions of PGT use that Germany allows, it seems unlikely that it will gain acceptance in this country.

The clinical assessment model

A third model for PGT regulation is to largely leave decisions about the use of PGT to the discretion and clinical judgment of relevant healthcare professionals, within broadly defined or more restrictive laws. Belgium and Italy do not have regulatory bodies in place that decide for which indications or individual couples PGT is allowed. In Spain, decisions within the legal framework are left to PGT providers and only request of PGT-HLA and exceptional requests of PGT need to be authorized by Spain's National Committee on Assisted Human Reproduction.

Belgium

In Belgium, there is no regulatory body and little directive legal regulation for PGT indications at the national level. Belgium's 2007 Law on Medically Assisted Reproduction and the Disposition of Supernumerary Embryos and Gametes states that a specific collaboration agreement between a fertility centre and centre for human genetics is necessary for centres to be allowed to offer and carry out PGT (*Belgisch Staatsblad*, 2007). The number of fertility centres that is allowed to perform PGT is decided by Royal Decree and can be no fewer than eight centres (*Belgisch Staatsblad*, 2007). The 2007 law prohibits PGT for so-called 'eugenic selection' as well as non-medical sex selection (*Belgisch Staatsblad*, 2007; Pennings, 2007). Other than that, no specific requirements for PGT-SR, PGT-M, and PGT-A are provided in the law (*Belgisch Staatsblad*, 2007; Pennings, 2007; Pennings and Mertes, 2019). There is also no list indicating for which conditions PGT is seen as accepted (Pennings, 2007; Council of Europe, 2015). The Belgian law does thus not indicate what the aim of PGT is or should be, but specifies what applications of PGT are not allowed (*Belgisch Staatsblad*, 2007; Pennings and Mertes, 2019). PGT-HLA matching is allowed, but according to Article 68 of the 2007 law, two criteria have to be met (*Belgisch Staatsblad*, 2007; Pennings, 2007). First, HLA typing is allowed for therapeutic benefit of an existing child of the parents, and second, it needs to be evaluated by the fertility centre and confirmed by the genetics centre that the wish for another child and use of PGT-HLA is not solely for the benefit of the existing child (*Belgisch Staatsblad*, 2007). In 2011, a proposal was submitted for amendment of the 2007 law to allow PGT for the detection of healthy carrier embryos that do not have the condition but can pass it to offspring, similar to the recent proposal about this topic in the Netherlands (Defraigne and Brotchi, 2011; Pennings and Mertes, 2019). This change of law was not accepted (Belgian Advisory Committee on Bioethics, 2009; Pennings and Mertes, 2019). The Advisory Committee on Bioethics issued a statement on the topic and concluded that it is allowed to deselect healthy carrier embryos if one is already involved in a PGT trajectory, but that is not justifiable to start a PGT cycle specifically to avoid carrier embryos

(Belgian Advisory Committee on Bioethics, 2009; Pennings and Mertes, 2019).

The Belgian government did not want to establish strict rules with their 2007 'reproductive medicine law' and left decision-making about the use of PGT to the discretion of centres and physicians (Pennings and Mertes, 2019; Arzheimer, 2020). This implies that the genetic and fertility centres in Belgium have the authority to make decisions regarding appropriateness of PGT (Pennings and Mertes, 2019). It is common practice that each centre evaluates individual PGT requests or indications with a multidisciplinary team and specific medical specialists when needed (De Rycke et al., 2020). Centres can also make use of local ethics boards (e.g. the ethical committee of the specific hospital) for evaluation of PGT requests (De Rycke et al., 2020). PGT-M, PGT-SR, and PGT-A are being performed in several centres across Belgium (Pennings and Mertes, 2019). PGT-P for medical indications seems to be allowed in Belgium, though healthcare professionals or centres would still need to evaluate its use.

Italy

Italy's regulation for PGT is based on interpretations of Law No. 40/2004 (2004). This law on reproductive medicine was initially interpreted to forbid PGT, as it enforced that only people diagnosed with infertility could use ARTs, that maximum three embryos could be created during IVF and that all viable embryos had to be transferred to the prospective parent's uterus at the same time (2004; Molinelli et al., 2012; Gianaroli et al., 2014; Vesto, 2017). However, the law did not explicitly prohibit PGT (2004; Vesto, 2017). The implicit ban on PGT was declared unconstitutional by the Regional Administrative Court of Latium in 2008 and the rule that only three embryos could be created and all embryos had to be transferred at once was declared unconstitutional by the Italian Constitutional Court in 2009 (Molinelli et al., 2012; Gianaroli et al., 2014; Bayefsky, 2016). By law, however, access to IVF/ICSI remained accessible only to couples with fertility problems. In 2012, the European Court of Human Rights ruled in favour of an Italian couple with no fertility issues wanting to access PGT to prevent the birth of a child with cystic fibrosis, stating that the Italian law was in violation of Article 8 'the right to respect for private and family life' (Gianaroli et al., 2014; Vesto, 2017; Ginoza and Isasi, 2020). The discrepancy of PGT not being allowed while prenatal testing and pregnancy determination were allowed for cystic fibrosis, played a role in the court's decision (Sándor, 2015). In 2015, the Italian Constitutional Court abandoned the requirement of infertility for PGT (sentence no. 96/2015) and stated that selective transfer of embryos for genetic diseases meeting severity criteria was no longer illegitimate (sentence no. 229/2015) (*La Corte Costituzionale*, 2015a,b). These criteria are the same criteria that are used to determine the permissibility of termination of pregnancy outside the legal period of 90 days and refer to 'significant anomalies or malformations of the unborn child' that could be a danger to the mental or physical health of the woman (Article 6b of Law no. 194/1978) (1978). PGT for 'eugenic purposes' is not allowed, as is stated in Article 13 of Law No. 40/2004 (2004; Vesto, 2017). One can opt for PGT through Centres for Medically Assisted Reproduction that are affiliated with the national health service or that are private in Italy (Granato, 2021). There is no regulatory body or mechanism to determine for which indications PGT is allowed in Italy (Bayefsky, 2016; Vesto, 2017).

Therefore, while PGT has been allowed in Italy for several years, it is not clear for which conditions PGT is used, or being seen as acceptable (Bayefsky, 2016; Vesto, 2017). According to recommendations of the Società Italiana di Genetica Umana, a

request for PGT needs to be evaluated by a multidisciplinary team consisting of specialized staff, including at least a geneticist and a gynaecologist (*Società Italiana di Genetica Umana*, 2017). The medical and psychological situation, safety, and chance of success for the ones requesting PGT have to be considered (*Società Italiana di Genetica Umana*, 2017). The recommendations of the *Società Italiana di Genetica Umana* suggest that PGT for late-onset conditions without full penetrance, such as breast and ovarian cancer associated with the BRCA1/2 genes, can be performed, following recommendations of the European Society of Human Reproduction and Embryology (Shenfield et al., 2003; *Società Italiana di Genetica Umana*, 2017). The recommendations of the *Società Italiana di Genetica Umana* also suggest that PGT-A and PGT for HLA typing can be performed in specific circumstances (*Società Italiana di Genetica Umana*, 2017). As criteria for approved PGT use do not seem to be clearly defined, this might leave open the possibility for PGT-P implementation for medical indications in Italy.

Spain

Spain introduced legislation for ARTs in 1988 (Act 35/1988) (1988; Corveleyn et al., 2007). The Act states that the use of ARTs for prevention or treatment of hereditary or genetic diseases is permitted (1988; Corveleyn et al., 2007). The 1988 national law did not yet include explicit regulation for PGT (Pavone and Arias, 2011; Pavone and Lafuente Funes, 2018). The National Committee on Assisted Human Reproduction (Comisión Nacional de Reproducción Humana Asistida/CNRHA), which was envisaged in the 1988 Act, was constituted in 1997 and tasked with providing guidance for assisted reproduction (1988; Corveleyn et al., 2007; Pavone and Arias, 2011; Pavone and Lafuente Funes, 2018). The CNRHA is the collegiate body of the Ministry of Health and consists among others of government representatives, a bioethics expert and representatives from national societies related to reproductive medicine (Comisión Nacional de Reproducción Humana Asistida 2023; 1988). Explicit regulation of PGT was not introduced until the National Assisted Reproduction Law was established in 2006 (Law 14/2006), in a reform of 1988 and 2003 laws (2006), even though PGT (except for HLA-matching) was already being performed at that time (Pavone and Arias, 2011; Pavone and Lafuente Funes, 2018). This new law permits the genetic selection of embryos for conditions that are 'serious, early-onset and for which no treatment exist' or 'to detect the alterations that may affect negatively the viability of the embryos' (2006). The National Assisted Reproduction Law thus provides a framework for the evaluation of the appropriateness of PGT but does not include a list of conditions for which PGT is permitted (Pavone, 2010; Pavone and Arias, 2011; Pavone and Lafuente Funes, 2018). In practice, PGT has been performed for adult-onset conditions, such as Huntington's disease and hereditary breast and ovarian cancer (BRCA1/2) (Pavone, 2010). It is said that the Spanish regulation and criteria for PGT are broadly defined in order to leave room for future PGT applications without needing to change the regulative framework (Pavone, 2010; Pavone and Lafuente Funes, 2018). PGT for non-medical indications, including non-medical sex selection, is not permitted in Spain (2006). Spain has also ratified the Oviedo convention, which indicates that sex selection for non-medical purposes is not allowed (Oviedo Convention, 1997). PGT-HLA matching for an ill sibling is permitted according to the law, but does need authorization from the CNRHA (2006).

If PGT requests meet the criteria of the 2006 law described above, PGT providers are expected to submit this information to the CHRHA through regional authorities. No authorization from

the CNRHA or regional authorities is required (Pavone and Lafuente Funes, 2018). If there is uncertainty about whether a PGT request falls within the criteria indicated in the regulation and in case of requests for PGT-HLA, specific authorization from the CNRHA via a case-by-case approach is needed (Comisión Nacional de Reproducción Humana Asistida 2023; 2006; Pavone and Arias, 2011; Pavone and Lafuente Funes, 2018). The use of PGT-A falls within the criteria of the 2006 law and therefore does not need authorization of the CNRHA (Pavone and Arias, 2011; Pavone and Lafuente Funes, 2018). Spain seems to leave open the possibility for permitting PGT-P, in so far as it would be used for medical indications only.

Discussion

In this article, we have analysed how various European countries regulate PGT, in order to evaluate the legal possibilities for implementation of PGT-P. We found that there are large differences between PGT regulation, regarding the focus on the PGT indication or on the specific case of the people requesting PGT, regarding the formulation of strict rules or broadly interpretable laws, and regarding having a clear regulatory and decision-making framework in place or leaving more room to PGT providers to make decisions. We captured these differences in three models: the medical indication model, the individual requests model, and the clinical assessment model. New developments within PGT, in particular the recent introduction of PGT-P, make it crucial to look critically at policy and regulation. We therefore assess the potential acceptability of PGT-P within these European countries, based on our analysis of current regulatory frameworks for PGT. With the exception of the UK, who have issued a statement on PGT-P, this analysis is therefore based on an extrapolation of current PGT regulations.

Opportunities for PGT-P within the analysed European countries

In countries regulating PGT with the medical indication model (The Netherlands, the UK), PGT-P would need to meet the criteria in place for PGT indications and would need to be approved by a national authority. In the UK, the HFEA already indicated that PGT-P does not meet their current PGT requirements and highlighted that there is no scientific consensus for PGT-P at the moment. They furthermore stated that 'Before this could be introduced into UK law it would need a shift in public policy, and a range of ethical concerns to be addressed, which would require parliamentary approval and the backing of society more generally' (Blakely, 2022). In the Netherlands, potential requests for PGT-P would likely not be approved by the LIC at this moment, since PGT-P does not meet PGT requirements relating to high risk and potentially to seriousness and lack of medical treatments. The fact that PGT-A is not performed in the Netherlands and receives different ratings of effectiveness from the HFEA in the UK depending on its use (Geraedts et al., 2016; Calhaz-Jorge et al., 2020; PGT Nederland, 2020; Human Fertilisation & Embryology Authority, 2023c), also implies that enough evidence of validity and utility is required before a screening or testing option is performed in these countries—evidence which is currently lacking for PGT-P (Karavani et al., 2019; Lázaro-Muñoz et al., 2021; Lencz et al., 2021; Turley et al., 2021). The proposal of offering PGT for genetic conditions without full penetrance (e.g. hereditary cancers) in 2008 already caused a lot of political debate in the Netherlands. Especially Christian parties were not in favour of allowing this and were afraid of a 'slippery slope' with regards to embryo selection if this would be allowed. The discussion around

this topic led to the 2009 Dutch PGT regulation, which outlined the evaluation framework and mechanism for PGT requests (Winter et al., 2012; Ministerie van Volksgezondheid, Welzijn en Sport 2019). Based on the existing requirements, evidence, and cultural contexts, a change to PGT regulation that would allow PGT-P in the countries with this model thus seems improbable in the foreseeable future.

Similarly, the countries with the individual requests model have criteria in place to decide for which indications PGT is allowed (France, Germany), relating to the knowledge that the parents are likely to have a child with a severe genetic condition or experience stillbirth or miscarriage. Knowledge of a high risk or heritability of a specific genetic condition is thus necessary, which would not necessarily be present with PGT-P, especially if PGT-P would be performed for a panel of various conditions regardless of known familial risk. PGT-P would thus not fall within the authorized use of PGT in the included countries following the individual requests model, and requests to regional authority bodies for PGT-P would likely not be approved. Especially considering Germany's initial prohibition of PGT, which can be said to be related to the level of protection attached to the embryo, to anti-discrimination against disabilities, and to a particular carefulness after the eugenic practices of Germany's Nazi past (Krones and Richter, 2004; Fahrenkrog, 2006), it seems unlikely that the law would be changed in order to authorize PGT-P.

With the clinical assessment model, decision-making about the appropriateness of PGT for specific indications is left to PGT providers within the limits of the law (Belgium, Italy, Spain). Especially in countries where the regulation in place with regards to PGT requirements is broadly defined, this would leave space for potential implementation of PGT-P. These broadly defined regulatory approaches can for example be related to Belgian's interest in the development of new reproductive technologies and its liberal position to the status of and research on human embryos (Pennings and Mertes, 2019; Arzheimer, 2020) and to the historical influence of private centres on PGT regulation in Spain (Pavone and Arias, 2011). However, the healthcare system and associated regulation also play a role: in Belgium, for example, PGT may only be performed in recognized fertility and genetics centres. Currently, no commercial entities are recognized to offer PGT in Belgium. In Italy and Spain, by contrast, PGT is largely offered privately, and commercial clinics are said to influence the field of PGT and its regulatory development in Spain (Pavone and Arias, 2011). Potentially, there might be more room for new commercial PGT offers in countries where space is already provided to private offerings of PGT and where the law is framed flexibly. However, considering that Italy's legal framework for PGT does refer to criteria regarding severity and considering the influence of the Catholic Church and their position on the moral status of the human embryo in Italy, which led to the strict 2004 law on assisted reproductive technologies that made PGT practically impossible (Parfenchuk, 2018), the space for PGT-P in Italy might be limited despite the medical assessment model and private offerings of PGT.

PGT for non-medical sex selection (e.g. for family balancing) or non-medical selection in general is not allowed in all countries included in the analysis. Even when PGT-P would be offered in any of the included countries in the future, it seems reasonable to believe that it would be limited to medical indications only and not offered for non-medical traits.

Importantly, our discussion concerns the use of PGT-P as a stand-alone tool for reproductive decision-making and embryo selection. In practice, it is also possible that PGT-P will be offered

as an add-on to people making use of PGT-A (potentially in addition to PGT-M/SR), or that PRS could potentially be used to improve the accuracy of risk predictions for monogenic conditions without full penetrance. While the former would likely meet the same regulatory roadblocks as we have described in this article, the latter might be evaluated differently within current legal frameworks. This is however beyond the scope of this article.

Perspectives of scientific organizations and healthcare professionals

Besides current legal frameworks, the stances of European scientific organizations such as the European Society of Human Genetics and the European Society of Human Reproduction and Embryology need to be considered. Both these societies have warned about PGT-P's prematurity with regards to its validity, utility, and ethics and the European Society of Human Genetics has described PGT-P as 'an unproven, unethical practice' (Forzano et al., 2021; European Society of Human Reproduction and Embryology, 2022). These recommendations likely impact local stances on potential implementation of PGT-P, regardless of legal possibilities. Furthermore, national, European, and global guidelines generally have not covered PGT-P specifically, but on the basis of PGT-M guidelines, the support for PGT-P appears to be limited at the moment (Siermann et al., 2022b). Furthermore, according to empirical research, the majority of healthcare professionals in Europe in this study are negative about current clinical implementation of PGT-P, due to concerns about the scientific validity and utility of PGT-P, limited benefits for prospective parents, difficulties of choices and counselling, and larger socio-ethical concerns (Siermann et al., 2023, 2024). Concluding, even in countries following the clinical assessment model, where we have shown that offering PGT-P might be possible within these countries' existing legislation, implementation of PGT-P is unlikely.

PGT regulation and opportunities for PGT-P in other countries

For our analysis, we have focussed on seven European countries, which means that certain geographical diversity is missing. We were able to find only limited information regarding the regulatory processes for PGT in other European countries. Lithuania's Law on Assisted Reproduction of 2016 limits PGT to preventing serious or severely disabling diseases, the criteria for which are determined by the Minister of Health (Grubliauskienė, 2022). In Denmark and Finland, PGT-M/SR can be provided in cases where prenatal diagnosis is seen as acceptable (i.e. high risk of a severe genetic disease), and in Sweden, in case one of the prospective parents carries a mutation or structural chromosomal aberration associated with a high risk of having a child with a serious condition (Hreinsson et al., 2020). In Finland and Iceland, PGT-A can be offered unconditionally, whereas in Denmark and Sweden, it is allowed only within a research protocol that is approved by a local ethics committee (Hreinsson et al., 2020). PGT-HLA needs to be approved on a case-by-case basis by a national board of health and welfare in Denmark, Norway, and Sweden (Hreinsson et al., 2020). In Norway, all couples requesting PGT need approval from the National Office for Health Service Appeals (Hreinsson et al., 2020). Norway thus falls within the individual requests model for all offered forms of PGT, on a national level. The same applies to Denmark and Sweden, but only for PGT-HLA. However, while some information is present regarding what PGT at large is allowed for in Sweden, Denmark, Finland, and Lithuania, it is unclear how approval decisions are made and by whom. It is uncertain whether this means that these countries fit with the

clinical assessment model. For the countries that specify criteria for approval of PGT, it seems that these criteria largely exclude PGT-P.

The regulation for PGT in the included European countries can be contrasted with other countries, outside of Europe, where there is more lenient or no regulation at all regarding PGT, such as the USA or Mexico (Soini, 2007; Bayefsky, 2016; Kettering, 2019). The USA for example do not have state or federal laws for PGT and the evaluation of appropriateness of PGT is left to recommendations of professional organizations (such as the American Society of Reproductive Medicine) and decisions of individual centres, teams, or healthcare professionals (Bayefsky, 2016). The limited regulation in the USA can be linked to the fact that there is no government-funded healthcare system, but a market system for fertility care (Bayefsky, 2016). PGT for any condition or trait, including non-medical sex selection for example, is allowed in the USA (Ethics Committee of the American Society for Reproductive Medicine, 2022). PGT-P is allowed and currently offered. The differences between European countries and countries such as the USA could potentially cause cross-border reproductive tourism among European citizens seeking more extensive PGT. In general, reproductive tourism is associated with health and legal risk as well as psychological and financial stress (Corveleyn et al., 2007; Soini, 2007; Ferraretti et al., 2010; Bayefsky, 2016).

Conclusion

The regulation of PGT differs per country, for which we identified three models. Some countries require central approval of PGT for each new indication (the medical indication model: the UK, the Netherlands), some countries evaluate each individual PGT request at the local level (the individual requests model: France, Germany), and other countries largely leave decision-making about clinical application of PGT to healthcare professionals (the clinical assessment model: Belgium, Italy, Spain). Although the three regulatory models are different—including the latitude each would offer for implementation of PGT-P—that latitude seems limited in all three models. Only in countries following the clinical assessment model, does there seem to be fewer barriers to the introduction of PGT-P in fertility clinics. However, in countries where centres providing PGT have to be licenced and where PGT is not offered commercially (Belgium), there seem to be few opportunities for the provision of PGT-P. The fact that European scientific organizations have voiced concerns about implementation of PGT-P and that healthcare professionals in Europe are generally negative about clinical introduction of PGT-P due to scientific and socio-ethical concerns demonstrates that, even if legally possible, the chance that PGT-P would be offered in European countries in the near future appears to be low. Whether legal and regulatory frameworks would be adapted to allow PGT-P in the future, likely depends on whether the scientific and socio-ethical barriers of PGT-P are ever resolved.

Data availability

No new data were generated or analysed in support of this research.

Authors' roles

M.S., P.B., E.M.B., and V.v.d.S. designed the study. The search and analysis were executed by M.S. The manuscript was drafted by

M.S. and was critically discussed and revised by E.M.B., V.v.d.S., and P.B. All the authors have approved the final version.

Funding

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant (agreement no. 813707).

Conflict of interest

The authors declare no conflict of interest.

References

- Arzheimer K. Secular citizens, pious MPs: why German attitudes about genetic testing are much more permissive than German laws. *Political Res Exch* 2020;**2**:1765693.
- Barlevy D, Cenolli I, Campbell T, Furrer R, Mukherjee M, Kostick-Quenet KM, Carmi S, Lencz T, Lazaro-Munoz G, Pereira S. Patient interest in and clinician reservations on polygenic embryo screening: a qualitative study of stakeholder perspectives. *J Assist Reprod Genet* 2024. <https://doi.org/10.1007/s10815-024-03074-0>
- Bayefsky MJ. Comparative preimplantation genetic diagnosis policy in Europe and the USA and its implications for reproductive tourism. *Reprod Biomed Soc Online* 2016;**3**:41–47.
- Belgian Advisory Committee on Bioethics. Opinion no. 49 of 20 April 2009 on the use of preimplantation genetic diagnosis (PGD) to detect healthy carriers of a mutation causing a severe hereditary disease for which offspring can have an increased risk. 2009.
- Belgisch Staatsblad. Wet betreffende de medisch begeleide voortplanting en de bestemming van de overtallige embryo's en de gameten, bl. 38575. 2007.
- Blakely R. Patients warned against American clinics offering 'unethical' embryo screening. *The Times* 2022. <https://www.thetimes.co.uk/article/patients-warned-against-american-clinics-offering-unethical-embryo-screening-zkb76zqq7> (26 September 2023, date last accessed).
- Bock von Wulfinden B. Contested change: how Germany came to allow PGD. *Reprod Biomed Soc Online* 2016;**3**:60–67.
- Bundesgerichtshof. Pressemitteilung Nr. 137/2010: Die Präimplantationsdiagnostik zur Entdeckung schwerer genetischer Schäden des extrakorporal erzeugten Embryos ist nicht strafbar. 2010.
- Calhaz-Jorge C, De Geyter CH, Kupka MS, Wyns C, Mocanu E, Motrenko T, Scaravelli G, Smeenk J, Vidakovic S, Goossens V. Survey on ART and IUI: legislation, regulation, funding and registries in European countries: The European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). *Hum Reprod Open* 2020;**2020**:hoz044.
- Chial H. Mendelian genetics: patterns of inheritance and single-gene disorders. *Nat Educ* 2008;**1**. <https://geneticamedicala.wordpress.com/2013/03/30/mendelian-genetics-patterns-of-inheritance-and-single-gene-disorders/>.
- Comisión Nacional de Reproducción Humana Asistida. 2023. <https://cnrha.sanidad.gob.es/> (18 September 2023, date last accessed).
- Corveleyn A, Morris MA, Dequeker E, Sermon K, Davies JL, Antinolo G, Schmutzler A, Vanecek J, Nagels N, Zika E et al. Provision and quality assurance of preimplantation genetic diagnosis in Europe. *Eur J Hum Genet* 2008;**16**:290–299.
- Corveleyn A, Zika E, Morris M, Dequeker E, Lawford Davies J, Sermon K, Antiñolo G, Schmutzler A, Vanecek J, Palau F et al.

