UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Evidence for treatment strategies in PCOS

Bordewijk, E.M.

Publication date
2022
Document Version
Final published version

Link to publication

Citation for published version (APA):
Bordewijk, E. M. (2022). *Evidence for treatment strategies in PCOS*.

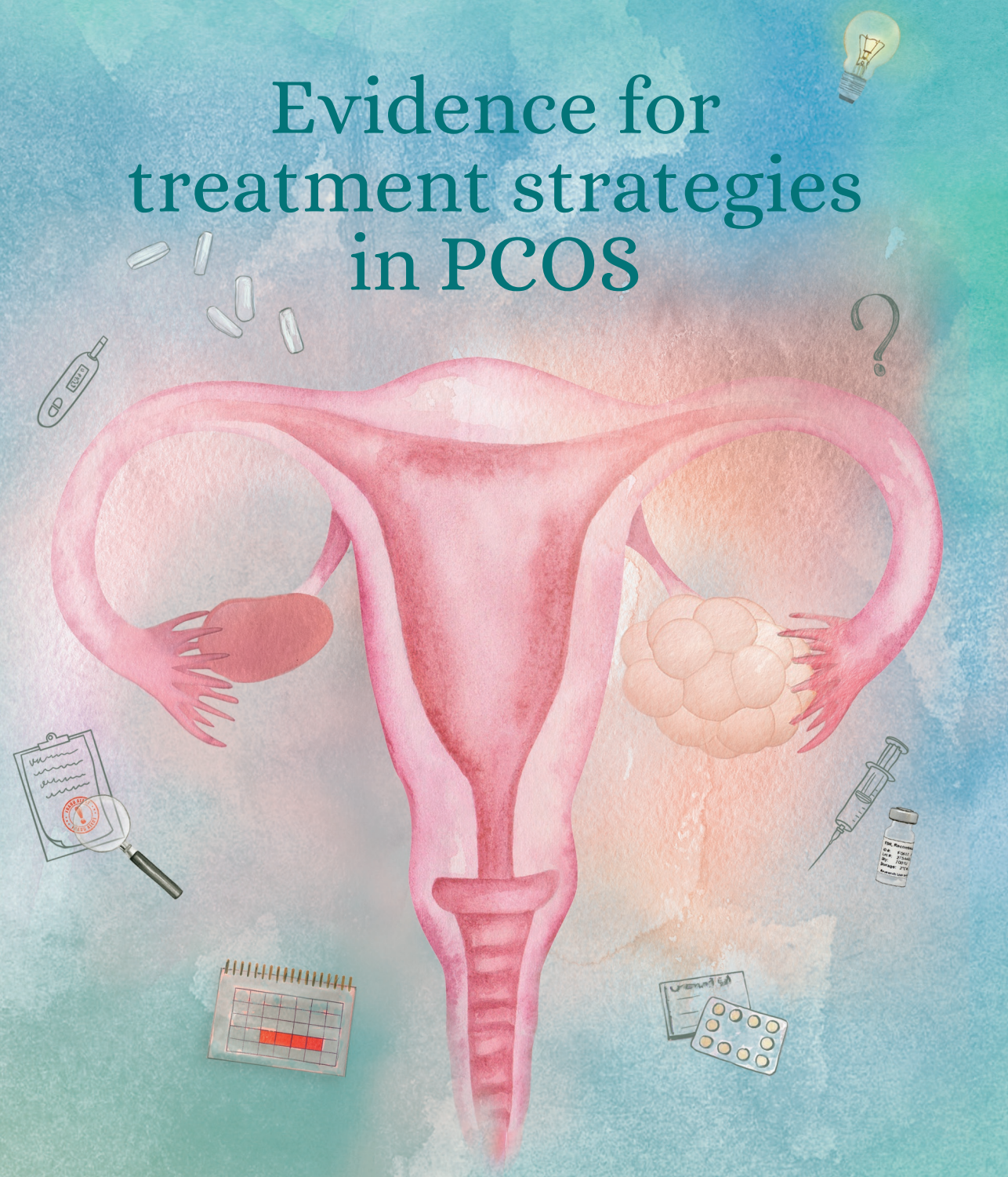
General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:11 Nov 2022

Evidence for treatment strategies in PCOS



Esmée Bordewijk

Evidence for treatment strategies in PCOS

Esmée Bordewijk,
Amsterdam, the Netherlands, 2022

Evidence for treatment strategies in PCOS
Esmée Bordewijk, Amsterdam, the Netherlands, 2022

Part of this research was supported by a grant of ZonMw (the Netherlands Organisation for Health Research and Development)

Publication of this thesis was financially supported by: Stichting Gynaecologische Endocrinologie en Kunstmatige Humane Voortplanting and Amsterdam Reproduction and Development (AR&D) Research Institute.

ISBN: 978-94-6469-071-2
Cover: Marga van den Heuvel
Lay-out: Bregje Jaspers | ProefschriftOntwerp.nl
Printing: ProefschriftMaken | www.proefschriftmaken.nl

Copyright – Esmée Marilijn Bordewijk, 2022
All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without prior permission of the author or, when appropriate, of the publisher of the publication or illustration material.

Evidence for treatment strategies in PCOS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 25 november 2022, te 10.00 uur

door Esmée Marilijn Bordewijk
geboren te Rotterdam

PROMOTIECOMMISSIE

Promotor:	prof. dr. M. Goddijn	AMC-UvA
Copromotor:	dr. M. van Wely	AMC-UvA
Overige leden:	prof. dr. L.M. Bouter	Vrije Universiteit Amsterdam
	prof. dr. A. Hoek	Rijksuniversiteit Groningen
	prof. dr. J.A.M. van der Post	AMC-UvA
	dr. M.H. Mochtar	AMC-UvA
	prof. dr. M.G.W. Dijkgraaf	AMC-UvA
	prof. dr. R.C. Painter	Vrije Universiteit Amsterdam Faculteit der Geneeskunde

CONTENTS

CHAPTER 1	General introduction and outline of this thesis	9
CHAPTER 2	Gonadotrophins versus clomifene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomifene failure (M-OVIN): a randomised, two-by-two factorial trial	23
CHAPTER 3	Gonadotrophins versus clomiphene citrate with or without IUI in women with normogonadotropic anovulation and clomiphene failure: a cost-effectiveness analysis.	43
CHAPTER 4	Gonadotrophins or clomiphene citrate in women with normogonadotropic anovulation and CC failure: does the endometrium matter?	65
CHAPTER 5	Long-term outcomes of using gonadotrophins versus clomiphene citracte with or without intrauterine insemination in women with normogonadotropic anovulation and clomiphene failure: follow-up study of a factorial randomised clinical trial	81
CHAPTER 6	First-line ovulation induction for polycystic ovary syndrome: an individual participant data meta-analysis.	101
CHAPTER 7	To share or not to share data: how valid are trials evaluating first-line ovulation induction for polycystic ovary syndrome?	137
CHAPTER 8	Methods to assess research misconduct in health-related research: A scoping review.	171
CHAPTER 9	Summary, implications for practice and future research	197
CHAPTER 10	Nederlandse samenvatting, implicaties voor de klinische praktijk en toekomstig onderzoek	211
APPENDICES	List of co-authors and affiliations	229
	Author contributions	233
	PhD Portfolio	235
	Dankwoord	238
	About the author	245

Chapter 1

General introduction and outline of this thesis

Polycystic ovarian syndrome (PCOS) affects multiple aspects of a woman's overall health. PCOS results often in fertility problems and is associated with an increased risk for anxiety disorders, depression, insulin resistance, type 2 diabetes mellitus, coronary heart disease, atherogenic dyslipidemia, and cerebrovascular morbidity.[1-6] If pregnant, these women have substantially increased risk for developing gestational diabetes, pre-eclampsia, fetal macrosomia, small-for-gestational age infants, and perinatal mortality.[7-9] Hospital admissions for women with PCOS are twice as high as for the general population.[10] All these PCOS related problems cause a lot of uncertainty and emotional distress. This is why an increasing group of women are asking more awareness for the still not fully understand syndrome.

Ovulation disorders

Reproduction is one of the key elements of life and failing to achieve the creation of offspring may lead to lifelong mental and physical health problems.[11] Nevertheless, approximately one out of seven couples with a child-wish are confronted with fertility problems. Infertility is defined by the failure to conceive after at least 12 months of regular unprotected sexual intercourse.[12,13]

A quarter of infertile cases are due to ovulation disorders. Ovulation disorders, presenting as menstrual disturbance, causes a woman's ovulation to be infrequent or absent which is defined as oligo- or anovulation. The classification of the World Health Organization (WHO) is used to distinguishes three classes of ovulation disorders based upon gonadotrophin and estrogen levels.[13]

This thesis focused on the WHO type II ovulation disorder, which affects around 85% of women with ovulation disorders making it the most common endocrine disorder of reproductive age women. WHO type II ovulation disorder results from absent or irregular ovulation due to hypothalamic-pituitary-ovarian dysfunction and is also known as normogonadotropic anovulation.[13-15] The majority of women with WHO type II ovulation disorder have PCOS. PCOS is a heterogeneous syndrome comprising of at least two of the following clinical characteristics according to the Rotterdam diagnostic criteria: oligo-/anovulation, clinical and/or biochemical hyperandrogenism or polycystic ovaries morphology based on ultrasound assessment.[16]

Treatments

Anovulation is one of the key features that women with PCOS are confronted with. First-line treatments for ovulation induction in these women are clomiphene citrate (CC) or letrozole, with or without metformin.[1] For a long period of time, CC has been used as the first-line ovulation induction agent.[1,13,17-19] CC is an effective primary treatment option in therapy-naive women suffering from normogonadotropic anovulation and PCOS.[20-22] In the

last decade several randomized controlled trials have found letrozole to be more effective than CC, on average leading to 10% more live births within 6 months.[22,23] Consequently, guidelines nowadays recommend letrozole as first-line treatment[1]. However, as letrozole is off-label medication, CC is still being used world-wide.[22,23] Six months of treatment with CC leads to conception in about 50% of these women.[21,24] Women not conceiving after six ovulatory cycles are defined as having CC-failure.[25]

The National Institute for Health and Care Excellence (NICE) guideline recommends not to extend treatment with CC for more than six cycles, but this recommendation is not underpinned by any evidence.[13] Second-line treatments are ovulation induction with gonadotrophins and surgery procedures of which laparoscopic ovarian drilling is being performed most commonly. In vitro fertilization (IVF) therapy is considered a third-line therapy.[26]

Randomised controlled trials and meta-analyses

Randomized controlled trials (RCTs) are scientific investigations of which the findings are considered to be top-level of evidence for clinical practice and have the least potential for bias when evaluating the effects of interventions. Systematic reviews and meta-analyses combine the results of multiple RCTs and have an even higher reliability when executed adequately. However, systematic reviews and meta-analyses are limited by design. Firstly, they do not allow combining evidence of multiple comparisons and secondly, they have limited power to evaluate effectiveness and safety of interventions within subgroups. Furthermore, such subgroup analyses may be at risk of ecological bias due to within-study interactions and are sometimes impossible to perform due to heterogeneous reporting of subgroup data in the primary trials.[27]

These deficiencies in meta-analyses can potentially be overcome by using individual participant data (IPD). IPD uses the raw data of the individual trials, such that interactions can be studied and inferences of subgroups can be made. As it allows derivation of the information beyond the primary publication, standardization of inclusion criteria, outcomes and analyses across trials and investigations of subgroup effects and time-to-event outcomes, IPD meta-analysis has been described as the gold standard in evidence synthesis.[27,28] However, IPD meta-analyses are rare compared to the other meta-analyses. To compare first-line ovulation induction interventions in women with PCOS many head-to-head meta-analyses and a few network meta-analysis have been performed.[21,22] When starting the research described in this thesis, no IPD meta-analysis on ovulation induction had been performed yet. As women with PCOS represent a heterogeneous population according to the diagnostic criteria, it is important to identify which individuals benefit mostly from a particular treatment in order to enable clinicians providing personalized care.[29]

Availability of IPD

A common issue that troubles IPD meta-analyses is the lack of availability of IPD from individual trials,[30] and this may limit the validity and precision of the results of evidence synthesis. The concerns expressed and reasons given by teams of investigators that refuse or are unable to share IPD include the heterogeneity and complexity of the contributing RCTs, administrative difficulties, lack of agreement on the purpose of sharing, the need for confidentiality and secure storage of data, data quality issues, ethical or ownership restrictions, failure to retain trial data and other personal considerations.[31,32] Also the willingness to share IPD may depend on the study's type of funding, sample size, risk of bias and the magnitude of the estimated treatment effect.[33] The attitude of researchers' institutions towards data sharing is also important.

Little is known about the comparability between shared and non-shared IPD RCTs with regard to quality and integrity. RCTs with IPD sharing are usually performed better compared to non-shared RCTs on a quantitative 'risk of bias' assessment.[34] As PCOS is very common and the prospect for patients are quite good this is a field worthwhile to look into quality and integrity world-wide.

Research misconduct

While updating a Cochrane review in the field of PCOS we detected several cases of suspected fraud.[35,36] Most of these studies included women with PCOS or unexplained infertility and were cited by many articles including meta-analyses, resulting in false data in meta-analyses and guidelines on PCOS.[1] This made us start focusing on research misconduct, as research integrity is of utmost importance for science, not only in fertility research, but in all fields.

Science relies on the integrity of findings reported. Remarkably, it was found that approximately 2% of scientists admitted to having fabricated, falsified, or modified data or results at least once, and on average over 14% of scientists observed these behaviors among their colleagues.[37] Research misconduct may result in a waste of financial and human resources and, more importantly, it might pose an immediate risk to human health.[37]

Research misconduct is defined as fabrication, falsification, or plagiarism. Fabrication is making up data, results, or recordings, and reporting them. Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not represented accurately in the research record. Plagiarism is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit. Research misconduct does not include honest error or differences of opinion.[38]

According to the Committee on Publication Ethics (COPE) code of conduct,editors have the obligation to take action in case of suspected misconduct.[39,40] After receiving a post-publication critique, the first step for editors is to follow their journals transparency policy for considering critiques. If such a policy is not available, journals should create this it. Often, the second step is that the editor or peer reviewer of the journal judges whether the claim needs more investigation, and when data support the claims made, original authors are invited to provide a response within a deadline.[41]

Only a third of top-ranking peer-reviewed journals have publicly available definitions of misconduct and less than half describe editorial procedures for handling suspected misconduct.[42] Investigating research misconduct is usually not straightforward, and therefore dealing with possible misconduct is not an easy task. Failure to adequately investigate possible misconduct may perpetuate unreliable research findings in the literature. Moreover, when researchers who commit fraud go unchecked, they may continue to practice misconduct.[40] There should be more awareness among publishers, editors and researchers that not all researchers work honestly, and science should not just rely only on trust anymore. Peer review should focus more on integrity issues and there is a need for an overview of methods to investigate research integrity issues in articles.

BACKGROUND AND SCOPE OF THE THESIS

In this thesis, the following of these knowledge gaps will be addressed.

Firstly, in daily practice, women with normogonadotropic anovulation and CC failure usually switch to ovulation induction with gonadotrophins, and intrauterine insemination (IUI) is often added.[18] Knowledge is lacking on the effectiveness of this switch to gonadotrophins and IUI compared with continued treatment with CC. This has never been studied in an RCT. Furthermore, as health care resources are limited, treatment costs are relevant. CC is relatively inexpensive due to the low price of the tablets and limited monitoring requirements. Ovulation induction with gonadotrophins is more expensive due to the medication price and the need for strict ultrasound monitoring.[43-46] We need knowledge on the relative costs and effectiveness of these interventions with and without IUI.

Secondly, several studies have pointed towards a negative antiestrogenic effect of CC on the endometrial development/receptivity, cervical mucus, and uterine blood flow.[47,48] Furthermore, a larger endometrial thickness (EMT) has been associated with increased live birth rates in previous studies. It is unknown whether EMT can be used as a biomarker to distinguish between women who would benefit from switching to gonadotropins and those who could continue using CC.

Thirdly, for women with normogonadotropic anovulation, more knowledge on subsequent treatment decisions and success rates in such a well-mapped population, is relevant for patients, fertility doctors and policy-makers. It would provide additional information for patients on their chances over time to fulfil their child-wish, for physicians to inform their patients better, and for policy-makers to create a realistic profile for budget-impact analysis. For women with both normogonadotropic anovulation and CC failure, there is a knowledge gap concerning the long-term cumulative chance for delivering at least one live birth.

Fourthly, women with PCOS represent a heterogeneous population according to the diagnostic criteria. Knowledge lacking on which individuals benefit mostly from a particular treatment impede clinicians to provide personalized care.[28] This knowledge gap should preferably be investigated by using individual participant data (IPD). The availability of IPD and the willingness to share data may be an indicator of quality, methodological soundness and integrity of trials when being considered for inclusion in systematic reviews. There has not been a head-to-head comparison between shared and non-shared RCTs contributing data to the same set of IPD meta-analyses in terms of methodological issues, which potentially endangers the robustness of synthesized evidence.

Fifthly, methods that investigate research misconduct accumulate and evolve. The literature needs to be reviewed for articles that mention, describe, validate, or apply methods for screening or assessing research misconduct in health-related research. A complete overview of these methods and their applicability is presently lacking.

OUTLINE OF THIS THESIS

Chapter 2 shows the results of a multicenter RCT comparing gonadotrophins with CC both with and without IUI in 666 women with normogonadotropic anovulation not being pregnant after six ovulatory cycles of CC (CC failure). The primary outcome was conception leading to birth of a live child within eight months after randomization.

Chapter 3 reports the cost-effectiveness analysis of the RCT presented in chapter 2. For each of the treatment strategies, we calculated the mean direct medical costs, effectiveness, and incremental cost-effectiveness ratios.

Chapter 4 is a post hoc analysis of the RCT in chapter 2. This study evaluates whether endometrial thickness, during the sixth ovulatory cycle of ovulation induction with CC, can be used as biomarker to select between women with normogonadotropic anovulation and CC failure who are better off switching to gonadotropins and those who could continue CC.

Chapter 5 presents the long-term outcomes of switching to gonadotrophins versus continuing treatment with CC. The study population comprised all women who participated in the RCT (chapter 2). The main outcome is cumulative chance for delivering at least one live birth in women with normogonadotropic anovulation and CC failure.

Chapter 6 provides the results of an IPD meta-analysis aiming to evaluate the effectiveness of different ovulation induction agents, as compared to CC, as the first-line choice for ovulation induction in women with PCOS and infertility. The main outcome is live birth and the study also explores interactions between treatment and participant-level baseline characteristics. We included RCTs comparing the following interventions with each other or placebo/no treatment in women with PCOS and infertility: CC, metformin, CC plus metformin, letrozole, gonadotrophin and tamoxifen.

Chapter 7 elucidates if RCTs without IPD sharing have lower quality and more methodological issues than those with IPD sharing. We included all RCTs evaluating first-line ovulation induction for PCOS included in the IPD meta-analyses in chapter 6. The studies are assessed and compared in the shared and non-shared groups on the following criteria: Risk of Bias (RoB 2.0), GRADE approach, adequacy of trial registration; description of statistical methods and reproducibility of univariable statistical analysis; excessive similarity or difference in baseline characteristics that is not compatible with chance; and other miscellaneous methodological issues.

Chapter 8 gives an overview of the methods available to investigate research misconduct in health-related research. We included papers that mentioned and/or described methods for screening or assessing research misconduct in health-related research. We categorized identified methods into four groups according to their scopes: overall concern, textual concern, image concern, and data concern.

Chapter 9 summarizes this thesis, provides implications for clinical practice, and suggestions for future research.

REFERENCE LIST

1. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Human reproduction*. 2018;33(9):1602-18.
2. Chaudhari, AP, Mazumdar, K, Mehta, PD. Anxiety, depression, and quality of life in women with polycystic ovarian syndrome. *Indian J Psychol Med* 2018;40:239–46.
3. DeUgarte CM, Bartolucci AA, Azziz R. Prevalence of insulin resistance in the polycystic ovary syndrome using the homeostasis model assessment. *Fertil Steril* 2005;83:1454–60.
4. Livadas S, Anagnostis P, Bosdou JK, Bantouna D, Paparodis R. Polycystic ovary syndrome and type 2 diabetes mellitus: A state-of-the-art review. *World J Diabetes*. 2022 Jan 15;13(1):5-26.
5. Osibogun O, Ogunmoroti O, Michos ED. Polycystic ovary syndrome and cardiometabolic risk: Opportunities for cardiovascular disease prevention. *Trends Cardiovasc Med*. 2020 Oct; 30(7):399-404.
6. Legro RS, Kunselman AR, Dunaif A. Prevalence and predictors of dyslipide- mia in women with polycystic ovary syndrome. *Am J Med* 2001;111:607–13. 12.
7. Boomsma CM, Eijkemans MJ, Hughes EG, Visser GH, Fauser BC, Macklon NS. A meta-analysis of pregnancy outcomes in women with polycystic ovary syndrome. *Hum Reprod Update* 2006;12:673–83.
8. Qin JZ, Pang LH, Li MJ, Fan XJ, Huang RD, Chen HY. Obstetric complications in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Reprod Biol Endocrinol* 2013;11:56.
9. Kjerulff LE, Sanchez-Ramos L, Duffy D. Pregnancy outcomes in women with polycystic ovary syndrome: a metaanalysis. *Am J Obstet Gynecol* 2011;204:558.e1-6.
10. Hart R, Doherty DA. The potential implications of a PCOS diagnosis on a woman's long-term health using data linkage. *J Clin Endocrinol Metab* 2015;100:911–9.
11. Ovarian Stimulation TEGGO, Bosch E, Broer S, Griesinger G, Grynberg M, Humaidan P, Kolibianakis E, Kunicki M, La Marca A, Lainas G, Le Clef N, Massin N, Mastenbroek S, Polyzos N, Sunkara SK, Timeva T, Töyli M, Urbancsek J, Vermeulen N, Broekmans F. ESHRE guideline: ovarian stimulation for IVF/ICSI†. *Hum Reprod Open*. 2020 May 1;2020(2):hoaa009.
12. World Health Organization (WHO). International Classification of Diseases, 11th Revision (ICD-11) Geneva: WHO 2018.
13. NICE. Fertility problems: assessment and treatment. London: National Institute for Health and Care Excellence (NICE); 2017 Sep. PMID: 32134604.
14. National Collaborating Centre for Ws, Children's H. National Institute for Health and Clinical Excellence: Guidance. Fertility: Assessment and Treatment for People with Fertility Problems. London: Royal College of Obstetricians & Gynaecologists National Collaborating Centre for Women's and Children's Health.; 2013.
15. Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril* 2016;106:6–15.

16. 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.
17. Norman RJ, Dewailly D, Legro RS, Hickey TE. Polycystic ovary syndrome. *Lancet* 2007; 370(9588): 685-97.
18. Thessaloniki EA-SPCWG. Consensus on infertility treatment related to polycystic ovary syndrome. *Hum Reprod* 2008; 23(3): 462-77.
19. Balen AH, Morley LC, Misso M, et al. The management of anovulatory infertility in women with polycystic ovary syndrome: an analysis of the evidence to support the development of global WHO guidance. *Hum Reprod Update* 2016; 22(6): 687-708.
20. Tang T, Lord JM, Norman RJ, Yasmin E, Balen AH. Insulin-sensitising drugs (metformin, rosiglitazone, pioglitazone, D-chiro-inositol) for women with polycystic ovary syndrome, oligo amenorrhoea and subfertility. *Cochrane Database Syst Rev* 2012; (5): CD003053.
21. Brown J, Farquhar C. Clomiphene and other antioestrogens for ovulation induction in polycystic ovarian syndrome. *Cochrane Database Syst Rev* 2016; 12: CD002249.
22. Wang R, Kim BV, van Wely M, et al. Treatment strategies for women with WHO group II anovulation: systematic review and network meta-analysis. *BMJ* 2017; 356: j138.
23. Franik S, Eltrop SM, Kremer JA, Kiesel L, Farquhar C. Aromatase inhibitors (letrozole) for subfertile women with polycystic ovary syndrome. The Cochrane database of systematic reviews. 2018;5: Cd010287.
24. Homburg R. Clomiphene citrate--end of an era? A mini-review. *Hum Reprod* 2005; 20(8): 2043-51.
25. Veltman-Verhulst SM, Fauser BC, Eijkemans MJ. High singleton live birth rate confirmed after ovulation induction in women with anovulatory polycystic ovary syndrome: validation of a prediction model for clinical practice. *Fertil Steril* 2012; 98(3): 761-8 e1.
26. Neven ACH, Laven J, Teede HJ, Boyle JA. A Summary on Polycystic Ovary Syndrome: Diagnostic Criteria, Prevalence, Clinical Manifestations, and Management According to the Latest International Guidelines. Seminars in reproductive medicine. 2018;36(1):5-12.
27. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
28. Broeze KA, Opmeer BC, van der Veen F, Bossuyt PM, Bhattacharya S, Mol BW. Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. *Hum Reprod Update* 2010;16:561–567.
29. Wang R, Mol BW. The Rotterdam criteria for polycystic ovary syndrome: evidence-based criteria? *Hum Reprod* 2017;32:261–264.
30. Bergeris A, Tse T, Zarin DA. Trialists' intent to share individual participant data as disclosed at ClinicalTrials.gov. JAMA 2018;319: 406–408.
31. Mbuagbaw L, Foster G, Cheng J, Thabane L. Challenges to complete and useful data sharing. Trials 2017;18:71.
32. Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time

and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. *BMJ* 2017;357:j1390.

33. Veroniki AA, Ashoor HM, Le SPC, Rios P, Stewart LA, Clarke M, Mavridis D, Straus SE, Tricco AC. Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. *J Clin Epidemiol* 2019;113:176–188.
34. International Weight Management in Pregnancy (i-WIP) Collaborative Group. Effect of diet and physical activity based interventions in pregnancy on gestational weight gain and pregnancy outcomes: meta-analysis of individual participant data from randomised trials. *BMJ* 2017;358:j3119.
35. Bordewijk EM, Ng KYB, Rakic L, Mol BWJ, Brown J, Crawford TJ, van Wely M. Laparoscopic ovarian drilling for ovulation induction in women with anovulatory polycystic ovary syndrome. *Cochrane Database Syst Rev*. 2020 Feb 11;2(2):CD001122.
36. Bordewijk EM, Wang R, Askie LM, Gurrin LC, Thornton JG, van Wely M, Li W, Mol BW. Data integrity of 35 randomised controlled trials in women' health. *Eur J Obstet Gynecol Reprod Biol.* 2020 Jun;249:72-83.
37. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 2009;4 e5738.
38. Office of Research Integrity. Definition of Research Misconduct. https://ori.hhs.gov/definition-research-misconduct
39. Committee on Publication Ethics (COPE). Code of conduct,https://publicationethics.org/files/Code%20of%20Conduct.pdf
40. Wager E. Coping with scientific misconduct. Bmj 2011;343 d6586.
41. COPE Council. 2021. *COPE Flowcharts and Infographics — Handling of Post-publication Critiques.* doi:https://doi.org/10.24318/o1VgCAih.
42. Bosch X, Hernández C, Pericas JM, Doti P, Marušić A. Misconduct policies in high-impact biomedical journals. *PloS one* 2012;7e51928-e.
43. Homburg R, Hendriks ML, Konig TE, et al. Clomifene citrate or low-dose FSH for the first-line treatment of infertile women with anovulation associated with polycystic ovary syndrome: a prospective randomized multinational study. *Hum Reprod* 2012; 27(2): 468-73.
44. Practice Committee of the American Society for Reproductive M. Use of clomiphene citrate in infertile women: a committee opinion. *Fertil Steril* 2013; 100(2): 341-8.
45. Moolenaar LM, Nahuis MJ, Hompes PG, van der Veen F, Mol BW. Cost-effectiveness of treatment strategies in women with PCOS who do not conceive after six cycles of clomiphene citrate. *Reprod Biomed Online* 2014; 28(5): 606-13.
46. Legro RS. Ovulation induction in polycystic ovary syndrome: Current options. *Best Pract Res Clin Obstet Gynaecol* 2016; 37: 152-9.
47. Weiss NS, van Vliet MN, Limpens J, Hompes PGA, Lambalk CB, Mochtar MH, van der Veen F, Mol BWJ, van Wely M. Endometrial thickness in women undergoing IUI with ovarian stimulation. How thick is too thin? A systematic review and meta-analysis. *Hum Reprod* 2017;32:1009–1018.
48. Gadalla MA, Huang S, Wang R, Norman RJ, Abdullah SA, El Saman AM, Ismail AM, van Wely M,

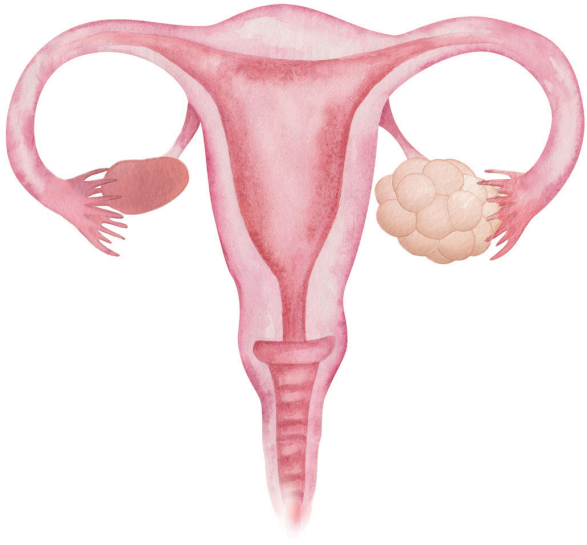
Mol BWJ. Effect of clomiphene citrate on endometrial thickness, ovulation, pregnancy and live birth in anovulatory women: systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2018;51:64–76.

Chapter 2

Gonadotrophins versus clomifene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomifene failure (M-OVIN): a randomised, two-by-two factorial trial.

Weiss NS, Nahuis MJ, Bordewijk EM, Oosterhuis JE, Smeenk JM, Hoek A, Broekmans FJ, Fleischer K, de Bruin JP, Kaaijk EM, Laven JS, Hendriks DJ, Gerards MH, van Rooij IA, Bourdrez P, Gianotten J, Koks C, Lambalk CB, Hompes PG, van der Veen F, Mol BWJ, van Wely M.

Lancet. 2018 Feb



ABSTRACT

Background

In many countries, clomifene citrate is the treatment of first choice in women with normogonadotropic anovulation (ie, absent or irregular ovulation). If these women ovulate but do not conceive after several cycles with clomifene citrate, medication is usually switched to gonadotrophins, with or without intrauterine insemination. We aimed to assess whether switching to gonadotrophins is more effective than continuing clomifene citrate, and whether intrauterine insemination is more effective than intercourse.

Methods

In this two-by-two factorial multicentre randomised clinical trial, we recruited women aged 18 years and older with normogonadotropic anovulation not pregnant after six ovulatory cycles of clomifene citrate (maximum of 150 mg daily for 5 days) from 48 Dutch hospitals. Women were randomly assigned using a central password-protected internet-based randomisation programme to receive six cycles with gonadotrophins plus intrauterine insemination, six cycles with gonadotrophins plus intercourse, six cycles with clomifene citrate plus intrauterine insemination, or six cycles with clomifene citrate plus intercourse. Clomifene citrate dosages varied from 50 to 150 mg daily orally and gonadotrophin starting dose was 50 or 75 IU daily subcutaneously. The primary outcome was conception leading to livebirth within 8 months after randomisation defined as any baby born alive after a gestational age beyond 24 weeks. Primary analysis was by intention to treat. We made two comparisons, one in which gonadotrophins were compared with clomifene citrate and one in which intrauterine insemination was compared with intercourse. This completed study is registered with the Netherlands Trial Register, number NTR1449.

Findings

Between Dec 8, 2008, and Dec 16, 2015, we randomly assigned 666 women to gonadotrophins and intrauterine insemination (n=166), gonadotrophins and intercourse (n=165), clomifene citrate and intrauterine insemination (n=163), or clomifene citrate and intercourse (n=172). Women allocated to gonadotrophins had more livebirths than those allocated to clomifene citrate (167 [52%] of 327 women *vs* 138 [41%] of 334 women, relative risk [RR] 1·24 [95% CI 1·05–1·46]; p=0·0124). Addition of intrauterine insemination did not increase livebirths compared with intercourse (161 [49%] *vs* 144 [43%], RR 1·14 [95% CI 0·97–1·35]; p=0·1152). Multiple pregnancy rates for the two comparisons were low and not different. There were three adverse events: one child with congenital abnormalities and one stillbirth in two women treated with

clomifene citrate, and one immature delivery due to cervical insufficiency in a woman treated with gonadotrophins.

Interpretation

In women with normogonadotropic anovulation and clomifene citrate failure, a switch of treatment to gonadotrophins increased the chance of livebirth over treatment with clomifene citrate; there was no evidence that addition of intrauterine insemination does so.

Funding

The Netherlands Organization for Health Research and Development.

RESEARCH IN CONTEXT

Evidence before this study

We searched PubMed on Sept 15, 2008, before the trial started to identify all previous studies investigating women with clomifene failure with the following search terms: ''ovulation induction'', ''polycystic ovary syndrome'', ''clomiphene citrate'' (CC), ''gonadotrophins'', and ''intrauterine insemination''.

We identified only non-randomised studies suggesting that continued treatment with clomifene citrate and a treatment switch to gonadotrophins were both effective options for these women. Whether intrauterine insemination increases pregnancy rates in women with clomifene citrate failure is unknown.

In view of this research gap, we aimed to assess whether, in women who have failed to conceive after six ovulatory cycles with clomifene citrate, ovulation induction with gonadotrophins leads to higher livebirth rates than continued ovulation induction with clomifene citrate and whether intrauterine insemination leads to more livebirths than intercourse.

Added value of this study

The M-OVIN (Modified Ovulation Induction) study compared in anovulatory women with clomifene citrate failure two types of medication as well as addition of intrauterine insemination with intercourse. We found that a switch to gonadotrophins significantly increased the livebirth rate compared with continued treatment with clomifene citrate and that the addition of intrauterine insemination to gonadotrophins or clomifene citrate did not increase livebirth rates.

Implications of all the available evidence

Our findings imply that, for normogonadotropic anovulatory women with clomifene citrate failure who wish to conceive, continued treatment with clomifene citrate or a treatment switch to gonadotrophins are both effective options in terms of livebirth rates, whereas we could not prove this for intrauterine insemination. The choice between clomifene citrate and gonadotrophins should be made based on women's preferences, costs, and reimbursement. Considering recent randomised research suggesting that letrozole gives higher livebirth rates than clomifene citrate in the first six cycles, future research should establish whether continuing letrozole is also effective and safe if women have not conceived within the first 6 months of treatment.

INTRODUCTION

Women with normogonadotropic anovulation have absent or irregular ovulation due to hypothalamic-pituitary-ovarian dysfunction associated with normal concentrations of endogenous oestradiol.[1] In these women wishing to conceive, clomifene citrate has long been used as a first-line ovulation induction agent.[2,3] Findings of systematic reviews and meta-analyses have shown that clomifene citrate is an effective primary treatment option in therapynaive women with normogonadotropic anovulation and polycystic ovary syndrome.[4–6] Although ovulation is restored in about 75% of women starting ovulation induction with clomifene citrate, 6 months of treatment leads to conception in only about half of these women.[5,7] Women not conceiving after six ovulatory cycles are defined as having clomifene citrate failure.[8] The National Institute for Health and Care Excellence (NICE) guideline recommends not to extend treatment with clomifene citrate for more than six cycles, but this recommendation is not underpinned by any evidence.[9] In daily practice, these women usually switch to ovulation induction with gonadotrophins and intrauterine insemination is often initiated instead of relying on regular intercourse.[10] However, the effectiveness of a switch to gonadotrophins and intrauterine insemination compared with continued treatment with clomifene citrate has never been studied in randomised clinical trials.

To address this research gap, we aimed to compare, in women who had six ovulatory cycles with clomifene citrate but did not conceive, the effectiveness of a switch to gonadotrophins compared with continued treatment with clomifene citrate and the effectiveness of adding intrauterine insemination to either clomifene citrate or gonadotrophins.

METHODS

Study design and participants

The Modified Ovulation Induction (M-OVIN) study was a multicentre randomised clinical trial done in 48 Dutch hospitals within the infrastructure of the Dutch Consortium for Healthcare Evaluation and Research in Obstetrics and Gynaecology. Eligible women were subfertile, aged 18 years and older with WHO type II anovulation (menstrual cycle >35 days, normogonadotropic, normo-oestrogenic, oligo-anovulation or anovulation), and had been ovulatory for six cycles on clomifene citrate treatment, with a maximum of 150 mg daily for 5 days, but had not conceived. Presence of ovulation was assessed by a basal body temperature curve, midluteal progesterone (>16 nmol/L), detection of a urinary luteinising hormone surge, or transvaginal sonography, depending on the local protocol. All women had undergone a basic fertility work-up including a semen analysis and endocrinology screening to rule out hyperprolactinaemia and uncorrected thyroid

dysfunction. Couples with male subfertility could not participate. Women with abnormal prolactin (0·05–0·80 IU/L) or thyroid-stimulating hormone (0·4–4·0 mU/L) were also not eligible. Tubal pathology had to be ruled out by either a negative Chlamydia antibody titre (CAT) or hysterosalpingography, transvaginal hydrolaparoscopy, or diagnostic laparoscopy showing at least one patent fallopian tube. Women with side-effects in previous clomifene citrate cycles were also not eligible. All women provided written informed consent. The study was granted approval by the Medical Ethical Committee of the Medical Spectrum Twente Enschede (Netherlands) and from the Central Committee on Research involving Human Subjects (CCMO, Netherlands). The board of directors of each of the participating centres approved local execution of the study. The protocol was published previously.[11] Two major adjustments to the protocol were made: in April, 2014, a change was made to the primary outcome from "ongoing pregnancy" to "livebirth". The second regarded the sample size. Both adjustments were approved by the Medical Ethical Committee.

Randomisation and masking

Eligible women were informed about the study during or immediately after their sixth treatment cycle either by their doctor or by a dedicated research nurse. Women were randomly assigned using a central password protected internet based randomisation program. The randomisation list had been prepared by an independent statistician with a variable block size and a maximum block size of 8. There was no masking.

We used a two-by-two factorial design to compare two pairs of interventions: a switch to ovulation induction with gonadotrophins versus continuing clomifene citrate and intrauterine insemination versus intercourse. Women were randomly assigned to six cycles with gonadotrophins plus intrauterine insemination, six cycles with gonadotrophins plus intercourse, six cycles with clomifene citrate plus intrauterine insemination, or six cycles with clomifene citrate plus intercourse.

Procedures

In women allocated to ovulation induction with gonadotrophins, a transvaginal ultrasound was usually done on the third day of a menstrual bleed and medication was started on that same day, but women were allowed to start medication up to day 5. Treatment was not started if ultrasound showed ovarian cysts bigger than 25 mm in mean diameter. According to local protocol, urinary or recombinant gonadotrophins were used with a starting dose of 50 or 75 IU daily. Follicular growth was strictly monitored by transvaginal ultrasound and we aimed for mono-follicular growth. When at least one follicle with a diameter of at least 16 mm was present, ovulation was triggered with 5000 IU or 10 000 IU of human chorionic gonadotrophin. If four or more dominant follicles (≥18 mm) developed, the cycle was cancelled - ie, couples were advised not to have intercourse and the planned intrauterine insemination was not done. In women allocated to intrauterine insemination,

semen samples were processed within 1 h of ejaculation according to the local protocol and women were inseminated 36–40 h after human chorionic gonadotrophin injection. Intrauterine insemination was done once per cycle.

In women allocated to ovulation induction with clomifene citrate, treatment was started on the third to fifth day of a menstrual bleed, in the same dosage as used in the last ovulatory cycle, varying between 50 mg and 150 mg daily, for 5 days. Ovulation was monitored by a basal body temperature curve, midluteal progesterone (>16 nmol/L), a urinary lutenising hormone surge, or transvaginal ultrasound, depending on the local protocol. Women undergoing ovulation induction with clomifene citrate plus intrauterine insemination were monitored by ultrasound; women assigned to clomifene citrate with intercourse were usually monitored by basal body temperature curve, midluteal progesterone measurement, or urinary lutenising hormone surge. In case of ovulation not followed by pregnancy, women continued taking the same dose of clomifene citrate until pregnancy occurred, or until the end of the study (8 months after randomisation). If ovulation did not occur, the dosage was increased in increments of 50 mg to a maximum of 150 mg daily in the next cycles. Follow-up started at the day of randomisation and ended on the first day of the last menstruation before a positive pregnancy test within six treatment cycles or at 8 months after randomisation, whichever came first. If pregnant, women had an ultrasound at 7 and 11 weeks of gestation and were followed up until delivery of their baby. If they miscarried or had an ectopic pregnancy within 8 months after randomisation, couples were advised to continue their allocated treatment.

Data were collected by trained research nurses and doctors. They used a structured case record form to register the actual interventions, the reproductive outcomes, the occurrence of gestational diabetes, hypertensive disorders, stillbirths, preterm labour, and fetal birthweight as well as the course and outcome of subsequent pregnancies. If the women's medical records did not give the necessary information, women were contacted by telephone to ask about their outcomes.

We expected some couples to drop out of the study as per usual clinical practice, particularly in this protocol in which women had already had six ovulatory treatment cycles before inclusion. Women who dropped out of the study were managed according to their preferences.

Outcomes

The primary outcome measure was conception leading to livebirth within 8 months after randomisation, defined as any baby born alive with a gestational age beyond 24 weeks. Secondary outcome measures were ongoing pregnancy, multiple pregnancy, miscarriage (defined as loss of an intrauterine pregnancy confirmed by ultrasound or histological examination before the 20th week of pregnancy), ectopic pregnancy, time from randomisation to the birth of a live child, fetal birthweight, and pregnancy complications - ie, hypertensive disorders, gestational diabetes, and preterm labour.[11] We did not monitor adverse drug events because these are already widely known for both types of medication. We do not report on all outcomes mentioned in the statistical analysis plan here. Outcomes such as clinical pregnancy rate, ovulation rate, and gestational age will be reported elsewhere.

Statistical analysis

When we first planned our study, we designed the trial as a two-by-two factorial superiority trial. After recruiting 136 women, we received governmental funding that allowed enlargement of our trial. To assess whether either switching to ovulation induction with gonadotrophins or addition of intrauterine insemination would increase the livebirth rate from 40% to 55%,[12,13] we needed to include 600 women (alpha of 5% and a power of 88% at three degrees of freedom). We decided to include a total of 660 women because 10% of women became pregnant after randomisation but before starting the trial. With these 660 women we would have sufficient power to find a difference in livebirth rate for the two comparisons that we have made. A detailed description of all steps in establishing the sample size is provided in the appendix. A statistical analysis plan was established before data lock.

The primary analysis was on an intention-to-treat basis. For the livebirth rates and other binary outcome measures, we calculated absolute risks, relative risks, and 95% confidence intervals. Chi-square test statistics were used to assess statistical significance. We reported categorical data as absolute numbers and percentages. We summarised normally distributed continuous variables as means with standard deviations, and non-normally distributed continuous variables as medians with IQRs. We formally tested for interaction between the two comparisons. We constructed Kaplan-Meier curves for time to conception leading to livebirth for gonadotrophins versus clomifene citrate, for intrauterine insemination versus intercourse, and for all four treatment arms separately. They were compared with a logrank test. Two-sided p values of less than 0·05 were considered to indicate statistical significance. We assessed whether there was interaction between treatment effect and body-mass index (BMI) at cut-off of 25 kg/m^2 as this was the mean BMI of our population. We also did a per-protocol analysis in which we only

included women that were treated according to the predefined protocol. SPSS software (version 23.0; IBM Corp, USA) was used for statistical analysis.

This study is registered with the Netherlands Trial Register, number NTR1449.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

Between Dec 8, 2008, and Dec 16, 2015, 762 women were registered as eligible. 96 women declined randomisation and 666 were randomly assigned. 166 women were allocated to ovulation induction with gonadotrophins combined with intrauterine insemination, 165 to ovulation induction with gonadotrophins, 163 to ovulation induction with clomifene citrate combined with intrauterine insemination, and 172 to continued ovulation induction with clomifene citrate (figure 1). We excluded five women after randomisation because they did not fulfil the inclusion criteria. None of these women became pregnant. The baseline characteristics were similar across the four groups (table 1).

Women allocated to gonadotrophins with intrauterine insemination underwent 540 cycles, women allocated to gonadotrophins only underwent 570 cycles, women allocated to clomifene citrate with intrauterine insemination underwent 612 cycles, and women allocated to clomifene citrate only underwent 681 cycles. Of these cycles, 65 (12%) were cancelled in the gonadotrophins with intrauterine insemination group and 61 (11%) in the gonadotrophins only group. Of these cancelled cycles, 35 (28%) were due to anovulation; the other cycles were cancelled because of multiple follicular growth (table 2).

Women allocated to gonadotrophins had significantly more livebirths than women allocated to clomifene citrate (167 [52%] of 327 women vs 138 [41%] of 334, relative risk [RR] 1·24 [95% CI 1·05–1·46]; p=0·0124; absolute difference 10·2% [95% CI 2·4–17·9]; table 3). The mean time to conception leading to a livebirth was 5 months (95% CI 4·7–5·4) following gonadotrophins and 5·5 months (5·1–5·8) following clomifene citrate (log-rank test; p=0·028; figure 2). Seven women (2%) allocated to gonadotrophins conceived a twin pregnancy versus eight women (2%) allocated to clomifene citrate (RR 0·89 [95% CI 0·33–2·4]; p=0·8262; absolute difference 0%).