- Preimplantation genetic diagnosis in Europe. JRC Scientific and Technical Reports. Luxembourg: Office for Official Publications of the European Communities, 2007.
- Council of Europe. Background document on preimplantation and prenatal genetic testing: clinical situation, legal situation. 2015.
- Creux H, Diaz M, Grynberg M, Papaxanthos-Roche A, Chansel-Debordeaux L, Jimenez C, Frantz S, Chevalier N, Takefman J, Hocké C. National survey on the opinions of French specialists in assisted reproductive technologies about social issues impacting the future revision of the French Bioethics laws. *J Gynecol Obstet Hum Reprod* 2020;**49**:101902.
- De Rycke M, Berckmoes V, De Vos A, Van De Voorde S, Verdyck P, Verpoest W, Keymolen K. Clinical experience of preimplantation genetic testing. *Reproduction* 2020;**160**:A45–A58.
- Defraigne C, Brotchi J. Wetsvoorstel tot wijziging van de wet van 6 juli 2007 betreffende de medisch begeleide voortplanting en de bestemming van de overtallige embryo's en de gameten teneinde een uitbreiding van het toepassingsgebied van de pre-implantatiediagnostiek mogelijk te maken. *Belgische Senaat* 2011. <https://www.senate.be/www/?Mlval=publications/viewPub&COLL=S&PUID=83887501&TID=83889051&POS=1&LANG=nl>.
- Duguet A-M, Boyer-Beviere B. Preimplantation genetic diagnosis: the situation in France and in other European countries. *Eur J Health Law* 2017;**24**:160–174.
- Eccles J, Marin D, Duffy L, Chen SH, Treff NR. Rate of patients electing for polygenic risk scores in preimplantation genetic testing. *Fertil Steril* 2021;**116**:e267–e268.
- Ethics Committee of the American Society for Reproductive Medicine. Use of reproductive technology for sex selection for nonmedical reasons: an Ethics Committee opinion. *Fertil Steril* 2022;**117**:720–726.
- European Society of Human Reproduction and Embryology. ESHRE supports the position of ESHG on embryo selection based on polygenic risk scores. 2022. <https://www.eshre.eu/Europe/Position-statements/PRS> (27 September 2023, date last accessed).
- Fahrenkrog AR. A comparison of international regulation of preimplantation genetic diagnosis and a regulatory suggestion for the United States. *Transnatl Law Contemp Probs* 2006;**15**:757–782.
- Ferraretti AP, Pennings G, Gianaroli L, Natali F, Magli MC. Cross-border reproductive care: a phenomenon expressing the controversial aspects of reproductive technologies. *Reprod Biomed Online* 2010;**20**:261–266.
- Forzano F, Antonova O, Clarke A, de Wert G, Hentze S, Jamshidi Y, Moreau Y, Perola M, Prokopenko I, Read A et al.; Public and Professional Policy Committee of the European Society of Human Genetics. The use of polygenic risk scores in pre-implantation genetic testing: an unproven, unethical practice. *Eur J Hum Genet* 2021;**30**:493–495.
- France. LOI no. 2004-800. 2004.
- France. LOI no. 2011-814 relative à la bioéthique. 2011.
- France. Article L2131-4. 2021.
- France. LOI no. 2021-1017 relative à la bioéthique. 2021.
- Furrer RA, Barlevy D, Pereira S, Carmi S, Lencz T, Lazaro-Munoz G. Polygenic embryo screening: high approval despite substantial concerns from the U.S. public. medRxiv. 2023, preprint: not peer reviewed. <https://doi.org/10.1101/2023.10.13.23297022>.
- Geraedts J, Coonen E, Dreesen J, Paulussen A, Die-Smulders C. Präimplantationsdiagnostik in den Niederlanden. *Medizinische Genetik* 2016;**28**:310–319.
- Germany. Embryo Protection Act. 1990.
- Germany. Embryo Protection Act (amendment). Federal Law Gazette, Part I. 2011.
- Gianaroli L, Crivello AM, Stanghellini I, Ferraretti AP, Tabanelli C, Magli MC. Reiterative changes in the Italian regulation on IVF: the effect on PGD patients' reproductive decisions. *Reprod Biomed Online* 2014;**28**:125–132.
- Ginoza MEC, Isasi R. Regulating preimplantation genetic testing across the world: a comparison of international policy and ethical perspectives. *Cold Spring Harb Perspect Med* 2020;**10**:a036681.
- Granato C. Diagnosi Genetica Pre-Implanto. Una Tecnica Preziosa. 2021. <https://chiaragranato.it/diagnosi-genetica-preimpianto-una-tecnica-preziosa/> (10 August 2023, date last accessed).
- Grubliauskienė N. Legal regulation of preimplantation genetic diagnosis: a comparative analysis of the Baltic Sea region and the Nordic countries. *Teisė* 2022;**125**:72–85.
- Hreinsson J, Lundin K, Iwarsson E, Hausken J, Einarsson S, Grondahl ML, Hyden-Granskog C, Ingerslev HJ. Preimplantation genetic testing legislation and accessibility in the Nordic countries. *Acta Obstet Gynecol Scand* 2020;**99**:716–721.
- Human Fertilisation & Embryology Authority. Code of Practice 9.0. 2021.
- Human Fertilisation & Embryology Authority. Approved PGT-M and PTT conditions. 2023a. <https://www.hfea.gov.uk/treatments/embryo-testing-and-treatments-for-disease/approved-pgt-m-and-ptt-conditions/> (9 August 2023, date last accessed).
- Human Fertilisation & Embryology Authority. Our authority, committees and panels. 2023b. <https://www.hfea.gov.uk/about-us/our-authority-committees-and-panels/> (9 August 2023, date last accessed).
- Human Fertilisation & Embryology Authority. Pre-implantation genetic testing for aneuploidy (PGT-A). 2023c. <https://www.hfea.gov.uk/treatments/treatment-add-ons/pre-implantation-genetic-testing-for-aneuploidy-pgt-a/> (23 October 2023, date last accessed).
- Italy. Legge no. 194. 1978.
- Italy. Norme in materia di procreazione medicalmente assistita. 2004.
- Karavani E, Zuk O, Zeevi D, Barzilai N, Stefanis NC, Hatzimanolis A, Smyrnis N, Avramopoulos D, Kruglyak L, Atzmon G et al. Screening human embryos for polygenic traits has limited utility. *Cell* 2019;**179**:1424–1435.e8.
- Kettering KR. "Is down always out?": the right of Icelandic parents to use preimplantation genetic diagnosis to select for a disability. *The George Washington Int Law Rev* 2019;**51**:1–29.
- Kim NK. Legislation on genetic diagnosis: comparison of South Korea and Germany: - with focus on the application and communication structure. *Dev Reprod* 2015;**19**:111–118.
- Koehler U, Schoen U, Mayer V, Holinski-Feder E. Preimplantation genetic diagnosis for monogenic disorders and chromosomal rearrangements—the German perspective. *J Reprodukt Endokrinol* 2013;**10**:38–44.
- Krones T, Richter G. Preimplantation genetic diagnosis (PGD): European perspectives and the German situation. *J Med Philos* 2004;**29**:623–640.
- L'Agence de la biomédecine. Conseil D'Orientation Seance du Jeudi 21 Novembre 2013. 2013.
- L'Agence de la biomédecine. Extension du DPI (DPI-HLA). 2022a.
- L'Agence de la biomédecine. Le diagnostic préimplantatoire et vous. 2022b.
- La Corte Costituzionale. Sentenza 96/2015: procreazione medicalmente assistita e diagnosi genetica preimpianto. 2015a.
- La Corte Costituzionale. Sentenza 229/2015: procreazione medicalmente assistita e diagnosi genetica preimpianto. 2015b.
- Lázaro-Muñoz G, Pereira S, Carmi S, Lencz T. Screening embryos for polygenic conditions and traits: ethical considerations for an emerging technology. *Genet Med* 2021;**23**:432–434.

- Lello L, Hsu M, Widen E, Raben TG. Sibling variation in polygenic traits and DNA recombination mapping with UK Biobank and IVF family data. *Sci Rep* 2023;**13**:376.
- Lello L, Raben TG, Hsu SDH. Sibling validation of polygenic risk scores and complex trait prediction. *Sci Rep* 2020;**10**:13190.
- Lello L, Tellier LCAM. Polygenic risk scores, PGT-P, and IVF: sibling and non-European ancestry validations. *Fertil Steril* 2021;**116**:E20.
- Lemke T, Rüppel J. Social dimensions of preimplantation genetic diagnosis: a literature review. *New Genet Soc* 2018;**38**:80–112.
- Lencz T, Backenroth D, Granot-HersHKovitz E, Green A, Gettler K, Cho JH, Weissbrod O, Zuk O, Carmi S. Utility of polygenic embryo screening for disease depends on the selection strategy. *Elife* 2021;**10**:e64716.
- Meyer MN, Tan T, Benjamin DJ, Laibson D, Turley P. Public views on polygenic screening of embryos. *Science* 2023;**379**:541–543.
- Ministerie van Volksgezondheid, Welzijn en Sport. Pre-implantatie Genetische Diagnostiek om dragerschap te voorkomen. 2019.
- Molinelli A, Bonsignore A, Darretta V, Anserini P. Results and unsolved problems following the amendment to the Italian Law on assisted reproduction brought about by the constitutional court. *Eur J Obstet Gynecol Reprod Biol* 2012;**163**:1–4.
- Nakasato K, Yamamoto BA, Kato K. Evaluating standards for 'serious' disease for preimplantation genetic testing: a multi-case study on regulatory frameworks in Japan, the UK, and Western Australia. *Hum Genomics* 2022;**16**:16.
- Nederlandse Overheid. Wet op bijzondere medische verrichtingen. 2021.
- Neuhausser WM, Fouks Y, Lee SW, Macharia A, Hyun I, Adashi EY, Penzias AS, Hacker MR, Sakkas D, Vaughan D. Acceptance of genetic editing and of whole genome sequencing of human embryos by patients with infertility before and after the onset of the COVID-19 pandemic. *Reprod Biomed Online* 2023;**47**:157–163.
- Nippert I. Genetics in contemporary Germany: regulatory responses to genetic testing options. In: eLS. Chichester: John Wiley & Sons, Ltd, 2016,1–4.
- Oviedo Convention. 1997. <https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treaty-num=164> (7 August 2023, date last accessed).
- Parfenchuk V. The right to preimplantation genetic diagnosis: biological citizenship and the challenge to the Italian law on medically assisted reproduction. *Eur J Legal Stud* 2018;**10**:127–164.
- Paternotte JM, Hermans STM. MEMORIE VAN TOELICHTING—Voorstel van wet van de leden Paternotte en Hermans tot wijziging van de Embryowet en de Wet op bijzondere medische verrichtingen in verband met het mogelijk maken van preimplantatie genetische test op dragerschap bij ernstige erfelijke aandoeningen. 2023.
- Pavone V. Genetic testing, geneticisation and social change: insights from genetic experts in Spain. In: Wieser B, Berger W (eds). *Assessing Life: On the Organisation of Genetic Testing*. München, Wien: Profil, 2010,101–132.
- Pavone V, Arias F. Beyond the geneticization thesis: the political economy of PGD/PGS in Spain. *Sci Technol Hum Values* 2011;**37**:235–261.
- Pavone V, Lafuente Funes S. Selecting what? Pre-implantation genetic diagnosis and screening trajectories in Spain. In: Wahlberg A, Gammeltoft TM (eds). *Selective Reproduction in the 21st Century*. Palgrave Macmillan, 2018,123–148.
- Pennings G. Belgian law on medically assisted reproduction and the disposition of supernumerary embryos and gametes. *Eur J Health Law* 2007;**14**:251–260.
- Pennings G, Mertes H. The Belgian experience: cultural and political factors shaping preimplantation genetic diagnosis during in vitro fertilization. In: Sills ES, Palermo GD (eds). *Human Embryos and Preimplantation Genetic Technologies*. Cambridge, MA: Academic Press, 2019,113–118.
- PGT Nederland. Jaarverslag. 2020.
- PGT Nederland. Landelijke indicatiecommissie. 2023. <https://www.pgt nederland.nl/wetenschap-ethiek-maatschappij/landelijke-indicatiecommissie> (9 August 2023, date last accessed).
- Sándor J. The ethical and legal analysis of embryo preimplantation testing policies in Europe. In: Sills ES (ed). *Screening the Single Euploid Embryo: Molecular Genetics in Reproductive Medicine*. Cham Heidelberg New York Dordrecht London: Springer, 2015,353–366.
- Shenfield F, Pennings G, Devroey P, Sureau C, Tarlatzis B, Cohen J, Force EET; ESHRE Ethics Task Force. Taskforce 5: preimplantation genetic diagnosis. *Hum Reprod* 2003;**18**:649–651.
- Siermann M, Claesen Z, Pasquier L, Raivio T, Tšuiiko O, Vermeesch JR, Borry P. A systematic review of the views of healthcare professionals on the scope of preimplantation genetic testing. *J Community Genet* 2022a;**13**:1.
- Siermann M, Tšuiiko O, Vermeesch JR, Raivio T, Borry P. A review of normative documents on preimplantation genetic testing: recommendations for PGT-P. *Genet Med* 2022b;**24**:1165–1175.
- Siermann M, Valcke O, Vermeesch JR, Raivio T, Tšuiiko O, Borry P. Limitations, concerns and potential: attitudes of healthcare professionals toward preimplantation genetic testing using polygenic risk scores. *Eur J Hum Genet* 2023;**31**:1133–1138.
- Siermann M, Valcke O, Vermeesch JR, Raivio T, Tšuiiko O, Borry P. "Are we not going too far?": socio-ethical considerations of preimplantation genetic testing using polygenic risk scores according to healthcare professionals. *Soc Sci Med* 2024;**343**:116599.
- Simpson JL, Kuliev A, Rechitsky S. Overview of preimplantation genetic diagnosis (PGD): historical perspective and future direction. In: Levy B (ed). *Prenatal Diagnosis*. New York: Humana Press, 2019,23–43.
- Società Italiana di Genetica Umana. Diagnosi Genetica Preimpianto—PGT Raccomandazioni SIGU 2017 per la pratica clinica. 2017.
- Soini S. Preimplantation genetic diagnosis (PGD) in Europe: diversity of legislation a challenge to the community and its citizens. *Med Law* 2007;**26**:309–324.
- Soini S, Ibarreta D, Anastasiadou V, Ayme S, Braga S, Cornel M, Coviello DA, Evers-Kiebooms G, Geraedts J, Gianaroli L et al.; ESHRE. The interface between assisted reproductive technologies and genetics: technical, social, ethical and legal issues. *Eur J Hum Genet* 2006;**14**:588–645.
- Spain. Ley 35 sobre Técnicas de Reproducción Asistida. 1988.
- Spain. Ley 14 sobre técnicas de reproducción humana asistida. 2006.
- The Netherlands. Embryowet. 2002 (revised in 2021).
- The Netherlands. Regeling preimplantatie genetische diagnostiek. 2009.
- The Netherlands. Regeling van de Minister van Volksgezondheid, Welzijn en Sport van 25 maart 2013, kenmerk PG/E 3158620, houdende wijziging van de Regeling preimplantatie genetische diagnostiek in verband met het vervallen van de beperking van de exclusietest bij de ziekte van Huntington. 2013.
- The United Kingdom. Human Fertilisation and Embryology Act. 1990.
- The United Kingdom. Human Fertilisation and Embryology Act. 2008.
- Treff NR, Eccles J, Lello L, Bechor E, Hsu J, Plunkett K, Zimmerman R, Rana B, Samoilenko A, Hsu S et al. Utility and first clinical application of screening embryos for polygenic disease risk reduction. *Front Endocrinol (Lausanne)* 2019a;**10**:845.
- Treff NR, Eccles J, Marin D, Messick E, Lello L, Gerber J, Xu J, Tellier L. Preimplantation genetic testing for polygenic disease relative risk reduction: evaluation of genomic index performance in 11,883 adult sibling pairs. *Genes (Base)* 2020;**11**:648.

- Treff NR, Zimmerman R, Bechor E, Hsu J, Rana B, Jensen J, Li J, Samoilenko A, Mowrey W, Van Alstine J et al. Validation of concurrent preimplantation genetic testing for polygenic and monogenic disorders, structural rearrangements, and whole and segmental chromosome aneuploidy with a single universal platform. *Eur J Med Genet* 2019b;**62**:103647.
- Tuffs A. Court allows preimplantation genetic diagnosis in Germany. *BMJ* 2010;**341**:c3741.
- Turley P, Meyer MN, Wang N, Cesarini D, Hammonds E, Martin AR, Neale BM, Rehm HL, Wilkins-Haug L, Benjamin DJ et al. Problems with using polygenic scores to select embryos. *N Engl J Med* 2021;**385**:78–86.
- Vesto A. Pre-implantation genetic diagnosis in accordance with Italian regulations. *J Clin Obstet Gynecol Infertil* 2017;**1**:1013.
- Widen E, Lello L, Raben TG, Tellier LCAM, Hsu SDH. Polygenic health index, general health, and pleiotropy: sibling analysis and disease risk reduction. *Scientific Rep* 2022;**12**:18173.
- Winter HB, Dondorp WJ, Ploem MC, Woestenburg NOM, Akerboom CPM, Legemaate J, de Wert GMWR. Evaluatie Embryowet en Wet donorgegevens kunstmatige bevruchting. 2012.
- Zhang S, Johnson RA, Novembre J, Freeland E, Conley D. Public attitudes toward genetic risk scoring in medicine and beyond. *Soc Sci Med* 2021;**274**:113796.

Reproductive genetics

Systematic molecular analyses for 115 karyotypically normal men with isolated non-obstructive azoospermia

Yuki Muranishi^{1,2}, Yoshitomo Kobori³, Yuku Katoh-Fukui¹, Satoshi Tamaoka¹, Atsushi Hattori¹, Akiyoshi Osaka³, Hiroshi Okada³, Kazuhiko Nakabayashi⁴, Kenichiro Hata⁴, Tomoko Kawai⁴, Hiroko Ogata-Kawata⁴, Toshiyuki Iwahata³, Kazuki Saito⁵, Masafumi Kon², Nobuo Shinohara², and Maki Fukami^{1,*}

¹Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, Japan

²Department of Renal and Genitourinary Surgery, Hokkaido University Graduate School of Medicine, Sapporo, Japan

³International Center for Reproductive Medicine, Dokkyo Medical University Saitama Medical Center, Koshigaya, Japan

⁴Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan

⁵Department of Comprehensive Reproductive Medicine, Graduate School, Tokyo Medical and Dental University, Tokyo, Japan

*Correspondence address. Department of Molecular Endocrinology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya, Tokyo, Japan. E-mail: fukami-m@ncchd.go.jp <https://orcid.org/0000-0001-9971-4035>

ABSTRACT

STUDY QUESTION: Do copy-number variations (CNVs) in the azoospermia factor (AZF) regions and monogenic mutations play a major role in the development of isolated (non-syndromic) non-obstructive azoospermia (NOA) in Japanese men with a normal 46, XY karyotype?

SUMMARY ANSWER: Deleterious CNVs in the AZF regions and damaging sequence variants in eight genes likely constitute at least 8% and approximately 8% of the genetic causes, respectively, while variants in other genes play only a minor role.

WHAT IS KNOWN ALREADY: Sex chromosomal abnormalities, AZF-linked microdeletions, and monogenic mutations have been implicated in isolated NOA. More than 160 genes have been reported as causative/susceptibility/candidate genes for NOA.

STUDY DESIGN, SIZE, DURATION: Systematic molecular analyses were conducted for 115 patients with isolated NOA and a normal 46, XY karyotype, who visited our hospital between 2017 and 2021.

PARTICIPANTS/MATERIALS, SETTING, METHODS: We studied 115 unrelated Japanese patients. AZF-linked CNVs were examined using sequence-tagged PCR and multiplex ligation-dependent probe amplification, and nucleotide variants were screened using whole exome sequencing (WES). An optimized sequence kernel association test (SKAT-O), a gene-based association study using WES data, was performed to identify novel disease-associated genes in the genome. The results were compared to those of previous studies and our in-house control data.

MAIN RESULTS AND THE ROLE OF CHANCE: Thirteen types of AZF-linked CNVs, including the hitherto unreported gr/gr triplication and partial AZFb deletion, were identified in 63 (54.8%) cases. When the gr/gr deletion, a common polymorphism in Japan, was excluded from data analyses, the total frequency of CNVs was 23/75 (30.7%). This frequency is higher than that of the reference data in Japan and China (11.1% and 14.7%, respectively). Known NOA-causative AZF-linked CNVs were found in nine (7.8%) cases. Rare damaging variants in known causative genes (DMRT1, PLK4, SYCP2, TEX11, and USP26) and hemizygous/multiple-heterozygous damaging variants in known spermatogenesis-associated genes (TAF7L, DNAH2, and DNAH17) were identified in nine cases (7.8% in total). Some patients carried rare damaging variants in multiple genes. SKAT-O detected no genes whose rare damaging variants were significantly accumulated in the patient group.

LIMITATIONS, REASONS FOR CAUTION: The number of participants was relatively small, and the clinical information of each patient was fragmentary. Moreover, the pathogenicity of identified variants was assessed only by *in silico* analyses.

WIDER IMPLICATIONS OF THE FINDINGS: This study showed that various AZF-linked CNVs are present in more than half of Japanese NOA patients. These results broadened the structural variations of AZF-linked CNVs, which should be considered for the molecular diagnosis of spermatogenic failure. Furthermore, the results of this study highlight the etiological heterogeneity and possible oligogenicity of isolated NOA.

STUDY FUNDING/COMPETING INTEREST(S): This study was supported by Grants from the Japan Society for the Promotion of Science (21K19283 and 21H0246), the Japan Agency for Medical Research and Development (22ek0109464h0003), the National Center for Child Health and Development, the Canon Foundation, the Japan Endocrine Society, and the Takeda Science Foundation. The results of this study were based on samples and patient data obtained from the International Center for Reproductive Medicine, Dokkyo Medical University Saitama Medical Center, Koshigaya, Japan. The authors have no conflicts of interest to disclose.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: whole exome sequencing / MLPA / spermatogenic failure / copy-number variation / mutation / non-obstructive azoospermia

Received: June 13, 2023. Revised: February 21, 2024. Editorial decision: February 29, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved. For permissions, please email: journals.permissions@oup.com

Introduction

Non-obstructive azoospermia (NOA) is the most severe form of spermatogenic failure that accounts for 10–15% of cases of male infertility (Wosnitzer et al., 2014). NOA occurs either as an isolated disorder or as a component of congenital syndromes such as Kallmann and Klinefelter syndromes (Kasak and Laan, 2021). Isolated (non-syndromic) NOA is a multifactorial condition caused by several genetic and environmental factors (Cervan-Martin et al., 2020). Reportedly, ~15% of cases with NOA result from Klinefelter syndrome or other chromosomal abnormalities, whereas ~7% carry microdeletions in the azoospermia factors a, b, and c (AZFa-c) regions at Yq11.2 and ~5% have monogenic mutations (Cioppi et al., 2021). Microdeletions in the AZF regions are assumed to impair sperm production by altering copy numbers of spermatogenic genes and increasing the risk of mosaic Y chromosomal loss in testicular cells (Cioppi et al., 2021). In the guidelines of the European Academy of Andrology (EAA) and the European Molecular Genetics Quality Network (EMQN), complete deletions in the AZFa or AZFb regions and b2/b4 deletion in the AZFc region are described as definite causes and a strong risk factor for NOA, respectively (Krausz et al., 2024). However, the statistics do not include duplications and complex copy-number variations (CNVs) in the AZF regions, although these variants have also been implicated in NOA (Saito et al., 2015). This is due to technical difficulties in detecting copy-number gains. Moreover, the frequency of monogenic mutations may be underestimated; although more than 160 genes have been reported as causative/susceptibility/candidate genes for NOA in humans or other animals (Cioppi et al., 2021; Jiao et al., 2021; Kasak and Laan, 2021; Krausz and Cioppi, 2021; Kherraf et al., 2022), several NOA-associated genes may remain unidentified. Indeed, whole exome sequencing (WES) by Ghieh et al. (2022) identified rare biallelic damaging variants in eight of 19 non-consanguineous patients with NOA due to maturation arrest. Similarly, Malcher et al. (2022) detected potentially deleterious variants in 29 of 39 patients with NOA. Thus, systematic molecular analyses are needed to clarify the genetic basis of isolated NOA.

Multiplex ligation-dependent probe amplification (MLPA) is a useful method to screen various CNVs at disease-associated loci (Stuppia et al., 2012). In addition, the optimized sequence kernel association test (SKAT-O), a gene-based association study using WES data, serves to identify causative/susceptibility genes for multifactorial disorders (Lee et al., 2012). In the present study, we conducted these methods, together with conventional sequence-tagged site-PCR (STS-PCR) and WES, to characterize the genetic variants in 115 men with isolated NOA and a normal 46, XY karyotype.

Materials and methods

Patients

The patient group consisted of 115 unrelated Japanese men, who visited our hospital between 2017 and 2021 because of infertility (Supplementary Table S1). These individuals met the following criteria: (i) clinically diagnosed with non-syndromic NOA; (ii) normal male-type external genitalia of Tanner Stage 5 with normal or relatively small testes; (iii) absence of sperm in microscopic-assisted testicular sperm extraction (TESE); (iv) 46, XY karyotype without apparent chromosomal translocation or inversion; (v) normal blood testosterone levels; and (vi) a lack of chronic disorders that may affect spermatogenesis. Most of the patients exhibited mildly or moderately elevated follicle-stimulating hormone

levels, which were assumed to reflect impaired negative feedback of inhibin on gonadotropin secretion (Kong et al., 2021).

Ethical approval and sample preparation

This study was approved by the Institutional Review Board Committee at the National Center for Child Health and Development and performed after obtaining written informed consent from participants. Genomic DNA was extracted from the peripheral leukocytes. Five control samples for MLPA were prepared in our previous study (Ogiwara et al., 2021). In-house WES data of eight healthy individuals were used as controls for mutation screening. As the control group for SKAT-O, we used in-house WES data obtained from 202 healthy Japanese men who had fathered at least one child.

Copy-number analyses of the AZF regions

First, we performed STS-PCR for all patients to detect microdeletions in the AZF regions. We examined six loci (sY84, sY86, sY127, sY134, sY254, and sY255) recommended as first-choice markers in the EAA/EMQN guidelines (Krausz et al., 2024), together with additional four other loci in the AZFc regions (sY1189, sY1191, sY1192, and sY1291). The methods have been described previously (Saito et al., 2015; Krausz et al., 2024).

Next, MLPA was conducted for patients who had no STS-PCR-detectable microdeletions. In addition, we analyzed patients with deletions in the AZFa or AZFb regions, to determine the sizes and structures of the CNVs. MLPA was performed using a commercially available kit (catalog number, P360-B2; MRC Holland, Amsterdam, Netherlands) and ABI 3500 Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) as described previously (Saito et al., 2015; Ogiwara et al., 2021). The results were analyzed using Coffalyser (MRC Holland). The relative peak area of each MLPA probe of a patient was calculated by dividing the peak area by the average of five control samples. The pathogenicity of microdeletions was assessed according to the EAA/EMQN guidelines (Krausz et al., 2024). The results of this study were compared to those of two previous studies conducted in Japan and China (Saito et al., 2015; Zhou et al., 2019).

Mutation screening

WES was performed for all patients. Libraries were constructed using the SureSelect Human All Exon Kit (version 6, Agilent Technologies, Santa Clara, CA, USA) and sequenced on NovaSeq 6000 (Illumina, San Diego, CA, USA) as 150-bp paired-end reads. Sequence reads were mapped against the human reference genome (GRCh37.p13) using BWA-MEM (version 0.7.12; <http://github.com/lh3/bwa>). Local realignment, base quality score recalibration, and variant calling were performed using the Genome Analysis Toolkit (version 3.7, Broad Institute Cambridge, MA, USA). Nucleotide variants were annotated using ANNOVAR (<http://annovar.openbioinformatics.org>). We focused on protein-truncating variants, splice-site variants, and damaging missense substitutions. To this end, functional consequences of missense substitutions were predicted using Combined Annotation Dependent Depletion (CADD; <http://cadd.gs.washington.edu>); scores of > 20 were considered damaging. Using the databases of the 1000 Genomes Project (<https://www.genome.gov/27528684/1000-genomes-project>), Tohoku Medical Megabank Organization (<https://www.megabank.tohoku.ac.jp>), or gnomAD (<https://gnomad.broadinstitute.org>), we excluded homozygous and heterozygous/hemizygous variants whose allele frequencies in the general population were more than 1.0% and 0.1%, respectively. Variants identified in our eight in-house WES data were also excluded. The pathogenicity of each variant was assessed

according to the American College of Medical Genetics (ACMG) guidelines (Richards et al., 2015). Variants of interest were confirmed by Sanger sequencing. The effects of splice site variants on mRNA structures were predicted using Alternative Splice Site Predictor (<http://wangcomputing.com/assp/>), Splice Site Prediction by Neural Network (https://www.fruitfly.org/seq_tools/splice.html), and Human Splicing Finder (version 3.1, <http://umd.be/Redirect.html>).