Table 1. Baseline characteristics of the participating couples.

	Gonado-trophins + IUI n = 164	**Gonado-trophins + intercourse n = 163**	**CC + IUI n = 163**	**CC + inter-course n = 171**
Age of women (years)	29·5 ± 3·7	29·9 ± 3·7	30·0 ± 3·6	29·9 ± 4·0
Ethnicity				
White	131 (85%)	134 (88%)	133 (86%)	141 (89%)
Non-white	24 (15%)	18 (12%)	21 (14%)	18 (11%)
BMI (kg/m^2)*	25·4 ± 5·1	25·6 ± 5·6	25·0 ± 4·9	25·4 ± 5·0
BMI >25.0 kg/m^2	76 (46%)	81 (49%)	64 (39%)	81 (47%)
Current smoker	29 (18%)	20 (12%)	22 (13%)	22 (13%)
Diabetes	1	1	3	2
Previous livebirth	32 (20%)	35 (21%)	36 (22%)	34 (20%)
Duration of subfertility (months)	26·3 ± 14·9	24·5 ± 12·5	24·5 ± 15·5	25·9 ± 19·0
Cycle pattern prior to treatment #				
Amenorrhea	124 (76%)	125 (77%)	115 (71%)	120 (70%)
Oligomenorrhea	21 (13%)	25 (15%)	27 (16%)	32 (19%)
Unknown	19 (11%)	13 (8%)	21 (13%)	19 (11%)
Median TMC *10^6	52 (20-106)	43 (16-113)	53 (15-132)	38 (16-99)
Polycystic ovaries on ultrasound ##	110 (67%)	103 (63%)	109 (67%)	117 (68%)
Mean serum biochemical values				
FSH (IU/L)	5·7 ± 2·1	5·7 ± 1·7	6·2 ± 2·2	6·0 ± 2·2
LH (IU/L)	9·7 ± 7·4	10·6 ± 7·8	10·6 ± 7·6	10·9 ± 10·8
Estrogen (pmol/L)	255 ± 295	239 ± 217	201 ± 159	271 ± 460
Total testosterone (nmol/L)	1·6 ± 1·7	1·6 ± 2·0	1·8 ± 2·2	1·8 ± 1·8

Data are mean (SD), n (%) or median (IQR). BMI = body-mass index. TMC = total motile sperm count. FSH = follicle stimulating hormone. LH = luteinizing hormone. CC = clomiphene citrate. IUI = intrauterine insemination.

*BMI was missing for 24 women; data were imputed by using multiple imputation.

Amenorrhea: absence of menstrual bleeding for >6 months. Oligomenorrhea: irregular menstrual bleedings with intervals of >35 days but ≤6 months

Defined as the presence of 12 or more follicles in each ovary measuring 2–9 mm in diameter

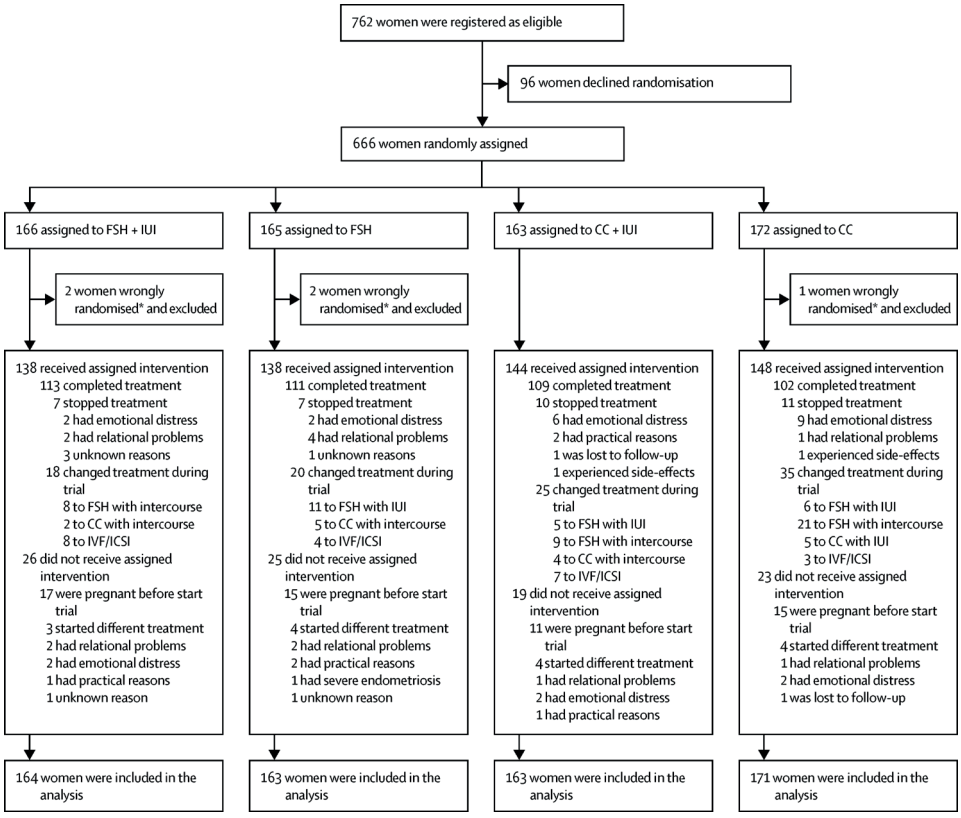


Figure 1. Trial profile
FSH=follicle-stimulating hormone. CC=clomifene citrate. IUI=intrauterine insemination. IVF=in-vitro fertilisation. ICSI=intracytoplasmic sperm injection.*2 women had thyroid disease, 1 woman had bilateral tubal pathology, 1 male partner had azoospermia, 1 woman only had two cycles with clomifene citrate before randomisation.

Table 2. Cycle results

	Gonado-trophins + IUI n=164	**Gonado-trophins + intercourse n=163**	**CC + IUI n=163**	**CC + in-tercourse n=171**
Total nr of cycles	540	570	612	681
Mean nr of cycles per woman	3·3 ± 2·0	3·5 ± 2·1	3·8 ± 1·8	4·0 ± 1·9
Mean nr of IUIs per woman	3·2 ± 2·2	0·04 ± 0·3	3·5 ± 2·2	0·05 ± 0·4
Total nr of cancelled cycles	65 (12%)	61 (11%)	4*	2*
Total units of gonadotrophins per woman	2594 ± 2439	2640 ± 2577	153 ± 823*	223 ± 823*
Total mg of CC per woman	4·5 ± 43·4 #	18·2 ± 128 #	1401 ± 1152	1255 ± 1139

Data are n (%) or mean (SD)
*After switching to gonadotrophins
# After switching to CC
CC = clomiphene citrate. IUI = intrauterine insemination

Women allocated to intrauterine insemination had more livebirths than women allocated to intercourse, but this difference was not statistically different (161 [49%] of 327 women vs 144 [43%] of 334 women, RR 1·14 [95% CI 0·97–1·35]; p=0·1152; absolute difference 6·1% [95% CI –1·71 to 13·8; table 3). The mean time to conception leading to a livebirth was 5·2 months (95% CI 4·8–5·5) with intrauterine insemination and 5·3 months (5·0–5·7) with intercourse (log-rank test; p=0·27; figure 2). There were 11 (3%) twin pregnancies after intrauterine insemination and four (1%) after intercourse (RR 2·8 [95% CI 0·90–8·7]; p=0·0743; absolute difference 2·0%). There were no high order pregnancies.

The number of miscarriages was higher after treatment with gonadotrophins (n=24 [7%]) than after clomifene citrate (n=11 [3%]; RR 2·2 [95% CI 1·11–4·5]; p=0·0243; absolute difference 4·0%). The number of ectopic pregnancies was similar between all groups. We found no differences in mean birthweights and pregnancy complications (table 3). We noted no interaction between the two comparisons (p=0·932). Also, there was no interaction of BMI and treatment effect for both comparisons.

We included 563 women in the per-protocol analysis. We noted more livebirths after gonadotrophins compared with clomifene citrate (123 [44%] of 279 women after gonadotrophins vs 90 [32%] of 284 women after clomifene citrate, RR 1·38 [95% CI 1·11–1·72]; p=0·0027; absolute difference 13%). Addition of intrauterine insemination did not increase livebirths compared with intercourse: 113 (41%) of 277 women had a livebirth after intrauterine insemination versus 100 (35%) of 286 women after intercourse (RR 1·17 [95% CI 0·94–1·44]; p=0·1548; absolute difference 13%).

There were three adverse events: one woman treated with clomifene citrate conceived a child with congenital abnormalities resulting in second trimester pregnancy termination, one woman treated with gonadotrophins with intrauterine insemination delivered at a gestational age of 20 weeks due to cervical insufficiency, and one woman treated with clomifene citrate had a stillbirth at a gestational age of 19 weeks.

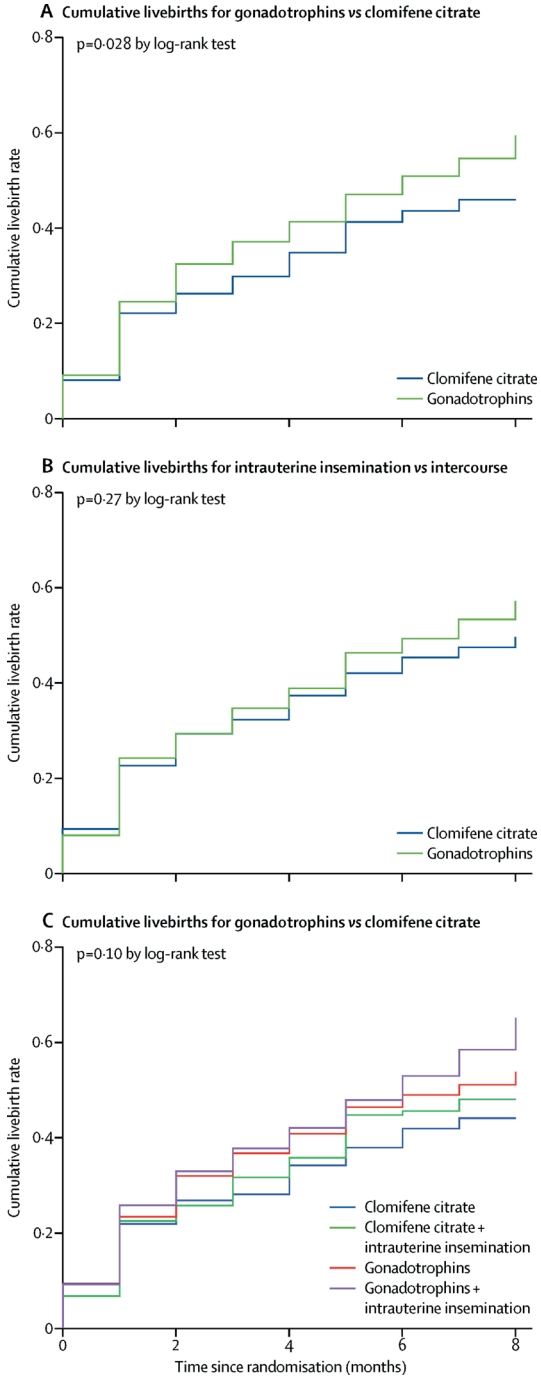


Figure 2. Time to conception leading to livebirth for the comparison gonadotrophins versus clomifene citrate, and intrauterine insemination versus intercourse.

Table 3. Primary and secondary outcomes

	Gonadotrophins plus intrauterine insemination (n=164)	Gonadotrophins plus intercourse (n= 163)	Clomifene citrate plus intrauterine insemination (n=163)	Clomifene citrate plus intercourse (n=171)	Gonadotrophins *vs* clomifene citrate RR (95% CI)	Gonadotrophins *vs* clomifene citrate p value	Intrauterine insemination *vs* intercourse RR (95% CI)	Intrauterine insemination *vs* intercourse p value
Livebirth	89 (54·3%)	78 (47·9%)	72 (44·2%)	66 (38·6%)	1·24 (1·05–1·46)	0·0124	1·14 (0·97–1·35)	0·12
Ongoing pregnancy	90 (54·9%)	80 (49·1%)	72 (44·2%)	66 (38·6%)	1·26 (1·07–1·48)	0·0063	1·14 (0·97–1·34)	0·13
Multiple pregnancy* per woman	4 (2·4%)	3 (1·8%)	7 (4·3%)	1 (0·6%)	0·89 (0·33–2·40)	0·82	2·8 (0·90–8·70)	0·07
Miscarriages per woman	15 (9·1%)	9 (5·5%)	8 (4·9%)	3 (1·8%)	2·2 (1·11–4·50)	0·02	1·96 (0·99–3·90)	0·05
Ectopic pregnancy per woman	1 (0·6%)	1 (0·6%)	3 (1·8%)	1 (0·6%)	*		*	
Birthweight (g)	3279 (695)	3302 (769)	3178 (714)	3408 (491)		0·96		0·14
Pregnancy complications								
Hypertensive disorders	4 (2%)	6 (4%)	5 (2%)	2 (1%)	*		*	
Gestational diabetes	3 (2%)	5 (3%)	3 (2%)	3 (2%)				
Preterm labour	6 (4%)	2 (1%)	0	1 (1%)				

Data are mean (SD) or n (%) unless otherwise stated. All multiple pregnancies were twin pregnancies. RR=relative risk.

*No RR calculated as proportions are low.

DISCUSSION

In this multicentre randomised trial, we found that, among normogonadotropic anovulatory women not pregnant after six ovulatory cycles with clomifene citrate, a switch to gonadotrophins with strict cycle monitoring increased the livebirth rate compared with continued treatment with clomifene citrate. The addition of intrauterine insemination did not increase livebirth rates. All four treatment groups resulted in acceptable pregnancy rates and low complication rates.

A strength of our study is the two-by-two factorial design. This design allowed us to dissect the effect of gonadotrophins and clomifene citrate and of intrauterine insemination versus intercourse. The per-protocol analysis limited to women that received the allocated treatment did not alter our results, suggesting that the treatment switches did not have a large effect on livebirth chances. A weakness could be that we allowed participating hospitals to use their local protocols for ovulation induction and intrauterine insemination. Alternatively, this pragmatic approach might increase the generalisability of the results. Plausible biological explanations for the finding of more livebirths with gonadotrophins than clomifene citrate may be the following. First, treatment with gonadotrophins requires strict cycle monitoring whereas treatment with clomifene citrate does not. Therefore, women given gonadotrophins have more specific knowledge on the timing of their ovulation, which might lead to a better timing of their intercourse. Second, clomifene citrate might have negative effects on the endometrium; however, studies assessing this effect in relation to pregnancy rates show conflicting results.[14–16] Third, clomifene citrate might induce cervical factor subfertility by influencing the cervical mucus.[17–19]

We do not know whether the differential monitoring in the women that underwent ovulation induction with clomifene citrate affected the outcomes, but it is not something we expect. The addition of intrauterine insemination, in which monitoring was more strict, did not result in significantly higher pregnancy chances. We believe one of the merits of our study is that even with minimal monitoring good results can be obtained with continued ovulation induction with clomifene citrate.

We found a small, not statistically significant effect of intrauterine insemination on livebirth rates. Apparently, intrauterine insemination does not contribute to pregnancy chances in women with anovulatory subfertility. We reported 4% multiple pregnancies after gonadotrophins versus 6% after clomifene citrate, which can be explained by the very purpose of ovulation induction in women with anovulation, which is to induce mono-follicular growth with low doses of gonadotrophins.[9,11] There has traditionally been reluctance to continue treatment with clomifene citrate because of safety issues.[9] However, direct evidence that cancer risks are increased after six cycles of clomifene citrate is lacking.

In our study, women given gonadotrophins had more miscarriages than women given clomifene citrate. Our study was not powered to detect a difference in miscarriage rate, hence this finding needs to be confirmed in future studies. We recorded only one second trimester miscarriage in the whole study population, which is very low and in contrast to the miscarriage rate seen after in-vitro fertilisation in a fresh transfer cycle in women with polycystic ovary syndrome.[20] This is probably due to the fact that ovulation induction aims to generate only one follicle in contrast to superovulation in in-vitro fertilisation, resulting in a thinner endometrium in ovulation induction. The cumulative livebirth rate after clomifene citrate in cycles 7–12 is similar to a previous observational study.[21] Similarly, the cumulative livebirth rate after gonadotrophins is in line with a previous prospective cohort study.[8] This underpins the reliability of our results.

Recent randomised trials and network meta-analyses reported that letrozole is associated with higher livebirth rates compared with clomifene citrate.[6,22] We therefore suggest that future research should aim to establish whether letrozole is also effective and safe if women have not conceived within the first 6 months of treatment. Based on our current finding that continued treatment with clomifene citrate is effective, one might hypothesise even higher livebirth rates for continued treatment with letrozole.
Our results can be used by couples treated with first-line ovulatory drugs who weigh the pros and cons of switching to gonadotrophins and addition of intrauterine insemination. Clomiphene citrate is known to cause more side-effects than gonadotrophins, whereas gonadotrophins necessitate daily injections combined with ultrasound monitoring of follicular development and are more expensive.[23] Findings of a recent patient preference study of women with anovulation wishing to conceive showed that just over half of these women chose treatment with the least medical interference and lowest burden whereas less than 50% preferred a treatment with the highest success rates irrespective of the burden.[24] To evaluate cost differences we have planned a cost-effectiveness analysis that will be reported elsewhere.

Our study shows that subfertile women with anovulation who are given clomifene citrate or gonadotrophins with or without intrauterine insemination reach acceptable pregnancy rates and low complication rates even until their 12th treatment cycle. This means that, in contrast to the recommendation of the NICE guideline for unexplained subfertility, switching to in-vitro fertilisation after six failed ovulation induction cycles is not necessary. The choice between these alternatives should therefore be made based on couples' preferences, costs, and reimbursement.

DECELERATIONS OF INTEREST

BWJM is supported by a NHMRC Practitioner Fellowship (GNT1082548). BWJM reports consultancy for Merck, ObsEva, and Guerbet. The Department of Obstetrics and Gynecology of the UMCG receives an unrestricted educational grant from Ferring Pharmaceutical BV, Netherlands**.** IAJvR reports personal fees from Ferring for an advisory board, outside the submitted work. CL reports grants from Ferring NV and Merck NV, outside the submitted work. JMJS reports grants and personal fees from Ferring, grants and personal fees from Merck Serono, personal fees from TEVA, outside the submitted work. FJMB is on the advisory board of Merck, Gideon Richter, and Ferring, outside the submitted work. FJMB reports grants from Merck Serono and Ferring, and consultancy work for Roche, outside the submitted work. KF reports a personal fee from Ferring and an unrestricted grant from Merck Serono.

REFERENCE LIST

1. Group ECW. Health and fertility in World Health Organization group 2 anovulatory women. *Hum Reprod Update* 2012; 18(5): 586-99.
2. Balen AH, Morley LC, Misso M, et al. The management of anovulatory infertility in women with polycystic ovary syndrome: an analysis of the evidence to support the development of global WHO guidance. *Hum Reprod Update* 2016; 22(6): 687-708.
3. Norman RJ, Dewailly D, Legro RS, Hickey TE. Polycystic ovary syndrome. *Lancet* 2007; 370(9588): 685-97.
4. Tang T, Lord JM, Norman RJ, Yasmin E, Balen AH. Insulin-sensitising drugs (metformin, rosiglitazone, pioglitazone, D-chiro-inositol) for women with polycystic ovary syndrome, oligo amenorrhoea and subfertility. *Cochrane Database Syst Rev* 2012; (5): CD003053.
5. Brown J, Farquhar C. Clomiphene and other antioestrogens for ovulation induction in polycystic ovarian syndrome. *Cochrane Database Syst Rev* 2016; 12: CD002249.
6. Wang R, Kim BV, van Wely M, et al. Treatment strategies for women with WHO group II anovulation: systematic review and network meta-analysis. *BMJ* 2017; 356: j138.
7. Homburg R. Clomiphene citrate--end of an era? A mini-review. *Hum Reprod* 2005; 20(8): 2043-51.
8. Veltman-Verhulst SM, Fauser BC, Eijkemans MJ. High singleton live birth rate confirmed after ovulation induction in women with anovulatory polycystic ovary syndrome: validation of a prediction model for clinical practice. *Fertil Steril* 2012; 98(3): 761-8 e1.
9. NICE. Fertility: Assessment and Treatment for People with Fertility Problems. 2017.
10. Thessaloniki EA-SPCWG. Consensus on infertility treatment related to polycystic ovary syndrome. *Hum Reprod* 2008; 23(3): 462-77.
11. Nahuis MJ, Weiss NS, van der Veen F, et al. The M-OVIN study: does switching treatment to FSH and / or IUI lead to higher pregnancy rates in a subset of women with world health organization type II anovulation not conceiving after six ovulatory cycles with clomiphene citrate - a randomised controlled trial. *BMC Womens Health* 2013; 13: 42.
12. Hammond MG, Halme JK, Talbert LM. Factors affecting the pregnancy rate in clomiphene citrate induction of ovulation. *Obstet Gynecol* 1983; 62(2): 196-202.
13. Kousta E, White DM, Franks S. Modern use of clomiphene citrate in induction of ovulation. *Hum Reprod Update* 1997; 3(4): 359-65.
14. Kolibianakis EM, Zikopoulos KA, Fatemi HM, et al. Endometrial thickness cannot predict ongoing pregnancy achievement in cycles stimulated with clomiphene citrate for intrauterine insemination. *Reprod Biomed Online* 2004; 8(1): 115-8.
15. De Geyter C, Schmitter M, De Geyter M, Nieschlag E, Holzgreve W, Schneider HP. Prospective evaluation of the ultrasound appearance of the endometrium in a cohort of 1,186 infertile women. *Fertil Steril* 2000; 73(1): 106-13.
16. Weiss NS, van Vliet MN, Limpens J, et al. Endometrial thickness in women undergoing IUI with ovarian stimulation. How thick is too thin? A systematic review and meta-analysis. *Hum Reprod* 2017; 32(5): 1009-18.

17. Gelety TJ, Buyalos RP. The effect of clomiphene citrate and menopausal gonadotropins on cervical mucus in ovulatory cycles. *Fertil Steril* 1993; 60(3): 471-6.
18. Hessel M, Brandes M, de Bruin JP, et al. Long-term ongoing pregnancy rate and mode of conception after a positive and negative post-coital test. *Acta Obstet Gynecol Scand* 2014; 93(9): 913-20.
19. Nahuis MJ, Weiss NS, Van der Velde M, et al. Does the postcoital test predict pregnancy in WHO II anovulatory women? A prospective cohort study. *Eur J Obstet Gynecol Reprod Biol* 2016; 199: 12731.
20. Chen ZJ, Shi Y, Sun Y, et al. Fresh versus Frozen Embryos for Infertility in the Polycystic Ovary Syndrome. *N Engl J Med* 2016; 375(6): 523-33.
21. Weiss NS, Braam S, Konig TE, et al. How long should we continue clomiphene citrate in anovulatory women? *Hum Reprod* 2014; 29(11): 2482-6.
22. Legro RS, Zhang H, Eunice Kennedy Shriver NRMN. Letrozole or clomiphene for infertility in the polycystic ovary syndrome. *N Engl J Med* 2014; 371(15): 1463-4.
23. Legro RS. Ovulation induction in polycystic ovary syndrome: Current options. *Best Pract Res Clin Obstet Gynaecol* 2016; 37: 152-9.
24. Weiss NS, Schreurs AMF, van der Veen F, et al. Women's perspectives on ovulation induction with or without IUI as treatment for normogonadotropic anovulation; A discrete choice experiment. *Human Reproduction Open* 2017; 2017(3)

Also on behalf of:
Bourdrez P, Broekmans FJM, de Bruin JP, Cohlen BJ, Fleischer K, Gerards MH, Gianotten J, Hendriks DJ, Hoek A, Kaaijk EM, Klijn NF, Koks CA, Kwee J, van de Laar-van Asseldonk TAM, Lambeek AF, Laven JSE, Nap AW, van Rijn-van Weert JM, Rijnsaardt-Lukassen HGM, Sluijmer AV, Smeenk JMJ, van Unnik GA, Vollebergh JHA, Vrouenraets FPJM

Chapter 3

Gonadotrophins versus clomiphene citrate with or without IUI in women with normogonadotropic anovulation and clomiphene failure: a cost-effectiveness analysis.

Bordewijk EM, Weiss NS, Nahuis MJ, Bayram N, van Hooff MHA, Boks DES, Perquin DAM, Janssen CAH, van Golde RJT, Lambalk CB, Goddijn M, Hompes PG, van der Veen F, Mol BWJ, van Wely M; M-ovin study group.

Human Reproduction. 2019 Feb

ABSTRACT

Study question

Are six cycles of ovulation induction with gonadotrophins more cost-effective than six cycles of ovulation induction with clomiphene citrate (CC) with or without IUI in normogonadotropic anovulatory women not pregnant after six ovulatory cycles with CC?

Summary answer

Both gonadotrophins and IUI are more expensive when compared with CC and intercourse, and gonadotrophins are more effective than CC.

What is known already

In women with normogonadotropic anovulation who ovulate but do not conceive after six cycles with CC, medication is usually switched to gonadotrophins, with or without IUI. The cost-effectiveness of these changes in policy is unknown.

Study design, size, duration

We performed an economic evaluation of ovulation induction with gonadotrophins compared with CC with or without IUI in a two-by-two factorial multicentre randomized controlled trial in normogonadotropic anovulatory women not pregnant after six ovulatory cycles with CC. Between December 2008 and December 2015 women were allocated to six cycles with gonadotrophins plus IUI, six cycles with gonadotrophins plus intercourse, six cycles with CC plus IUI or six cycles with CC plus intercourse. The primary outcome was conception leading to a live birth achieved within 8 months of randomization.

Participants/materials, setting, methods

We performed a cost-effectiveness analysis on direct medical costs. We calculated the direct medical costs of ovulation induction with gonadotrophins versus CC and of IUI versus intercourse in six subsequent cycles. We included costs of medication, cycle monitoring, interventions, and pregnancy leading to live birth. Resource use was collected from the case report forms and unit costs were derived from various sources. We calculated incremental cost-effectiveness ratios (ICER) for gonadotrophins compared to CC and for IUI compared to intercourse. We used non-parametric bootstrap resampling to investigate the effect of uncertainty in our estimates. The analysis was performed according to the intention-to-treat principle.

Main results and the role of chance

We allocated 666 women in total to gonadotrophins and IUI (*n* = 166), gonadotrophins and intercourse (*n* = 165), CC and IUI (*n* = 163), or CC and intercourse (*n* = 172). Mean direct medical costs per woman receiving gonadotrophins or CC were €4495 versus €3006 (cost difference of €1475 (95% CI: €1457–€1493)). Live birth rates were 52% in women allocated to gonadotrophins and 41% in those allocated to CC (relative risk (RR) 1.24:95% CI: 1.05–1.46). The ICER was €15 258 (95% CI: €8721 to €63 654) per additional live birth with gonadotrophins. Mean direct medical costs per woman allocated to IUI or intercourse were €4497 versus €3005 (cost difference of €1510 (95% CI: €1492–€1529)). Live birth rates were 49% in women allocated to IUI and 43% in those allocated to intercourse (RR = 1.14:95% CI: 0.97–1.35). The ICER was €24 361 (95% CI: €–11 290 to €85 172) per additional live birth with IUI.

Limitations, reasons for caution

We allowed participating hospitals to use their local protocols for ovulation induction and IUI, which may have led to variation in costs, but which increases generalizability. Indirect costs generated by transportation or productivity loss were not included. We did not evaluate letrozole, which is potentially more effective than CC.

Wider implications of the findings

Gonadotrophins are more effective, but more expensive than CC, therefore, the use of gonadotrophins in women with normogonadotropic anovulation who have not conceived after six ovulatory CC cycles depends on society's willingness to pay for an additional child. In view of the uncertainty around the cost-effectiveness estimate of IUI, these data are not sufficient to make recommendations on the use of IUI in these women. In countries where ovulation induction regimens are reimbursed, policy makers and health care professionals may use our results in their guidelines.

Study funding/competing interest(s)

This trial was funded by the Netherlands Organization for Health Research and Development (ZonMw number: 80-82310-97-12067). The Eudract number for this trial is 2008-006171-73. The Sponsor's Protocol Code Number is P08-40. CBLA reports unrestricted grant support from Merck and Ferring. BWM is supported by a NHMRC Practitioner Fellowship (GNT1082548) and reports consultancy for Merck, ObsEva and Guerbet.

Trial registration number NTR1449.

INTRODUCTION

In women with normogonadotropic anovulation who wish to conceive, clomiphene citrate (CC) has long been used as first line treatment for ovulation induction.[1-4] Women not conceiving after six ovulatory cycles are defined as having CC failure.[5] In daily practice, these women often switch to ovulation induction with gonadotrophins and intrauterine insemination (IUI) is often initiated instead of relying on regular intercourse.[2]

The evidence for such a policy change has long been lacking. We recently reported the results of the Modified Ovulation Induction (M-ovin) study, a two-by-two factorial multicentre randomised controlled trial (RCT) comparing ovulation induction with gonadotrophins to CC with or without IUI in normogonadotropic anovulatory women with CC failure.[6] In that study, we randomly assigned women to either six cycles with gonadotrophins plus IUI, six cycles with gonadotrophins plus intercourse, six cycles with CC plus IUI or six cycles with CC plus intercourse. The primary outcome was a live birth achieved within 8 months of randomisation. We made two comparisons, one in which gonadotrophins were compared with CC and one in which IUI was compared with intercourse. This trial showed that a switch of treatment to gonadotrophins led to an absolute increase in live birth of 10% over treatment with CC. IUI did not lead to an increase in live births compared with intercourse. In view of limited health care resources, costs are also important in deciding which treatment should be advised to patients. In contrast to CC, which is relatively cheap due to the low price of the tablets and limited monitoring requirements, ovulation induction with gonadotrophins is expensive due to the price of medication and the need for strict ultrasound monitoring.[7-10] Knowledge on the relative cost and effectiveness of these interventions with or without IUI is lacking. The aim of this study was to provide an economic evaluation of ovulation induction with gonadotrophins compared to CC with or without IUI in women with CC failure.

MATERIALS AND METHODS

Study design

This economic evaluation was performed alongside the M-ovin study, a two-by-two factorial RCT in 48 Dutch hospitals that compared ovulation induction with gonadotrophins with CC with or without IUI in normogonadotropic anovulatory women with CC failure. Details about the study design, sample size calculation, study procedures and outcomes have been described previously.[11,6] Ethical approval was obtained by the Medical Ethical Committee of the Medical Spectrum Twente Enschede (Netherlands) and from the Central Committee on Research involving Human Subjects (CCMO, Netherlands). The board of directors of each of the participating centres approved local execution of the study.

In short, sub-fertile women of at least 18 years of age with normogonadotropic anovulation who had been ovulatory for six cycles on CC, but who had not conceived, were eligible for the trial. Couples with male subfertility and double sided tubal pathology could not participate. Women were randomly assigned using a central password protected internetbased randomisation programme. The randomisation list had been prepared by an independent statistician with a variable block size and a maximum block size of 8. There was no masking. Consenting women were randomly allocated to any of four treatments on a 1:1:1:1 basis, i.e. six cycles of gonadotrophins plus IUI, six cycles of gonadotrophins plus intercourse, six cycles of CC plus IUI or six cycles of CC plus intercourse. We used a two-bytwo factorial design to compare two pairs of interventions: a switch to ovulation induction with gonadotrophins versus continuing CC and IUI versus intercourse.

Ovulation induction, cycle monitoring, semen preparation and insemination were performed according to local hospital protocols. The starting dose of gonadotrophins was 50 or 75 IU daily and participating clinics used either urinary or recombinant gonadotrophins depending on their local protocol. Follicular growth was monitored by transvaginal ultrasound. We used 5000 IU of human chorionic gonadotrophin (hCG) to trigger ovulation. The dosage of CC was a minimum of 50 mg to a maximum of 150 mg daily, for five days. If ovulation did not occur, the dosage was increased with steps of 50 mg with a maximum of 150 mg daily in the next cycles. Women undergoing ovulation induction with CC plus IUI underwent monitoring by ultrasound, women undergoing CC plus intercourse were usually monitored by basal body temperature curve, mid luteal progesterone measurement or urinary luteal hormone surge depending on the local protocol. In the case of IUI, a single insemination per cycle was performed.

The primary outcome measure was conception leading to a live birth within eight months after randomisation. A live birth was defined as any baby that was born alive after a gestational age beyond 24 weeks. Secondary outcomes included multiple pregnancy rate, ongoing pregnancy rate, miscarriage, and ectopic pregnancy.

Economic evaluation

The economic evaluation was performed as a cost-effectiveness analysis from a health care perspective, thus focusing on direct medical costs during treatment.

Resource use

Data on resource use were collected from the individual case report forms of the RCT. For each woman, we registered the medication, cycle monitoring (number of ultrasounds), and interventions (cycles with IUI, cycles with IVF) they received within six subsequent cycles or until a live birth occurred within a time horizon of 8 months. If women changed their treatment to IVF/ICSI, resource use was estimated on the basis of previously published

data on resource costs for IVF/ICSI.[12] Within the M-ovin study, 21 women switched to treatment with IVF or ICSI during their study period i.e. before finishing their allocated treatment (8 women who were allocated to FSH+IUI, 4 women who were allocated to FSH, 7 women who were allocated to CC+IUI, 3 women who were allocated to CC). Because of the intention to treat principle that we have used, the pregnancies resulting from treatment with IVF/ICSI were included in the main analysis of our RCT.

Unit costs

Direct unit costs included the costs of medication, cycle monitoring, interventions, and the costs of pregnancy leading to live birth. The costs for medication and the unit costs of cycle monitoring and interventions were obtained from the costs as retrieved by an expert panel on cost-effectiveness from the Dutch Consortium for Research in Women's Health. The expert panel, consisting of gynecologists, economists and a methodologist, collected the actual total medical costs per cost unit from resources that are being used in fertility studies within our Consortium from two university hospitals and one general hospital. For our final calculation we used the average costs of the three Dutch hospitals.

We derived costs for pregnancy and delivery from a cost analysis of singleton versus twin pregnancies, in which the costs for a singleton and twin pregnancies up until 6 weeks after delivery was described.[13] The costs of a miscarriage with or without curettage, ectopic pregnancy and stillbirth were obtained from the pricelist of one general hospital. All costs were expressed in 2017 euros (€) and corrected for inflation or deflation whenever necessary using the consumer pricing index.[14]

Statistical analysis

For each of the four treatments we calculated the mean costs and effectiveness on the basis of the intention-to-treat principle. For effectiveness we calculated absolute risks, relative risks and corresponding 95% boundaries. Costs were calculated by multiplying the quantity of resource use and unit costs. For each treatment we calculated the mean cost per woman. For costs we calculated mean cost differences and 95% boundaries as estimated on the basis of bootstrapping by taking 1000 random samples. Costs were combined with effectiveness by calculating Incremental Cost-Effectiveness Ratios (ICER) for gonadotrophins compared with CC and for IUI compared with intercourse. The ICER was defined as the ratio between the differences in costs and the differences in effects between two interventions. We used a non-parametric bootstrap resampling to investigate the effect of uncertainty in our estimates. The uncertainty was visualized by plotting a cost-effectiveness plane. CC and intercourse were the reference strategies (in the origin of the cost-effectiveness plane).

We drew a cost-effectiveness acceptability curve, expressing the probability that a strategy will be cost-effective at a specific willingness-to-pay for an additional child, given the uncertainty. The range was from 0 to 135 000 euros.

In view of the factorial design, we investigated the interaction between IUI and ovulation induction with costs. We first evaluated if factors have a multiplicative effect and used a general linear model in transformed cost data.

Per protocol and sensitivity analyses

We did a per-protocol analysis in which we included women who were actually treated according to the predefined protocol.[11] We performed four one way sensitivity analyses to explore the impact of key factors in the cost-effectiveness analyses. In the first analysis we excluded IVF cycles (Model 1), in the second we used ongoing pregnancy as main measure of effectiveness (Model 2), in the third we calculated with unit costs used in the United Kingdom which were collected from a NHS hospital (Model 3), in the fourth we assumed that all CC-cycles were monitored by ultrasound (Model 4) and in the fifth that none of the CC-cycles were monitored by ultrasound (Model 5). All statistical analyses were performed using SPSS (version 23.0; IBM Corp., USA) and Microsoft Excel (version 2016) for the bootstrapping.

RESULTS

Study population and eff ectiveness outcomes

Between December 2008 and December 2015, we randomised 666 women: 166 women were allocated to ovulation induction with gonadotrophins combined with IUI, 165 to ovulation induction with gonadotrophins, 163 to ovulation induction with CC combined with IUI, and 172 to continued ovulation induction with CC. Five women were excluded since they had been erroneously randomised. The baseline characteristics of the participating women can be found in appendix 1.

Effectiveness outcomes are summarized in Table 1. Live birth rates were 52% after gonadotrophins versus 41% after CC, RR 1.24 (95% CI 1.05-1.46); absolute difference 10.2% (95% CI 2.4–17.9). Live birth rates were 49% after IUI versus 43% after intercourse, RR 1.14 (95% CI 0.97-1.35); absolute difference 6.1% (95% CI –1.71 to 13.8). There was no interaction between CC or gonadotrophins and presence of IUI on live birth (p=0.0124). Multiple pregnancy rates were low and did not differ significantly for both comparisons. The mean time to pregnancy was 0.5 months shorter after ovulation induction with gonadotrophins compared to ovulation induction with CC (log rank p=0.028) whereas the mean time to pregnancy was the same when comparing IUI with intercourse (log rank p=0.27).

Table 1. Primary and secondary outcomes

	Gonado-trophins + IUI n = 164	Gonado-trophins n = 163	CC + IUI n = 163	CC n = 171	Gonado-trophins vs CC Rate difference RR (95% CI)	IUI vs inter-course Rate difference RR (95% CI)
Live birth	89 (54.3)	78 (47.9)	72 (44.2)	66 (38.6)	1.24 (1.05-1.46)	1.14 (0.97-1.35)
Ongoing pregnancy	90 (54.9)	80 (49.1)	72 (44.2)	66 (38.6)	1.26 (1.07-1.48)	1.14 (0.97-1.34)
Multiple pregnancy	4 (2.4)	3 (1.8)	7 (4.3)	1 (0.6)	0.89 (0.33-2.40)	2.8 (0.90-8.70)
Miscarriages*	15 (9.1)	9 (5.5)	8 (4.9)	3 (1.8)	-	-
Ectopic pregnancy*	1 (0.6)	1 (0.6)	3 (1.8)	1 (0.6)	-	-
Stillbirth*	1 (0.6)	2 (1.2)	0 (0.0)	0 (0.0)	-	-

Data are n (%) unless otherwise stated.
All multiple pregnancies were twin pregnancies and live births.
* Secondary outcomes.

Economic evaluation

Resource use and unit costs

The mean resource use per woman is summarized in Table 2. The number of ultrasounds were higher in the women who received gonadotrophins, which resulted in more hospital visits. Women who received CC were also monitored with basal body temperature curve, mid luteal progesterone measurement or urinary LH surge, which resulted in less monitory ultrasounds and therefore less hospital visits compared to gonadotrophins. Women allocated to gonadotrophins with or without IUI and CC plus IUI received a HCG-trigger. No HCG-trigger was given to the women allocated to CC plus intercourse. Unit costs are listed in Table 3.

Table 2. Resource use per woman*

	Gonadotrophins Gonadotrophins + IUI		CC + IUI	CC
Cycle monitoring/Intervention				
- Ultrasound (N)	15.87 (10.53)	16.67 (10.66)	12.69 (7.72)	8.31 (5.98)
- IUI (N)	3.22 (2.26)	0.15 (0.71)	3.57 (2.30)	0.12 (0.55)
- IVF (N)	0.04 (0.20)	0.02 (0.16)	0.06 (0.36)	0.03 (0.25)
Medication				
- CC(50mg)	0.18 (1.27)	0.48 (2.91)	28.37 (22.74)	26.16 (21.02)
- FSH (75 IU)	36.00 (32.76)	39.94 (37.29)	2.87 (12.35)	4.84 (14.15)
- HCG (5000 IU)	3.27 (2.32)	3.42 (2.27)	3.69 (2.32)	0.49 (1.26)

Data are mean (SD).

Table 3. Unit costs

Cost item	Unit	Unit costs (Euros)	Reference
Cycle monitoring/Interventions			
- Ultrasound	1	62.50	Dutch Consortium*
- IUI	1	320.54	Dutch Consortium*
- IVF	1	1365.84	Dutch Consortium*
Medication			
- CC	50mg	0.53	Dutch Consortium*
- FSH	75 IU	24.75	Dutch Consortium*
- HCG	5000 IU	5.83	Dutch Consortium*
Pregnancy and delivery			
- Singleton	1	3107.00	Lukassen *et al* 2004
- Twin	1	16 419.00	Lukassen *et al* 2004
- Miscarriage	1	1494.76	One general hospital
- Ectopic pregnancy	1	4295.65	One general hospital
- Stillbirth	1	3107.00	One general hospital

Unit costs are based on Dutch price levels in 2017.
*Costs are derived from the expert panel Dutch Consortium for Research in Women's Health.

Costs

The mean costs per woman eight months after randomisation were €4984 for gonadotrophins plus IUI, €4003 for gonadotrophins plus intercourse, €4006 for CC plus IUI for €2045 with CC plus intercourse (Fig 1A).

For the comparison gonadotrophins versus CC we found mean costs per woman of €4495 with gonadotrophins and €3007 with CC (cost difference was €1475 (95% CI €1457 to €1493)) (Fig 1B). For the comparison IUI versus intercourse we found mean costs per woman of €4497 with IUI and €3005 with intercourse (cost difference was €1510 (95% CI €1492 to €1529)) (Fig 1C).

Cost-effectiveness

The ICER for ovulation induction with gonadotrophins compared with ovulation induction with CC was €15 258 (95% CI €8721 to €63 654) reflecting the additional costs necessary to achieve one additional live birth in women treated with gonadotrophins compared with CC. The majority of the bootstrap samples were located in the northeastern quadrant, reflecting higher costs with higher effectiveness for gonadotrophins versus CC (Fig. 2).

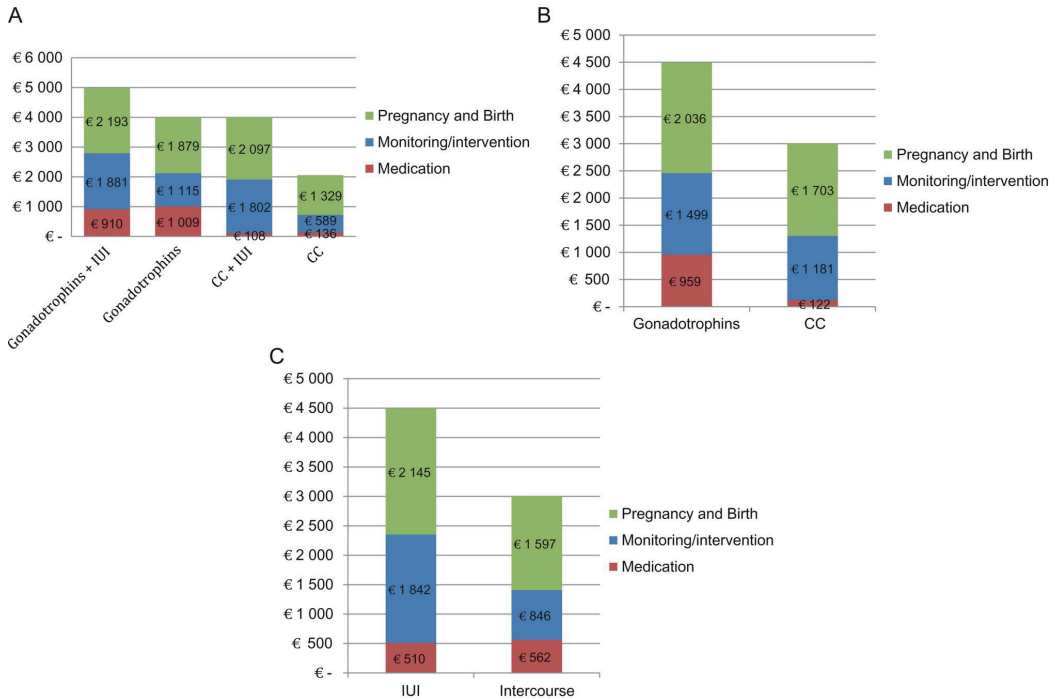


Figure 1. Mean costs per woman

A. Mean costs per woman for gonadotrophins plus IUI, gonadotrophins plus intercourse, CC plus IUI and CC plus intercourse split into the mean costs of medication, cycle monitoring/interventions (number of ultrasounds, cycles with IUI, use of IVF) and pregnancy leading to live birth. All costs are expressed in euros.

B. Mean costs per woman for the comparison gonadotrophins versus CC split into the mean costs of medication, cycle monitoring/interventions (number of ultrasounds, cycles with IUI, use of IVF) and pregnancy leading to live birth. All costs are expressed in euros.

C. Mean costs per woman for the comparison IUI versus intercourse split into the mean costs of medication, cycle monitoring/interventions (number of ultrasounds, cycles with IUI, use of IVF) and pregnancy leading to live birth. All costs are expressed in euros.