First, we screened rare damaging variants in 162 known causative/susceptibility/candidate genes for NOA (Supplementary Table S2). These genes were selected based on recent review articles (Cioppi et al., 2021; Jiao et al., 2021; Kasak and Laan, 2021; Krausz and Cioppi, 2021; Kherraf et al., 2022) and classified into five categories: (i) 29 genes whose pathogenic variants have been identified in multiple unrelated patients with isolated NOA and whose mutations lead to spermatogenic failure in non-human animal species (definite causative genes), (ii) 28 genes whose pathogenic variants have been identified only in a patient(s) from one family and whose mutations lead to spermatogenic failure in non-human species (putative causative genes), (iii) 24 genes that have been linked to the risk of azo/oligozoospermia in humans (susceptibility genes), (iv) 64 genes expressed in the human testis and associated with spermatogenic failure in non-human species (candidate genes), and (v) 17 genes associated with syndromic NOA in humans (causative genes for syndromic NOA).

Next, we attempted to identify biallelic or heterozygous damaging variants across the genome. To this end, we searched for homozygous, multiple-heterozygous, and hemizygous variants assessed as damaging by *in silico* prediction. When such variants were detected, we examined whether the affected genes are expressed in the testis and have been linked to any disorders. In particular, we searched for genes involved in spermatogenesis in both humans and other animals.

Gene-based genome-wide association test

We conducted SKAT-O to identify NOA-associated genes in the genome. To increase the sensitivity of SKAT-O, we focused on patients who had no known NOA-causative or -associated variants. Exome data of the selected patients and 202 control individuals were processed through the DRAGEN pipeline (version 4.0.3; Illumina). The results of the patient and control groups were compared using the SNP and Variation Suite Software (version 8.4.1; Golden helix, Bozeman, MT, USA). SKAT-O was performed using the very small-sample algorithm with the standard grid of rho values. We searched for Refseq genes whose rare damaging variants were significantly more frequent in the patient group than in the control group. Missense substitutions assessed as benign/tolerated by five or all six programs in dbNSFP (version 3.0, <http://database.liulab.science/dbNSFPconn>) and common variants whose allele frequency in the gnomAD database was

Table 1. Frequency of copy number variations (CNVs) in the azoospermia factor (AZF) regions.

Participants	The present study (Japan) Patient (n = 115) ^a	Saito et al. (Japan)		Zhou et al. (China)	
		Patient (n = 56) ^b	Control (n = 65) ^c	Patient (n = 197) ^d	Control (n = 402) ^e
Total	63 (54.8%)	33 (58.9%)	25 (38.5%)	62 (31.5%)	76 (18.9%)
Simple copy number variations					
Simple deletions					
gr/gr deletion	40 (34.8%)	23 (41.0%)	20 (30.8%)	14 (7.1%)	20 (5.0%)
b2/b4 deletion ^f	7 (6.1%)	1 (1.8%)	1 (1.5%)	9 (4.6%)	0
b1/b3 deletion	1 (0.8%)	0	1 (1.5%)	0	2 (0.5%)
b2/b3 deletion	2 (1.7%)	0	0	8 (4.1%)	24 (6.0%)
CDY1 deletion	0	1 (1.8%)	0	0	0
complete AZFa deletion ^f	0	0	0	2 (1.0%)	0
partial AZFb deletion	1 (0.8%)	0	0	0	0
complete AZFb deletion ^f	1 (0.8%)	0	0	1 (0.5%)	0
complete AZFb+c deletion ^f	0	0	0	7 (3.6%)	0
Simple amplifications ^g					
gr/gr duplication	1 (0.8%)	3 (5.3%)	0	7 (3.6%)	5 (1.2%)
gr/gr triplication	1 (0.8%)	0	0	0	0
b2/b4 duplication	1 (0.8%)	0	0	0	2 (0.5%)
b2/b3 duplication	0	0	0	3 (1.5%)	1 (0.2%)
partial AZFa duplication	1 (0.8%)	0	0	1 (0.5%)	4 (1.0%)
complete AZFb+c duplication	0	1 (1.8%)	0	0	0
Complex rearrangements					
Multiple deletions					
complete AZFa deletion with gr/gr deletion ^f	1 (0.8%)	0	0	0	0
Multiple amplifications ^g					
multiple duplications in AZFc	0	1 (1.8%)	0	2 (1.0%)	4 (1.0%)
multiple duplications in AZFb+c	0	0	1 (1.5%)	0	0
Deletion plus amplification ^g					
gr/gr deletion with b2/b4 duplication	5 (4.3%)	1 (1.8%)	2 (3.1%)	0	0
gr/gr deletion with b2/b3 triplication	0	0	0	0	2 (0.5%)
two gr/gr deletions with b2/b4 duplication	0	1 (1.8%)	0	3 (1.5%)	2 (0.5%)
two gr/gr deletions with b2/b4 triplication	1 (0.8%)	1 (1.8%)	0	0	0
b2/b3 deletion with gr/gr duplication	0	0	0	2 (1.0%)	6 (1.5%)
b2/b3 deletion with gr/gr triplication	0	0	0	0	2 (0.5%)
b2/b3 deletion with DAZ1/2 duplication	0	0	0	3 (1.5%)	2 (0.5%)

^a Patients with non-obstructive azoospermia (NOA).

^b Patients with spermatogenic failure.

^c Men with one or more children.

^d Men with azoospermia, small testes, and elevated follicle-stimulating hormone levels.

^e Men with normal sperm count, concentration, motility, and morphology who had fathered at least one child without assisted reproductive technologies.

^f Known NOA-causative CNVs described in the EAA/EMQN guidelines.

^g CNVs detected only by multiplex ligation-dependent probe amplification.

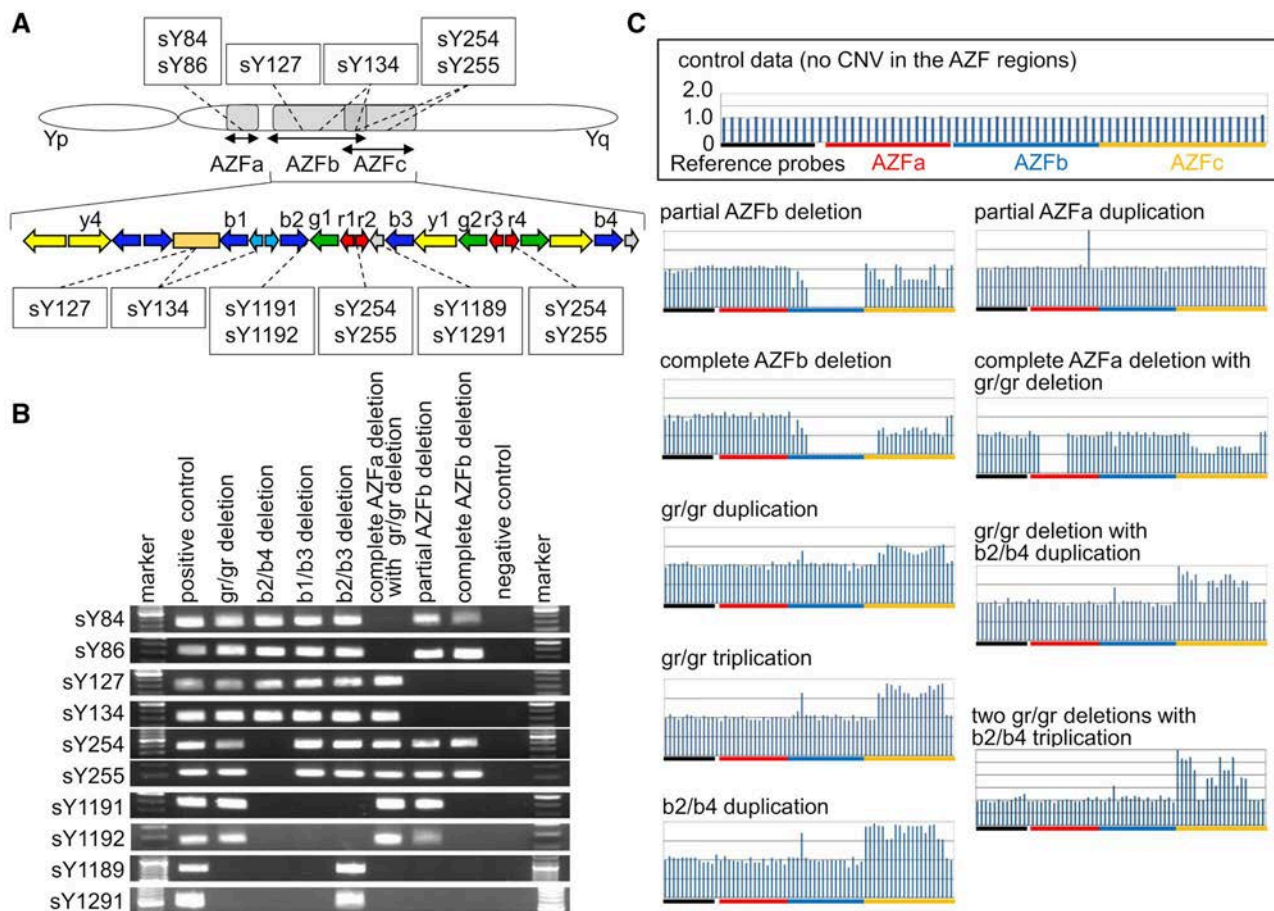


Figure 1. Copy-number analyses of the azoospermia factor (AZF) regions. (A) The genomic structure of the AZFa, b, and c regions. The colored arrows depict repetitive sequences. (B) Representative results of sequence-tagged site-PCR for ten loci. (C) Representative results of multiplex ligation-dependent probe amplification. Only duplications and complex rearrangements are shown. The peak value of 1.0 indicates copy-number neutral probes. CNV, copy number variation.

more than 0.1% were excluded. Bonferroni-corrected *P*-values of <0.05 were considered significant.

Results

Copy-number analyses of the AZF regions

STS-PCR detected seven types of microdeletions, and subsequent MLPA detected six additional types of CNVs in the AZF regions. Altogether, 13 types of CNVs were identified in 63 of the 115 patients (Table 1, Fig. 1 and Supplementary Table S1). Of these, gr/gr deletion was found in 40 patients, while other simple deletions, including partial and complete deletions in the AZFb region, were shared by 12 patients. Four types of simple amplifications were found in four patients. Three types of complex rearrangements, including multiple deletions in the AZFa and AZFc regions, were identified in seven patients. Notably, known NOA-causative microdeletions, described in the EAA/EMQN guidelines (complete AZFa/AZFb deletions and b2/b4 deletion) (Krausz et al., 2024), were identified in nine patients.

We compared the results with those of previous studies in Japan and China (Saito et al., 2015; Zhou et al., 2019) (Table 1). The total frequency of AZF-linked CNVs in our patient group, as well as the frequencies of individual CNVs, was comparable to that of the previously reported Japanese patient group and higher than that of the Japanese control group. Likewise, our results were similar to the data of the Chinese patient group, with the

exception of the high frequency of gr/gr deletion in Japanese patients (34.8% vs 7.1%). Of note, we detected a gr/gr triplication, a partial AZFb deletion, and multiple deletions involving the AZFa and AZFc regions, which have not been previously reported (Table 1, Supplementary Fig. S1).

Mutation screening

Sequence analysis of the 162 target genes identified 14 rare damaging variants in 15 patients (Table 2). These variants were present in heterozygous or hemizygous states. Of these, three and two variants affected the definite causative genes for autosomal dominant (DMRT1, PLK4, and SYCP2) and X-linked NOA (TEX11 and USP26), respectively (Fig. 2). The c.925+1G>T variant in TEX11 was assessed as pathogenic according to the ACMG guidelines (PVS1+PM1+PM2+PM4+PP4), whereas the remaining substitutions were categorized as variants of uncertain significance (VUS). The p. R344W variant in USP26 has previously been identified in two NOA patients (Ma et al., 2016). Furthermore, a variant in a putative causative gene SYCP3 was found in one patient, and rare damaging variants in causative genes for syndromic NOA (AR, CHD7, and KISS1R) were identified in eight patients. In addition, 69 VUS were detected in the other genes (Supplementary Table S3).

Next, we analyzed putative biallelic/hemizygous variants in the genome. WES detected 73 homozygous, 204 multiple-heterozygous, and 165 hemizygous variants (Table 3 and

Table 2. Rare variants in known causative/susceptibility/candidate genes for non-obstructive azoospermia.

Gene		Variant											
Name	ID	Mode of inheritance ^a	cDNA	Protein	Zygosity	CADD_phred ^b	1000G_all ^c	ToMMo ^d	gnomAD ^e	Previously identified in patients ^f	Amino acid conservation ^g	ACMG ^h	Case
Variants in definite causative genes													
DMRT1	NM_021951	AD	c.204G>T	p.K68N	Heterozygote	27.6	–	–	–	No	Medium	VUS	13
PLK4	NM_014264	AD	c.323T>C	p.F108S	Heterozygote	28.5	–	–	–	No	Medium	VUS	15
SYCP2	NM_014258	AD	c.188T>G	p.I63S	Heterozygote	–	–	–	–	No	Medium	VUS	10
			c.1526_1528del	p.R509del	Heterozygote	–	–	–	1.63 × 10 ^{−4}	No	Low	VUS	11
TEX11	NM_031276	XL	c.925 + 1G>T	–	Hemizygote	–	–	–	–	No	–	Pathogenic	14
USP26	NM_031907	XL	c.1030C>T	p.R344W	Hemizygote	24.1	–	–	6.63 × 10 ^{−5}	Yes	Low	VUS	12
Variants in putative causative genes													
SYCP3	NM_00117948	AD	c.647T>C	p.M216T	Heterozygote	25.9	–	–	–	No	Medium	VUS	16
Variants in a causative gene for syndromic non-obstructive azoospermia													
AR	NM_000044	XL	c.167_168insGCA	p.L56_L57insQ	Hemizygote	–	–	–	1.05 × 10 ^{−4}	No	–	VUS	72
			c.215_216insGCG	p.Q72_Q73insR	Hemizygote	–	–	–	6.97 × 10 ^{−5}	No	–	VUS	69
			c.528C>A	p.S176R	Hemizygote	27.4	5.30 × 10 ^{−4}	–	6.61 × 10 ^{−4}	No	Medium	VUS	73
			c.1208C>T	p.A403V	Hemizygote	25.8	–	–	2.70 × 10 ^{−5}	No	Medium	VUS	68
CHD7	NM_017780	AD	c.5815C>T	p.R1939W	Heterozygote	35.0	–	–	–	No	High	VUS	70
			c.6745G>A	p.D2249N	Heterozygote	27.2	–	9.00 × 10 ^{−4}	1.14 × 10 ^{−5}	No	High	VUS	1,74
KISS1R	NM_032551	AD, oligogenic	c.316C>T	p.P106S	Heterozygote	24.5	–	–	6.38 × 10 ^{−6}	No	Medium	VUS	71

ACMG, the American College of Medical Genetics guidelines; AD, autosomal dominant; VUS, variant of unknown significance; XL, X-linked.

^a Mode of inheritance of spermatogenic failure.^b Functional prediction using CADD (Combined).^c Allele frequency in the 1000 Genomes Project (<https://www.genome.gov/27528684/1000-genomes-project>). Scores above 20 are indicative of deleterious effects.^d Allele frequency in the Tohoku Medical Megabank Organization Database (ToMMo, <https://www.megabank.tohoku.ac.jp/>).^e Allele frequency in the Genome Aggregation Database (gnomAD, <https://gnomad.broadinstitute.org/>).^f Patients with non-obstructive azoospermia reported in PubMed or the Human Gene Mutation Database.^g Sequence conservation of the affected amino acid across species: high, conserved from fish to human; medium, conserved from mouse to human; low, conserved only among primates.^h ACMG, American College of Medical Genetics; VUS, variants of uncertain significance.

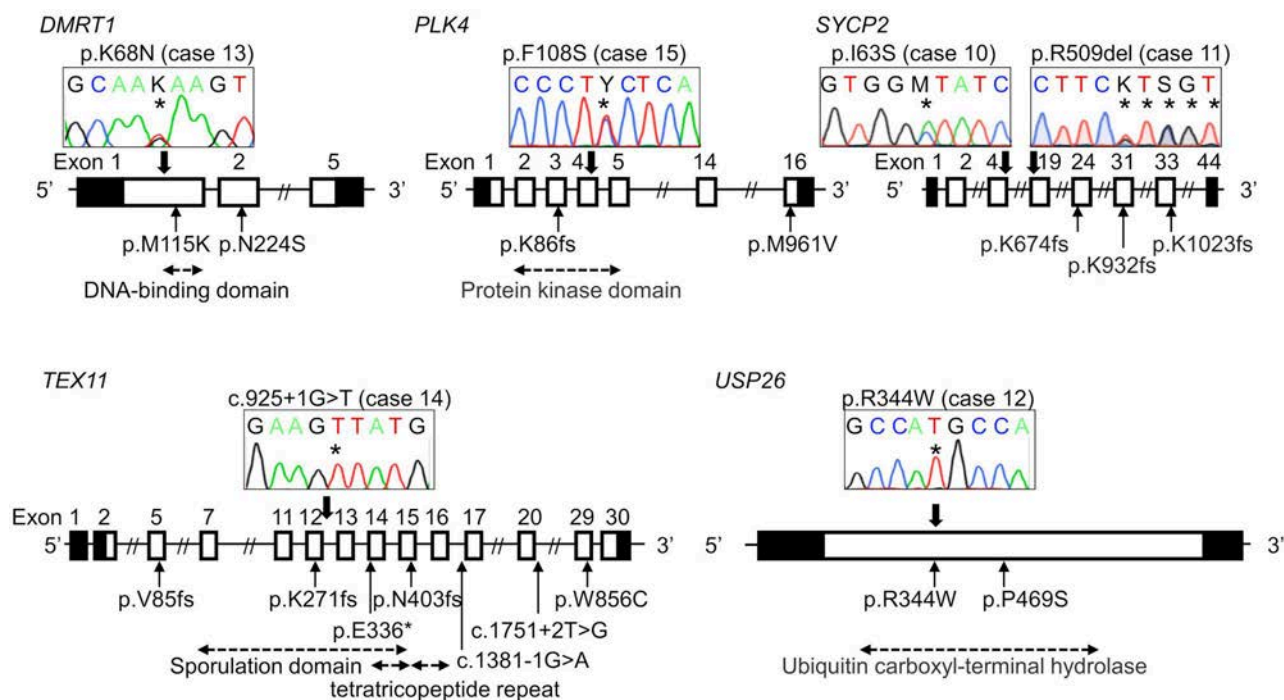


Figure 2. Variants identified in definite causative genes for autosomal dominant or X-linked non-obstructive azoospermia (NOA). The variants identified in our cases (thick arrows) and known pathogenic variants submitted to the Human Gene Mutation Database (HGMD) (<https://www.hgmd.cf.ac.uk/ac/index.php>) are shown. Asterisks in the chromatographs indicate mutated nucleotides. The horizontal broken arrows depict sequences encoding specific domains. The black and white boxes indicate the noncoding and coding regions, respectively. The exons and introns are not drawn to scale.

Supplementary Table S4). Of these, nine variants were located in genes known to be associated with spermatogenic failure in humans and other animals (Table 3). The nine variants included a hemizygous variant in *TAF7L* and multiple-heterozygous variants in *DNAH2* and *DNAH17*. Case 19 carried hemizygous variants in two genes. The remaining 433 variants were unlikely to be associated with NOA; most of the affected genes have been implicated in disorders other than spermatogenic failure (Supplementary Table S4).

Gene-based association test

SKAT-O was performed for 93 patients who had no known NOA-causative CNVs or monogenic variants (cases 23–115). The results showed no genes whose rare damaging variants were significantly more common in the patient group than in the control group (Fig. 3).

Discussion

This study aimed to elucidate the genetic basis of isolated NOA in karyotypically normal men. As the first step, we examined submicroscopic CNVs in the AZF regions. Thirteen types of CNVs were identified in 63 of the 115 (54.8%) patients. The results provided three notable findings. First, 40 (34.8%) patients carried *gr/gr* deletion. This deletion is a hallmark of the D-haplotype allele, which accounts for ~30% of the Y chromosomes of Japanese men (Iijima et al., 2020), although the deletion occasionally occurs on other haplotype alleles (Yang et al., 2010). Reportedly, *gr/gr* deletion on the D-haplotype alleles barely exerts negative effects on spermatogenesis, whereas the same deletion on other haplotype alleles mildly increases the risk of azo/oligozoospermia (Rozen et al., 2012; Saito et al., 2015;

Krausz et al., 2024). Thus, the population-dependent effects of *gr/gr* deletion should be considered during molecular diagnosis of spermatogenic failure. Given that the frequency of *gr/gr* deletion in our patient group was slightly higher than that in a Japanese control group (34.8% vs 30.8%), this deletion may increase the risk of NOA in Japanese men with non-D-haplotype alleles. Second, when the *gr/gr* deletion was excluded from data analysis, the total frequency of AZF-linked CNVs in our patient group was 23/75 (30.7%). This frequency is higher than that of control groups from Japan and China (11.1% and 14.7%, respectively) and comparable to that of a Japanese patient group (30.3%). These findings suggest that several CNVs of our patients other than *gr/gr* deletion are associated with the risk of NOA. Indeed, known NOA-causative microdeletions described in the EAA/EMQN guidelines (Krausz et al., 2024) were shared by nine patients (7.8%). These findings imply that damaging CNVs in the AZF regions account for at least 8% of the cases with isolated NOA. Lastly, we identified a *gr/gr* triplication, a partial AZFb deletion, and multiple deletions in the AZFa and AZFc regions, which have not been identified in the previous study (Saito et al., 2015; Zhou et al., 2019). The partial AZFb deletion had a unique genomic structure indicative of two inversions and one deletion (Supplementary Fig. S2). The results of this study provide further evidence for the structural diversity of the AZF regions. Our data confirmed the usefulness of MLPA in the detection and characterization of AZF-linked CNVs, because MLPA was capable of detecting six CNVs that were overlooked by STS-PCR. The outcome of TESE of each AZF-linked CNV needs to be clarified in future studies.

As a second step, we performed mutation screening for 162 causative/susceptibility/candidate genes. Fourteen rare damaging variants were identified in 15 patients. Of these, the hitherto

Table 3. Multiple-heterozygous and hemizygous variants identified in genes associated with spermatogenic failure in both humans and other animals.

Gene	Variant														
	Name	ID	Mode of inheritance	Major expression site	Human patients		Gene-disrupted animal	cDNA	Protein	CADD_phred ^a	1000G_all ^b	ToMMo ^c	gnomAD ^d	Zygosity	Case
Major phenotype					Fertility										
AKAP4 CFAP47	NM_003886	XL	Testis	Testis	Asthenozoospermia	Infertile	Infertile	c.1167C>A	p.N389K	23.8	2.65 × 10 ⁻⁴	–	1.10 × 10 ⁻⁵	Hemizygote	17
	NM_001304548	XL	Testis, lung, thyroid	Testis, lung, thyroid	Asthenoteratozoospermia	Infertile	Infertile	c.177G>A	p.G6D	21.3	2.60 × 10 ⁻⁴	–	9.80 × 10 ⁻⁵	Hemizygote	18
								c.2819C>T	p.A940V	20.8	–	–	1.64 × 10 ⁻⁵	Hemizygote	19
DNAH17	NM_173628	AR	Testis, brain	Testis, brain	Asthenozoospermia	Infertile	Infertile	c.9317T>G	p.M3106R	23.6	–	5.50 × 10 ⁻³	7.00 × 10 ⁻⁵	Hemizygote	20
								c.3178G>A	p.G1060S	23.1	–	9.00 × 10 ⁻³	1.18 × 10 ⁻⁴	Multiple-	21
								c.6518C>T	p.P2173L	24.0	2.00 × 10 ⁻⁴	–	1.61 × 10 ⁻⁵	heterozygote	22
DNAH2	NM_020877	AR	Testis, lung	Testis, lung	Asthenozoospermia	Infertile	Infertile	c.5507A>G	p.K1836R	28.4	2.00 × 10 ⁻⁴	1.00 × 10 ⁻³	8.14 × 10 ⁻⁵	Multiple-	22
								c.9291G>T	p.E3097D	23.0	2.00 × 10 ⁻⁴	1.00 × 10 ⁻³	7.44 × 10 ⁻⁵	heterozygote	19
TAF7L	NM_024885	XL	Testis	Testis	Oligoasthenozoospermia	Infertile	Reduced litter size	c.666dupT	p.T223fs	–	–	–	–	Hemizygote	19

AR, autosomal recessive; XL, X-linked.

^a CADD, Combined Annotation Dependent Depletion. Scores above 20 are indicative of deleterious effects (<https://cadd.gs.washington.edu/>).^b 1000G_all, The 1000 Genomes Project (<https://www.genome.gov/27528684/1000-genomes-project>).^c ToMMo, Tohoku Medical Megabank Organization Database (<https://www.megabank.tohoku.ac.jp/>).^d gnomAD, Genome Aggregation Database (<https://gnomad.broadinstitute.org/>).