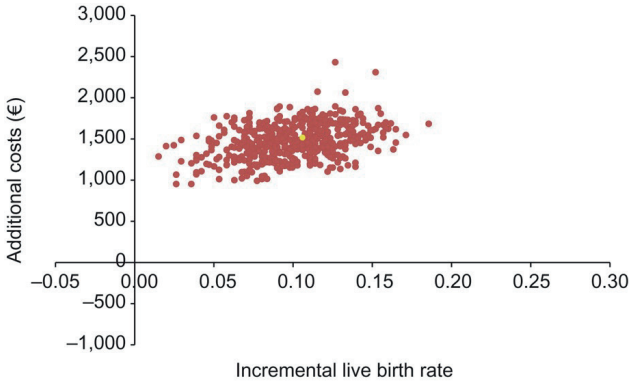


Figure 2. Cost-effectiveness plane gonadotrophins compared with CC

Cost-effectiveness plane: gonadotrophins versus CC. Each point in the cost-effectiveness plane represents the uncertainty of the additional costs and effect of gonadotrophins compared with CC after nonparametric bootstrap resampling (1000 random samples). The light grey dot in the middle represents the cost-effectiveness rate.

The ICER for IUI compared with intercourse was €24 361 (95% CI €-11 290 to €85 172) reflecting the additional costs necessary to achieve one additional live birth in the IUI group, compared with intercourse. The majority of the bootstrap samples were located in the north eastern quadrant (95%), reflecting higher costs with comparable effectiveness for IUI versus intercourse (Fig. 3).

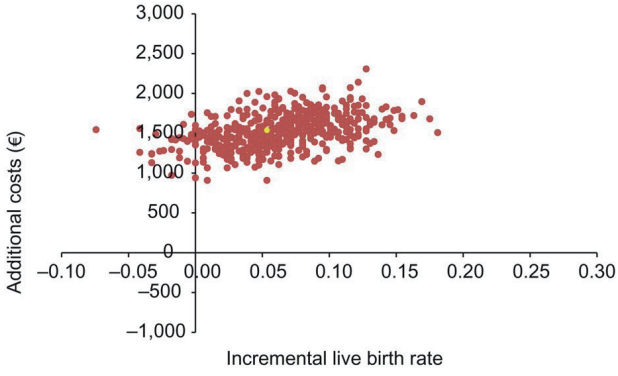


Figure 3. Cost-effectiveness plane IUI compared with intercourse

Cost-effectiveness plane: IUI versus intercourse. Each point in the cost-effectiveness plane represents the uncertainty of the additional costs and effect of IUI compared with intercourse after nonparametric bootstrap resampling (1000 random samples). The light grey dot in the middle represents the cost-effectiveness rate.

For both comparisons we drew a cost-effectiveness acceptability curve (appendix 2). For a willingness-to-pay of €15 000 for an additional live birth, there is 51% chance that gonadotrophins is cost-effective compared with CC and this was 96% for a willingness to pay of €30 000. For a willingness-to-pay of €15 000 for an additional live birth, there is 15% chance that IUI is cost-effective compared with intercourse and this was 61% for a willingness to pay of €30 000.

Costs increased as more and more complex interventions were ordered, i.e. from CC, gonadotrophins, CC plus IUI, to gonadotrophins plus IUI. This implies costs were additive. The general linear model analysis did not indicate presence of interaction between IUI and ovulation induction on costs (p=0.62).

Per protocol and sensitivity analyses

Of the 666 women, 566 women were treated according to protocol and were included in the analysis. We noted more livebirths after gonadotrophins compared with CC, 125 (46%) of 274 women after gonadotrophins versus 95 (33%) of 292 women after CC (RR 1.39 (95% CI 1.10 – 1.57) absolute difference 13%). We found mean costs per woman of €4550 with gonadotrophins and €2596 with CC (cost difference was €2056 (95% CI €2040 - €2072)). The ICER for ovulation induction with gonadotrophins compared with ovulation induction with CC was €15 582 (95% CI €10 013 – €37 323) which is higher compared to the intention-totreat ICER.

Addition of IUI did not significantly increase livebirths compared with intercourse: 118 (42%) of 281 women had a livebirth after IUI versus 102 (36%) of 285 women after intercourse (RR 1.14 (95% CI 0.96–1.36) absolute difference 6%). We found mean costs per woman of €4282 with IUI and €2578 with intercourse. The cost difference was €1586 (95% CI €1568 - €1604). The ICER for IUI compared with intercourse was €25 628 (95% CI €-11 870 – €72 340) which is higher compared to the intention-to-treat ICER.

For the comparison of gonadotrophins versus CC the results of the sensitivity analyses are shown in Table IV a. If we excluded IVF cycles (Model 1), the ICER was €15 426. When ongoing pregnancy was the main measure of effectiveness (Model 2) the ICER was €11 157. Calculating with unit costs of the United Kingdom (Model 3) resulted in a ICER was £19 744. If all CC-cycles were 100% monitored by ultrasound (Model 4) the ICER would lower to €13 460 and if none of the CC-cycles were monitored by ultrasound (Model 5) the ICER would increase to €17 222.

Table 4. One way sensitivity analyses in Euro

Model	Description	Mean cost gonadotrophins (SD)	Mean cost CC (SD)	Difference (95% CI#)	ICER (95% CI#)
0	Base case	4536 (2501)	2996 (2735)	1475 (1457 to 1493)	15 258 (8721 to 63 654)
1	Excluded IVF	4504 (4504)	3020 (2791)	1507 (1490 to 1525)	15 426 (8852 to 64 210)
2	Endpoint ongoing pregnancy	2495 (1858)	1356 (1283)	1190 (1180 to 1201)	11 157 (5567 to 43 736)
3	Costs UK*	5410 (3033)	3429 (2824)	1918 (1898 to 1938)	19 744 (11 036 to 86 114)
4	All CC cycles monitored with ultrasound	4609 (2699)	3195 (2586)	1311 (1293 to 1329)	13 460 (7592 to 55 704)
5	CC cycles not monitored with ultrasound	4496 (3109)	2662 (2830)	1677 (1659 to 1695)	17 222 (9923 to 72 383)

A. Gonadotrophins compared with CC.
Model 0: Base case, live birth as effectiveness outcome, Model 1: Excluded all IVF cycles; effectiveness outcome live birth remained fixed, Model 2: The costs of pregnancy and birth were excluded (costs for miscarriage and ectopic are still included), effectiveness outcome was changed to ongoing pregnancy, Model 3: Effectiveness outcome live birth remained fixed, and costs from a UK (NHS) were used as input, Model 4: All CC cycles monitored with ultrasound; effectiveness outcome live birth remained fixed, Model 5: None of the CC cycles are monitored with ultrasound but with basal body temperature curve, mid luteal progesterone measurement or urinary LH surge. #Non-parametric confidence interval based on 1000 bootstrap replications.
* Costs UK are in pounds.

For the comparison of IUI versus intercourse the results of the sensitivity analyses are shown in Table IV b. If we excluded IVF cycles (Model 1), the ICER was €23 786. When ongoing pregnancy was the main measure of effectiveness (Model 2) the ICER was €17 531.

Calculating with unit costs of the United Kingdom (Model 3) resulted in a ICER of £34 420.

DISCUSSION

We performed an economic evaluation alongside a two-by-two factorial multicentre RCT comparing ovulation induction with gonadotrophins with CC, and IUI with intercourse in women with normogonadotropic anovulation and CC failure. Women allocated to gonadotrophins had significantly more live births than those allocated to CC, but at higher costs. These higher costs were generated by more ultrasound monitoring and higher costs of medication in the gonadotrophin group. The additional cost necessary to achieve one additional live birth was €15,258 (95% CI €8721 to €63,654).

Women allocated to IUI did not have significantly more live births than those allocated to intercourse. The costs were significantly higher for women assigned to IUI compared with intercourse. The additional cost necessary to achieve one additional live birth was €24,361 (95% CI €-11.290 to €85.172). The wide confidence interval, crossing unity, implicates a large degree of uncertainty around the cost-effectiveness.

The present study has several strengths. First, we designed the study to assess live birth rates which is the most important outcome from the patient's perspective. Second, this economic evaluation was based on a randomised study with prospective registration of resource use. We incorporated all interventions and associated costs that took place in eight months, closely reflecting daily practice. Third, by performing several sensitivity analyses, we showed that our outcomes were robust making the results applicable to other hospitals. Finally, in the per-protocol analysis and the four sensitivity analyses CC and intercourse remained less costly, indicating that our results are robust when varying several treatment details.

A weakness of our study is that we allowed participating hospitals to use their local protocols for ovulation induction and IUI, which resulted in heterogeneous data on cycle monitoring and that we did not take into account indirect costs generated by transportation or productivity loss.

Our finding that continuing CC is less costly than switching to gonadotrophins matches the results of a cost-effectiveness study in women with PCOS using fictional treatment scenarios.[9] In that study, continuing CC for another six cycles followed by six or twelve cycles with gonadotrophins, followed by IVF was more cost-effective than a direct switch to gonadotrophins followed by IVF. The cost-effectiveness of IUI was not included in that study.

Several recent studies have shown that first line treatment with the aromatase inhibitor Letrozole is associated with higher live birth rates than with CC as was summarized in

a network meta-analysis.[15] Letrozole tablets are only slightly more expensive than CC tablets.[16] A cost-effectiveness analysis comparing Letrozole with gonadotrophins in women with CC failure could result in a smaller cost difference than with gonadotrophins, but this needs to demonstrated before conclusions are drawn.

Since our cost-effectiveness analysis used a health care perspective, we focused on direct medical costs during treatment. From a societal perspective, indirect costs generated by transportation or productivity loss can also contribute to the costs of the ovulation induction treatments. Treatment with gonadotrophins plus IUI leads to more visits to the clinic in view of cycle monitoring and interventions and would thus result in more indirect costs. As a consequence, including societal costs would enlarge the cost difference between gonadotrophins and CC, and IUI and intercourse. On the other hand, due to the higher live birth rates after gonadotrophins, fewer cycles would need to be performed. Thus, over the treatment period of eight months, this potential difference in costs may disappear.

The unit costs of the interventions vary between countries. Country-specific prices and assumptions need to be considered before generalizing these results to other countries. When using prices from a NHS teaching hospital in the United Kingdom, we found that the mean costs were higher for both gonadotrophins and IUI, leading to more costs per additional live birth for gonadotrophins compared with CC and for IUI compared with intercourse. In countries where the unit costs are higher, such as the United States, it is likely that gonadotrophins and IUI will be even more expensive.

Cost-effectiveness of interventions have to be known, but are -in themselves- not decisive in finalizing the optimal treatment policy. Decisive is the 'willingness to pay' i.e. the monetary value that society is willing to pay for higher live birth rates, but the problem is that there is no consensus on the level of costs per extra live birth that is acceptable. The NICE Fertility Guideline suggests a threshold of £30.000 per quality adjusted life year (QALY), but also highlights that QALYs cannot be derived from live births arising from assisted reproduction as QALYs are intended to capture improvements in health among patients and not in creating life. Patient preference studies in subfertile women reveal that couples are willing to pay €100-€500 extra to increase pregnancy rates by a few percent.[17,18]

In conclusion, in women with normogonadotropic anovulation who have not conceived after six ovulatory CC cycles, gonadotrophins are more effective, but generate higher costs compared to CC. In countries where ovulation induction regimens are reimbursed, policy makers and health care professionals may use our results in their guidelines. Importantly, apart from the costs, couples must be counseled that CC is known to cause more side

effects than gonadotrophins, whereas gonadotrophins require daily injections combined with ultrasound monitoring of follicular development.[10]

In view of the uncertainty around the cost-effectiveness estimate of IUI, we cannot make recommendations on the use of IUI in these women and more data are needed.

Acknowledgements
We thank all couples that participated in the trial, the hospitals and their staff, the research nurses and the staff of the Dutch Consortium of Healthcare Evaluation and Research in Obstetrics and Gynaecology for logistic support.

Funding
The Netherlands Organization for Health Research and Development (ZonMw number: 80-82310-97-12067). The Eudract number for this trial is 2008-006171-73. The Sponsor's Protocol Code Number is P08-40.

Conflict of interest
C.B.L.A. reports unrestricted grant support from Merck and Ferring. B.W.M. is supported by a NHMRC Practitioner Fellowship (GNT1082548) and reports consultancy for Merck, ObsEva and Guerbet.

SUPPLEMENTAL DATA

Appendix 1. Baseline characteristics of the participating couples

	Gonadotrophins + IUI n = 164	**Gonadotrophins + intercourse n = 163**	**CC + IUI n = 163**	**CC + intercourse n = 171**
Age of women (years)	29·5 ± 3·7	29·9 ± 3·7	30·0 ± 3·6	29·9 ± 4·0
Ethnicity				
White	131 (85%)	134 (88%)	133 (86%)	141 (89%)
Non-white	24 (15%)	18 (12%)	21 (14%)	18 (11%)
BMI (kg/m^2)*	25·4 ± 5·1	25·6 ± 5·6	25·0 ± 4·9	25·4 ± 5·0
BMI >25.0 kg/m^2	76 (46%)	81 (49%)	64 (39%)	81 (47%)
Current smoker	29 (18%)	20 (12%)	22 (13%)	22 (13%)
Diabetes	1	1	3	2
Previous livebirth	32 (20%)	35 (21%)	36 (22%)	34 (20%)
Duration of subfertility (months)	26·3 ± 14·9	24·5 ± 12·5	24·5 ± 15·5	25·9 ± 19·0
Cycle pattern prior to treatment #				
Amenorrhea	124 (76%)	125 (77%)	115 (71%)	120 (70%)
Oligomenorrhea	21 (13%)	25 (15%)	27 (16%)	32 (19%)
Unknown	19 (11%)	13 (8%)	21 (13%)	19 (11%)
Median TMC *10^6	52 (20-106)	43 (16-113)	53 (15-132)	38 (16-99)
Polycystic ovaries on ultrasound ##	110 (67%)	103 (63%)	109 (67%)	117 (68%)
Mean serum biochemical values		5·7 ± 1·7		
FSH (IU/L)	5·7 ± 2·1	10·6 ± 7·8	6·2 ± 2·2	6·0 ± 2·2
LH (IU/L)	9·7 ± 7·4	239 ± 217	10·6 ± 7·6	10·9 ± 10·8
Estrogen (pmol/L)	255 ± 295	1·6 ± 2·0	201 ± 159	271 ± 460
Total testosterone (nmol/L)	1·6 ± 1·7		1·8 ± 2·2	1·8 ± 1·8

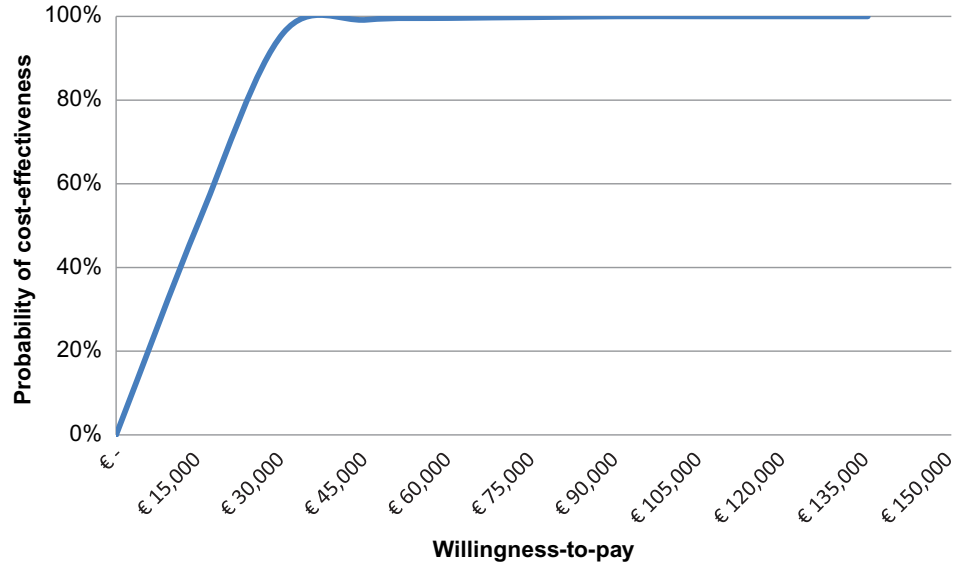
Data are mean (SD), n (%) or median (IQR). BMI = body-mass index. TMC = total motile sperm count. FSH = follicle stimulating hormone. LH = luteinizing hormone. CC = clomiphene citrate. IUI = intrauterine insemination.

*BMI was missing for 24 women; data were imputed by using multiple imputation.

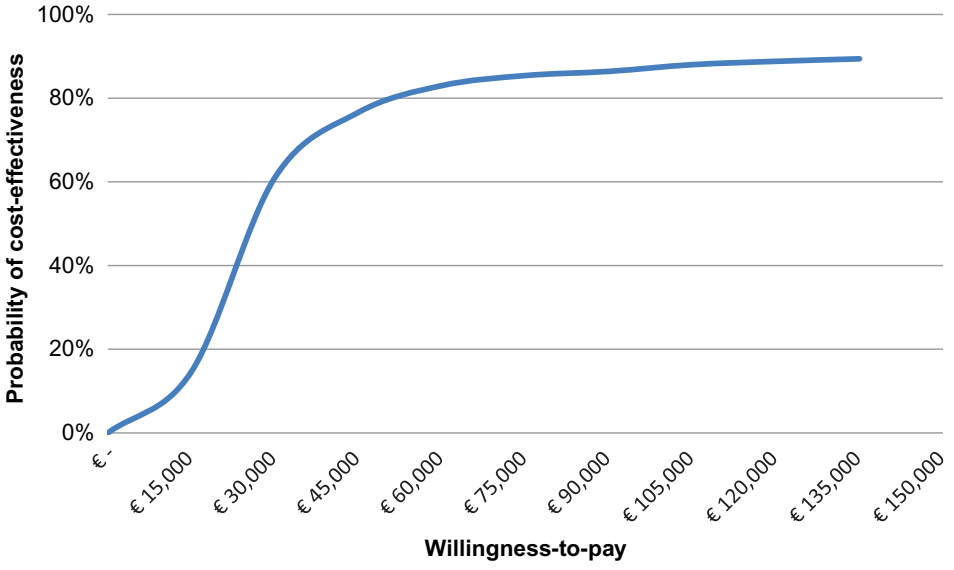
Amenorrhea: absence of menstrual bleeding for >6 months. Oligomenorrhea: irregular menstrual bleedings with intervals of >35 days but ≤6 months

Defined as the presence of 12 or more follicles in each ovary measuring 2–9 mm in diameter

Appendix 2. Supplementary Figure S1 Cost-effectiveness acceptability curve gonadotrophins compared with CC.



Appendix 3. Supplementary Figure S2 Cost-effectiveness acceptability curve IUI compared with intercourse.



REFERENCE LIST

1. Norman RJ, Dewailly D, Legro RS, Hickey TE. Polycystic ovary syndrome. *Lancet* 2007; 370(9588): 685-97.
2. Thessaloniki EA-SPCWG. Consensus on infertility treatment related to polycystic ovary syndrome. *Hum Reprod* 2008; 23(3): 462-77.
3. NICE. Fertility: Assessment and Treatment for People with Fertility Problems. 2013.
4. Balen AH, Morley LC, Misso M, et al. The management of anovulatory infertility in women with polycystic ovary syndrome: an analysis of the evidence to support the development of global WHO guidance. *Hum Reprod Update* 2016; 22(6): 687-708.
5. Veltman-Verhulst SM, Fauser BC, Eijkemans MJ. High singleton live birth rate confirmed after ovulation induction in women with anovulatory polycystic ovary syndrome: validation of a prediction model for clinical practice. *Fertil Steril* 2012; 98(3): 761-8 e1.
6. Weiss NS, Nahuis MJ, Bordewijk E, et al. Gonadotrophins versus clomifene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomifene failure (M-OVIN): a randomised, two-by-two factorial trial. *Lancet (London, England)* 2018; 391(10122): 758-65.
7. Homburg R, Hendriks ML, Konig TE, et al. Clomifene citrate or low-dose FSH for the first-line treatment of infertile women with anovulation associated with polycystic ovary syndrome: a prospective randomized multinational study. *Hum Reprod* 2012; 27(2): 468-73.
8. Practice Committee of the American Society for Reproductive M. Use of clomiphene citrate in infertile women: a committee opinion. *Fertil Steril* 2013; 100(2): 341-8.
9. Moolenaar LM, Nahuis MJ, Hompes PG, van der Veen F, Mol BW. Cost-effectiveness of treatment strategies in women with PCOS who do not conceive after six cycles of clomiphene citrate. *Reprod Biomed Online* 2014; 28(5): 606-13.
10. Legro RS. Ovulation induction in polycystic ovary syndrome: Current options. *Best Pract Res Clin Obstet Gynaecol* 2016; 37: 152-9.
11. Nahuis MJ, Weiss NS, van der Veen F, et al. The M-OVIN study: does switching treatment to FSH and / or IUI lead to higher pregnancy rates in a subset of women with world health organization type II anovulation not conceiving after six ovulatory cycles with clomiphene citrate - a randomised controlled trial. *BMC Womens Health* 2013; 13: 42.
12. van Tilborg TC, Oudshoorn SC, Eijkemans MJC, et al. Individualized FSH dosing based on ovarian reserve testing in women starting IVF/ICSI: a multicentre trial and cost-effectiveness analysis. *Hum Reprod* 2017; 32(12): 2485-95.
13. Lukassen HG, Schonbeck Y, Adang EM, Braat DD, Zielhuis GA, Kremer JA. Cost analysis of singleton versus twin pregnancies after in vitro fertilization. *Fertil Steril* 2004; 81(5): 1240-6.
14. Statistics N. Statline Consumers Pricing Index. 2017. http://statline.cbs.nl.
15. Wang R, Kim BV, van Wely M, et al. Treatment strategies for women with WHO group II anovulation: systematic review and network meta-analysis. *BMJ* 2017; 356: j138.
16. Pharmacotherapeutic Compass. 2017. www.farmacotherapeutischkompas.nl - 2017.

17. Palumbo A, De La Fuente P, Rodriguez M, et al. Willingness to pay and conjoint analysis to determine women's preferences for ovarian stimulating hormones in the treatment of infertility in Spain. *Hum Reprod* 2011; 26(7): 1790-8.
18. Weiss NS, Schreurs AMF, van der Veen F, et al. Women's perspectives on ovulation induction with or without IUI as treatment for normogonadotropic anovulation; A discrete choice experiment. *Hum Reprod Open* 2017*;* issue 3.

Also on behalf of:
Smeenk JMJ, Hoek A, Broekmans FJM, Fleischer K, de Bruin JP, Kaaijk EM, Laven JSE, Hendriks DJ, Gerards MH, Bourdrez P, Gianotten J, Koks C, Bayram N, van Hooff M, Boks DES, Perquin DAM, Janssen CAH, van Golde RJT, Nap AW, van Rijn-van Weert JM, Vollebergh JHA, Klijn NF, Rijnsaardt-Lukassen HGM, Sluijmer AV

Chapter 4

Gonadotrophins or clomiphene citrate in women with normogonadotropic anovulation and CC failure: does the endometrium matter?