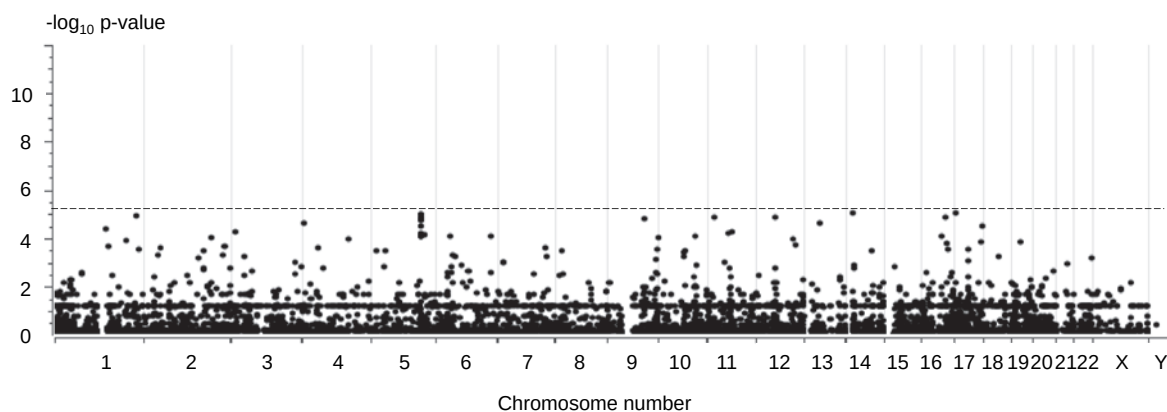


Figure 3. The Manhattan plot of SKAT-O. The broken line indicates the cut-off value of significance (Bonferroni-corrected P-values of 0.05).

unreported variants in *DMRT1* and *TEX11* appear to be associated with NOA in cases 13 and 14, respectively, because (i) the p. K68N variant in *DMRT1* had a high CADD score and the c.925 + 1G>T variant in *TEX11* was predicted to disrupt a splice donor site (Fig. 2, Supplementary Table S5), (ii) these variants were absent from all public databases examined, and (iii) monoallelic pathogenic variants of these genes are known to cause isolated NOA (Yatsenko et al., 2015; Emich et al., 2023). These findings support the importance of *DMRT1* and *TEX11* as causative genes for isolated NOA (Tang et al., 2022; Wyrwoll et al., 2023) and broaden the mutation spectrum of these genes. Similarly, the p. R344W variant of *USP26* is likely to underlie NOA in case 12, because the same variant has been reported as the cause of NOA in two patients (Ma et al., 2016). The damaging variants in the definite causative genes *PLK4* and *SYCP2* may also be associated with NOA, while the pathogenicity of the variant in the putative causative gene *SYCP3* remains to be clarified. Although rare damaging variants in *AR*, *CHD7*, and *KISS1R* were identified in seven cases, the pathogenicity of these variants was questionable, because the variant-positive patients exhibited no clinical features indicative of androgen insensitivity syndrome or gonadotropin deficiency. The other 70 heterozygous variants were unlikely to be the cause of NOA; none of the affected genes have been implicated in autosomal dominant NOA.

Next, we searched for putative biallelic/hemizygous variants in the genome. WES identified nine damaging variants in genes involved in spermatogenesis both in humans and animals. Of these, the frameshift variant in *TAF7L* appears to underlie NOA in case 19, because haploinsufficiency of this gene is known to cause oligoasthenoteratozoospermia (Bai et al., 2022). In addition, two cases carried multiple-heterozygous variants in *DNAH2* or *DNAH17*, which are known as causative genes for autosomal recessive asthenozoospermia (Li et al., 2019; Whitfield et al., 2019). Considering that biallelic pathogenic variants in other *DNAH* family genes cause isolated NOA (Gershoni et al., 2017; Malcher et al., 2022), biallelic variants of *DNAH2* and *DNAH17* may also lead to this phenotype. Recent studies have suggested that multiple morphological abnormalities of sperm flagella (MMAF) and NOA arise from common monogenic abnormalities (Tian et al., 2023). Our results further support the etiological overlap between MMAF and NOA. However, because we could not confirm the compound heterozygosity of the variants, this hypothesis awaits further validation. Likewise, the pathogenicity of variants in *AKAP4* and *CFAP47* and other genes needs to be examined in future studies. Our data highlight the importance of biallelic/hemizygous variants in the etiology of isolated NOA, but also suggest

the rarity of these variants among patients. Notably, several patients in our cohort carried multiple rare variants. For example, case 19 had rare damaging variants in two putative causative genes. These findings provide support for the possible oligogenicity of isolated NOA (Nakamura et al., 2017).

As the last step, we performed a genome-wide screening of NOA-associated genes. SKAT-O detected no genes whose rare damaging variants significantly accumulated in the patient group. These data exclude the possibility that a few specific genes play a major role in the development of isolated NOA. The negative results of SKAT-O, together with the rarity of known monogenic variants in our patients, indicate that the etiology of isolated NOA is highly diverse among patients. Several unexamined factors, such as nucleotide alterations in non-coding regions, submicroscopic chromosomal rearrangements outside the AZF regions, and environmental stress, may contribute to this disorder (Aitken and Baker, 2020; Cerván-Martín et al., 2022).

This study has several limitations. First, the number of participants was relatively small and consisted of only Japanese men. Second, the pathogenicity of identified variants was assessed only by *in silico* analyses. Since parental samples were not analyzed in this study, we could not determine whether the variants were *de novo* mutations. Hence, most variants identified in the patients were assessed as VUS according to the ACMG guidelines. Lastly, we did not analyze detailed clinical features of the variant-positive patients. Thus, it remains unknown whether each monogenic mutation is associated with specific complications or hormonal abnormalities. Further studies are needed to validate our results and to clarify the clinical characteristics of each causative variant.

Conclusion

This study revealed the etiological heterogeneity of isolated NOA. AZF-linked CNVs are present in more than half of Japanese patients, and of these, deleterious variants likely account for at least 8% of the genetic causes. It appears that damaging variants in *DMRT1*, *TEX11*, *USP26*, *DNAH2*, *DNAH17*, and three other genes jointly constitute approximately 8% of the etiology of NOA, while the role of other genetic variants is limited. Further studies are necessary to elucidate the overall picture of the molecular basis of this disorder.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The main data underlying this article are available in the article and its [Supplementary Material](#). Other data underlying this article will be shared on reasonable request to the corresponding author.

Authors' roles

Y.M., Y.K., Y.K.-F., and M.F. designed the study. M.K. and N.S. supervised the study. Y.K., A.O., H.O., and T.I. recruited the patients and collected clinical data. Y.M., S.T., A.H., K.N., K.H., T.K., and H.O.-K. performed molecular analyses. Y.M., Y.K.-F., S.T., A.H., K.S., and M.F. analyzed the data. Y.M., Y.K.-F., and M.F. drafted the manuscript. All authors revised the manuscript for important intellectual content. M.F. obtained the funding.

Funding

This study was supported by grants from the Japan Society for the Promotion of Science (21K19283 and 21H0246), the Japan Agency for Medical Research and Development (22ek0109464h0003), the National Center for Child Health and Development, the Canon Foundation, the Japan Endocrine Society, and the Takeda Science Foundation. The results of this study were based on samples and patient data obtained from the International Center for Reproductive Medicine, Dokkyo Medical University Saitama Medical Center, Koshigaya, Japan.

Conflict of interest

The authors declare that there is no conflicts of interest.

References

- Aitken RJ, Baker MA. The role of genetics and oxidative stress in the etiology of male infertility – a unifying hypothesis? *Front Endocrinol (Lausanne)* 2020;**11**:581838.
- Bai H, Sha Y, Tan Y, Li P, Zhang Y, Xu J, Xu S, Ji Z, Wang X, Chen W et al. Deleterious variants in TAF7L cause human oligoasthenoteratozoospermia and its impairing histone to protamine exchange inducing reduced in vitro fertilization. *Front Endocrinol (Lausanne)* 2022;**13**:1099270.
- Cerván-Martín M, Bossini-Castillo L, Guzmán-Jiménez A, Rivera-Egea R, Garrido N, Lujan S, Romeu G, Santos-Ribeiro S, Castilla JA, Gonzalvo MDC et al.; Lisbon Clinical Group. Common genetic variation in KATNAL1 non-coding regions is involved in the susceptibility to severe phenotypes of male infertility. *Andrology* 2022;**10**:1339–1350.
- Cervan-Martin M, Castilla JA, Palomino-Morales RJ, Carmona FD. Genetic landscape of nonobstructive azoospermia and new perspectives for the clinic. *J Clin Med* 2020;**9**:300.
- Cioppi F, Rosta V, Krausz C. Genetics of azoospermia. *Int J Mol Sci* 2021;**22**:3264.
- Dai X, Shi F, Cheung CKY, Li J, Lin S. Abnormal Y chromosome detection in infertile males using multiplex ligation-dependent probe amplification. *Andrologia* 2022;**54**:e14316.
- Emich J, Gaikwad AS, Stallmeyer B, Fietz D, Schuppe HC, Oud MS, Kliesch S, Gromoll J, Friedrich C, Tüttelmann F. Functional assessment of DMRT1 variants and their pathogenicity for isolated male infertility. *Fertil Steril* 2023;**119**:219–228.
- Gershoni M, Hauser R, Yogev L, Lehavi O, Azem F, Yavetz H, Pietrokovski S, Kleiman SE. A familial study of azoospermic men identifies three novel causative mutations in three new human azoospermia genes. *Genet Med* 2017;**19**:998–1006.
- Ghieh F, Barbotin AL, Swierkowski-Blanchard N, Blanchard N, Leroy C, Fortemps J, Gerault C, Hue C, Mambu Mambuani H, Jaillard S et al. Whole-exome sequencing in patients with maturation arrest: a potential additional diagnostic tool for prevention of recurrent negative testicular sperm extraction outcomes. *Hum Reprod* 2022;**37**:1334–1350.
- Iijima M, Shigehara K, Igarashi H, Kyono K, Suzuki Y, Tsuji Y, Kobori Y, Kobayashi H, Mizokami A. Y chromosome microdeletion screening using a new molecular diagnostic method in 1030 Japanese males with infertility. *Asian J Androl* 2020;**22**:368–371.
- Jiao SY, Yang YH, Chen SR. Molecular genetics of infertility: loss-of-function mutations in humans and corresponding knockout/mutated mice. *Hum Reprod Update* 2021;**27**:154–189.
- Kasak L, Laan M. Monogenic causes of non-obstructive azoospermia: challenges, established knowledge, limitations and perspectives. *Hum Genet* 2021;**140**:135–154.
- Kherraf ZE, Cazin C, Bouker A, Fourati Ben Mustapha S, Hennebicq S, Septier A, Coutton C, Raymond L, Nouchy M, Thierry-Mieg N et al. Whole-exome sequencing improves the diagnosis and care of men with non-obstructive azoospermia. *Am J Hum Genet* 2022;**109**:508–517.
- Kong X, Ye Z, Chen Y, Zhao H, Tu J, Meng T, Xiong C, Li H, Gong Y, Zheng L et al. Clinical application value of Inhibin B alone or in combination with other hormone indicators in subfertile men with different spermatogenesis status: a study of 324 Chinese men. *J Clin Lab Anal* 2021;**35**:e23882.
- Krausz C, Navarro-Costa P, Wilke M, Tüttelmann F. EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: State of the art 2023. *Andrology* 2024;**12**:487–504.
- Krausz C, Cioppi F. Genetic factors of non-obstructive azoospermia: consequences on patients' and offspring health. *J Clin Med* 2021;**10**:4009.
- Lee S, Ermond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani DC, Wurfel MM, Lin X; NHLBI GO Exome Sequencing Project—ESP Lung Project Team. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 2012;**91**:224–237.
- Li Y, Sha Y, Wang X, Ding L, Liu W, Ji Z, Mei L, Huang X, Lin S, Kong S et al. DNAH2 is a novel candidate gene associated with multiple morphological abnormalities of the sperm flagella. *Clin Genet* 2019;**95**:590–600.
- Ma Q, Guo H, Li Y, Li C, Chen J, Luo M, Jiang Z, Li H, Gui Y. A novel missense mutation in USP26 gene is associated with nonobstructive azoospermia. *Reprod Sci* 2016;**23**:1434–1441.
- Malcher A, Stokowy T, Berman A, Olszewska M, Jedrzejczak P, Sielski D, Nowakowski A, Rozwadowska N, Yatsenko AN, Kurpisz MK. Whole-genome sequencing identifies new candidate genes for nonobstructive azoospermia. *Andrology* 2022;**10**:1605–1624.
- Nakamura S, Miyado M, Saito K, Katsumi M, Nakamura A, Kobori Y, Tanaka Y, Ishikawa H, Yoshida A, Okada H et al. Next-generation sequencing for patients with non-obstructive azoospermia: implications for significant roles of monogenic/oligogenic mutations. *Andrology* 2017;**5**:824–831.
- Ogiwara Y, Miyado M, Suzuki E, Niida S, Ozaki K, Fukami M. Structural and numerical Y chromosomal variations in elderly men identified through multiplex ligation-dependent probe amplification. *J Hum Genet* 2021;**66**:1181–1184.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E et al.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus

- recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;**17**:405–424.
- Rozen SG, Marszalek JD, Irenze K, Skaletsky H, Brown LG, Oates RD, Silber SJ, Ardlie K, Page DC. AZFc deletions and spermatogenic failure: a population-based survey of 20,000 Y chromosomes. *Am J Hum Genet* 2012;**91**:890–896.
- Saito K, Miyado M, Kobori Y, Tanaka Y, Ishikawa H, Yoshida A, Katsumi M, Saito H, Kubota T, Okada H et al. Copy-number variations in Y-chromosomal azoospermia factor regions identified by multiplex ligation-dependent probe amplification. *J Hum Genet* 2015;**60**:127–131.
- Stuppia L, Antonucci I, Palka G, Gatta V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci* 2012;**13**:3245–3276.
- Tang D, Li K, Geng H, Xu C, Lv M, Gao Y, Wang G, Yu H, Shao Z, Shen Q et al. Identification of deleterious variants in patients with male infertility due to idiopathic non-obstructive azoospermia. *Reprod Biol Endocrinol* 2022;**20**:63.
- Tian S, Tu C, He X, Meng L, Wang J, Tang S, Gao Y, Liu C, Wu H, Zhou Y et al. Biallelic mutations in CFAP54 cause male infertility with severe MMAF and NOA. *J Med Genet* 2023;**60**:827–834.
- Whitfield M, Thomas L, Bequignon E, Schmitt A, Stouvenel L, Montantin G, Tissier S, Duquesnoy P, Copin B, Chantot S et al. Mutations in DNAH17, encoding a sperm-specific axonemal outer dynein arm heavy chain, cause isolated male infertility due to asthenozoospermia. *Am J Hum Genet* 2019;**105**:198–212.
- Wosnitzer M, Goldstein M, Hardy MP. Review of Azoospermia. *Spermatogenesis* 2014;**4**:e28218.
- Wyrwoll MJ, Köckerling N, Vockel M, Dicke AK, Rotte N, Pohl E, Emich J, Wöste M, Ruckert C, Wabschke R et al. Genetic architecture of azoospermia-time to advance the standard of care. *Eur Urol* 2023;**83**:452–462.
- Yang Y, Ma M, Li L, Su D, Chen P, Ma Y, Liu Y, Tao D, Lin L, Zhang S. Differential effect of specific gr/gr deletion subtypes on spermatogenesis in the Chinese Han population. *Int J Androl* 2010;**33**:745–754.
- Yatsenko AN, Georgiadis AP, Ropke A, Berman AJ, Jaffe T, Olszewska M, Westernstroer B, Sanfilippo J, Kurpisz M, Rajkovic A et al. X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *N Engl J Med* 2015;**372**:2097–2107.
- Zhou R, Cheng J, Ma D, Tan J, Wang Y, Hu P, Xu Z. Identifying novel copy number variants in azoospermia factor regions and evaluating their effects on spermatogenic impairment. *Front Genet* 2019;**10**:427.

Reproductive genetics

Genomic alterations in ovarian endometriosis and subsequently diagnosed ovarian carcinoma

A. Linder¹ , S. Westbom-Fremer^{2,3}, C. Mateoiu⁴, A. Olsson Widjaja¹, T. Österlund^{5,6} , S. Veerla^{2,3}, A. Ståhlberg^{5,6,7},
B. Ulfenborg⁸ , I. Hedenfalk^{2,3,*†} , and K. Sundfeldt^{1,9,*†} 

¹Department of Obstetrics and Gynecology, Sahlgrenska Center for Cancer Research, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

²Division of Oncology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden

³Lund University Cancer Centre (LUCC), Lund University, Lund, Sweden

⁴Department of Pathology, Sahlgrenska University Hospital, Gothenburg, Sweden

⁵Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden

⁶Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, Gothenburg, Sweden

⁷Department of Laboratory Medicine, Sahlgrenska Center for Cancer Research, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

⁸Department of Biology and Bioinformatics, Systems Biology Research Center, School of Bioscience, University of Skövde, Skövde, Sweden

⁹Region Västra Götaland, Sahlgrenska University Hospital, Gothenburg, Sweden

*Correspondence address. Department of Obstetrics and Gynecology, Sahlgrenska Center for Cancer Research, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, SE-405 30 Gothenburg, Sweden. E-mail: karin.sundfeldt@gu.se (K.S.)  <https://orcid.org/0000-0002-7135-3132>; Division of Oncology, Department of Clinical Sciences Lund, Lund University, SE-223 81 Lund, Sweden. E-mail: ingrid.hedenfalk@med.lu.se (I.H.)  <https://orcid.org/0000-0002-6840-3397>

†These authors contributed equally to this work.

ABSTRACT

STUDY QUESTION: Can the alleged association between ovarian endometriosis and ovarian carcinoma be substantiated by genetic analysis of endometriosis diagnosed prior to the onset of the carcinoma?

SUMMARY ANSWER: The data suggest that ovarian carcinoma does not originate from ovarian endometriosis with a cancer-like genetic profile; however, a common precursor is probable.

WHAT IS KNOWN ALREADY: Endometriosis has been implicated as a precursor of ovarian carcinoma based on epidemiologic studies and the discovery of common driver mutations in synchronous disease at the time of surgery. Endometrioid ovarian carcinoma and clear cell ovarian carcinoma are the most common endometriosis-associated ovarian carcinomas (EAOs).

STUDY DESIGN, SIZE, DURATION: The pathology biobanks of two university hospitals in Sweden were scrutinized to identify women with surgically removed endometrioma who subsequently developed ovarian carcinoma (1998–2016). Only 45 archival cases with EAO and previous endometriosis were identified and after a careful pathology review, 25 cases were excluded due to reclassification into non-EAO (n = 9) or because ovarian endometriosis could not be confirmed (n = 16). Further cases were excluded due to insufficient endometriosis tissue or poor DNA quality in either the endometriosis, carcinoma, or normal tissue (n = 9). Finally 11 cases had satisfactory DNA from all three locations and were eligible for further analysis.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Epithelial cells were collected from formalin-fixed and paraffin-embedded (FFPE) sections by laser capture microdissection (endometrioma n = 11) or macrodissection (carcinoma n = 11) and DNA was extracted. Normal tissue from FFPE sections (n = 5) or blood samples collected at cancer diagnosis (n = 6) were used as the germline controls for each included patient. Whole-exome sequencing was performed (n = 33 samples). Somatic variants (single-nucleotide variants, indels, and copy number alterations) were characterized, and mutational signatures and kataegis were assessed. Microsatellite instability and mismatch repair status were confirmed with PCR and immunohistochemistry, respectively.

MAIN RESULTS AND THE ROLE OF CHANCE: The median age for endometriosis surgery was 42 years, and 54 years for the subsequent ovarian carcinoma diagnosis. The median time between the endometriosis and ovarian carcinoma was 10 (7–30) years. The data showed that all paired samples harbored one or more shared somatic mutations. Non-silent mutations in cancer-associated genes were frequent in endometriosis; however, the same mutations were never observed in subsequent carcinomas. The degree of clonal dominance, demonstrated by variant allele frequency, showed a positive correlation with the time to cancer diagnosis (Spearman's rho 0.853, $P < 0.001$). Mutations in genes associated with immune escape were the most conserved between paired samples, and regions harboring these genes were frequently affected by copy number alterations in both sample types. Mutational burdens and mutation signatures suggested faulty DNA repair mechanisms in all cases.

LARGE SCALE DATA: Datasets are available in the supplementary tables.

LIMITATIONS, REASONS FOR CAUTION: Even though we located several thousands of surgically removed endometriomas between 1998 and 2016, only 45 paired samples were identified and even fewer, 11 cases, were eligible for sequencing. The observed high level

Received: September 18, 2023. Revised: January 25, 2024. Editorial decision: February 1, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

of intra- and inter-heterogeneity in both groups (endometrioma and carcinoma) argues for further studies of the alleged genetic association.

WIDER IMPLICATIONS OF THE FINDINGS: The observation of shared somatic mutations in all paired samples supports a common cellular origin for ovarian endometriosis and ovarian carcinoma. However, contradicting previous conclusions, our data suggest that cancer-associated mutations in endometriosis years prior to the carcinoma were not directly associated with the malignant transformation. Rather, a resilient ovarian endometriosis may delay tumorigenesis. Furthermore, the data indicate that genetic alterations affecting the immune response are early and significant events.

STUDY FUNDING/COMPETING INTEREST(S): The present work has been funded by the Sjöberg Foundation (2021-01145 to K.S.; 2022-01-11:4 to A.S.), Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (965552 to K.S.; 40615 to I.H.; 965065 to A.S.), Swedish Cancer Society (21-1848 to K.S.; 21-1684 to I.H.; 22-2080 to A.S.), BioCARE—A Strategic Research Area at Lund University (I.H. and S.W.-F.), Mrs Berta Kamprad's Cancer Foundation (FBKS-2019-28, I.H.), Cancer and Allergy Foundation (10381, I.H.), Region Västra Götaland (A.S.), Sweden's Innovation Agency (2020-04141, A.S.), Swedish Research Council (2021-01008, A.S.), Roche in collaboration with the Swedish Society of Gynecological Oncology (S.W.-F.), Assar Gabrielsson Foundation (FB19-86, C.M.), and the Lena Wäpplings Foundation (C.M.). A.S. declares stock ownership and is also a board member in Tulebovaasta, SiMSen Diagnostics, and Iscaff Pharma. A.S. has also received travel support from EMBL, Precision Medicine Forum, SLAS, and bioMCC. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Keywords: ovarian endometriosis / clear cell ovarian carcinoma / endometrioid ovarian carcinoma / exome-sequencing / mutation / copy number alteration

Introduction

Endometriosis, a benign disease that affects 10–18% of young fertile women (Giudice, 2010; Zondervan et al., 2018), may transform into cancer in the ovaries later in life. There is substantial epidemiologic evidence supporting a link between ovarian endometriosis (OE) and epithelial ovarian carcinoma (OC). Women with endometriosis have a lifetime risk of OC of 2–3% compared to a risk of 1–1.5% in the overall female population (Pearce et al., 2012; Hermens et al., 2020). There are no known predictive molecular markers or signs indicating which women with endometriosis are at risk of developing OC. If they were known, active surveillance or prophylactic surgery could be indicated to save lives.

Endometrioid ovarian carcinoma (E-OC) and clear cell ovarian carcinoma (CC-OC) are the most common endometriosis-associated ovarian carcinomas (EAOs). The genomic landscape of endometriosis and EAO has been explored in recent years. Mechanisms involved in endometriosis development include epigenetic defects, chronic inflammation, and aberrations in the regulation of estrogen and progesterone receptor pathways, with complex interactions between stromal, inflammatory, and epithelial cells (Bulun et al., 2019; Mari-Alexandre et al., 2019).

Several findings of known cancer-driving mutations in deep infiltrating endometriosis (DIE), OE, and iatrogenic endometriosis with benign histology have been reported (Anglesio et al., 2015, 2017; Lac et al., 2018; Suda et al., 2018). Frequently affected genes in endometriosis include *ARID1A*, *PIK3CA*, *KRAS*, *PTEN*, and *TP53*, which are also frequently mutated in EAO (Kuo et al., 2009; Wiegand et al., 2010; Lu et al., 2015; Murakami et al., 2017).

Endometriosis is defined as the presence of endometrial stroma and epithelium outside of the uterine corpus. The lesions are most typically located in the pelvis area, on the peritoneal surface or infiltrating nearby organs, but can be found throughout the abdominal cavity and at distant sites such as the lungs. OE, DIE, adenomyosis, and peritoneal endometriosis are suggested to be distinct subtypes. Endometriosis lesion in the vicinity of, or in continuity with, EAO is a common histopathologic finding, in which shared somatic mutations have been demonstrated (Sato et al., 2000; Anglesio et al., 2015; Er et al., 2016; Wiegand et al., 2010; Yamamoto et al., 2011). Multiple mutations in cancer-associated genes in endometrial (eutopic) tissue samples and multifocal endometriosis (ectopic) lesions have previously been described, showing that *KRAS* and *PIK3CA* mutations are shared across multiple regions of endometriosis but not

between eutopic and ectopic endometrial tissue of individual subjects (Suda et al., 2018). Conserved mutations in cancer-associated genes have also been shown between epithelial endometrial, endometriotic, and cancer cells derived from a single ancestral clone (Suda et al., 2020). Based on these findings, it is appealing to imagine a shared mutational lineage between endometrial tissue, endometriotic lesions, and possibly metachronous EAO.

Investigations on EAO cases and endometriosis prior to cancer diagnosis have not yet been conducted and may shed light on causative mechanisms. In this study, we have explored the mutational and copy number landscapes in EAO and preceding specimens of OE in paired samples from 11 patients using archival material, macro- and laser capture micro dissection, and whole-exome sequencing.

Materials and methods

Patient selection and tissue samples

Formalin-fixed and paraffin-embedded (FFPE) tissue specimens from 45 archival cases with EAO and previous endometriosis, surgically removed prior to OC diagnosis, between 1998 and 2016 were identified at the Pathology Departments of Sahlgrenska University Hospital in Gothenburg (n = 13) and Skåne University Hospital in Lund (n = 32) (Fig. 1). After pathology review (C.M./S.W.-F.), 25 cases were excluded due to reclassification into non-EAO (n = 9) or because OE could not be confirmed (n = 16). Further cases (n = 9) had either insufficient endometriosis tissue or showed poor DNA quality in either normal or endometriosis tissue (Supplementary Table S1). Eleven cases had satisfactory DNA from all three locations and were eligible for further analysis. FFPE sections rich in non-neoplastic epithelium from the fallopian tube or uterine cervix (n = 5) or blood samples collected at cancer diagnosis (n = 6) were used as the germline reference.

Tissue dissection and DNA extraction

All tissue biopsies were assessed by a specialist in gynecologic pathology (C.M./S.W.-F.) on standard hematoxylin and eosin slides. OE epithelium was collected using the PALM MicroBeam system (Zeiss, Stockholm, Sweden) by Laser Capture Microdissection. Epithelium was collected from a median of one separate biopsy of OE tissue (range, 1–5) corresponding to a median of two (range, 1–5) disconnected OE loci per specimen

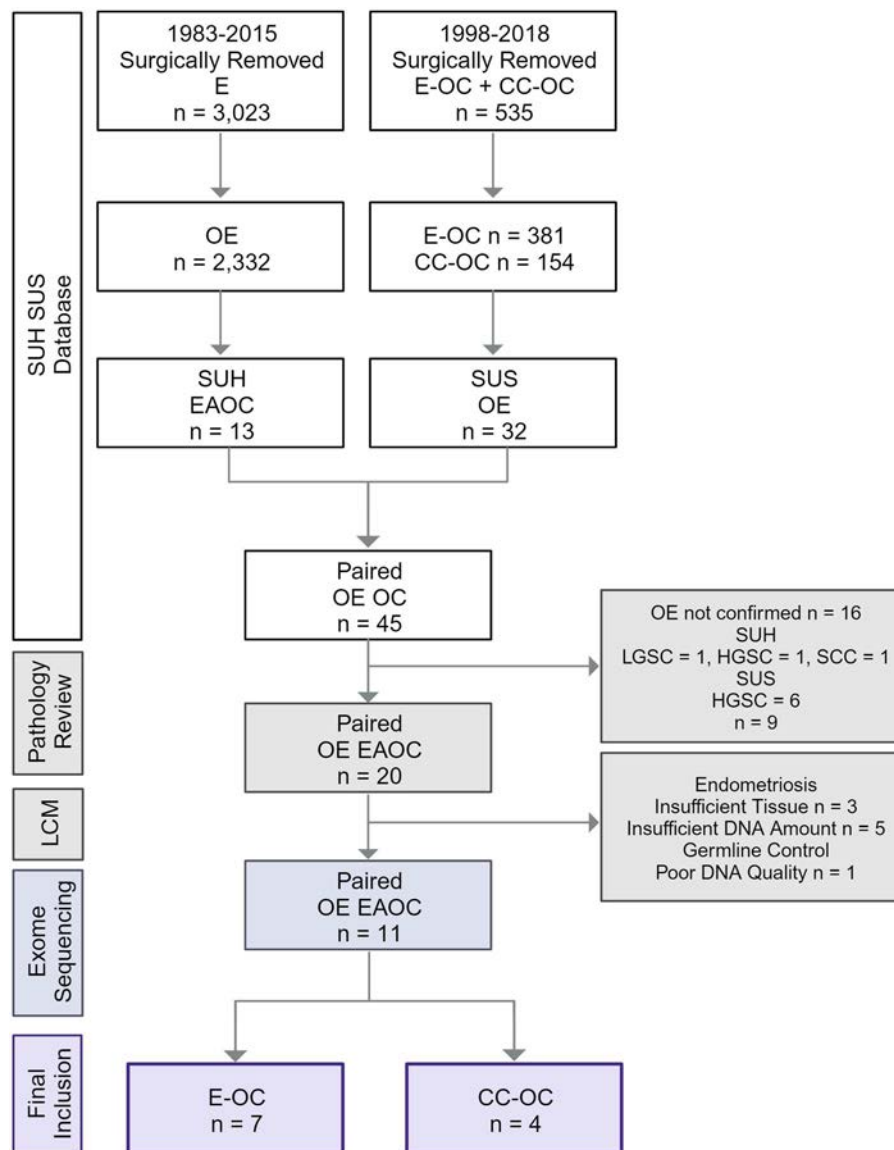


Figure 1. Patient flow chart and inclusion. CC-OC: clear cell ovarian carcinoma; E: endometriosis; EAOE: endometriosis-associated ovarian carcinoma; E-OC: endometrioid ovarian carcinoma; HGSC: high-grade serous carcinoma; LCM: laser capture microdissection; LGSC: low-grade serous carcinoma; OE: ovarian endometriosis; SCC: squamous cell carcinoma; SUH: Sahlgrenska University Hospital; SUS: Skåne University Hospital. Created with BioRender.com.

(Supplementary Fig. S1). DNA was extracted using the QIAamp DNA Micro kit (Qiagen) for Gothenburg cases and QIAamp FFPE kit (Qiagen) for Lund cases. Cell-rich carcinoma and normal tissue were macrodissected and DNA was extracted using the GeneRead DNA FFPE kit (Qiagen, Hilden, Germany) for Gothenburg cases and QIAamp FFPE kit (Qiagen, Hilden, Germany) for Lund cases. DNA from blood was extracted using QIAamp DNA Micro Kit according to the protocol (Qiagen). DNA concentration was assessed with the Qubit Fluorometer (ThermoFisher Scientific, MA, USA) and DNA quality was evaluated by TruSeq FFPE DNA Library Prep QC Kit where a ΔC_t value <7.0 was required for continuation to exome sequencing.

Exome sequencing

Sample and library preparation, enrichment, and elution were performed according to the manufacturer's protocols (Human Comprehensive exome, Twist Bioscience, Carlsbad, CA). For library preparation, 50ng of genomic DNA was fragmented by enzymatic digestion, followed by end repair and dA-tailing. The

library was enriched by seven PCR cycles using twist unique dual index primers. Then, 200ng of the library was hybridized to the Twist CoreExome EF Multiplex Complete hybridization probes in a six-sample multiplex reaction. The library was amplified with seven cycles, purified, quantified, and sequenced (Nextseq500, Illumina, San Diego, CA). Image analysis and base calling were performed (Illumina Real Time Analysis, Illumina) with default parameters.

Data processing and analysis of somatic mutations

Read mapping and data processing were performed using the Sarek pipeline (<https://nf-co.re/sarek>). FreeBayes, Mutect2, and Strelka were run for calling single-nucleotide variants (SNVs). Variants were accepted if they passed all three callers, or if they passed FreeBayes and Strelka or Mutect2. For FreeBayes, a cutoff of QUAL >20 was used. Accepted mutations were filtered to remove those with an observed allele frequency of $<0.5\%$, those with sequencing depth of <5 observed reads in samples, and

those present in the normal reference sample. Variants were annotated with SnpEff, Variant Effect Predictor (VEP), and Catalogue of Somatic Mutations in Cancer (COSMIC) cancer census genes and counts (COSMIC, v95). Function and clinical implications were assessed based on Sift, PolyPhen, and ClinVar according to the VEP annotations. Previously suggested spurious genes (Lawrence et al., 2013; Tamborero et al., 2013; Shyr et al., 2014) were excluded unless identified as endometriosis-associated genes (Joseph and Mahale, 2019). The Maftools R package was used for characterization and visualization of mutational patterns and kataegis analysis (Mayakonda et al., 2018). Gene ontology and gene set enrichment were assessed with GSEA (Subramanian et al., 2005; Liberzon et al., 2015) and DAVID (Dennis et al., 2003). Mutational profiles were compared to the COSMIC single-base substitution (SBS), doublet-base substitution (DBS), and small indel (ID) mutational signatures (Alexandrov et al., 2013), by calculating cosine similarities (lsa package, version 0.73.3). Signatures with a similarity <0.5 to all samples were excluded. All samples and remaining signatures were grouped by similarity using hierarchical clustering and visualized as a heatmap (pheatmap package, version 1.0.12).

Copy number analysis

Somatic copy number alteration (CNA) analysis was performed using SAMtools (Li et al., 2009) pileup of normal and tumor sample BAM files, followed by CNA calling with VarScan2 (Koboldt et al., 2012). Calls were normalized by the ratio of uniquely mapped reads in the normal sample to uniquely mapped reads in the tumor sample and adjusted for GC content. VarScan2 results were smoothed and segmented with DNACopy package version 1.72.0 (Seshan, 2022), and recentered genome-wide. CNA segments were merged by VarScan2's mergeSegments.pl script using default parameters, and mapped to genomic annotations with the annotatr package 1.24.0 (Cavalcante and Sartor, 2017).

Immunohistochemistry

Immunohistochemical (IHC) staining was performed for MLH1 (FLEX Monoclonal mouse Anti-Human Clone ES05, Dako, Glostrup, Denmark), MSH2 (FLEX Monoclonal Mouse Anti-Human Clone FE11, Dako), MSH6 (FLEX Monoclonal Rabbit Anti-Human Clone EP49, Dako), and PMS2 (FLEX Monoclonal Rabbit Anti-Human 2 Clone EP51, Dako) using a Dako Autostainer (Autostainer Link 48) and EnVision FLEX, High pH (Link) kit. Loss of mismatch repair (MMR) protein expression was defined as granular or absence of nuclear staining for any of the MMR proteins in all epithelial cells or a distinct area of OE or tumor, with normal immunoreactivity in the internal positive control (i.e. lymphocytes and stromal cells). IHC for ARID1A (Monoclonal Rabbit Anti-Human clone EPR13501, Abcam) was performed with Ventana Benchmark Ultra. Complete loss of ARID1A expression in all tumor or OE cells was considered negative and for a subclonal pattern, evidence of complete negativity in sections of the tumor or OE was required. IHC was assessed independently by two observers (C.M., S.W.-F.). One of the EAO blocks with inadequate fixation was replaced by an omental metastasis from the same case.

Truemark microsatellite instability assay

For microsatellite instability (MSI) analysis, remnant DNA from the exome sequencing was used when available (Lund) or freshly prepared (Gothenburg). Tissue samples were macrodissected and DNA was extracted from the FFPE (QIAamp DSP DNA FFPE Tissue Kit, Qiagen) and corresponding blood (QIAamp DNA Micro Kit). PCR reactions were set up with 2 ng DNA and prepared for

electrophoresis (TrueMark™ MSI Assay, ThermoFisher Scientific). The fragment analysis reactions were diluted 1:100, or adjusted after initial assessment, before data collection (Applied Biosystems 3500 Genetic Analyzer). Data were analyzed with the TrueMark™ MSI Analysis Software (v1.0).

Statistical analysis

Statistical analysis was performed using the IBM SPSS Statistics version 28. Groups were compared with the Mann-Whitney U or Kruskal-Wallis test for two or more groups, respectively. Correlation was evaluated with Spearman's correlation. Two-sided tests were used and $P \leq 0.05$ was considered statistically significant. Two OC samples (CC-OC-04, E-OC-10) were excluded from statistics and calculations when the data were presented on OC-group level. Data in the text are presented as median with minimum and maximum values in parenthesis. A permutation test was performed to assess whether common missense mutations between paired samples were expected by chance. Permutations were performed for 1000 iterations, where mutations were randomly seeded in the exomes of 10 *in silico*-paired samples. The number of mutations per sample was assumed to follow a negative binomial distribution fitted to the mutation data in this study (except for E-OC-10 that harbored a much larger number of mutations). To make the permuted data comparable to the real data, mutations in each *in silico* sample were seeded in the exons affected by mutations in the corresponding sample.

Ethical approval

The study was approved by the ethics committee of the Sahlgrenska University Hospital (Dnr 201-15 and T522-17).

Results

Patient data

Eleven women with surgically removed OE who later developed OC were included in the present study. Seven developed E-OC and four developed CC-OC. The median age at diagnosis of OE and OC was 42 and 54 years, respectively (Supplementary Table S2). There was no difference in age distribution between the patients at the time of OE (Mann-Whitney, $P=0.315$) or OC diagnosis (Mann-Whitney, $P=0.649$) considering the two histotypes. The median time between OE and E-OC was 10 years (range, 4–27) and CC-OC 9 years (range, 9–30). There was no correlation between age at OE and time to cancer diagnosis (Spearman's rho -0.321 , $P=0.482$). Four patients had bilateral OE, and of them, one patient developed E-OC in both ovaries. Three patients had OE and OC on opposite sides, of which two underwent unilateral salpingo-oophorectomy at the time of OE surgery (E-OC and CC-OC), and the third patient developed an ultramutated E-OC. Laterality of the OE was unknown for four patients (Supplementary Table S2).

Whole-exome analysis of OE

Whole-exome sequencing of the 11 OE samples and their matched germline control was performed for the identification of endometriosis-associated somatic mutations. The median sequencing depth was 20 reads per identified variant (range, 5–726). A total of 4248 somatic mutations were detected with a median of 204 mutations per sample (range, 78–1453) or 22.6 (range, 5.5–281.8) somatic mutations per sequenced Mb (mut/mB) (Supplementary Fig. S2). The median fraction of non-silent mutations per sample was 15% (3.5–26), while 15% (10–40) and 69% (33–87) were silent or located outside transcribed region, respectively (Fig. 2A). In all, 763 non-silent mutations were identified in

624 genes (Supplementary Table S3). This corresponded to a median of 2.8 non-silent mut/Mb per sample (range, 0.2–73.9) with a median variant allele frequency (VAF) of 0.209 (range, 0.18–0.38). There was no correlation between age at OE diagnosis and the total mut/Mb (Spearman's $\rho = -0.409$, $P = 0.212$) or non-silent mut/Mb

in cancer-associated genes (Spearman's $\rho = -0.482$, $P = 0.133$). Median VAF was ≥ 0.2 in 89.5% of the variants (Fig. 2B). For endometriosis that later developed into E-OC and CC-OC, the median VAF was 0.29 (range, 0.27–0.44) and 0.30 (range, 0.27–0.32), respectively. The median VAF showed a positive correlation with time to OC diagnosis (Spearman's $\rho = 0.615$, $P < 0.05$). The relationship was emphasized when including the VAFs for non-silent mutations only (Fig. 2C, Spearman's $\rho = 0.853$, $P < 0.001$). The time between diagnoses showed a trend for positive correlation with cancer-associated mut/Mb that did not reach statistical significance (Spearman's $\rho = 0.550$, $P = 0.079$), and a corresponding sub-analysis of non-silent mutations showed a similar result (Spearman's $\rho = 0.495$, $P = 0.121$). There was a positive correlation between mut/Mb and number of individual biopsies with OE locus per specimen (Spearman's $\rho = 0.641$, $P < 0.05$), but not considering the number of spatially separate loci (Spearman's $\rho = 0.344$, $P = 0.301$). Hypermutated regions, kataegis, were observed in OE-03 and OE-05 (Supplementary Fig. S3A and B). They carried mutations possibly affecting apolipoprotein B mRNA editing enzyme catalytic polypeptide (APOBEC) genes, with an intronic variant in OE-03 (APOBEC3F c.567-50G>A) and an intergenic variant in OE-05 (RP11-444J21.2-APOBEC1 n.7781293T>C). These two samples also demonstrated a significantly high median VAF (median 0.36 and 0.44) (Fig. 2B).

Non-silent mutations in OE

Twenty-nine (29) genes were identified as affected by non-silent mutations in more than one patient (Fig. 3A). Eight of the samples displayed one or more non-silent mutations in a cancer-associated gene (Fig. 3B). In all, 136 mutations, corresponding to 37 cancer-associated genes were observed (Supplementary Table S3). This included a recurrent mutation in KRAS, represented by two likely pathogenic mutations, c.35G>C and c.35G>A. Of the 37 mutated cancer-associated genes, 17 (46%) were associated with maintenance of genome integrity such as DNA damage, methylation and chromosomal rearrangement (Fig. 3B), and six samples displayed mutations likely to contribute to genomic instability (Supplementary Table S4). Five samples harbored mutations affecting the MAPK signaling pathway, including mutations in MOS and CHUK (OE-03), FGFR4 (OE-05), KRAS (OE-06, OE-07), CACNA1B (OE-07), MAP3K13, and PPP3CA (OE-09) (Supplementary Table S3). A comparison with previously published exome-sequencing data of endometriosis, without any known association with cancer, showed that 52% of the mutated genes in the present study were mutated in OE (Li et al., 2014), while only three mutated genes were shared with endometriosis from other sites (Anglesio et al., 2017) (Fig. 3C). Of the cancer-associated genes identified in the current study, 68% (25/37) were also mutated in OE (Fig. 3D). A sub-analysis comparison to genes that were found mutated in $>10\%$ (2/17) of the samples in the external OE cohort showed that 47% (17/36) of the cancer-associated genes in our dataset also harbored a mutation in the dataset from OE in women without any known cancer association (Li et al., 2014). The corresponding comparison with data from non-OE showed that only two of the cancer-associated genes overlapped with OE: one from each data set.

Whole-exome analysis of the subsequent OC

Whole-exome sequencing of the 11 OC samples and their matched germline control was performed for identification of cancer-associated somatic mutations. The median read depth was 21 (range, 5–633). A total of 33 232 somatic mutations, corresponding to 6909 affected genes were detected (Supplementary Fig. S4). The median number of detected mutations per OC

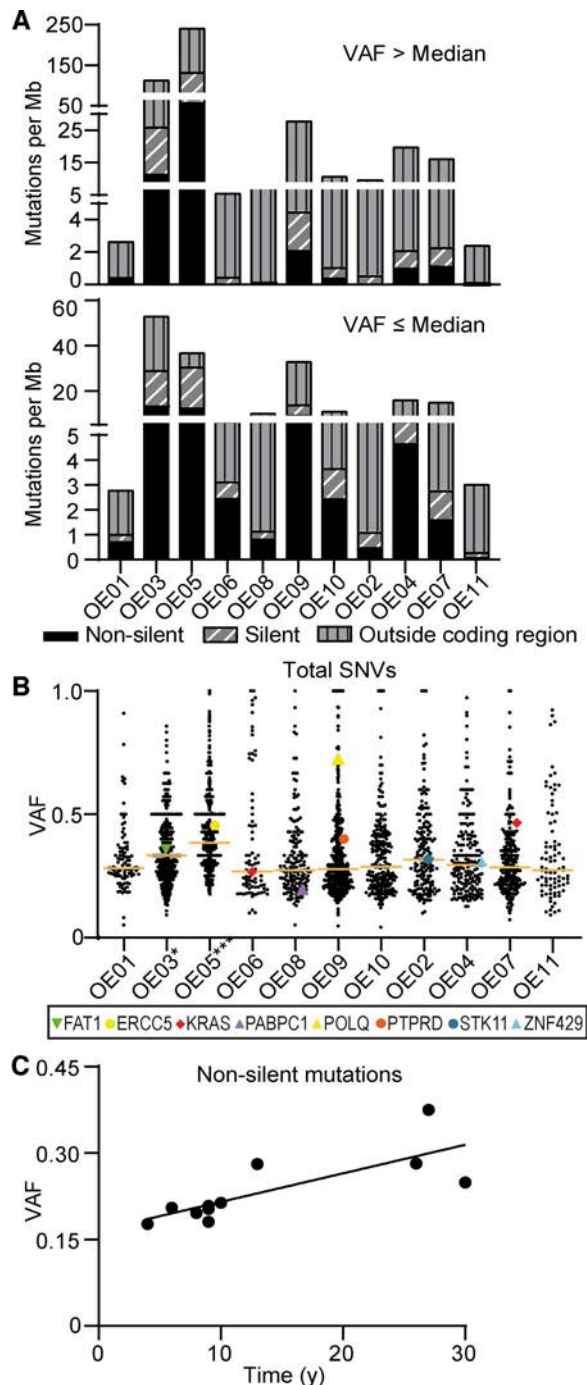


Figure 2. Evaluation of the mutational landscape of ovarian endometriosis by whole-exome sequencing. (A) Distribution of SNVs per Mb, separated by impact. Data were dichotomized by median VAF (≤ 0.353). (B) All mutations detected per specimen, horizontal line marks median VAF. Symbols annotate the non-silent SNV in the cancer-associated gene with the highest VAF. The VAF of OE-03 was higher than that of all samples except OE-05, and OE-05 had the significantly highest VAF (Kruskal–Wallis, $*P < 0.05$, $**P < 0.001$). (C) Correlation between median VAF of non-silent mutations, and time between OE and OC diagnoses. Mb: megabase; OE: ovarian endometriosis; SNV: single-nucleotide variant; VAF: variant allele frequency; Y: year.

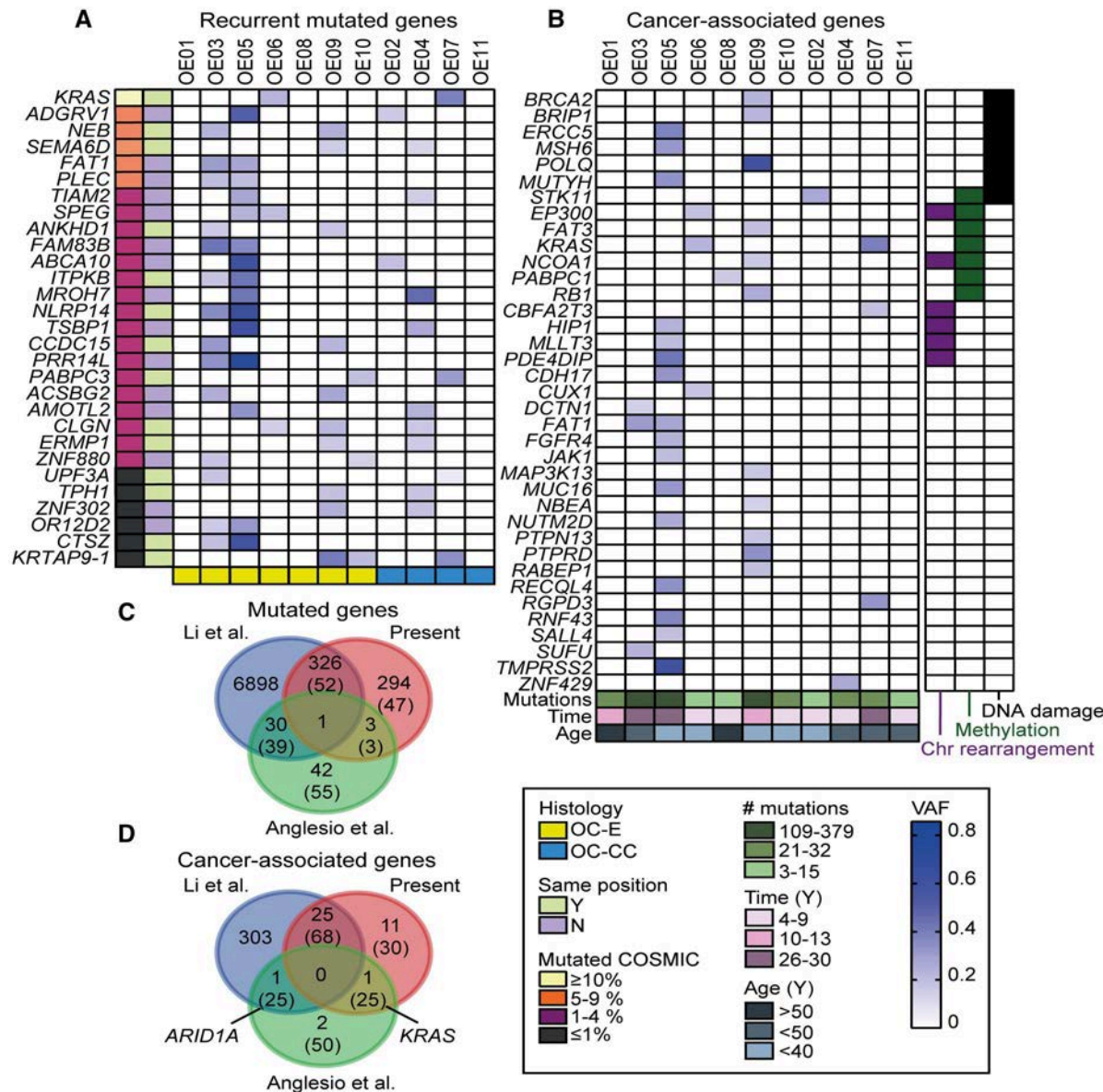


Figure 3. Non-silent mutations in ovarian endometriosis. (A) Genes harboring a mutation in more than one sample. Mutated genes assessed per sample and displayed with the histology of subsequent malignancy. In cases where the same basepair is altered in more than one specimen, this is marked in the figure. Additionally displayed is the reported mutation-frequency in COSMIC (left-hand side-panel). Color and density indicate the VAF of each somatic mutation. Reported spurious genes are excluded. (B) Mutated cancer-associated genes per sample. Patients ordered by histology of subsequent malignancy, OC-E (first seven columns) OC-CC (last four columns). Marked in the figure are the number of mutations according to low, medium, and high. Also shown is time between the two diagnoses and age at diagnosis. Overlapping functional annotations from GSEA are displayed (right-hand side-panel). (C) Venn diagram, comparison of the current dataset with previously published data. OE (Li et al.) and endometriosis in other loci (Anglesio et al.). (D) Equivalent comparison of cancer-associated genes. The number of overlapping genes is displayed together with the percentage within parenthesis. Percentage was calculated from the cohort with the least mutated genes. Chr: chromosome; COSMIC: catalogue of somatic mutations in cancer; N: No; OC-CC: clear cell ovarian carcinoma; OC-E: endometrioid ovarian carcinoma; OE: ovarian endometriosis; VAF: variant allele frequency; Y: Yes; (Y): time in years.

sample was 501 (range, 6–27 722). Two OCs were excluded from the statistical analysis and calculations; E-OC-10 had an excessive number of mutations, 27 722, of which 9666 were non-silent, and CC-OC-04 had an insufficient coverage of 1944 Mb at cutoff and only six mutations were identified, of which two were non-silent. After this exclusion, the range was 45–2097 mutations per sample, corresponding to a median of 32.9 mut/Mb (range, 3.95–328.9). Considering all mutations, 95.8% had a VAF ≥ 0.2 , with a median VAF per sample of 0.294 (range, 0.22–0.44) (Table 1) and a median VAF for non-silent mutations of 0.269 (range, 0.18–0.42). All OCs except E-OC-03 displayed an above-median VAF mutation in a cancer-associated gene (Fig. 4A). In all, 9996 non-silent

mutations were identified in 6561 genes (Supplementary Table S4). A median of 20% (range, 5–36%) of the mutations per sample were identified as non-silent (Fig. 4B), corresponding to a median number of non-silent mutations of 63 (range, 2–9666), or range of 3–295, after exclusion as above. The latter corresponded to a median of 4.8 mutations per Mb (range, 0.3–119.9).