Bordewijk EM, Weiss NS, Nahuis MJ, Kwee J, Lambeek AF, van Unnik GA, Vrouenraets FPJ, Cohlen BJ, van de Laar-van Asseldonk TAM, Lambalk CB, Goddijn M, Hompes PG, van der Veen F, Mol BWJ, van Wely M; M-ovin study group.

Human Reproduction. 2020 Jun

ABSTRACT

Study question

Is endometrial thickness (EMT) a biomarker to select between women who should switch to gonadotropins and those who could continue clomiphene citrate (CC) after six failed ovulatory cycles?

Summary answer

Using a cut-off of 7 mm for EMT, we can distinguish between women who are better off switching to gonadotropins and those who could continue CC after six earlier failed ovulatory CC cycles.

What is already known

For women with normogonadotropic anovulation, CC has been a long-standing first-line treatment in conjunction with intercourse or intrauterine insemination (IUI). We recently showed that a switch to gonadotropins increases the chance of live birth by 11% in these women over continued treatment with CC after six failed ovulatory cycles, at a cost of €15 258 per additional live birth. It is unclear whether EMT can be used to identify women who can continue on CC with similar live birth rates without the extra costs of gonadotropins.

Study design, size, duration

Between 8 December 2008 and 16 December 2015, 666 women with CC failure were randomly assigned to receive an additional six cycles with a change to gonadotropins ($n = 331$) or an additional six cycles continuing with CC ($n = 335$), both in conjunction with intercourse or IUI. The primary outcome was conception leading to live birth within 8 months after randomisation. EMT was measured mid-cycle before randomisation during their sixth ovulatory CC cycle. The EMT was available in 380 women, of whom 190 were allocated to gonadotropins and 190 were allocated to CC.

Participants/materials, setting, methods

EMT was determined in the sixth CC cycle prior to randomisation. We tested for interaction of EMT with the treatment effect using logistic regression. We performed a spline analysis to evaluate the association of EMT with chance to pregnancy leading to a live birth in the next cycles and to determine the best cut-off point. On the basis of the resulting cut-off point, we calculated the relative risk and 95% CI of live birth for gonadotropins versus CC at EMT values below and above this cut-off point. Finally, we calculated incremental cost-effectiveness ratios (ICER).

Main results and the role of chance

Mid-cycle EMT in the sixth cycle interacted with treatment effect ($P < 0.01$). Spline analyses showed a cut-off point of 7 mm. There were 162 women (45%) who had an EMT ≤ 7 mm in the sixth ovulatory cycle and 218 women (55%) who had an EMT > 7 mm. Among the women with EMT ≤ 7 mm, gonadotropins resulted in a live birth in 44 of 79 women (56%), while CC resulted in a live birth in 28 of 83 women (34%) (RR 1.57, 95% CI 1.13–2.19). Per additional live birth with gonadotropins, the ICER was €9709 (95% CI: €5117 to €25 302). Among the women with EMT > 7 mm, gonadotropins resulted in a live birth in 53 of 111 women (48%) while CC resulted in a live birth in 52 of 107 women (49%) (RR 0.98, 95% CI 0.75–1.29).

Limitations, reasons for caution

This was a post hoc analysis of a randomised controlled trial (RCT) and therefore mid-cycle EMT measurements before randomisation during their sixth ovulatory CC cycle were not available for all included women.

Wider implications of the findings

In women with six failed ovulatory cycles on CC and an EMT ≤ 7 mm in the sixth cycle, we advise switching to gonadotropins, since it improves live birth rate over continuing treatment with CC at an extra cost of €9709 to achieve one additional live birth. If the EMT > 7 mm, we advise to continue treatment with CC, since live birth rates are similar to those with gonadotropins, without the extra costs.

Study funding/competing interest(s)

The original MOVIN trial received funding from the Dutch Organization for Health Research and Development (ZonMw number: 80-82310-97-12067). C.B.L.A. reports unrestricted grant support from Merck and Ferring. B.W.M. is supported by a NHMRC Practitioner Fellowship (GNT1082548) and reports consultancy for Merck, ObsEva, IGENOMIX and Guerbet. All other authors have nothing to declare.

Trial registration number

Netherlands Trial Register, number NTR1449

INTRODUCTION

Normogonadotropic anovulation WHO type II is one of the most common conditions in reproductive aged women with a prevalence of 8–13%.[1-3] Clomiphene citrate (CC) has long been used as first-line treatment for ovulation induction in these women. Although ovulation is restored in about 75% of women starting ovulation induction with CC, six cycles of treatment leads to conception in only about half of these women. Women who ovulate on CC but have not conceived after six ovulatory cycles of intercourse or intrauterine insemination (IUI) are traditionally defined as having CC failure. For these women, gonadotropins are second-line pharmacological agents (moderate quality of evidence).[2,4] We recently performed a randomised controlled trial in 666 women with normogonadotropic anovulation and CC failure, in which we compared a switch to gonadotropins with continuing treatment with CC for another six cycles.[5] Switching to gonadotropins resulted in a cumulative live birth rate of 52% and continued ovulation induction with CC in a cumulative live birth rate of 41% (RR 1.24 [95% CI 1.05–1.46]; *P* = 0.0124). There were seven multiple pregnancies in the women treated with gonadotropins (2.1%) and eight in the women treated with CC (2.4%). The additional costs necessary to achieve one additional live birth in women treated with gonadotropins compared with CC was €15 258 (95% CI €8721 to €63 654).[6] It is unclear whether endometrial thickness (EMT) can be used to identify women who can continue on CC with similar live birth rates without the extra costs of gonadotropins. EMT seems a logical candidate biomarker since many studies have found evidence for negative effects of CC on the endometrium.[7,8] The aim of this study was thus to evaluate whether EMT, during the sixth ovulatory cycle of ovulation induction with CC, can be used as biomarker to select between women with normogonadotropic anovulation and CC failure who are better off switching to gonadotropins and those who could continue CC.

MATERIALS AND METHODS

Study design

We conducted a secondary analysis of the M-ovin study, a two-by-two factorial RCT, in 48 Dutch hospitals that compared live birth rates after ovulation induction with gonadotropins or CC with or without IUI in normogonadotropic anovulatory women with CC failure.

Ethical approval

Ethical approval was obtained by the Medical Ethical Committee of the Medical Spectrum Twente Enschede (Netherlands) and from the Central Committee on Research involving Human Subjects (CCMO, Netherlands). The board of directors of each of the participating

centres approved local execution of the study. The original study was registered in the Netherlands Trial Register, number NTR1449.

Procedures

Details about the study design, sample size calculation, study procedures and outcomes have been described previously.[5,9] In summary, subfertile women of at least 18 years of age with normogonadotropic anovulation who had been ovulatory for six cycles on CC, but who had not conceived, were eligible for the trial. Couples with male subfertility and double-sided tubal pathology were not eligible. Consenting women were randomly allocated to six cycles of gonadotropins plus IUI, six cycles of gonadotropins plus intercourse, six cycles of CC plus IUI or six cycles of CC plus intercourse, on a 1:1:1:1 basis. We used a two-by-two factorial design to compare two pairs of interventions: a switch to ovulation induction with gonadotropins versus continuing CC and IUI versus intercourse.

Ovulation induction, cycle monitoring, semen preparation and insemination were performed according to local hospital protocols. The starting dose of gonadotropins was 50 or 75 IU daily, and participating clinics used either urinary or recombinant follicle stimulating hormone depending on their local protocol. Follicular growth was monitored by transvaginal ultrasound. We used 5000 IU of human chorionic gonadotrophin (hCG) to trigger ovulation. The dosage of CC ranged between 50 and 150 mg daily, for 5 days. If ovulation did not occur, the dosage was increased in steps of 50 mg with a maximum of 150 mg daily in the next cycles. Women undergoing ovulation induction with CC plus IUI were monitored by ultrasound, while women undergoing CC plus intercourse were monitored by basal body temperature curve, mid-luteal progesterone measurement or urinary luteal hormone surge depending on the local protocol.

The primary outcome measure was conception leading to live birth within eight months after randomisation. A live birth was defined as any baby born alive after a gestational age of 24 weeks. During the study, the data was collected by research nurses and after the last live birth, we closed the database. A secondary outcome was cost.

Mid-cycle EMT before randomisation

We collected data on mid-cycle EMT measured in the sixth cycle of ovulation induction with CC before randomisation. The ultrasound was planned to be pre-ovulatory according to local protocol. We started collecting this data after an amendment to the protocol. This amendment started after including 286 patients. The data was collected from the individual case report forms of the RCT. The EMT was measured by transvaginal ultrasound.

Statistical analysis

All analyses were performed for the outcome live birth rate. First, we tested for interaction of EMT with the treatment effect using logistic regression. Second, we performed a spline analysis to evaluate the association of EMT with chance of pregnancy leading to a live birth and to determine the best cut-off point. On basis of the resulting cut-off point, we calculated the relative risk and 95% CI of a live birth for gonadotropins versus CC at EMT values below and above this value. Third, we constructed Kaplan–Meier curves for time to conception leading to live birth for gonadotropins versus CC in relation to the cut-off value. The curves were compared with a log-rank test. A *P* value of less than 0.05 was considered to indicate statistical significance. Fourth, we analysed by logistic regression the EMT values over time for cycles 7 until 12 relative to the mid-cycle EMT of Cycle 6 determined in the women who received gonadotropins and the women who received CC. Fifth, we examined whether there was an association between live birth rates and different doses of CC. Sixth, we performed an economic analysis to determine the difference in costs between gonadotropins and CC below and above the cut-off value using previously determined cost data.[6] Costs were combined with effectiveness by calculating Incremental Cost-Effectiveness Ratios (ICER). SPSS software (version 23.0; IBM Corp., USA) was used for statistical analysis. STATA (Version 14.2; Stata Corp) was used for the spline analysis.

RESULTS

Mid-cycle EMT before randomisation

Between 8 December 2008 and 16 December 2015, 666 women had been allocated to receive an additional six cycles with a change to gonadotropins (*N* = 331) or additional 6 cycles continuing with CC (*N* = 335). The mid-cycle EMT of the sixth CC cycle prior to randomisation was available in 380 women (57%), of whom 190 were allocated to gonadotropins and 190 were allocated to CC. The baseline characteristics of the women in whom EMT had been measured were similar and are summarised in Table 1. The values of EMT ranged from 2.0 to 20.4 mm.

Table 1. Baseline characteristics of the women.

	Gonadotrophins ***n*** **= 190**	**CC** ***n*** **= 190**
Age of women (years)	29.8 (3.8)	29.8 (3.8)
Ethnicity		
White	157 (83%)	159 (82%)
Non-white	23 (12%)	21 (11%)
BMI (kg/m^2)	25.2 (5.4)	25.3 (4.7)
BMI > 25.0 kg/m^2	80 (42%)	81 (43%)
Current smoker	24 (13%)	20 (11%)
Diabetes	1 (0.5%)	4 (2.1%)
Previous live birth	42 (22%)	36 (19%)
Duration of subfertility (months)	25.3 (14.2)	24.3 (16.4)
Cycle pattern before treatment†		
Amenorrhoea	141 (74%)	145 (76%)
Oligomenorrhoea	32 (17%)	30 (16%)
Unknown	17 (8.9%)	15 (7.9%)
TMC (×10^6)	98 (158)	89 (127)
Polycystic ovaries on ultrasound‡	130 (68%)	138 (73%)
Mean serum biochemical values		
FSH (IU/L)	5.7 (1.9)	5.9 (2.1)
LH (IU/L)	10.5 (7.4)	10.9 (9.1)
Oestrogen (pmol/L)	259 (295)	229 (241)
Total testosterone (nmol/L)	1.6 (2.2)	1.6 (1.4)
Mid-cycle EMT	8.0 (2.3)	8.0 (2.5)
EMT > 7 MM	111 (58%)	107 (56%)

Data are mean (SD) or *n* (%). BMI = body mass index. TMC = total motile sperm count. FSH = follicle-stimulating hormone. LH = luteinising hormone. †Amenorrhoea: absence of menstrual bleeding for >6 months. Oligomenorrhoea: irregular menstrual bleedings with intervals of >35 days but ≤6 months. ‡Defined as the presence of 12 or more follicles in each ovary measuring 2–9 mm in diameter.

There was an interaction between treatment and EMT on live birth ($P < 0.01$). The spline function clearly visualises the interaction between EMT and treatment on the outcome live birth and points towards a cut-off point at an EMT of 7 mm (Fig. 1). Among 162 women (45%) with EMT ≤ 7 mm, gonadotropins resulted in a live birth in 44 of 79 women (56%) and CC resulted in a live birth in 28 of 83 (34%) (RR 1.57, 95% CI 1.13–2.19). Among 218 women (55%) with EMT > 7 mm, gonadotropins resulted in a live birth in 53 of 111 women (48%) and CC resulted in a live birth in 52 of 107 women (49%) (RR 0.98, 95% CI 0.75–1.29).

For an EMT of ≤7, the mean time to conception leading to a live birth was 4.7 months (95% CI 4.0–5.4) following gonadotropins and 6.0 months (95% CI 5.4–6.6) following CC (log-rank test; $P = 0.008$; Fig. 2). For an EMT of >7, the mean time to conception leading to a live birth was 5.4 months (95% CI 4.9–6.0) following gonadotropins and 5.0 months (95% CI 4.4–5.6) following CC (log-rank test; $P = 0.56$; Fig. 3). The insemination method used, intercourse or IUI did not have any impact on either EMT values or live birth rates.

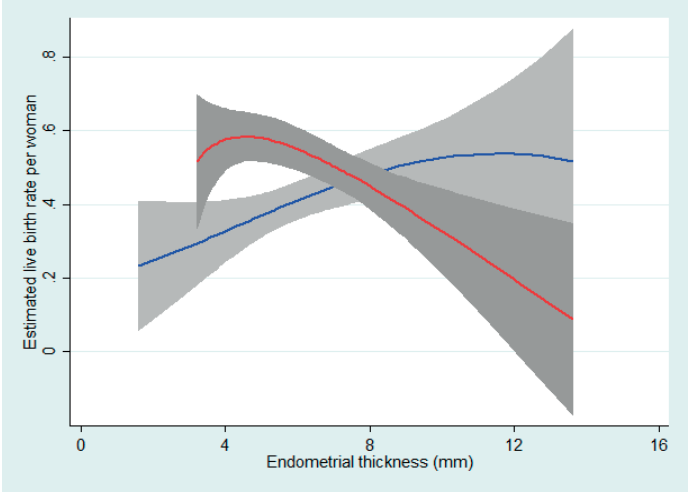


Figure 1. Spline function: interaction between endometrial thickness and treatment (gonadotropins and clomiphene citrate) on live birth.

Red line: gonadotropins and blue line: clomiphene citrate.

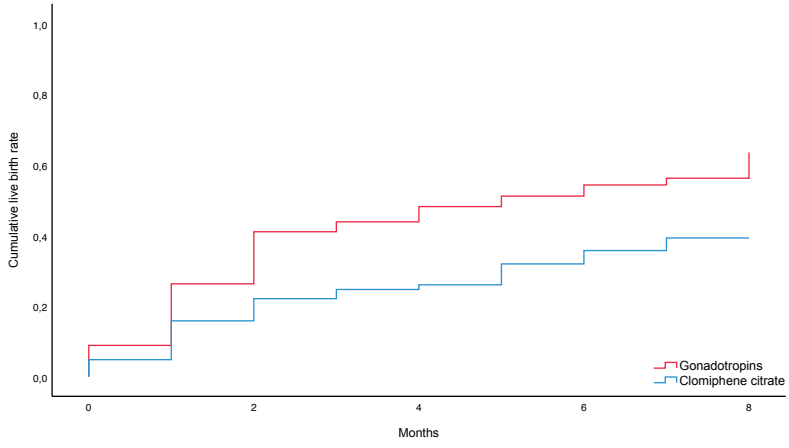


Figure 2. Kaplan–Meijer curve for live birth rate in women with an endometrial thickness ≤ 7.

Time to conception leading to live birth for the comparison gonadotropins versus clomiphene citrate, and endometrial thickness ≤ 7 mm.

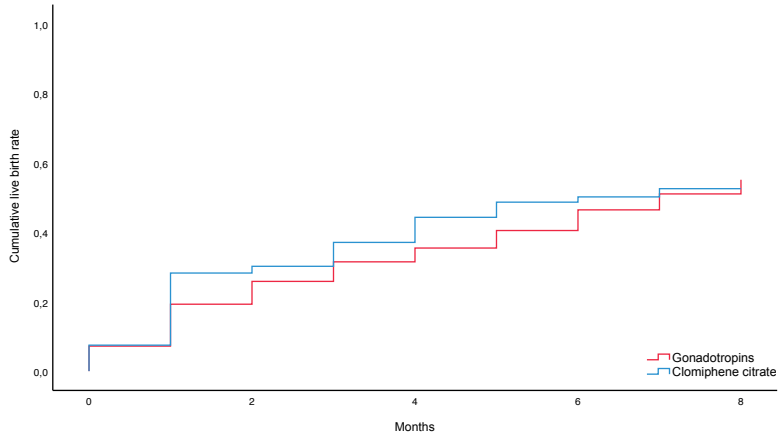


Figure 3. Kaplan–Meijer curve for live birth rate in women with an endometrial thickness > 7 mm. Time to conception leading to live birth for the comparison gonadotropins versus clomiphene citrate and an endometrial thickness EMT > 7 mm.

EMT during the course of treatment

Over time, the EMT increased slightly in the women who received gonadotropins. The EMT remained stable from Cycle 7 to Cycle 12 in the women who received CC (Fig. 4).

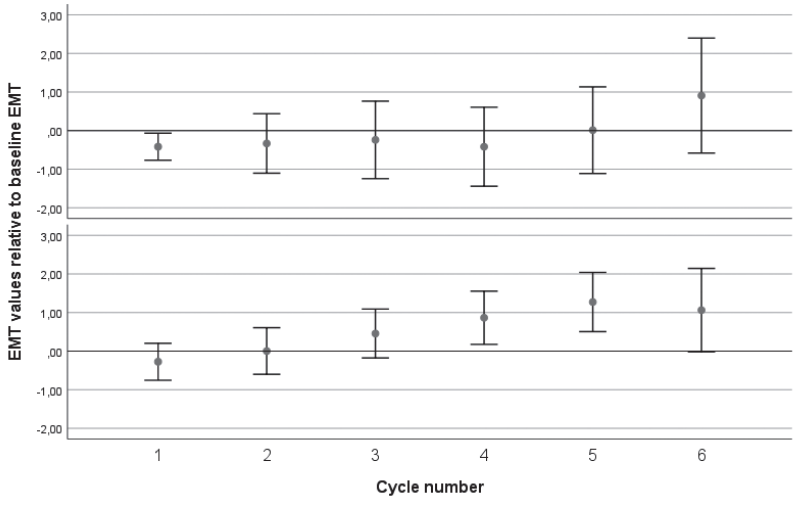


Figure 4. Endometrial thickness values of cycles 7 until 12 relative to the mid-cycle endometrial thickness of Cycle 6 over 6 cycles for gonadotropins and clomiphene citrate. Upper blot gonadotropins and lower blot clomiphene citrate. The dots represent the endometrial thickness mean difference. The lines represent the 95% confidence interval.

Dose of CC

There was no difference in live birth between women who received <100 mg CC (34/88) and women who received ≥100 mg CC (35/85) (RR 0.95, 95% CI 0.70–1.29). There was also no association between the dose of CC and the EMT values (Fig. 5).

Incremental cost-effectiveness

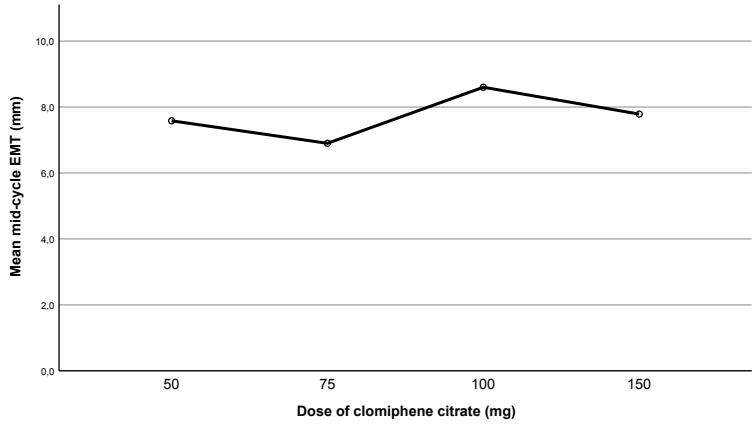


Figure 5. Dose of clomiphene citrate associated with the mean mid-cycle endometrial thickness.

Mean direct medical costs per woman with EMT ≤ 7 mm receiving gonadotropins versus CC were €4873 versus €2778 (cost difference of €2251 (95% CI: €2231–€2272)). The ICER was €9709 (95% CI: €5117 to €25 302) per additional live birth with gonadotropins (Supplementary Fig. S1).

Mean direct medical costs per woman with EMT > 7 mm receiving gonadotropins or CC were €4741 versus €3100 (cost difference of €1463 (95% CI: €1446–€1480)). The ICER estimates from the bootstrap analysis reflected greater costs for gonadotropins for similar effectiveness (Supplementary Fig. S2).

DISCUSSION

In this study, we evaluated whether EMT can be used as a biomarker to select women with normogonadotropic anovulation who should switch to ovulation induction with gonadotropins and those who could continue on CC after six failed ovulatory cycles. A cut-off value of 7 mm was able to do so. In women with an EMT ≤ 7 mm in the sixth cycle, switching to gonadotropins improved live birth rate over continuing treatment with CC, at an extra cost of €9709 to achieve one additional live birth. In women with an EMT > 7

mm in the sixth cycle, continuing with CC produced similar live birth rates without the extra costs of gonadotropins.

A strength of our study is that the EMT measurements were performed in the context of a RCT by many different doctors performing the ultrasound measurements, enhancing both the internal validity and generalisability of the results. We investigated whether we had adequate power to perform this secondary study by calculating the power on basis of logistic regression with a binary independent variable (treatment) and a binary interacting variable (EMT). We had 83% power with our sample size of 380 women and 1:1 distributions of the variables.

A weakness of our study is that mid-cycle EMT measurements before randomisation during the sixth ovulatory CC cycle were not available for 286 of the women (43%) included in the original RCT. When the study started, we did not collect data on EMT in the case record forms, but the participating centres performed EMT measurements during treatment according to their local protocol as part of their routine monitoring. We added the EMT in the case record forms after the trial had included 286 women, and consequently, we only have EMT measurements of 380 women.