Non-silent mutations in OC

There were 52 genes identified as affected by non-silent mutations in cancer-associated genes (Fig. 4C). The most frequently mutated cancer-associated gene was TP53, which was mutated in five OC samples (three E-OC and two CC-OC). All TP53

mutations were predicted to be pathogenic. The genes *KRAS*, *PIK3CA*, *AFF3*, and *SPEN* were mutated in three carcinomas. Two tumors, both of the CC-OC histotype, had deleterious mutations in *ERBB3*. Two tumors, one of each OC histotype, had a predicted pathogenic mutation in *PIK3CA* hotspots (c.1624G>A and c.3140A>G). Three E-OC specimens harbored mutations in *KRAS*, of which two displayed a hotspot mutation in c.35G>A, while the hypermutated E-OC-10 harbored an atypical mutation, c.39C>T. Stop-gain mutations were detected in *PTEN* and *ARID1A* in one E-OC and one CC-OC, respectively. The *ARID1A* stop-gain mutation was confirmed with IHC, while the remaining specimens were positive for *ARID1A* (Supplementary Fig. S5). The hypermutated E-OC-10 sample carried non-silent mutations in 331 cancer-associated genes, including a deleterious missense mutation in the exonuclease domain of *POLE*; in all, it was analogous with the ultramutated genotype of endometrial carcinoma. This tumor also harbored mutations in several genes associated with maintenance of genome integrity (Supplementary Table S5). Additionally, E-OC-10 showed evidence of kataegis (Supplementary Fig. S3C) and had multiple mutations in kataegis genes (seven in *APOBEC* genes and three in *REV1*), one of which was a missense variant (*APOBEC3B* c.31C>T).

Shared mutations in paired endometriosis and OC

A total of 308 shared mutations were detected in paired OE and OC samples corresponding to a median of 19 (range, 0–134) mutations per sample. All samples displayed at least one shared SNV between paired OE and OC, apart from Case 04 (Supplementary Table S6). The median VAF per sample was 0.40 (range, 0.28–0.61) and 0.39 (range, 0.16–0.70) in the OE and corresponding OC. We detected 33 shared non-silent mutations in 31 genes (Fig. 5).

Eight cases shared 1–11 non-silent variants per case. Of the variants, 13 showed a higher VAF in OC versus OE, 13 had <0.05 difference in VAFs, and seven had a lower VAF in OC compared to OE. Except for one stop codon variant in *NEB*, all variants were of missense type. Of the common variants, 51% (17/33) were predicted tolerated and benign. Seven variants were assessed as being deleterious and possibly damaging, of these, two cases had higher VAF in OC than in OE: *DLGAP3* and *NUFIP1* (Case 01), and *WI2-3308P17.2* (Case 09). Six cases (55%) shared mutations associated with effects on the immune response. Mutations in *PIBF1* (Case 01), *HLA-B* (Case 02), and *HLA-DQA2* (Case 08) were predicted as damaging or deleterious, while the remaining mutations were tolerable or benign, or no prediction annotation was made. The permutation test showed that the probability of observing one or more common missense mutations by chance was <5%.

SBS, DBS, and indel signatures

Evaluation of the cosine similarity between the mutational profile of each sample and COSMIC SBS and DBS signatures suggested that DNA repair mechanisms were frequently represented by signatures with ≥ 0.5 similarity (Fig. 6). Signatures attributed to mismatch repair deficiency (MMRd) (SBS6, 15, 26, 44, and DBS7), were observed in nine OE and seven OC samples, respectively. IHC staining for MMRd confirmed aberrations in two OCs (OC-E-03 and OC-E-10) (Table 1, Supplementary Fig. S6). IHC for OC-CC-11 was inconclusive due to fixation artefacts, while the remaining samples showed no MMR aberrations. Analysis of MSI suggested that OE-02 displayed high instability, while data from OE-04 and OE-05 were inconclusive, and the remaining OEs did not display MSI. Four OCs displayed MSI, two with high (OC-E-05 and OC-E-10) and two with low MSI (OC-CC-02 and OC-E-08). The

Table 1. Data summary for all samples.

Group	Sample	SNVs				CNAs				IHC		
		Mut/Mb		VAF (median)		Total		Cancer-associated genes		ARID1A	Results (MMRd)	
		Total	Non-silent	Total	Non-silent	Loss	Gain	Loss	Gain		MMRd	MSI
Ovarian endometriosis	OE-01	5.6	1.2	0.29	0.28	4138	602	146	19	P	MMRp	MSS
	OE-02	18.0	0.6	0.32	0.21	128	0	2	0	P	MMRp	MSI-High
	OE-03	169.3	25.6	0.36	0.28	776	203	22	4	P	MMRp	MSS
	OE-04	36.9	5.8	0.3	0.2	6	8	0	0	P	MMRp	No call
	OE-05	281.8	73.5	0.44	0.38	1252	19	43	0	P	MMRp	No call
	OE-06	13.0	2.5	0.27	0.21	6	14	0	0	P	MMRp	MSS
	OE-07	32.1	2.8	0.3	0.25	164	1677	5	37	P	MMRp	MSS
	OE-08	19.3	1.0	0.27	0.18	17	0	1	0	P	MMRp	MSS
	OE-09	61.7	11.8	0.28	0.21	38	0	0	0	P	MMRp	MSS
	OE-10	22.6	3.0	0.29	0.2	28	0	1	0	P	MMRp	MSS
	OE-11	5.5	0.2	0.27	0.18	77	679	2	19	P	IR	MSS
Ovarian carcinoma	OC-E-01	58.0	4.5	0.3	0.27	3145	6631	107	198	P	MMRp	MSS
	OC-CC-02	45.2	7.4	0.36	0.39	1251	414	43	13	P	MMRp	MSI-low
	OC-E-03	3.95	0.3	0.18	0.18	3981	2218	110	70	P	MMRd ^a	MSS
	OC-CC-04	3086.4*	1028.8*	0.42	0.42	0	10 468	0	329	P	MMRp	MSS
	OC-E-05	328.9	119.9	0.27	0.25	3402	2140	104	161	P	MMRp	MSI-high
	OC-E-06	94.1	4.8	0.22	0.2	555	1075	25	39	P	MMRp	MSS
	OC-CC-07	10.3	1.9	0.34	0.32	2963	2809	92	91	P	MMRp	MSS
	OC-E-08	22.0	4.9	0.47	0.34	11	4831	0	150	P	MMRp	MSI-low
	OC-E-09	32.9	7.7	0.26	0.23	1477	4114	55	143	P	MMRp	MSS
	OC-E-10	1013.6**	353.4**	0.42	0.41	100	80	5	2	P	MMRd ^b	MSI-high
	OC-CC-11	6.7	1.2	0.26	0.24	4038	1476	189	49	N	IR	MSS

CNA: copy number alterations; IHC: immunohistochemistry; IR: inconclusive results; MMRd: mismatch repair deficiency; MMRp: mismatch repair proficient; MSI: microsatellite instable; MSS: microsatellite stable; Mut/Mb: mutations per megabase; N: negative; P: positive; SNV: single-nucleotide variants; VAF: variant allele frequency.

^a Subclonal MSH2/6.

^b Negative MSH6.

* Low coverage.

** Hypermutated.

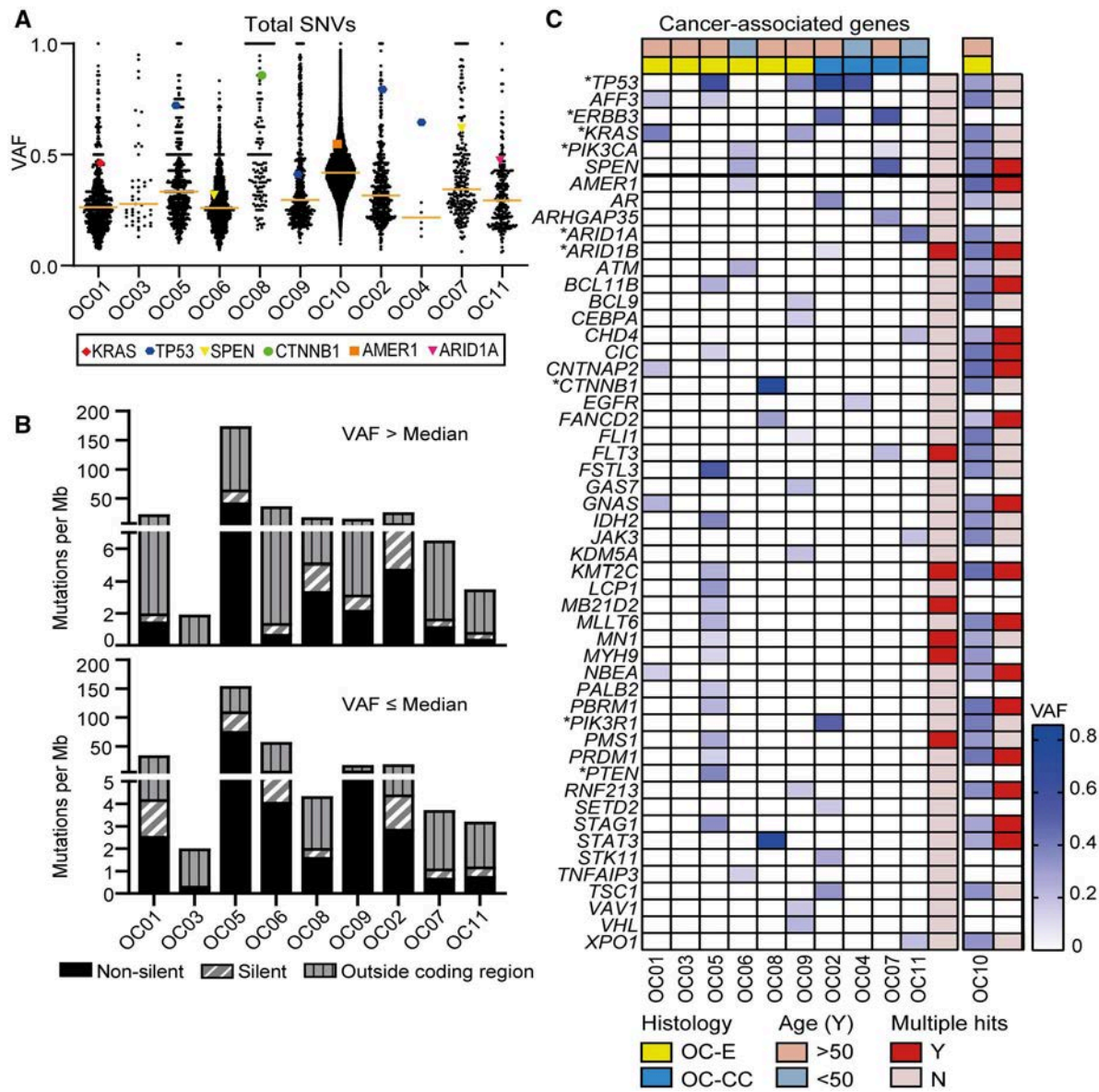


Figure 4. Characterization of mutations in endometriosis-associated ovarian carcinoma. (A) All mutations detected per specimen; the horizontal line marks the median VAF. Symbols annotate the non-silent SNV in the cancer-associated gene with the highest VAF. (B) Distribution of SNVs per Mb, separated by impact and location. Data were dichotomized by median VAF (≤ 0.402). (C) Heatmap of the 52 cancer-associated genes mutated in the 11 OC specimens. Color densities indicate the VAF of each somatic mutation. The hypermutated E-OC-10 was considered separately, and subsequently included for a complete data set. Horizontal demarcation indicates the six genes mutated in more than one sample, not including the contribution of E-OC-10. The asterisk (*) indicates genes expected to be affected by mutations in EAOC. EAOC: endometriosis-associated ovarian carcinomas; N: No; OC-CC: clear cell ovarian carcinoma; OC-E: endometrioid ovarian carcinoma; SNV: single-nucleotide variant; VAF: variant allele frequency; Y: Yes; (Y): Time in years.

signature associated with homologous repair deficiency (SBS3) was observed in eight OE and OC samples, respectively. The most frequent signature, suggestive of deficiency in transcription-coupled nucleotide excision repair (TC-NER) (SBS5) was seen in all samples, except OC-CC-04 and OC-E-10. Indel signatures (IDs), ID1 and ID2, suggestive of DNA slippage during replication, were the most frequently noted, in 77% (17/22) and 59% (13/22) of the samples. The OC-E-10, harboring missense mutations in *POLE*, *POLD1*, *MSH2*, *MSH6*, and *PMS2*, showed high similarity to additional signatures coupled to *POLE* or *POLD1* mutations (SBS20 and 14) and MMRd (SBS44 and DBS10).

CNAs in endometriosis and OCs

When comparing paired samples, OC showed more CNAs than OE, with a median of 256 (range, 17–521) events per tumor and

135 (range, 62–379) events per OE specimen. All samples in the present cohort displayed CNAs affecting multiple genes with a median of 128 (range, 14–4740), and 5542 (range, 181–10 468) genes affected in OE and OC, respectively (Table 1, Supplementary Table S7). Overall, the OEs displayed more chromosomal loss than gain of genetic material (Supplementary Fig. S7). Generally, chromosomal gains consisted of smaller chromosomal regions and partial gene sequences. Eight OEs displayed a CNA affecting one or more cancer-associated genes (Table 1), median 49.5 (range, 0–165), where a predicted impact could be estimated (i.e. deletion or amplification of transcribed regions or whole gene/s, respectively). OE-01 displayed the highest number of CNAs, including loss of a large fraction of chromosome X (ChrXp22.33-q28, LogR -0.89) (Fig. 7A), affecting numerous cancer-associated genes (145) where the entire or majority of the

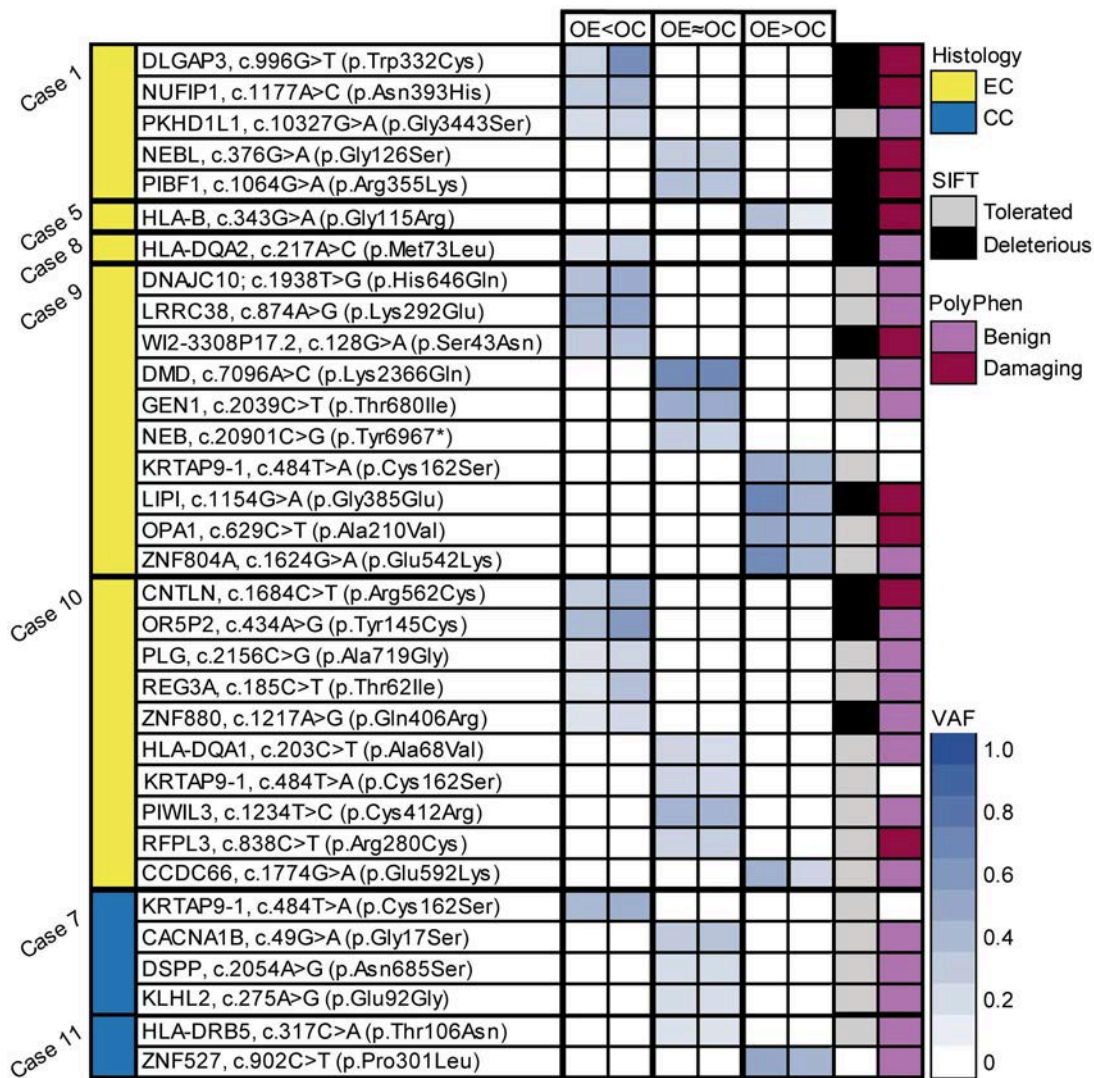


Figure 5. Shared non-silent mutations in paired endometriosis and carcinomas. Data are displayed by differences in variant allele frequency between paired samples of OE and OC. Data were trichotomized as increased VAF in OC (OE < OC), ≤ 0.05 difference (OE $\approx$ OC), or decreased in OC (OE > OC). Affected genes are displayed with base pair and amino acid alterations. Variants are predicted as tolerated (gray) or deleterious (black) in SIFT and as benign (pink) or possibly/probably damaging (purple) in PolyPhen. CC: clear cell ovarian carcinoma; EC: endometrioid ovarian carcinoma; OE: ovarian endometriosis; VAF: variant allele frequency.

gene displayed loss. Corresponding chromosomal loss was seen in the subsequent OC (ChrXp22.33-q28, LogR -0.44 to -1.16). The most telomeric region, with the most discrepant LogR, harbored two cancer-associated genes, *FLNA* and *RPL10*. Overlapping loss of chromosome X was also seen in Case 05, where both OE and OC displayed loss of multiple cancer-associated genes including *FLNA* and *RPL10*. Both patients developed E-OC. Two OEs that later developed into CC-OC displayed large gains on chromosome 16, 16p13.3-p13.13 and 16p13.3-p13.11 (Fig. 7B) in OE-07 and OE-11, respectively. This region contained six cancer-associated genes that were amplified in both specimens. For Case 07, overlap was found in OE and the subsequent CC-OC (16p13.3-p13.13), affecting eight cancer-associated genes. Data showed that, except for Cases 08 and 10, all samples shared common regions with CNAs, with a median of 10 affected genes per sample (0–540 and 0–2179 genes with gains and losses, respectively). Two pseudogenes, *DUX4L1/3*, displayed gain in three samples, and in total 137 genes displayed gain in ≥ 2 cases, including five cancer-associated genes (*CAMTA1*, *PRDM16*, *RPL22*, *SKI*, *TNFRSF1*). Functional annotation suggested enrichment of tumor necrosis

factor signaling (Cases 07 and 11). Case 07 also displayed enrichment of amplification of genes associated with negative regulation of HLA class I molecules. Shared chromosomal loss in ≥ 2 samples affected 683 genes, including 31 cancer-associated genes, where loss of *BGN* was found in three cases. Four OCs and two OEs displayed CNAs on Chromosome 6 affecting immune response genes. Loss of 6p21.33 (Fig. 7C) was seen in three samples (OE-03, OE-05, OC-E-06), where the smallest region of overlap contained four genes (*HLA-B*, *HLA-C*, *MICA*, *MICB*). OC-E-09 displayed chromosomal gain affecting an adjacent transcribed region of *HLA-G* (201 bp, LogR 1.01). Gain of 6p22.1-p21.33 was seen in OC-CC-02; this region contained the major HLA class I genes and two additional cancer-associated genes (*POU5F1*, *TRIM27*). OC-E-03 showed deletion of 6p21.32, containing a part of *HLA-DPA1* (13 965 bp, LogR -0.82).

Discussion

In the present study, we assessed the genomic landscapes of paired OE and EAOC; 11 samples from women with OE who later

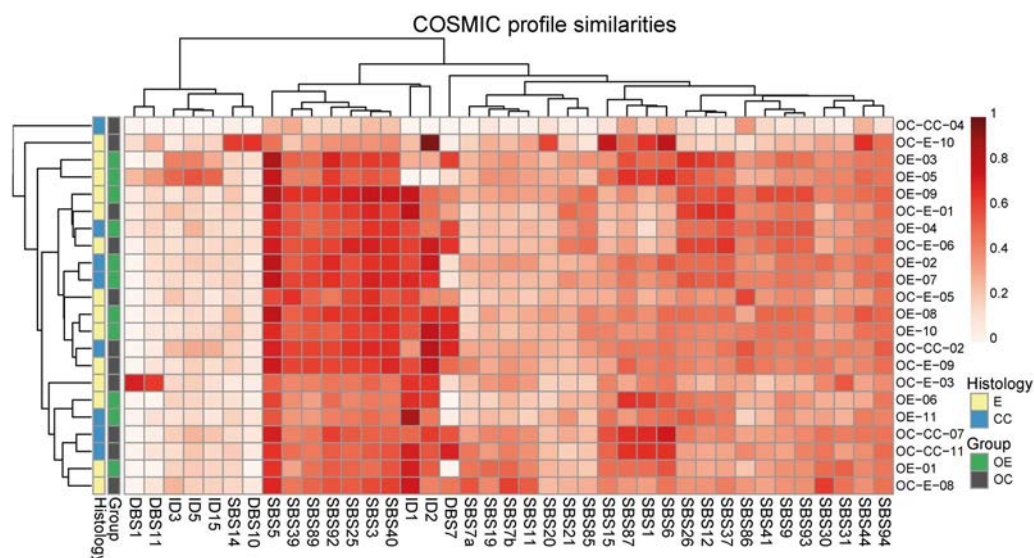


Figure 6. Profile similarities of SBS, DBS, and ID COSMIC signatures for all samples. Mutation profiles with a minimum of one sample with a signature similarity of ≥ 0.5 are included in the figure. The color density indicates the level of similarity (cosine similarity range 0.5–0.95). CC: clear cell; COSMIC: catalogue of somatic mutations in cancer; DBS: doublet-base substitution; E: endometrioid; ID: indel; OC: ovarian carcinoma; OE: ovarian endometriosis; SBS: single-base substitution.

developed either endometrioid or clear cell OC were analyzed. To our knowledge, this is the first study that has explored the mutation and copy number profiles of endometriosis with a confirmed subsequent malignant association.

We here identified conserved mutations between OE and paired OC, indicating a potential common ancestry. However, although mutations in cancer-driver genes have been suggested as the origin of malignant transformation of OE (Suda et al., 2020), no shared cancer-associated mutations were observed. Thus, although mutations in conventional cancer-driver genes were confirmed to occur frequently in OE, our original hypothesis could not be supported.

Noteworthy, the median VAFs of the shared mutations were higher than the overall median VAFs, suggesting a positive selection for clones harboring these mutations (Suda et al., 2018). Conserved mutations in genes associated with an immune response were the most prevalent, with affected genes in 5 of the 11 cases. Additionally, CNA analyses showed that aberrations involving immunological pathways were conserved. In total, seven cases exhibited genetic aberrations suggestive of an altered immune response. The single most frequently mutated type of genes was human leukocyte antigen (HLA) genes; four cases harbored a shared non-silent mutation in an HLA gene, three in class II, and one in HLA class I genes. Furthermore, the present data showed CNAs in four carcinomas and two endometriomas, mainly affecting HLA class I genes, emphasizing the importance of these mechanisms in the development of endometriosis and EAO. In line, chromosomal instability in the HLA super-locus on 6p21 has previously been described as a common alteration in EAO (Micci et al., 2014), and aberrant HLA expression has been described for both endometriosis (Semino et al., 1995; Matalliotakis et al., 2001) and OC (Hazini et al., 2021).

Additionally, the most frequently observed aberration in paired samples was chromosomal loss of BGN (encoding biglycan). Low BGN expression has been associated with decreased immunogenicity and poor patient outcome in KRAS-driven malignancies (Subbarayan et al., 2022). Furthermore, a deleterious mutation in *PIBF1* (encoding progesterone immunomodulatory

binding Factor 1) was seen in paired samples. This protein regulates the immune response during normal pregnancy, and has been suggested to facilitate tumor growth by suppressing the local antitumor response (Szekeres-Bartho and Polgar, 2010). Taken together, these genetic aberrations suggests that adaption to inflammation is an early and crucial event, in line with known mechanisms by which tumors evade the immune system (Jhunjhunwala et al., 2021). Furthermore, this finding suggests that immune-associated treatments may be beneficial not only for the treatment of endometriosis but may also have cancer-preventive effects.

Considering the chronic inflammation accompanying endometriosis (Zondervan et al., 2018), and the well-described association between inflammation and the genomic instability associated with tumor development (Coussens and Werb, 2002), it is reasonable to assume that the inflammatory environment contributes to the increased incidence of EAO. Based on other types of cancer and their association with inflammation (Balkwill and Mantovani, 2001), it could be expected that chronic inflammation would increase the incidence of cancer to an even higher extent than what is seen in association with endometriosis (Pearce, et al., 2012; Hermens et al., 2020). Furthermore, chronic inflammation is not exclusive to OE, but universal to endometriosis, and mutations in cancer-driver genes have been shown for different types of endometriosis, although no association with EAO has been described (Anglesio et al., 2015, 2017). Interestingly, the present data show a strong positive correlation between VAF and time between the two diagnoses. This implies that a dominant OE may delay tumor establishment; in other words, protect against an emerging malignancy. This is in line with the previously shown cancer-preventive properties of non-malignant epithelium in association with an accumulation of cancer-associated mutations (Colom et al., 2021). Moreover, the data show that there was no correlation between age at OE diagnosis and mutational burden, suggesting that mutations are selected, unlike mutations in normal endometrium that passively accumulate over time (Moore et al., 2020). This is in line with clonal evolution of cancer (Greaves and Maley, 2012), and

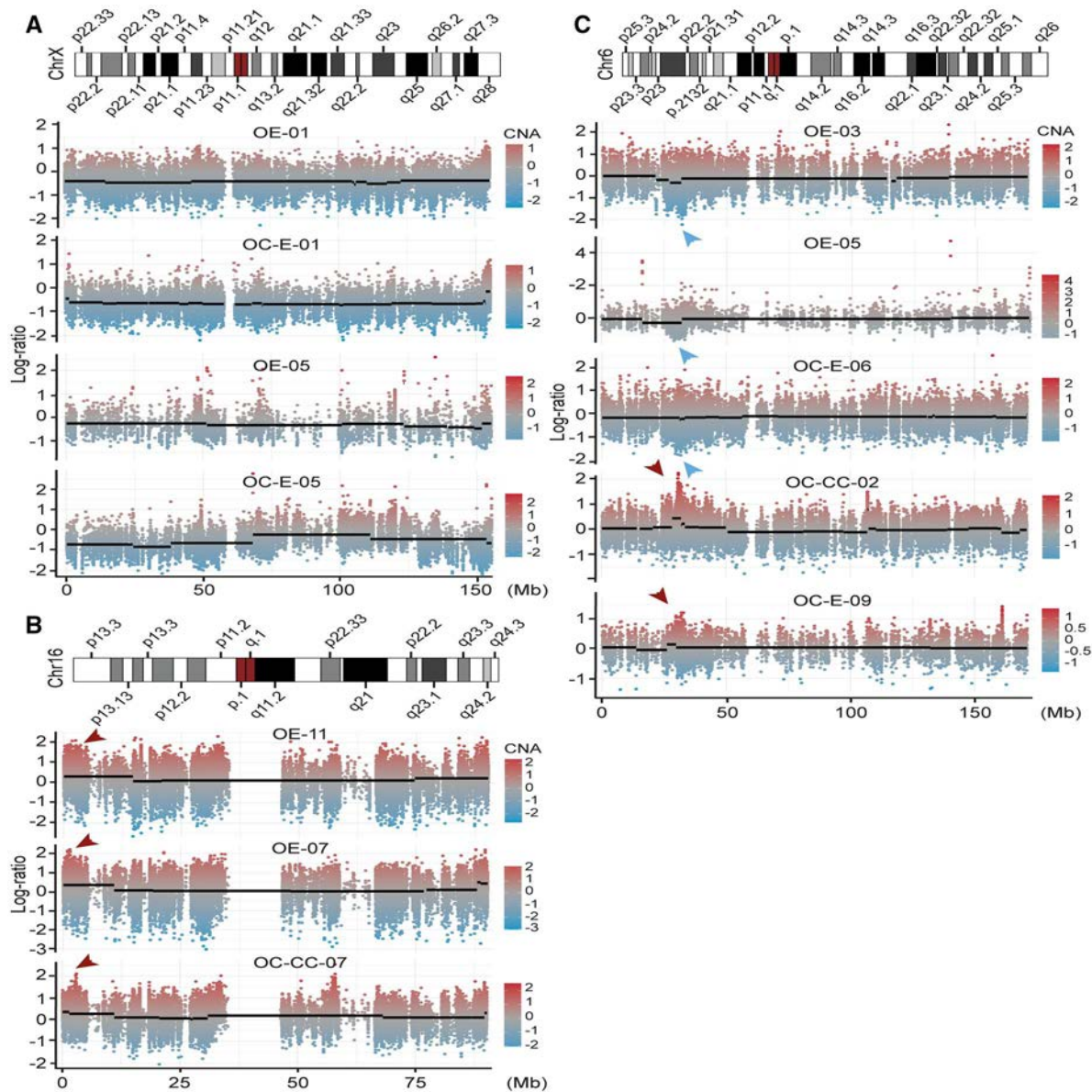


Figure 7. Visualization of recurrent large copy number alterations. (A) Extensive loss of chr X in two OE and paired OC with endometrioid histotype. (B) Gain of chr16 in OE that later developed OC of clear cell histotype, a corresponding gain was detected in paired OC for one of the cases. (C) Two OEs and four OCs displayed CNAs affecting chr 6, including loss and gain affecting HLA genes of class I and II. The y-axis represents the normalized median logR. chr: chromosome; CNA: copy number alteration; Mb: megabase; OC-CC: clear cell ovarian carcinoma; OC-E: endometrioid ovarian carcinoma; OE: ovarian endometriosis.

supports previous observations that clonal expansion of OE involves the accumulation of mutations in cancer-driver genes (Suda et al., 2018). However, the absence of conserved cancer-associated mutations between OE and OC indicates that the mutations associated with the malignant transformation are later events that occur in an originally less-potent precursor. There is sparse data on the genomic configuration of multiple and/or bilateral endometriotic lesions in the same individual. One case report described multiple shared mutations in cancer-associated genes in synchronous samples from the endometrium, different endometriosis lesions, and CC-OC (Suda et al., 2020) and in another cohort, the same group found shared KRAS or PIK3CA mutations between multiple unilateral OE lesions in six individuals, and between bilateral OE lesions in a single case (Suda et al., 2018). These findings suggest an inherent potential for progression to cancer in all OE lesions within these individuals. Hence, despite the discrepancy in laterality between OE and

OC in some of the cases in our study, the pair-wise comparisons made are relevant and interesting in relation to clonal evolution in cancer progression.

KRAS was the most frequently mutated cancer-associated gene in the current study. All affected specimens displayed the same hotspot mutation (c.35G) except the hypermutated OC-10, which carried a rarer mutation in an adjacent codon. It has been suggested that selection for specific c.35G mutations drives different tumor types (Ostrow et al., 2016). The most well-described KRAS mutation in OE (p.G12V, c. G35T) (Suda et al., 2018) was observed in one of the OE samples. This specific mutation has previously been linked to inflammation in OE (Yachida et al., 2021a). Of note, expected cancer-associated genes such as PIK3CA, ARID1A, PPP2R1A, and FBXW7 (Anglesio et al., 2015, 2017; Suda et al., 2018) did not harbor any non-silent mutations in OE. Moreover, the number of OE specimens with observed cancer-associated mutations in the present data was higher than what

has previously been reported (Anglesio et al., 2017; Suda et al., 2018). Comparative analyses suggested that OEs share several commonly affected genes (Li et al., 2014), including a large fraction (68%) of cancer-associated genes, compared to only 1% in data from endometriosis at other locations (Anglesio et al., 2017), suggesting that this genetic profile is associated with the OE subtype. A limitation of the present study was the small cohort, thus further studies are warranted to determine whether a specific genetic profile of OE contributes to the development of EAOC.

Considering OC, the mutation frequency of the conventional EAOC-associated genes differed from what has previously been described. The most frequently reported mutated gene in EAOC, ARID1A, described in 50% and 30% of CC-OC and E-OC tumors (Yachida et al., 2021), was only found to be mutated in one CC-OC. The effect of the mutation was confirmed by IHC as the single case displaying aberrant ARID1A protein expression. Thus, the frequency of aberrant ARID1A protein expression was also considerably lower than expected (Heinze et al., 2022). PIK3CA, with a described mutation frequency of 41% and 53% in E-OC and CC-OC (Yachida et al., 2021b), respectively, was found mutated in E-OC only. In contrast, the most frequently mutated gene in the tumor specimens was TP53: mutated in three E-OC and two CC-OC specimens, respectively. This suggests that in the present cohort TP53 mutations were overrepresented, particularly in CC-OC.

The data showed that both OE and OC displayed mutational profiles associated with deficiency in genes controlling genomic integrity. Analysis of cosine similarity to mutational signatures suggested that most of the specimens displayed a profile associated with deficiency in homologous repair as well as defective DNA mismatch repair. Although these findings could only be confirmed for two OCs using IHC, the MSI results suggested that at least one OE and four OCs displayed instabilities consistent with MMR deficiency. Furthermore, seven of the OEs and four of the OCs displayed genetic aberrations such as non-silent mutations or CNAs in genes associated with maintenance of genomic integrity or repair mechanisms. This included aberrations shared between paired samples such as mutations in the homologous repair gene *GEN1* (Lee et al., 2015) and deletion of *FLNA*, associated with the BRCA damage response pathway (Velkova et al., 2010). Additionally, the ultramutated OC and two of the OEs displayed signs of cancer-like kataegis (Alexandrov et al., 2013). However, only the OC specimen displayed the nucleotide substitution patterns typically associated with APOBEC-related kataegis. The OEs, on the other hand, displayed significantly high VAFs and prolonged time between diagnoses, suggesting an exceedingly strong clone.

Taken together, these data suggest that an innate defective DNA damage response and overall genomic instability, while central to the malignant transformation, also account for the resilience of OE.

The study has several limitations. The aim was to study possible causality in surgically resected OE prior to OC diagnosis in the same patient. We acknowledge that even though this is the first study of its kind, the sample size was small, the sequenced endometriosis was resected, sometimes on the contralateral side of subsequent OC, and no in-between surgery data was known, i.e. recurrence of endometrioma. However, 8 of the 11 OCs were at stages I and II at the time of diagnosis and second surgery, which might suggest that these patients were more closely surveilled since OCs are commonly diagnosed at stages III and IV.

Taken together, the present study suggests that, as a robust endometriosis lesion prevails, the accumulation of specific

cancer-driver mutations may ultimately lead to the development of a malignancy. The time required, and probability, for this to occur are associated with the genotype of the original endometriotic lesion and the environment where the precursor is located. The positive correlation between the number of distinct OE subpopulations and mutational burden, and the overall low number of recurrent mutations in the cohort, suggest intra- and inter-patient heterogeneity. This study provides a novel comprehensive understanding of the genetic diversity of OE associated with metachronous carcinoma. The complexity presented here prompts further investigative studies for an in-depth understanding of the risk of malignant transformation associated with OE.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article are available in the article and in its online [supplementary material](#).

Acknowledgements

We thank Birgitta Weijdegård, Maria Yhr, and Anna Ebbesson for excellent technical work, and Eva Rambech for MSI data evaluation. The authors would like to acknowledge the SciLifeLabs, Bioinformatics and Data Centre (BDC), Gothenburg University, and Center for Translational Genomics (CTG), Lund University for providing expertise and service with sequencing.

Authors' roles

Conceptualization: K.S., C.M., A.S., and T.Ö.; methodology: K.S., C.M., I.H., A.L., T.Ö., A.S., and B.U.; investigation: A.L., S.W.-F., C.M., B.U., S.V., and T.Ö.; validation: S.W.-F. and A.O.W.; visualization: A.L., B.U., A.O.W., S.W.-F., and S.V.; supervision: K.S., I.H., and A.L.; writing the original draft: A.L., S.W.-F., C.M., B.U., and K.S.; critical reviewing of data and draft: K.S., I.H., A.L., A.S., T.Ö., S.W.-F., A.O.W., S.V., and B.U.; and funding: K.S., I.H., A.S., S.W.-F., and C.M.. All authors read and approved the final manuscript.

Funding

The present work has been funded by the Sjöberg Foundation (2021-01145 to K.S.; 2022-01-11:4 to A.S.), Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (965552 to K.S.; 40615 to I.H.; 965065 to A.S.), Swedish Cancer Society (21-1848 to K.S.; 21-1684 to I.H.; 22-2080 to A.S.), BioCARE—a strategic research area at Lund University (I.H. and S.W.-F.), Mrs Berta Kamprad's Cancer Foundation (FBKS-2019-28, I.H.), Cancer and Allergy Foundation (10381, I.H.), Region Västra Götaland (A.S.), Sweden's Innovation Agency (2020-04141, A.S.), Swedish Research Council (2021-01008, A.S.), Roche in collaboration with the Swedish Society of Gynecological Oncology (S.W.-F.), Assar Gabrielsson Foundation (FB19-86, C.M.), and the Lena Wäpplings Foundation (C.M.).

Conflict of interest

A.S. declares stock ownership and is also a board member in Tulebovaasta, SiMSen Diagnostics, and Iscaff Pharma. A.S. has

also received travel support from EMBL, Precision Medicine Forum, SLAS, and bioMCC. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL et al.; ICGC PedBrain. Signatures of mutational processes in human cancer. *Nature* 2013;**500**:415–421.
- Anglesio MS, Bashashati A, Wang YK, Senz J, Ha G, Yang W, Aniba MR, Prentice LM, Farahani H, Li Chang H et al. Multifocal endometriotic lesions associated with cancer are clonal and carry a high mutation burden. *J Pathol* 2015;**236**:201–209.
- Anglesio MS, Papadopoulos N, Ayhan A, Nazeran TM, Noe M, Horlings HM, Lum A, Jones S, Senz J, Seckin T et al. Cancer-associated mutations in endometriosis without cancer. *N Engl J Med* 2017;**376**:1835–1848.
- Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet* 2001;**357**:539–545.
- Bulun SE, Wan Y, Matei D. Epithelial mutations in endometriosis: link to ovarian cancer. *Endocrinology* 2019;**160**:626–638.
- Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics* 2017;**33**:2381–2383.
- Colom B, Herms A, Hall MWJ, Dentro SC, King C, Sood RK, Alcolea MP, Piedrafita G, Fernandez-Antoran D, Ong SH et al. Mutant clones in normal epithelium outcompete and eliminate emerging tumours. *Nature* 2021;**598**:510–514.
- Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002;**420**:860–867.
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;**4**:P3.
- Er TK, Su YF, Wu CC, Chen CC, Wang J, Hsieh TH, Herreros-Villanueva M, Chen WT, Chen YT, Liu TC et al. Targeted next-generation sequencing for molecular diagnosis of endometriosis-associated ovarian cancer. *J Mol Med (Berl)* 2016;**94**:835–847.
- Giudice LC. Clinical practice. Endometriosis. *N Engl J Med* 2010;**362**:2389–2398.
- Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012;**481**:306–313.
- Hazini A, Fisher K, Seymour L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. *J Immunother Cancer* 2021;**9**:1–17.
- Heinze K, Nazeran TM, Lee S, Kramer P, Cairns ES, Chiu DS, Leung SCY, Kang EY, Meagher NS, Kennedy CJ et al. Validated biomarker assays confirm that ARID1A loss is confounded with MMR deficiency, CD8(+) TIL infiltration, and provides no independent prognostic value in endometriosis-associated ovarian carcinomas. *J Pathol* 2022;**256**:388–401.
- Hermens M, van Altena AM, Nieboer TE, Schoot BC, van Vliet H, Siebers AG, Bekkers RLM. Incidence of endometrioid and clear-cell ovarian cancer in histological proven endometriosis: the ENOCA population-based cohort study. *Am J Obstet Gynecol* 2020;**223**:107.e1–107.e11.
- Jhunjhunwala S, Hammer C, Delamarre L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat Rev Cancer* 2021;**21**:298–312.
- Joseph S, Mahale SD. Endometriosis knowledgebase: a gene-based resource on endometriosis. *Database (Oxford)* 2019;**2019**:1–9.
- Koboldt DC, Zhang QY, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;**22**:568–576.
- Kuo KT, Mao TL, Jones S, Veras E, Ayhan A, Wang TL, Glas R, Slamon D, Velculescu VE, Kuman RJ et al. Frequent activating mutations of PIK3CA in ovarian clear cell carcinoma. *Am J Pathol* 2009;**174**:1597–1601.
- Lac V, Verhoef L, Aguirre-Hernandez R, Nazeran TM, Tessier-Cloutier B, Praetorius T, Orr NL, Noga H, Lum A, Khattri J et al. Iatrogenic endometriosis harbors somatic cancer-driver mutations. *Hum Reprod* 2018;**34**:69–78.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;**499**:214–218.
- Lee SH, Princz LN, Klugel MF, Habermann B, Pfander B, Biertumpfel C. Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. *Elife* 2015;**4**:1–24.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–2079.
- Li X, Zhang Y, Zhao L, Wang L, Wu Z, Mei Q, Nie J, Li X, Li Y, Fu X et al. Whole-exome sequencing of endometriosis identifies frequent alterations in genes involved in cell adhesion and chromatin-remodeling complexes. *Hum Mol Genet* 2014;**23**:6008–6021.
- Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;**1**:417–425.
- Lu Y, Cuellar-Partida G, Painter JN, Nyholt DR, Morris AP, Fasching PA, Hein A, Burghaus S, Beckmann MW, Lambrechts D et al.; International Endogene Consortium (IEC). Shared genetics underlying epidemiological association between endometriosis and ovarian cancer. *Hum Mol Genet* 2015;**24**:5955–5964.
- Mari-Alexandre J, Carcelen AP, Agababayan C, Moreno-Manuel A, Garcia-Oms J, Calabuig-Farinas S, Gilabert-Estelles J. Interplay between microRNAs and oxidative stress in ovarian conditions with a focus on ovarian cancer and endometriosis. *Int J Mol Sci* 2019;**20**:1–24.
- Matalliotakis IM, Athanassakis I, Goumenou AG, Neonaki MA, Koumantakis EE, Vassiliadis S, Koumantakis EE. The possible anti-inflammatory role of circulating human leukocyte antigen levels in women with endometriosis after treatment with danazol and leuporelin acetate depot. *Mediators Inflamm* 2001;**10**:75–80.
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 2018;**28**:1747–1756.
- Micci F, Haugom L, Abeler VM, Davidson B, Trope CG, Heim S. Genomic profile of ovarian carcinomas. *BMC Cancer* 2014;**14**:315.
- Moore L, Leongamornlert D, Coorens THH, Sanders MA, Ellis P, Dentro SC, Dawson KJ, Butler T, Rahbari R, Mitchell TJ et al. The mutational landscape of normal human endometrial epithelium. *Nature* 2020;**580**:640–646.
- Murakami R, Matsumura N, Brown JB, Higasa K, Tsutsumi T, Kamada M, Abou-Taleb H, Hosoe Y, Kitamura S, Yamaguchi K et al. Exome sequencing landscape analysis in ovarian clear cell carcinoma shed light on key chromosomal regions and mutation gene networks. *Am J Pathol* 2017;**187**:2246–2258.
- Ostrow SL, Simon E, Prinz E, Bick T, Shentzer T, Nagawkar SS, Sabo E, Ben-Izhak O, Hershsberg R, Hershkovitz D. Variation in KRAS driver substitution distributions between tumor types is

- determined by both mutation and natural selection. *Sci Rep* 2016;**6**:21927.
- Pearce CL, Templeman C, Rossing MA, Lee A, Near AM, Webb PM, Nagle CM, Doherty JA, Cushing-Haugen KL, Wicklund KG et al.; Ovarian Cancer Association Consortium. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol* 2012;**13**:385–394.
- Sato N, Tsunoda H, Nishida M, Morishita Y, Takimoto Y, Kubo T, Noguchi M. Loss of heterozygosity on 10q23.3 and mutation of the tumor suppressor gene PTEN in benign endometrial cyst of the ovary: possible sequence progression from benign endometrial cyst to endometrioid carcinoma and clear cell carcinoma of the ovary. *Cancer Res* 2000;**60**:7052–7056.
- Semino C, Semino A, Pietra G, Mingari MC, Barocci S, Venturini PL, Ragni N, Melioli G. Role of major histocompatibility complex class I expression and natural killer-like T cells in the genetic control of endometriosis. *Fertil Steril* 1995;**64**:909–916.
- Seshan VE. DNACopy: DNA copy number data analysis. R package version 1.70.0, 2022.
- Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJ, van Karnebeek C, Wasserman WW. FLAGS, frequently mutated genes in public exomes. *BMC Med Genomics* 2014;**7**:64.
- Subbarayan K, Massa C, Leisz S, Steven A, Bethmann D, Biehl K, Wickenhauser C, Seliger B. Biglycan as a potential regulator of tumorigenicity and immunogenicity in K-RAS-transformed cells. *Oncoimmunology* 2022;**11**:2069214.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–15550.
- Suda K, Cruz Diaz LA, Yoshihara K, Nakaoka H, Yachida N, Motoyama T, Inoue I, Enomoto T. Clonal lineage from normal endometrium to ovarian clear cell carcinoma through ovarian endometriosis. *Cancer Sci* 2020;**111**:3000–3009.
- Suda K, Nakaoka H, Yoshihara K, Ishiguro T, Tamura R, Mori Y, Yamawaki K, Adachi S, Takahashi T, Kase H et al. Clonal expansion and diversification of cancer-associated mutations in endometriosis and normal endometrium. *Cell Rep* 2018;**24**:1777–1789.
- Szekeres-Bartho J, Polgar B. PIBF: the double edged sword. Pregnancy and tumor. *Am J Reprod Immunol* 2010;**64**:77–86.
- Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandath C, Reimand J, Lawrence MS, Getz G, Bader GD, Ding L et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep* 2013;**3**:2650.
- Velkova A, Carvalho MA, Johnson JO, Tavtigian SV, Monteiro ANA. Identification of filamin A as a BRCA1-interacting protein required for efficient DNA repair. *Cell Cycle* 2010;**9**:1421–1433.
- Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, Senz J, McConechy MK, Anglesio MS, Kalloger SE et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* 2010;**363**:1532–1543.
- Yachida N, Yoshihara K, Suda K, Nakaoka H, Ueda H, Sugino K, Yamaguchi M, Mori Y, Yamawaki K, Tamura R et al. Biological significance of KRAS mutant allele expression in ovarian endometriosis. *Cancer Sci* 2021a;**112**:2020–2032.
- Yachida N, Yoshihara K, Yamaguchi M, Suda K, Tamura R, Enomoto T. How does endometriosis lead to ovarian cancer? The molecular mechanism of endometriosis-associated ovarian cancer development. *Cancers (Basel)* 2021b;**13**:1–14.
- Yamamoto S, Tsuda H, Miyai K, Takano M, Tamai S, Matsubara O. Gene amplification and protein overexpression of MET are common events in ovarian clear-cell adenocarcinoma: their roles in tumor progression and prognostication of the patient. *Mod Pathol* 2011;**24**:1146–1155.
- Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Vigano P. Endometriosis. *Nat Rev Dis Primers* 2018;**4**:9.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Human Reproduction, 2024, 39, 1141–1154
<https://doi.org/10.1093/humrep/deae043>
Original Article

Understanding and addressing unexplained infertility: from diagnosis to treatment

Claudia Massarotti ^{1,2}, Juan J. Fraire-Zamora ³, George Liperis ^{4,5}, Julia Uraji ⁶, Kashish Sharma ⁷, Munevver Serdarogullari ⁸, Omar F. Ammar ^{9,10}, Sofia Makieva ¹¹, Zoya E. Ali¹², Daniela Romualdi ¹³, Edgardo Somigliana ¹⁴, Denny Sakkas ¹⁵, Eline Dancet ¹⁶, and Mina Mincheva ^{17,*}

¹IRCCS Ospedale Policlinico San Martino, Genova, Italy

²DINOGMI Department, University of Genova, Genova, Italy

³EUGIN Group, R&D, Barcelona, Spain

⁴Westmead Fertility Centre, Institute of Reproductive Medicine, University of Sydney, Westmead, NSW, Australia

⁵Embryorigin Fertility Centre, Larnaca, Cyprus

⁶MVZ Kinderwunsch am Seestern, Düsseldorf, Germany

⁷HealthPlus Fertility Center, HealthPlus Network of Specialty Centers, Abu Dhabi, United Arab Emirates

⁸Department of Histology and Embryology, Faculty of Medicine Cyprus International University, Northern Cyprus, Turkey

⁹Ar-Razzi Private Hospital, IVF Centre, Ramadi, Iraq

¹⁰Department of Obstetrics and Gynaecology, College of Medicine, University of Anbar, Ramadi, Iraq

¹¹Kinderwunschzentrum, Klinik für Reproduktions-Endokrinologie, Universitätsspital Zürich, Zurich, Switzerland

¹²Research & Development Department, Hertility Health Limited, London, UK

¹³Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario 'Agostino Gemelli' IRCCS, Rome, Italy

¹⁴Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico and Università degli Studi di Milano, Milan, Italy

¹⁵Boston IVF, Waltham, MA, USA

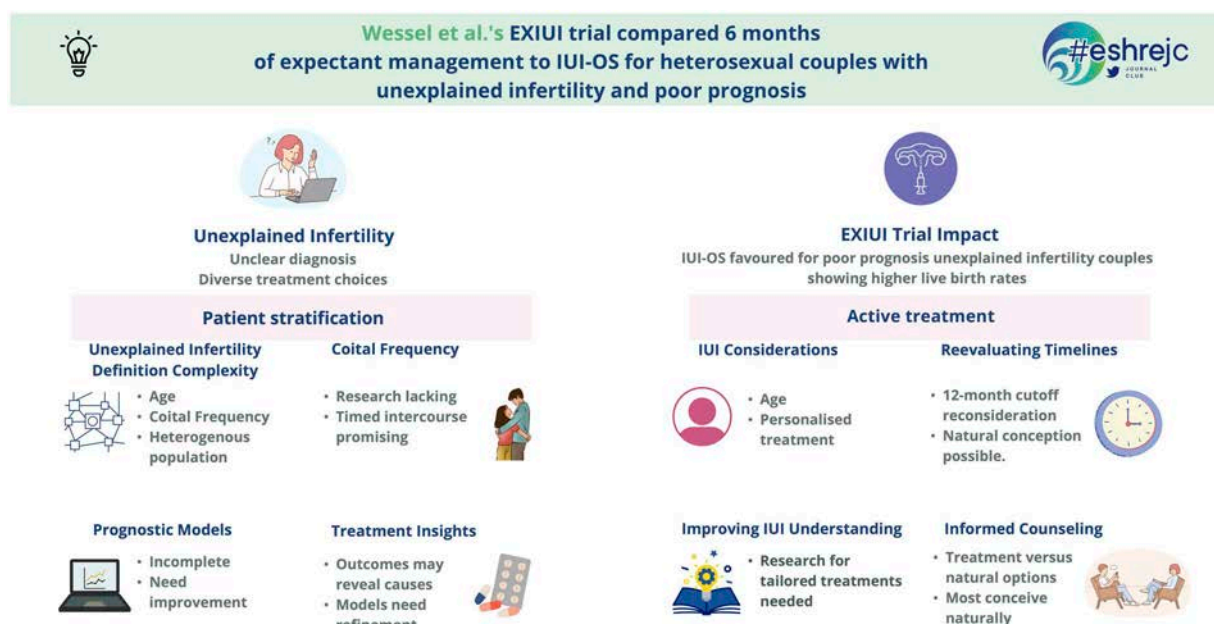
¹⁶Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

¹⁷Independent Researcher, London, UK

*Correspondence address. London, UK. E-mail: mina.mincheva1@gmail.com  <https://orcid.org/0000-0003-2917-2546>

Keywords: ESHRE Journal Club / unexplained infertility / subfertility / randomized controlled trial / expectant management / intra-uterine insemination / IUI / active treatment

GRAPHICAL ABSTRACT



The live edition of the ESHRE Journal Club at the 39th ESHRE annual meeting discussed a study by Wessel et al. (2022a) that compared 6 months of expectant management to IUI-OS in heterosexual couples with unexplained infertility and poor prognosis for natural conception, with main discussion focal points centred around the definition of unexplained infertility, prognostic models, and applicability of expectant management versus active treatment for couples with unexplained infertility. IUI-OS, intrauterine insemination with ovarian stimulation.