We believe this data has clinical implications, provided they are confirmed in future studies. Our original randomised trial showed that a switch of treatment to gonadotropins led to an absolute increase in live birth of 11% over treatment with CC in women with normogonadotropic anovulation and CC failure, while we here show that in women with an EMT ≤ 7 mm gonadotropins it leads to an absolute increase in live birth of 22% over continued treatment with CC. The additional cost necessary to achieve one additional live birth was calculated at €15 258 in the cost-effectiveness analysis of the original randomised trial.[5,6] Consequently, the costs necessary to achieve one additional live birth in women with an EMT ≤ 7 mm are now lower at €9709. In women with an EMT > 7 mm, live birth rates were similar for those treated with gonadotropins and those who continued treatment with CC. Since gonadotropins are more expensive, CC should be the dominant strategy in women with an EMT > 7 mm.

The lower live birth rate with CC in women with an EMT smaller than 7 mm might be explained by the anti-estrogenic effects of CC on the endometrial development/receptivity, cervical mucus and uterine blood flow.[8] All women in this study had already had 6 cycles of CC. We hypothesise that the women with an EMT ≤ 7 mm in the sixth ovulatory cycle are more sensitive to the negative anti-estrogenic effect of CC on the endometrial receptivity, cervical mucus and uterine blood flow, resulting in an endometrium of lesser thickness and lesser receptiveness. The women with an EMT > 7 mm might be less sensitive to the anti-estrogenic effect of CC and are therefore able to produce a thick and good quality

endometrium. The effect of CC appeared to be independent of CC dosage as higher doses had no effect on the measured EMT values, nor on the chance to a conception leading to a live birth. If women with an EMT ≤ 7 mm in the sixth ovulatory cycle switch to gonadotropins, there is no longer an anti-estrogenic of CC on the endometrium. From the spline analysis, no further inferences can be made on conception trends with decreasing or increasing EMT in the CC and FSH groups, due to the insecurity around the effect estimates at the more extreme EMT values. These previously described negative effects of CC may explain in part why first line treatment with CC results in lower live birth rates than the aromatase inhibitor letrozole, as was summarised in an IPD meta-analysis.[10]

In women with normogonadotropic anovulation and six failed ovulatory cycles on CC and an EMT ≤ 7 mm in the sixth cycle, we advise switching to gonadotropins, since it improves live birth rate over continuing treatment with CC, at an extra cost of €9709 to achieve one additional live birth. If the EMT > 7 mm, we advise continuing treatment with CC, since live birth rates are similar to those with gonadotropins, but without the extra costs. This information can be used to update the guideline (https://www.monash.edu/medicine/sphpm/mchri/pcos/guideline).

Acknowledgements

We thank all couples who participated in the trial, the hospitals and their staff, the research nurses and the staff of the Dutch Consortium of Healthcare Evaluation and Research in Obstetrics and Gynaecology for logistic support.

Funding

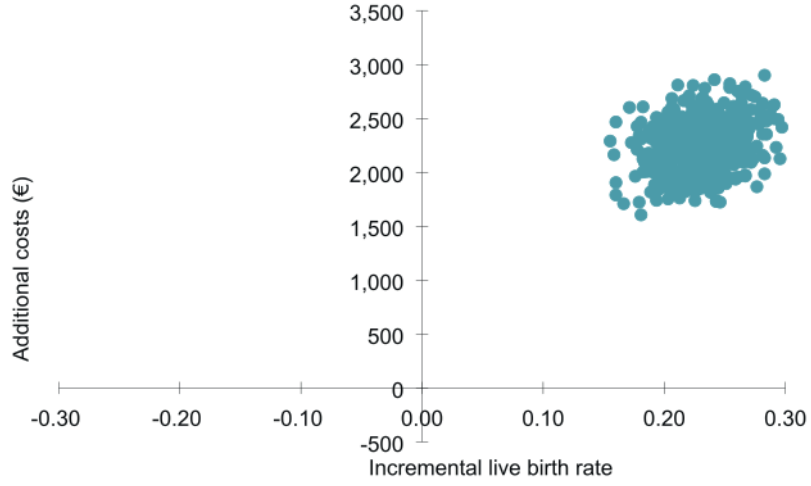
Netherlands Organization for Health Research and Development (ZonMw number: 80-82310-97-12067). The Eudract number for this trial is 2008-006171-73. The Sponsor's Protocol Code Number is P08-40.

Conflict of interest

C.B.L.A. reports unrestricted grant support from Merck and Ferring. B.W.M. is supported by a NHMRC Practitioner Fellowship (GNT1082548) and reports consultancy for Merck, ObsEva, iGENOMIX and Guerbet. All other authors have nothing to declare.

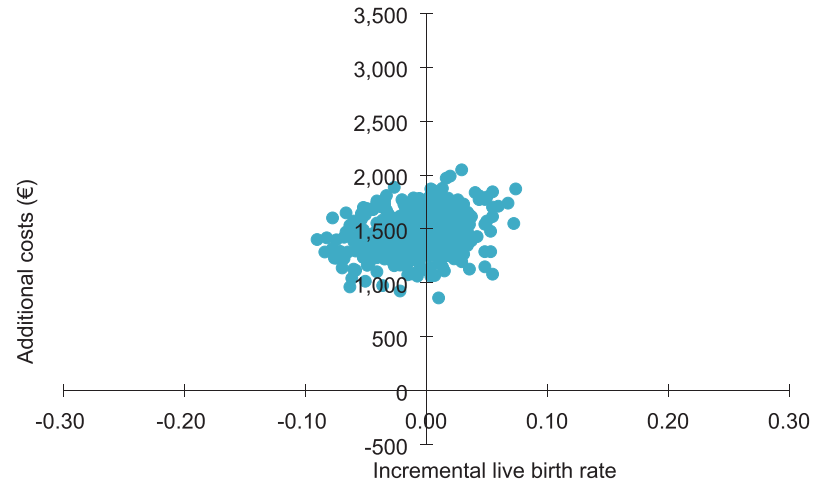
SUPPLEMENTAL DATA

Appendix 1



Supplementary Figure S1 Cost-effectiveness plane of ovulation induction with gonadotropins compared to clomiphene citrate for an endometrial thickness ≤7 mm Cost-effectiveness plane

Appendix 2



Supplementary Figure S2 Cost-effectiveness plane of ovulation induction with gonadotropins compared to clomiphene citrate for an endometrial thickness >7 mm Cost-effectiveness plane

REFERENCE LIST

1. Balen AH, Morley LC, Misso M, Franks S, Legro RS, Wijeyaratne CN, Stener-Victorin E, Fauser BC, Norman RJ, Teede H. The management of anovulatory infertility in women with polycystic ovary syndrome: an analysis of the evidence to support the development of global WHO guidance. *Hum Reprod Update* 2016;22: 687–708.
1. 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.
1. Costello MF, Garad RM, Hart R, Homer H, Johnson L, Jordan C, Mocanu E, Qiao J, Rombauts L, Teede HJ *et al.* A review of second- and third-line infertility treatments and supporting evidence in women with polycystic ovary syndrome. *Med Sci* 2019a;7:1–11.
1. Costello MF, Misso ML, Balen A, Boyle J, Devoto L, Garad RM, Hart R, Johnson L, Jordan C, Legro RS *et al.* Evidence summaries and recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome: assessment and treatment of infertility. *Human Reprod Open* 2019b;2019:1–24.
1. Weiss NS, Nahuis MJ, Bordewijk E, Oosterhuis JE, Smeenk JM, Hoek A, Broekmans FJ, Fleischer K, de Bruin JP, Kaaijk EM *et al.* Gonadotrophins versus clomifene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomifene failure (M-OVIN): a randomised, two-by-two factorial trial. *Lancet* 2018;391:758–765.
1. Bordewijk EM, Weiss NS, Nahuis MJ, Bayram N, van Hooff MHA, Boks. DES, Perquin DAM, Janssen CAH, van Golde RJT, Lambalk CB *et al.* Gonadotrophins versus clomiphene citrate with or without IUI in women with normogonadotropic anovulation and clomiphene failure: a cost-effectiveness analysis. *Hum Reprod* 2019;34: 276–284.
1. Weiss NS, van Vliet MN, Limpens J, Hompes PGA, Lambalk CB, Mochtar MH, van der Veen F, Mol BWJ, van Wely M. Endometrial thickness in women undergoing IUI with ovarian stimulation. How thick is too thin? A systematic review and meta-analysis. *Hum Reprod* 2017;32:1009–1018.
1. Gadalla MA, Huang S, Wang R, Norman RJ, Abdullah SA, El Saman AM, Ismail AM, van Wely M, Mol BWJ. Effect of clomiphene citrate on endometrial thickness, ovulation, pregnancy and live birth in anovulatory women: systematic review and meta-analysis. *Ultrasound Obstet Gynecol* 2018;51:64–76.
1. Nahuis MJ, Weiss NS, van der Veen F, Mol BW, Hompes PG, Oosterhuis J, Lambalk NB, Smeenk JM, Koks CA, van Golde RJ *et al.* The M-OVIN study: does switching treatment to FSH and / or IUI lead to higher pregnancy rates in a subset of women with world health organization type II anovulation not conceiving after six ovulatory cycles with clomiphene citrate a randomised controlled trial. *BMC Womens Health* 2013;13:42.
1. Wang R, Li W, Bordewijk EM, Legro RS, Zhang H, Wu X, Gao J, Morin-Papunen L, Homburg R, Konig TE *et al.* First-line ovulation induction for polycystic ovary syndrome: an individual participant data meta-analysis. *Hum Reprod Update* 2019;25:717–732.

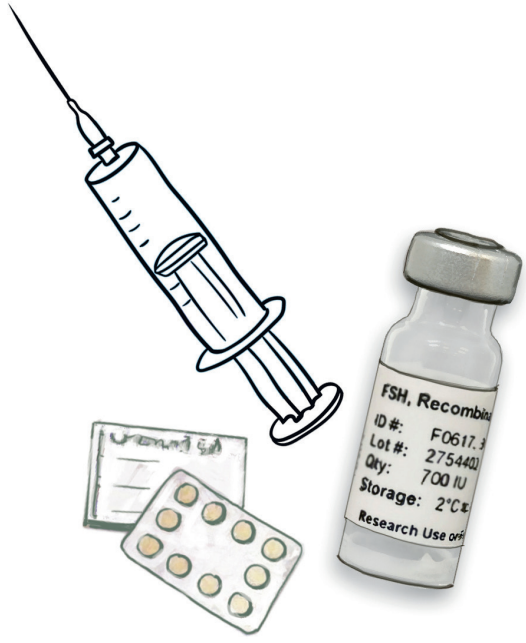
Also on behalf of:
M.J. Nahuis, J.M.J. Smeenk, F.J.M. Broekmans, K. Fleischer, J.P. de Bruin, E.M. Kaaijk, J.S.E. Laven, D.J. Hendriks, M.H. Gerards, P. Bourdrez, J. Gianotten, C. Koks, N. Bayram, M. van Hooff, D.E.S Boks, D.A.M. Perquin, C.A.H. Janssen, R.J.T. van Golde, J. Kwee, A.F. Lambeek, G.A. van Unnik, F.P.J. Vrouenraets, B.J. Cohlen, A.W. Nap, J.M. van Rijn-van Weert, J.H.A. Vollebergh, N.F. Klijn, H.G.M. Rijnsaardt-Lukassen, H.R. Verhoeve, E.A. Brinkhuis, T.K. Schukken, T.E.M. Verhagen, G.J.E. Oosterhuis, N.E.A. Vogel, D.A. Hoozemans, I.A.J van Rooij, C.B. Lambalk, M. Goddijn, P.G.A. Hompes.

Chapter 5

Long-term outcomes of using gonadotrophins versus clomiphene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomiphene failure: follow-up study of a factorial randomised clinical trial.

Bordewijk EM, Jannink TI, Weiss NS, de Vries T, Hoek A, Goddijn M, Mol BWJ, van Wely M; M-ovin study group.

Accepted for publication in Human reproduction 2022 Oct



ABSTRACT

Study question

What are the long-term outcomes after the use of gonadotrophins versus clomiphene citrate (CC) with or without intrauterine insemination (IUI) in women with normogonadotropic anovulation and clomiphene failure?

Summary answer

About four in five women with normogonadotropic anovulation and CC failure had a live birth, with no evidence of a difference in pregnancy outcomes between the allocated groups.

What is already known

CC has long been used as first line treatment for ovulation induction in women with normogonadotropic anovulation. Between 2009 and 2015, a two-by-two factorial multicentre RCT in 666 women with normogonadotropic anovulation and 6 cycles CC failure was performed (M-ovin trial). This study compared a switch to gonadotrophins with continued treatment with CC for another six cycles, with or without IUI within 8 months. Switching to gonadotrophins increased the chance of conception leading to live birth by 11% over continued treatment with CC after six failed ovulatory cycles, at a cost of €15 258 per additional live birth. The addition of IUI did not significantly increase live birth rates.

Study design, size, duration

In order to investigate the long-term outcomes of switching to gonadotrophins versus continuing treatment with CC and IUI versus intercourse we conducted a follow-up study. The study population comprised all women who participated in the M-ovin trial.

Participants/materials, setting, methods

The participating women were asked to complete a web-based questionnaire. The primary outcome of this study was cumulative live birth. Secondary outcomes included clinical pregnancies, multiple pregnancies, miscarriage, stillbirth, ectopic pregnancy, fertility treatments, neonatal outcomes and pregnancy complications.

Main results and the role of chance

We approached 564 women (85%), of whom 374 (66%) responded (184 allocated to gonadotrophins; 190 to CC). After a median follow-up time of 8 years, 154 women in the gonadotrophin group had a live birth (83.7%) versus 150 women in the CC group (78.9%) (RR 1.06, 95% CI 0.96 – 1.17). A second live birth occurred in 85 of 184 women (46.2%) in the gonadotrophin group and in 85 of 190 women (44.7%) in the CC group

(RR 1.03, 95% CI 0.83 – 1.29). Women allocated to gonadotrophins had a third live birth in 6 of 184 women (3.3%) and women allocated to CC had a third live birth in 14 of 190 women (7.4%). There were respectively 12 and 11 twins in the gonadotrophin and CC groups. The use of fertility treatments in the follow-up period was comparable between both groups. The addition of IUI resulted in a first live birth in 156 of 192 women (81.3%) and intercourse resulted in a live birth in 144 of 182 women (79.1%) (RR: 1.03 95% CI 0.93 - 1.13).

Limitations, reasons for caution:

We have complete follow-up results for 57% of the women. 185 women did not respond to the questionnaire, while 102 women had not been approached, due to missing contact details.

Wider implications of the findings

Women with normogonadotropic anovulation and CC failure have a high chance to reach at least one live birth. In terms of pregnancy rates, the long-term differences between switching to gonadotrophins are small compared to continued treatment with CC.

Study funding/competing interests

The original study received funding from the Dutch Organization for Health Research and Development (ZonMw number: 80-82310-97-12067).

Trial registration number

This follow-up study was registered in the OSF Register, https://osf.io/pf24m. The original M-ovin trial was registered in the Netherlands Trial Register, number NTR1449.

INTRODUCTION

Most women failing to achieve a pregnancy due to anovulation have WHO type II normogonadotropic anovulation. With a prevalence of 8 to 13% it is a common hormonal condition in women of reproductive age and most of these women have polycystic ovary syndrome.[1]

In these women, ovulation-induction with clomiphene citrate (CC) or letrozole is the first-line treatment. While CC has been used for decades, letrozole - used off-label - has recently shown to be superior over CC. In 60-85% of the women ovulation induction restores ovulation. Women on CC who have not conceived after six ovulatory cycles are defined as having CC failure. In these women, gonadotrophins could be used as second-line ovulation-induction treatment (moderate quality of evidence).[1-3]

Between 2009 and 2015, we conducted a two-by-two factorial multicentre randomised clinical trial (RCT),[4,5] in which 666 women with normogonadotropic anovulation and CC failure were included. In this RCT, a switch to gonadotrophins was compared with continued treatment with CC for another six cycles, while addition of IUI was also evaluated. The cumulative live birth rate was 52% after switching to gonadotrophins and in 41% after continuing ovulation induction with CC (RR 1.24; 95% CI 1.05-1.4, P = 0.012). The addition of IUI did not significantly increase the live birth rate (RR 1.14; 95% CI 0.97-1.35, P = 0.1152). The cost-effectiveness analyses showed that the extra costs of having one additional live birth in women treated with gonadotrophins compared with CC were €15 258 (95% CI €8721 to €63 654).[6] The treatment effect was specifically present in women with a thin endometrium at the initial CC treatment.[7]

The original trial included six ovulation induction cycles within 8 months. More knowledge on subsequent treatment decisions and success rates in such a well-mapped population, is relevant for patients, gynaecologists, fertility doctors and policymakers. Patients would like more information on their chances over time to fulfil their child-wish, doctors would like to inform their patients better and policymakers need such information to make a realistic profile for budget impact analysis.

The aim of the present study was to investigate the long-term reproductive outcomes in terms of a long-term cumulative chance for delivering at least one live birth in women who originally were allocated to gonadotrophins or CC.

MATERIALS AND METHODS

Study design and participants

This study is a follow-up study of the M-ovin trial, a two-by-two factorial RCT in 48 Dutch hospitals that compared live birth rates after ovulation induction with gonadotrophins or CC with or without IUI in normogonadotropic anovulatory women with CC failure. Between December 2009 and December 2015, a total of 666 women were included. Subfertile women of at least 18 years of age with normogonadotropic anovulation who had been ovulatory for six cycles on CC, but who had not conceived, were eligible for the trial. Couples with severe male subfertility or double-sided tubal pathology were not eligible.

After written informed consent, women were randomly allocated to six cycles of gonadotrophins plus IUI, six cycles of gonadotrophins plus intercourse, six cycles of CC plus IUI or six cycles of CC plus intercourse on a 1:1:1:1 basis. We used a two-by-two factorial design to compare two pairs of interventions: a switch to ovulation induction with gonadotrophins versus continuing CC and IUI versus intercourse. The primary outcome measure was conception leading to live birth within eight months after randomisation. A live birth was defined as any baby that was born alive after a gestational age beyond 24 weeks. During the study the data was collected by research nurses and after the last live birth we closed the database. We performed a cost-effectiveness analysis alongside the study. Further details about the study design, sample size calculation, study procedures and outcomes have been described previously.[4-7]

All previously included women, of whom we had contact details of, were asked by e-mail to participate in this follow-up study. They were all asked for informed consent and they received a digital questionnaire. The first contact was made by the principal investigators or representatives of the centres where the women were included. Women who did not respond were sent a second e-mail, followed by telephone contact. Collection of the follow up data occurred between 02-12-2020 and 18-03-2022.

Questionnaire

The web-based questionnaire included topics about pregnancies, fertility treatments and neonatal outcomes. We collected data on all pregnancies that occurred within the follow-up period, including live birth, multiple pregnancies, miscarriages, stillbirth and ectopic pregnancies. We also asked whether these pregnancies had occurred by natural conception, with or without ovulation induction (CC and FSH, with or without IUI), by any form of ART (IVF, ICSI, frozen embryo transfer), by ovarian drilling or by possible other methods.

Outcomes

The primary outcome of this study was cumulative first live birth (defined as any baby that was born alive after 24 weeks amenorrhea). Secondary outcomes were second live birth, third or more live births, clinical pregnancies, multiple pregnancies, miscarriage (all pregnancy losses until 20 weeks of gestation), stillbirth (all pregnancy losses after 20 weeks of gestation), ectopic pregnancy (defined as a pregnancy in which implantation takes place outside the uterine cavity), neonatal outcomes (such as fetal birthweight), and pregnancy complications. We used the consensus definitions as established by the *Core Outcome Measure for Infertility Trials* (COMMIT) initiative.[8]

Data handing

The data on fertility treatments, pregnancies, miscarriages and neonatal outcomes were retrieved via the web-based questionnaire device LimeSurvey (Version 2.6.7). This data was collected in the LimeSurvey web-based case record form and later transferred to a SPSS file. The data on pregnancy complications was collected from the PRN database and directly stored in a SPSS file. Data handling was performed with a coded set, with the participant code only available to members of the study group and research nurse of the participating hospitals.

Statistical analysis

We compared the outcomes after follow-up according to the randomisation groups, i.e. gonadotrophins versus CC and IUI versus intercourse.

Cumulative live birth was expressed as a relative ratio (RR) and risk difference (RD). Analyses for first, second and third live birth were performed for the women participating in the follow-up study. To account for loss-to follow we calculated a hazard rate (with corresponding 95% confidence intervals (CI) while including all women participating in the original trial and plotted Kaplan–Meier curves to visualise live birth rate over time. Analyses of first live birth was performed in the ITT population. Analyses for, second and third live birth were performed only in the women participating in the follow-up study.

Conceptions, pregnancy losses, twin pregnancies, miscarriages, still birth and ectopic pregnancies were presented descriptively in a flow chart OR with 95% CI.

For a subgroup analysis we calculated the relative risk and 95% CI of live birth for gonadotropins versus CC at EMT values below and above an EMT of 7mm.

To assess whether non-response bias may have affected results, we compared responders and non-responder baseline and live birth outcomes within the original RCT.

The analyses were performed using SPSS software (version 26.0; IBM Corp., USA).

Ethical considerations

The M-ovin trial was registered in the Netherlands Trial Register, number NTR1449, and approved by the Medical Ethical Committee of the Medical Spectrum Twente Enschede (Netherlands) and from the Central Committee on Research involving Human Subjects (CCMO, Netherlands). The board of directors of the participating centres approved local execution of the study. Informed consent was obtained and included permission to investigate long-term outcomes. For this follow-up study, the Medical Ethical Committee of the Amsterdam UMC, location AMC, approved and provided a non-WMO statement on October 22, 2020 (MEC no. 20.508). This study was registered in the OSF Register, https://osf.io/pf24m.

RESULTS

Between Dec 8, 2008 and Dec 16, 2015, 666 women had been randomised to receive an additional six cycles with a change to gonadotrophins (N=331) or additional six cycles continuing with CC (N=335). During the trial, five women did not start treatment such that a total of 661 women were eligible for this follow-up study. Between December 2020 and March 2022, we approached 564 women (85%), of whom 374 (66%) responded (184 allocated to gonadotrophins; 190 to CC) and completed the follow-up questionnaire. We had no follow-up data for 287 women (43%): 185 (28%) women were contacted but did not return the questionnaire while 102 (15%) women were not approached due to missing contact details (n=95) or non-participation of a local hospital (n=7).

Median follow-up time was 98 months (min 75 months; max 154 months) in the women allocated to gonadotrophins and 99 months (min 75 months; max 159 months) in the women allocated to CC. The mean age at follow-up was 37.7 (SD 4.0) in the gonadotrophin arm and 38.0 (SD 3.8) in the CC arm. Other baseline characteristics are described in Table 1 and Table 2.

The flow chart with first cumulative live birth number is expressed in Figure 1A. Second and third cumulative births are expressed in Figure 1B.