Received: January 21, 2024. Revised: February 25, 2024. Editorial decision: February 28, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved. For permissions, please email: journals.permissions@oup.com

Introduction

Unexplained infertility poses a significant challenge to reproductive medicine professionals, especially when they have to decide, in collaboration with the patients, on the optimal course of clinical management to achieve a pregnancy. The International Committee for Monitoring Assisted Reproductive Technologies (ICMART) defines unexplained infertility in heterosexual couples as 'infertility in couples with apparently normal ovarian function, fallopian tubes, uterus, cervix, and pelvis, along with adequate coital frequency, as well as apparently normal testicular function, genitourinary anatomy, and a normal ejaculate'. However, the definition also emphasizes that the 'potential for this diagnosis is dependent upon the methodologies used and/or those methodologies available' (Zegers-Hochschild et al., 2017). The definition of unexplained infertility highlights two crucial points that underlie the uncertainties in defining this highly heterogeneous patient group: unexplained infertility is a diagnosis by exclusion and it is expected that, with advances in diagnostic tests and interventions, some unexplained cases may become explained in the future. Furthermore, the diagnosis includes 'adequate coital frequency' as one of the criteria, which can only be evaluated through sensitive sexual anamnesis.

The recently published ESHRE consensus evidence-based guidelines on unexplained infertility (Romualdi et al., 2023) critically appraised the available evidence base about the diagnosis and treatment of couples with unexplained infertility, topics that are frequently controversial in clinical practice. In particular, when counselling on the best treatment plan, the shared decision between active treatment and expectant management is never an easy one. The guideline development group (GDG) gives a strong recommendation in favour of intrauterine insemination (IUI) with ovarian stimulation (OS) as a first-line treatment instead of IVF. However, the GDG also provides a good practice recommendation to base the decision about starting active treatment on the prognosis of the specific couple (Romualdi et al., 2023).

The live edition of the ESHRE Journal Club at the 39th ESHRE annual meeting discussed a paper by Wessel et al. (2022b) reporting on the results of the EXIUI trial that compared 6 months of expectant management to IUI with OS (IUI-OS) in heterosexual couples with unexplained infertility and a poor prognosis for natural conception as defined by using the prediction model developed by Hunault et al. (2004). Wessel et al. designed a non-inferiority multicentre randomized controlled trial (RCT) to test whether expectant management is not worse than IUI-OS with ongoing pregnancy leading to a live birth as the primary outcome. The authors considered inferiority if the absolute difference in expectant management live birth rate (LBR) would exceed minus 7%, using an expected LBR of 30% after 6 months of IUI-OS as a benchmark. The primary analysis was based on an intention-to-treat (ITT) (i.e. including all couples originally allocated) and a per-protocol (PP) analysis was also performed (i.e. including patients who completed the treatment) since 35 couples discontinued treatment during the study. Couples allocated to expectant management in the ITT analysis had a 40% decrease in live births compared to the IUI-OS group (LBR: 13% vs 33%, respectively) and had an absolute difference of minus 20% (90% CI: -30% to -9%), showing that expectant management was inferior to IUI-OS. The cumulative probabilities of conception leading to live birth over time were calculated together with a hazard ratio to assess differences between groups, showing that couples allocated to expectant management were less likely to experience

live birth at any time point when compared to couples allocated to IUI-OS. Similar results were observed in the PP analysis, confirming the inferiority of expectant management over IUI-OS.

The Wessel et al. study concludes in favour of active treatment with IUI-OS for couples with unexplained infertility and a poor prognosis for natural conception. These results challenge for example the guideline of the National Institute of Health and Care Excellence (NICE) that advises against the use of IUI based on cost-effectiveness, except in same-sex couples, serodiscordant couples, and couples who cannot have intercourse (National Institute for Health and Care Excellence, 2013). The live edition of the ESHRE Journal Club fostered a lively discussion between experts in the field and a large ESHRE audience which touched on all the contentious aspects about the definition of unexplained infertility, prognostic models, and whether expectant management or active treatment is more appropriate for couples with unexplained infertility.

Patient stratification

The study of Wessel et al. (2022) highlighted the two main challenges that complicate the management of patients with unexplained infertility—the definition of unexplained infertility and the reproductive prognosis of this patient group. It is estimated that 30–50% of couples assessed for infertility have normal ovulatory function, seminal values, and fallopian tubal patency (Esteves et al., 2015). Since several factors—such as age, ovarian reserve, gamete quality, uterine-cervical-tubal factors, genetics, immunology, endometriosis, sexual difficulties, and failed fertilization—can contribute to infertility, it is acknowledged that unexplained infertility encompasses a highly heterologous patient population (Buckett and Sierra, 2019).

With the aim of assisting practitioners to provide the best care for couples with unexplained infertility, the ESHRE guidelines on unexplained infertility have incorporated several novel aspects in the existing ICMART definition of unexplained infertility (Romualdi et al., 2023). Firstly, the guidelines introduce an age limit of up to 40 years for the woman's age based on mathematical modelling of decline in fecundity with age (Somigliana et al., 2016). A defined age cut-off would allow to disaggregate unexplained infertility from age-related infertility, thus avoiding inappropriate interventions in women of advanced reproductive age, who are misdiagnosed as having unexplained infertility. Indeed, the false positive rate of diagnosing a woman with unexplained infertility is around 80% at 40 years of age (Somigliana et al., 2016). The second novelty of the ESHRE guidelines on unexplained infertility is the requirement of 'adequate coital frequency' for the infertility to be considered unexplained (Romualdi et al., 2023). Defining whether a reported coital frequency is 'adequate' remains challenging. Coital frequency is likely to affect reproductive outcomes but research on what coital frequency can be expected among infertile couples is scarce. The frequency sexual intercourse among the general population varies depending on a plethora of socio-cultural, cyclic, and lifestyle factors and is related to sexual functioning, as well as physical and mental health. Limited evidence suggests that three in four couples trying to conceive in the USA or the UK have coitus at least four times per month (Gaskins et al., 2018; Wellings et al., 2019). Although not reported in the EXIUI article, an ESHRE poster published by the same group showed a median coital frequency per cycle of only four among the expectant management group, with 75% of the couples allocated to expectant

management having intercourse at least once per cycle (Wessel *et al.*, 2022a). A similar trial in New Zealand reported a mean sexual intercourse frequency of about three during the fertile window of the menstrual cycle but did not report interquartile ranges per month, cycle, or fertile window (Farquhar *et al.*, 2018). Furthermore, data from the same trial show that a small number of women had no sexual activity during the fertile window of an expectant management cycle. Similarly, there was no intercourse during the fertile window in 18% of the hormonally monitored cycles of Japanese women trying to conceive (Konishi *et al.*, 2020). These data stress the need for sensitive communication given the topic, the threat of social desirability response bias (Konishi *et al.*, 2020) and the documented likelihood of couples not having intercourse during the fertile window of an expectant management cycle. The importance of assessing the coital frequency is reflected in the ESHRE guideline on unexplained infertility which recommend taking a sexual history from both partners (The Guideline Group on Unexplained Infertility, 2023).

The timing of coitus has been related to the probability of conception among healthy women (Wilcox *et al.*, 1995). A recent Cochrane review indicates that timed intercourse using urine ovulation tests potentially improves clinical outcomes in women under 40 trying to conceive for <1 year compared to intercourse without ovulation prediction (Gibbons *et al.*, 2023). It remains uncertain whether the duty to have sex implied by timing intercourse results in distress or sexual dysfunctions among couples who have not attempted conception for long enough to be considered infertile (Gibbons *et al.*, 2023). An RCT among healthy individuals trying to conceive showed that being educated to time the intercourse or to have sex every other day did not affect the increase in distress and sexual dysfunctions over time, which was also seen in the control group (Martins *et al.*, 2022). Interview studies showed that timing intercourse aggravated the negative impact of the 'duty' of attempting natural conception on the sexual functioning and distress of infertile couples (Jafarzadeh-Kenarsari *et al.*, 2015; Zhuoran *et al.*, 2018; Dreischor *et al.*, 2022b). Further research on the efficacy of timed intercourse in infertile couples is warranted given the high variability in cycle length among some infertile women as well as the negative impact of infertility on distress and on sexual functioning (Ali *et al.* (2023)).

Effective treatment solutions for unexplained infertility are not easy to identify due to the inability to diagnose an underlying pathology. Predictive models for the chances of natural conception have the potential to become a routine tool in assisting the effective management of unexplained infertility (McLernon *et al.*, 2014). Though several prognostic models have been developed over the years, they have not been widely adopted by the fertility providers due to limitations pertinent to each of them. The Hunault model, which can be applied only once after fertility work-up, when the couple first presents with infertility, lacks the ability to provide revised prediction following a timeframe of expectant management of 12 months (Hunault *et al.*, 2004). In contrast, repeated estimates can be provided by a different model that can be used both when planning to undergo treatment and planning to conceive naturally (McLernon *et al.*, 2014, 2019). However, the McLernon *et al.* model also has its limitations as the model has been based on a dataset from a single centre and important prognostic factors such as ethnicity and BMI have not been included. Several other models that have been developed to predict chances of conception with different treatment types as well as natural conception (Shingshetty *et al.*, 2022) share similar drawbacks related to dataset size and validation on limited geographical populations of patients.

It appears that the most beneficial model for clinical practice would be one that can predict clinical outcomes with or without treatment, since it will allow to calculate the additional chance of achieving a pregnancy following treatment over that of expectant management only (McLernon *et al.*, 2014). The greatest barrier in refining the predictive parameters for prognostic models stems from the lack of diagnostic tests to make the unknown known. Therefore, in current practice the same treatment applied to patients with unexplained infertility may have different outcomes for different sub-groups of unexplained infertility couples who would be re-classified to explained infertility if appropriate diagnostic tools were at hand. Nonetheless, stepping up to an IVF/ICSI treatment itself can be a form of empirical diagnosis, where suboptimal outcomes of the treatment—such as low number of oocytes retrieved, poor fertilization, etc.—can shed light on the underlying causes of the couple's infertility. Apart from guiding the decision-making process of physicians for unexplained infertility management, providing prognoses to patients is in line with patient preferences and has been shown to temper the unrealistic high expectations of both partners in couples with a less than average IVF prognosis (Devroe *et al.*, 2022).

Active treatment: the value of IUI in clinical practice

With continuous efforts to improve definitions of unexplained infertility, diagnostic tests, and predictive models to assist clinical decision-making and patient counselling, the next step is to consider whether active treatment is required. The open-label RCT by Farquhar *et al.* (2018) showed higher LBRs in the IUI group versus expectant management (31% vs 9%), but it should be noted that 8 out of the 31 pregnancies in the IUI group resulted from natural conception before or between IUI cycles. Coherently, the paper by Wessel *et al.* (2022b) showed improved outcomes with IUI-OS over expectant management in couples with poor prognosis (clinical pregnancy: 43% vs 19%, respectively). These findings should not mean that IUI can be applied on a blanket level, which is also highlighted by recommendations issued by professional body guidelines (National Institute for Health and Care Excellence, 2013; Romualdi *et al.*, 2023). Several factors should be considered by clinicians before counselling individual patients in favour of active treatment with IUI.

First, as discussed, maternal age in particular affects fertility, and thus, women above a certain age threshold should no longer be classified as having unexplained infertility. While the ESHRE guidelines propose the age of 40 as a limit, the authors of the discussed study found (though underpowered through low sample size) that even above age 38 years, IUI no longer showed a positive effect on LBR, implying that age-related fertility issues may be masking the effectiveness of IUI (Wessel *et al.*, 2022b). While keeping in mind that applying cut-offs created through population-based data to individuals can be problematic (Tonelli, 1998), considerations for the use of IUI should be applied on a case-by-case basis.

A second factor that should be taken into account during decision-making for active management is the coital frequency identified by the sensitive sexual anamnesis. If couples are unable to have coitus the infertility should no longer be considered unexplained (The Guideline Group on Unexplained Infertility, 2023) and unstimulated IUI or self-insemination with a syringe correctly timed around ovulation can overcome the need for vaginal intercourse to conceive (National Institute for Health and Care Excellence, 2013). These couples could additionally be

referred to a sex therapist addressing their sexual health. Other couples who are able to have intercourse but are held back to do so by specific sexual dysfunctions (e.g. dyspareunia, erectile dysfunction) could also benefit from medicinal and/or sex therapeutic interventions in addition to fertility treatment. Couples with a suboptimal intercourse frequency who do not suffer from a specific sexual dysfunction may benefit from a self-administered sex counselling program while opting for different fertility treatment strategies. A recent RCT showed that the 'Pleasure&pregnancy' program improved certain dimensions of sexual functioning and limited the decrease in coital frequency over time among couples going through expectant management (Dreischor et al., 2022a).

A third important aspect is that unexplained infertility should be considered as a matter of time for many patients. Understandably, patients trying to conceive can have personal, social and/or medical time pressure, and would like to have results sooner rather than later. While the cut-off for infertility is defined as failure to conceive after 12 months of unprotected intercourse (Zegers-Hochschild et al., 2017), many couples with unexplained infertility may be able to conceive naturally after this period, as they are not necessarily sterile and their chance of conception is not static (Hull et al., 1985). Data suggests that true unresolved infertility or sterility occurs only in a minority (3–5%) of all couples (Greenhall and Vessey, 1990) and as such many couples are able to conceive without a medical intervention but it may simply take longer to do so, reflecting the degree of subfertility (McLernon et al., 2014). In fact, a recent 15-year follow-up study showed that over 90% of couples will eventually conceive naturally (Vaughan et al., 2022). Therefore, patients should be counselled about the benefits and costs of active treatment versus a longer time of trying to conceive naturally so that they can make informed decision on when to pursue IUI or other fertility treatments.

Overall, it is important to understand the mechanisms by which IUI should be better than expectant management for couples with unexplained infertility, so that professionals can counsel patients appropriately and that professionals do not opt for an IUI-for-all approach. For example, IUI cycles have a much more controlled nature as ovulation is tracked or even induced, while untracked natural cycles may be unknowingly anovulatory. Another difference between IUI and natural conception is that for IUI the sperm is washed and prepared in the laboratory prior to insemination. Sperm preparation is a fundamental technique in the provision of assisted reproduction technology with the purpose to yield high percentage of morphologically normal and motile cells, with reduced DNA damage, free of debris, non-germ cells, and dead spermatozoa (World Health Organization, 2021). There are three widely used sperm preparation techniques which are based on different principles to select sperm with good quality (wash, swim-up, density gradient). However, available data do not provide certainty whether there is a difference in pregnancy outcomes between the three sperm preparation techniques for infertile couples undergoing IUI (Boomsma et al., 2019). With more research, it may be possible to improve information for clinicians and patients with unexplained infertility alike on when it is time for active management with IUI or other treatments when no other fertility issues are diagnosed.

Conclusion

While indicating a trigger for initiation of a fertility diagnostic work-up, the current 1-year definition of infertility does not correctly reflect the biology of reproduction and introduces a misleading concept of infertility beyond 12 months. Consequently,

failure of conception within this period is interpreted as evidence of sterility by many couples, who then seek immediate treatment. As diagnosis of unexplained infertility is commonly made when clinical investigations fail to identify any obvious barriers to conception, this patient group naturally encompasses a population with highly heterogeneous conditions, some of them being female age-related infertility and inadequate coital frequency. Recommendations on clinical management of unexplained infertility are increasingly acknowledging these aspects (Romualdi et al., 2023) which can be pivotal for effective treatment. Therefore, it seems essential to investigate 'simple' factors like lifestyle, together with doubling research efforts in developing powerful diagnostic tools to identify the unknowns in the field of unexplained infertility aetiology (The Guideline Group on Unexplained Infertility, 2023). Coital frequency should always be assessed by a sensitive sexual anamnesis and managing the sexual health of couples should always be considered irrespective of the treatment decision. Moreover, it is important to take patient preferences into account when deciding on treatment options. Prognostic models can help the decision-making on when to treat and what treatment plan is best in couples with unexplained infertility. Nonetheless, further research needs to be undertaken to identify treatment predictive factors to evolve these models into accurate and reliable tools in routine practice.

Data availability

No datasets were generated or analysed in the current manuscript.

Acknowledgements

The authors thank the Editor-in-Chief of Human Reproduction, Prof. Nils Lambalk, and all the participants of the ESHRE live journal club at the ESHRE 2023 Annual Meeting.

Authors' roles

C.M., J.J.F.-Z., G.L., J.U., K.S., M.S., O.F.A., S.M., Z.E.A., and M.M. conceptualized, organized, and moderated the discussion; C.M., G.L., J.J.F.-Z., and M.M. led the discussion; D.R., E.S., D.S., and E.D. contributed to the discussion as experts; O.F.A. prepared the graphical abstract; C.M., J.J.F.-Z., G.L., J.U., and M.M. drafted the manuscript. All authors provided critical revision to the manuscript/graphical abstract/and approved the final version.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

M.M. is a Guideline Development Group member for the ESHRE guidelines on unexplained infertility. D.R. is the lead of the Guideline Development Group for the ESHRE guidelines on unexplained infertility. All the other authors have no conflicts of interest to declare.

References

Ali ZE, Liperis G, Mincheva M, Serdarogullari M, Uraji J, Ammar OF, Fraire-Zamora JJ. From stress to success: exploring methods

- addressing psychosocial stress in couples trying to conceive. *Hum Reprod* 2023;**38**:2277–2279.
- Boomsma CM, Cohlen BJ, Farquhar C. Semen preparation techniques for intrauterine insemination. *Cochrane Database Syst Rev* 2019;**10**:CD004507. <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004507.pub4/full?highlight=Abstract=iui>
- Buckett W, Sierra S. The management of unexplained infertility: an evidence-based guideline from the Canadian Fertility and Andrology Society. *Reprod Biomed Online* 2019;**39**:633–640.
- Devroe J, Peeraer K, D'Hooghe TM, Boivin J, Laenen A, Vriens J, Dancet EAF. Great expectations of IVF patients: the role of gender, dispositional optimism and shared IVF prognoses. *Hum Reprod* 2022;**37**:997–1006.
- Dreischor F, Dancet E, Goddijn M, Lambalk C, Van Wely M, Laan E, Custers I. O-092 The 'Pleasure&Pregnancy' eHealth programme and its effect on the sexual functioning and ongoing pregnancy rates of couples with unexplained subfertility: a randomized controlled trial. *Hum Reprod* 2022a;**37**:deac104.106.
- Dreischor F, Laan ETM, Peeters F, Peeraer K, Lambalk CB, Goddijn M, Custers IM, Dancet EAF. The needs of subfertile couples continuing to attempt natural conception: in-depth interviews. *Hum Reprod Open* 2022b;**2022**:hoac037.
- Esteves SC, Schattman GL, Agarwal A. Definitions and relevance of unexplained infertility in reproductive medicine. In: Schattman GL, Esteves SC, Agarwal A (eds). *Unexplained Infertility: Pathophysiology, Evaluation and Treatment*. New York, NY: Springer, 2015,3–5. https://doi.org/10.1007/978-1-4939-2140-9_1
- Farquhar CM, Liu E, Armstrong S, Arroll N, Lensen S, Brown J. Intrauterine insemination with ovarian stimulation versus expectant management for unexplained infertility (TUI): a pragmatic, open-label, randomised, controlled, two-centre trial. *Lancet* 2018;**391**:441–450.
- Gaskins AJ, Sundaram R, Buck Louis GM, Chavarro JE. Predictors of sexual intercourse frequency among couples trying to conceive. *J Sex Med* 2018;**15**:519–528.
- Gibbons T, Reavey J, Georgiou EX, Becker CM; Cochrane Gynaecology and Fertility Group. Timed intercourse for couples trying to conceive. *Cochrane Database Syst Rev* 2023;**9**:CD011345.
- Greenhall E, Vessey M. The prevalence of subfertility: a review of the current confusion and a report of two new studies Supported by grants from the Oxford Regional Health Authority, Oxford, United Kingdom; by ICI, Macclesfield, United Kingdom; and by Sandoz, Camberley, United Kingdom. *Fertil Steril* 1990;**54**:978–983.
- Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, Desai KM. Population study of causes, treatment, and outcome of infertility. *Br Med J (Clin Res Ed)* 1985;**291**:1693–1697.
- Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JLH, Velde ERT. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Hum Reprod* 2004;**19**:2019–2026.
- Jafarzadeh-Kenarsari F, Ghahiri A, Zargham-Boroujeni A, Habibi M. Exploration of the counseling needs of infertile couples: a qualitative study. *Iran J Nurs Midwifery Res* 2015;**20**:552–559.
- Konishi S, Saotome TT, Shimizu K, Oba MS, O'Connor KA. Coital frequency and the probability of pregnancy in couples trying to conceive their first child: a prospective cohort study in Japan. *Int J Environ Res Public Health* 2020;**17**:4985.
- Martins MV, Fernandes J, Pedro J, Barros A, Xavier P, Schmidt L, Costa ME. Effects of trying to conceive using an every-other-day strategy versus fertile window monitoring on stress: a 12-month randomized controlled trial. *Hum Reprod* 2022;**37**:2845–2855.
- McLernon DJ, Lee AJ, Maheshwari A, van Eekelen R, van Geloven N, Putter H, Eijkemans MJ, van der Steeg JW, van der Veen F, Steyerberg EW et al. Predicting the chances of having a baby with or without treatment at different time points in couples with unexplained subfertility. *Hum Reprod* 2019;**34**:1126–1138.
- McLernon DJ, Te Velde ER, Steyerberg EW, Mol BWJ, Bhattacharya S. Clinical prediction models to inform individualized decision-making in subfertile couples: a stratified medicine approach. *Hum Reprod* 2014;**29**:1851–1858.
- National Institute for Health and Care Excellence. *Fertility Problems: Assessment and Treatment*. NICE, 2013. <https://www.nice.org.uk/guidance/CG156> (20 January 2024, date last accessed).
- Romualdi D, Ata B, Bhattacharya S, Bosch E, Costello M, Gersak K, Homburg R, Mincheva M, Norman RJ, Piltonen T et al.; Guideline Group on Unexplained Infertility. Evidence-based guideline: unexplained infertility. *Hum Reprod* 2023;**38**:1881–1890.
- Shingshetty L, Maheshwari A, McLernon DJ, Bhattacharya S. Should we adopt a prognosis-based approach to unexplained infertility? *Hum Reprod Open* 2022;**2022**:hoac046.
- Somigliana E, Paffoni A, Busnelli A, Filippi F, Pagliardini L, Vigano P, Vercellini P. Age-related infertility and unexplained infertility: an intricate clinical dilemma. *Hum Reprod* 2016;**31**:1390–1396.
- The Guideline Group on Unexplained Infertility. *Evidence-Based Guideline: Unexplained Infertility*. ESHRE, 2023. <https://www.eshre.eu/guideline/UI> (20 January 2024, date last accessed).
- Tonelli MR. The philosophical limits of evidence-based medicine. *Acad Med* 1998;**73**:1234–1240.
- Vaughan DA, Goldman MB, Koniares KG, Nesbit CB, Toth TL, Fung JL, Reindollar RH. Long-term reproductive outcomes in patients with unexplained infertility: follow-up of the Fast Track and Standard Treatment Trial participants. *Fertil Steril* 2022;**117**:193–201.
- Wellings K, Palmer MJ, Machiyama K, Slaymaker E. Changes in, and factors associated with, frequency of sex in Britain: evidence from three National Surveys of Sexual Attitudes and Lifestyles (Natsal). *BMJ* 2019;**365**:l1525.
- Wessel J, Mochtar M, Custers I, Van Wely M, Van Eekelen R, Mol F. P-639 A randomised controlled trial comparing intrauterine-insemination or expectant management in couples with unexplained subfertility and a poor prognosis for natural conception: impact on coital frequency. *Hum Reprod* 2022a;**37**:deac107.588.
- Wessel JA, Mochtar MH, Besselink DE, Betjes H, De Bruin JP, Cantineau AEP, Groenewoud ER, Hooker AB, Lambalk CB, Kwee J et al. Expectant management versus IUI in unexplained subfertility and a poor pregnancy prognosis (EXIUI study): a randomized controlled trial. *Hum Reprod* 2022b;**37**:2808–2816.
- Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation – effects on the probability of conception, survival of the pregnancy, and sex of the baby. *N Engl J Med* 1995;**333**:1517–1521.
- World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 6th edn. Geneva: World Health Organization, 2021. <https://iris.who.int/bitstream/handle/10665/343208/9789240030787-eng.pdf?sequence=1> (20 January 2024, date last accessed).
- Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, Rienzi L, Sunde A, Schmidt L, Cooke ID et al. The International Glossary on Infertility and Fertility Care, 2017. *Fertil Steril* 2017;**108**:393–406.
- Zhuoran W, Wanpeng L, Tao P, Coates R. Qualitative research on infertile Chinese couples' understanding of sexuality. *Fam Pract* 2018;**35**:88–92.

Correction to: Identification of two hidden clinical subgroups among men with idiopathic cryptozoospermia

This is a correction to: Lena Charlotte Schülke, Joachim Wistuba, Verena Nordhoff, Hermann M Behre, Jann-Frederik Cremers, Sabine Kliesch, Sara Di Persio, Nina Neuhaus, Identification of two hidden clinical subgroups among men with idiopathic cryptozoospermia, Human Reproduction, 2024, deae013, <https://doi.org/10.1093/humrep/deae013>

In the originally published version of this article, the title was incorrectly given as “Andrology of two hidden clinical subgroups

among men with idiopathic cryptozoospermia”, due to an error that occurred during copyediting.

The Publisher apologizes for this.

The title has been corrected to read “Identification of two hidden clinical subgroups among men with idiopathic cryptozoospermia”.