Table 1. Baseline characteristics of the participating couples: gonadotrophins vs CC

	Gonadotrophins (N=184)	CC (N=190)
Age of women (years)	29.6 (3.6)	29.8 (3.6)
Ethnicity		
White	164 (89%)	170 (89%)
Non-white	12 (7%)	13 (7%)
BMI (kg/m2)	25.5 (5.3)	24.9 (4.6)
BMI>25.0 kg/m2	87 (47%)	74 (39%)
Current smoker	27 (15%)	22 (12%)
Diabetes	1	3
Previous livebirth	33 (18%)	36 (19%)
Duration of subfertility (months)	23.1 (11.5)	24.3 (17.4)
Cycle pattern before treatment†		
Amenorrhoea	30 (16%)	34 (18%)
Oligomenorrhoea	138 (75%)	133 (70%)
Unknown	16 (9%)	23 (12%)
TMC ($\times 10^6$)	109 (177)	81 (110)
Polycystic ovaries on ultrasound‡	122 (66%)	137 (72%)

Data are mean (SD) or n (%). BMI =body mass index. TMC= total motile sperm count. †Amenorrhoea: absence of menstrual bleeding for >6 months. Oligomenorrhoea: irregular menstrual bleedings with intervals of >35 days but ≤6 months. ‡Defined as the presence of 12 or more follicles in each ovary measuring 2–9 mm in diameter.

Table 2. Baseline characteristics of the participating couples: IUI vs intercourse

	IUI (N=192)	Intercourse (N=182)
Age of women (years)	29.6 (3.7)	29.9 (3.6)
Ethnicity		
White	171 (89%)	163 (90%)
Non-white	13 (7%)	12 (7%)
BMI (kg/m2)	24.9 (5.0)	25.5 (5.0)
BMI>25.0 kg/m2	74 (39%)	87 (48%)
Current smoker	28 (15%)	21 (12%)
Diabetes	1	3
Previous livebirth	33 (17%)	32 (18%)
Duration of subfertility (months)	22.9 (11.4)	24.6 (17.6)
Cycle pattern before treatment†		
Amenorrhoea	32 (17%)	32 (18%)
Oligomenorrhoea	140 (73%)	131 (72%)
Unknown	20 (10%)	19 (10%)
TMC ($\times 10^6$)	111 (171)	77 (115)
Polycystic ovaries on ultrasound‡	132 (69%)	127 (70%)

Data are mean (SD) or n (%). BMI =body mass index. TMC= total motile sperm count. †Amenorrhoea: absence of menstrual bleeding for >6 months. Oligomenorrhoea: irregular menstrual bleedings with intervals of >35 days but ≤6 months. ‡Defined as the presence of 12 or more follicles in each ovary measuring 2–9 mm in diameter.

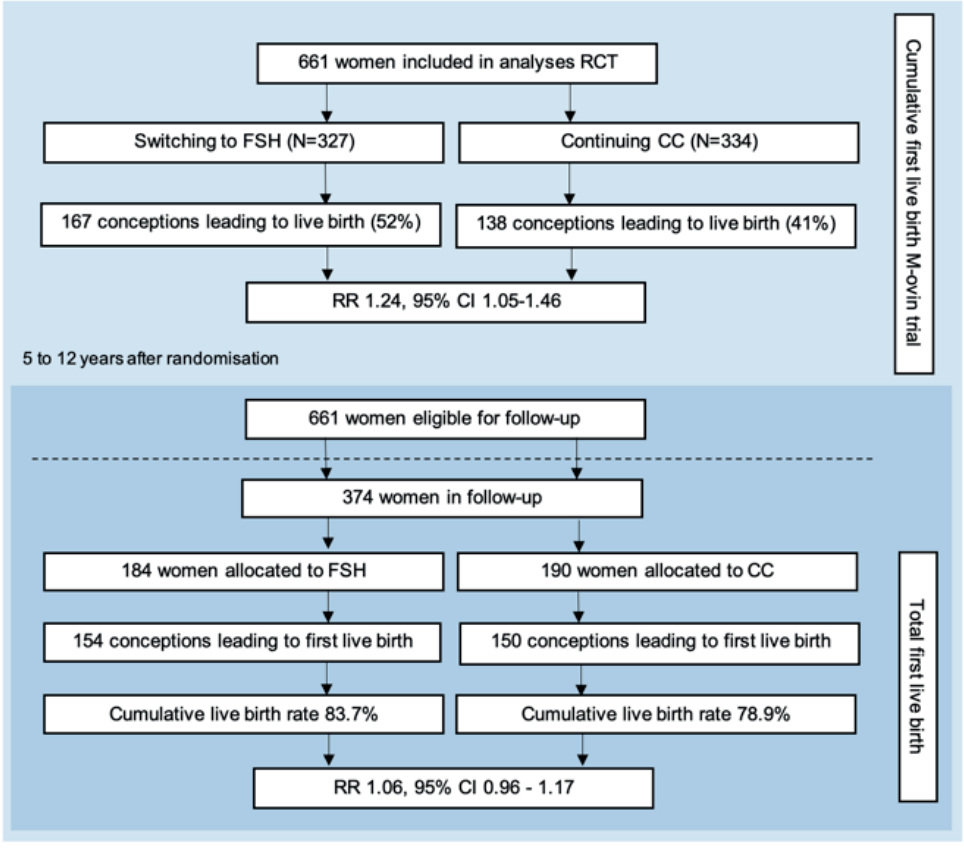


Figure 1A. Flowchart of outcomes of the first pregnancy leading to a live birth

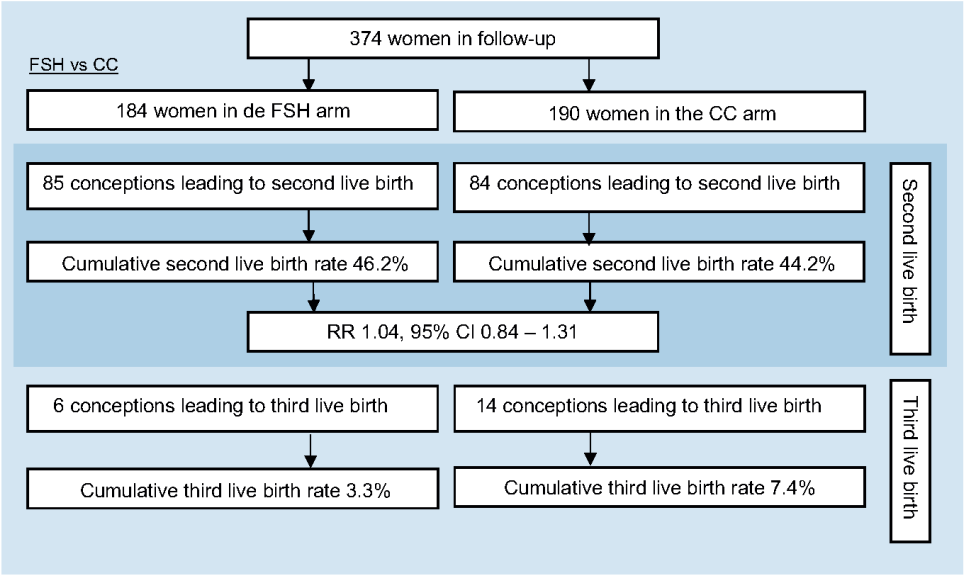


Figure 1B. Flowchart of outcomes of the second and third pregnancy leading to a live birth

First cumulative live birth

Among the 374 women, treatment with gonadotrophins resulted in a live birth in 154 of 184 women (83.7%) and treatment with CC resulted in a live birth in 150 of 190 women (78.9%) (RR: 1.06, 95% CI 0.96 – 1.17; RD 4.75%, 95% CI -3.13 – 12.63). Time to conception leading to a live birth is depicted in Figure 2. Including all 661 women, the hazard rate for live birth was 1.20 (95% CI 0.99 – 1.44). For the 374 women who participated in the follow-up period, the hazard rate for live birth was 1.14 (95% CI 0.91 – 1.43) Figure 2A visualised live birth over time for the complete population, the median time to conception leading to a first live birth was 6.7 months (95% CI 5.08 – 8.26) following gonadotrophins and 10.2 months (95% CI 7.76 – 12.62) following CC (log rank p=0.058). Figure 2B visualised live birth over time for the follow up population, the median time to conception leading to a first live birth was 8.09 months (95% CI 4.61 – 11.47) following gonadotrophins and 10.68 months (95% CI 7.59 – 13.76) following CC (log rank p=0.24).

We found no interaction between insemination method (IUI or intercourse) and treatment (gonadotrophins or CC) on live birth (p=0.91).

Table 3 shows all fertility treatments that resulted in a first live birth. In the gonadotrophin group 11 of the 154 live births (7.1%) were conceived by natural conception and in the CC group 15 of the 150 (10.0%) were conceived by natural conception.

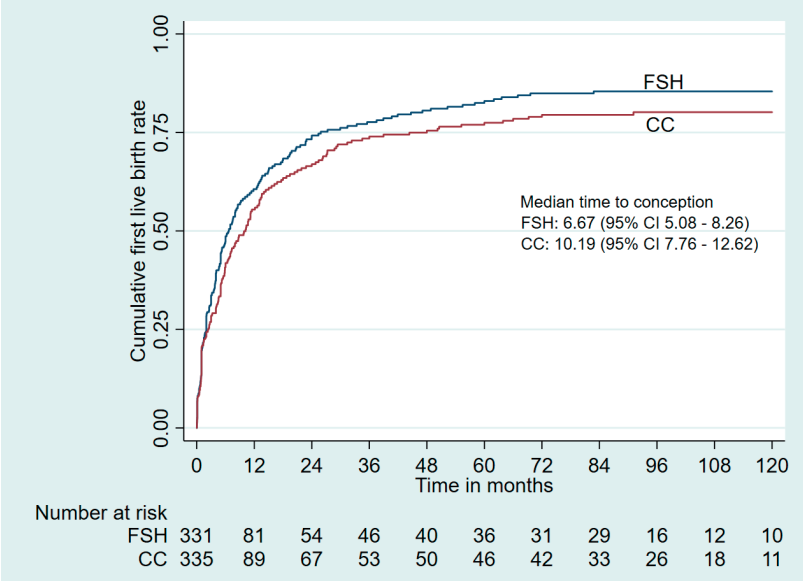


Figure 2A. Cumulative livebirths for the first live birth, gonadotrophins *vs* clomiphene citrate, including all women from the primary RCT and follow up period (log rank p=0.058).

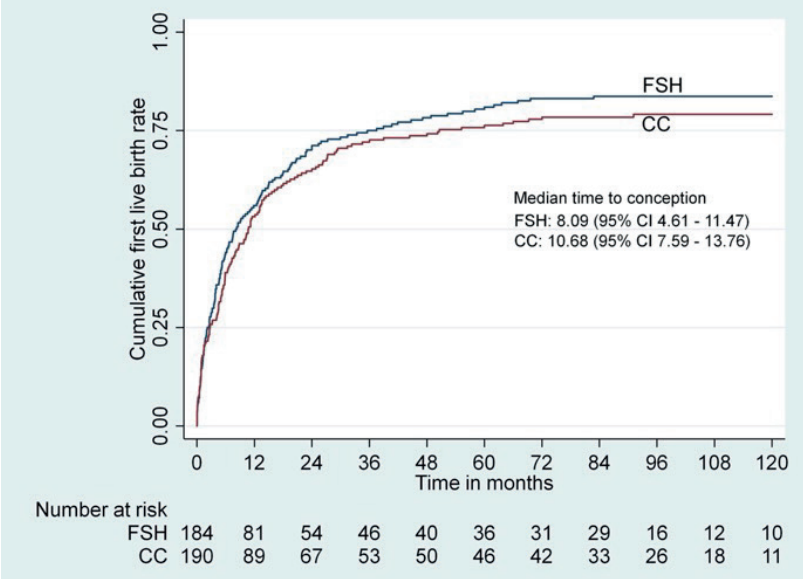


Figure 2B. Cumulative livebirths for the first live birth, gonadotrophins *vs* clomiphene citrate, including women in the follow up period (log rank p=0.24).

Table 3. Period of first live birth and fertility treatments for first live birth

First live birth		Gonadotrophins (N=184)	CC (N=190)
M-ovin trial (≤ 8 months)		92 (50.0)	78 (41.0)
	Natural conception	2	2
	CC	1	70
	Gonadotrophins	89	5
	IVF/ICSI	0	1
	Laparoscopic drilling	0	0
Follow-up (> 8 months)		62 (33.7)	72 (37.9)
	Natural conception	11	15
	CC	3	6
	Gonadotrophins	19	23
	IVF/ICSI	26	26
	Laparoscopic drilling	1	0
	Unknown	2	2
Prior to the study†		5 (2.7)	9 (4.7)
No first live birth		27 (14.7)	33 (17.4)

Data are n (%). †These women did not conceive during and after the study period but had at least one live birth prior to the study.

Second, third and fourth cumulative live birth after inclusion

Among the 374 women, a second live birth occurred in in 85 of 184 women (46.2%) in the gonadotrophin group and in 85 of 190 women (44.7%) in the CC group (RR 1.03, 95% CI 0.83 – 1.29; RD 1.46%, 95% CI -8.63 – 11.55). We found no interaction between insemination method (IUI or intercourse) and treatment (gonadotrophins or CC) on live birth (p=0.84).

In the gonadotrophin group 47 of the 85 live births were conceived by natural conception and in the CC group 44 of the 85 were conceived by natural conception.

Six of 184 women (3.3%) delivered a third child in the gonadotrophin group and 14 of 190 women (7.4%) delivered a third child in the CC group. In the gonadotrophin group four of the six live births were conceived by natural conception and in the CC group all of the 14 live births were conceived by natural conception.

One woman in the gonadotrophin group had a fourth live birth. Table 4 shows the fertility treatment that resulted in a second, third, or fourth live birth.

Table 4. Fertility treatments* for second and third live birth

		Gonadotrophins (N=184)	CC (N=190)
Second live birth		85 (46.2)	85 (44.7)
	Natural conception	47	44
	CC	5	13
	CC + IUI	1	5
	Gonadotrophins	6	2
	Gonadotrophins + IUI	10	6
	IVF	10	8
	ICSI	5	7
	Laparoscopic drilling	0	0
	Unknown	1	2
Third live birth		6 (3.3)	14 (7.4)
	Natural conception	5	14
	Gonadotrophins + IUI	1	0
Fourth live birth		1 (0.5)	0 (0)
	Natural conception	1	0

Data are mean n (%). *Women had more fertility treatments, but these ended not in a pregnancy or in a miscarriage and are not noted in this table.

Twin pregnancies

Twin pregnancies occurred in 12 women (7.9%) allocated to gonadotrophins, four within the study period of eight months after randomisation and eight women during follow-up.

Of the women allocated to CC, 11 women had a twin pregnancy (7.4%), four within the study period and seven women during follow-up. Of these 23 twin pregnancies 19 resulted in a live birth, three miscarried and one was a vanishing twin of which the remaining singleton was delivered.

Miscarriage, stillbirth and ectopic pregnancy

Of the women allocated to gonadotrophins, fourteen women had a miscarriage within the study period of eight months after randomisation and 63 (47 women) miscarriages occurred during follow-up. Over the follow-up period, in total 52 women had at least one miscarriage (28.3%)

Of the women allocated to CC, nine women had a miscarriage within the study period of eight months after randomisation and 58 (39 women) miscarriages occurred during follow-up. In total 45 women had at least one miscarriage (23.7%)

Before the MOVIN study 25 women had a miscarriage allocated to gonadotrophins (13.6%) and 24 women allocated to CC (12.6%).

In the gonadotrophin group three women had a stillbirth (1.6%); one within the study period of eight months after randomisation and two women during follow-up (one conceived with gonadotrophins and one with gonadotrophins and IUI). One pregnancy was preterm terminated due to congenital abnormalities (conceived with gonadotrophins).
In the CC group one woman had a stillbirth (0.5%), which had happened during follow-up (conceived with IVF).

In the gonadotrophin group four women had an ectopic pregnancy (2.2%); one within the study period of eight months after randomisation and three women during follow-up (conceived naturally, with CC and one with gonadotrophins).
In the CC group four women had an ectopic pregnancy (2.1%); one within the study period of eight months after randomisation and three women during follow-up (one conceived with CC and two with IVF).

IUI versus intercourse

Among the 374 women, addition of IUI resulted in a first live birth in 158 of 192 women (82.3%) and treatment with CC resulted in a live birth in 146 of 182 women (80.2%) (RR: 1.03 95% CI 0.93 - 1.13; 2.13%, 95% CI -5.95, 10.21).

A second live birth occurred in in 84 of 192 women (43.8%) in the IUI group and in 86 of 182 women (47.3%) in the intercourse group (RR 0.93 95% CI 0.74 - 1.16; RD -3.50%, 95% CI -13.59, 6.59). Ten of 182 women delivered a third child in the IUI group and 10 of 172 women delivered a third child in the intercourse group. One woman in the IUI group had a fourth live birth.

Subgroup Thin versus Thick endometrium

The endometrium thickness (EMT) in the sixth ovulatory cycle before randomisation was available in 235 women (116 allocated to gonadotropins and 119 allocated to CC) participating in this follow-up study. There were 90 women (38%) who had an EMT of ≤7mm and 145 women (62%) who had an EMT >7mm.

Among the women with an EMT ≤7 mm, gonadotropins resulted in a live birth of 39 of 41 women (95%), while CC resulted in a live birth of 39 of 49 women (80%) (RR 1.20, 95% CI 1.02-1.40). Among the women with an EMT >7mm, gonadotropins resulted in a live birth of 59 of 75 women (79%), while CC resulted in a live birth of 55 of 70 women (79%) (RR 1.00, 95% CI 0.84-1.19).

Non-responders

The 287 women of whom we failed to retrieve follow-up data had a similar mean age and BMI at start of the MOVIN trial as responders (see supplementary table 1). Of these non-responders, 136 women (47%) had a live birth after 8 months of randomization in the original RCT. The live birth rate in the trial was 45% for responders.

DISCUSSION

In the original MOVIN study in women with normogonadotropic anovulation and CC failure, switching to treatment with gonadotrophins resulted in a 11% higher live birth rate than continuing with CC. In this follow-up study we found that within a median follow-up time of 98 months (8.2 years), eight out of 10 women had at least one live birth. At long-term follow-up gonadotrophins resulted in 4.8% more first live births but this difference was no longer statistically significant. The median time to conception leading to a first live birth was 6.7 months following gonadotrophins and 10.2 months following CC for the entire population. The use of fertility treatments in the follow-up period (i.e. after the study period) were comparable between the two groups. Use of IUI or intercourse was not associated with cumulative live birth rate and there was no interaction. Respectively 46% and 45% of the women in the gonadotrophins and CC groups had a second live birth. More than half of the second live births were conceived in a natural cycle in both groups. The use of fertility treatments in the follow-up period were also comparable between the two groups for the second and third live birth. Women with a mid-cycle EMT ≤7 mm in the sixth ovulatory cycle with CC before randomization seem to have a higher first live birth rate with gonadotrophins compared to CC, respectively 95% versus 80%. While in women with an EMT >7mm the live birth rates are 79% in both groups.

A strength of this follow-up study is that our analyses are based on a strong and balanced two-by-two factorial RCT design. Certain shortcomings should also be acknowledged. The main limitation is that we had no follow-up data for 43% (N = 287) of the women participating in the randomised controlled trial such that selection bias can't be excluded. Of 287 women who did not participate in the follow-up, we know that 135 had a first live birth in the MOVIN study. On the other hand, these women had similar baseline profiles and live birth rates in the initial MOVIN trial were even slightly higher. In view of this, the population included in this study appears representative of the whole population.

Our study found a cumulative live birth rate of switching to gonadotrophins in women with CC-failure of 51% after 8 months and 83.7% after a follow-up period of 8.2 years. A previous follow up cohort found a cumulative live birth rate of 50% (42 out of 84 women) after a follow-up period of 2 years. The lower live birth rate was likely due to the inclusion of both women with CC failure and women who did not ovulate after 3 cycles of CC.[9]

The results of our study suggest that the live birth rate of gonadotrophins and CC are comparable over a longer period. This seems to be particularly true when the EMT is above 7 mm as was also previously shown for the pre follow-up data.[7] The original trial showed that women with normogonadotropic anovulation and CC failure had significantly more live births if they switched to gonadotrophins in comparison to continuing treatment with CC for another six cycles at costs of €15 258 for one extra live birth.[5,6] Since the long-term cumulative live birth rates are comparable in both arms, continuing treatment with six cycles of CC treatment is likely to reduce costs in the long run as well and will also reduce the use of injections and may therefore reduce the treatment burden in women.

The majority of the women who delivered their first live born, conceived with a fertility treatment. Almost half of all women in both the gonadotrophin and CC arm delivered a second child. In both arms, approximately 52-55% of the women conceived their second live birth in a natural cycle. This difference might be explained by a restored hormone balance due to ageing in women with normogonadotropic anovulation.[10]

We hypothesized that the women in the CC arm that did not conceive during the trial would first use gonadotrophins before they switch to IVF whereas the women that did not conceive in the gonadotrophin arm would start with IVF directly after 6 cycles of gonadotrophins. This hypothesis was rejected as in this study since 25 women conceived with IVF or ICSI in the gonadotrophins group and 26 in the CC group.

Nowadays, letrozole is considered as a more effective first line medication for ovulation induction. It would be of interest to compare continuous letrozole with gonadotrophins and evaluate both short and long-term outcomes.

In conclusion, our results show that differences in the long-term cumulative live birth rates in women with normogonadotropic anovulation and CC failure are small comparing switching to gonadotrophins and continuing with CC. Women with normogonadotropic anovulation and CC failure have a high chance to reach at least one live birth. The continued and repeated use of CC has been shown to reduce costs[6] and remains a good alternative for more invasive and expensive treatments.

Acknowledgments
We thank all couples that participated in the trial, the hospitals and their staff.

Funding
The original study received funding from the Dutch Organization for Health Research and Development (ZonMw number: 80-82310-97-12067).

Conflict of interest
AH reports consultancy for development and implementation of a lifestyle App MyFertiCoach developed by Ferring Pharmaceutical Company.
BWM is supported by a NHMRC Investigatorgrant (GNT1176437). BWM reports consultancy for ObsEva and Merck and travel support from Merck.
IAJR takes part of an advisory board from Ferring.
CBL is chief editor of Human Reproduction and his former department receives unresticted grants for scientific research from Ferring, Merck and Guerbet.
All other authors have nothing to declare.

SUPPLEMENTAL DATA

Appendix 1. Baseline characteristics of the non-responding couples

	Gonadotrophins (N=143)	CC (N=144)
Age of women (years)	29.9 (3.7)	30.1 (4.1)
Ethnicity		
White	103 (72%)	105 (73%)
Non-white	26 (18%)	26 (18%)
BMI (kg/m2)	25.8 (5.2)	25.6 (5.2)
BMI>25.0 kg/m2	64 (45%)	63 (44%)
Current smoker	22 (15%)	22 (15%)
Diabetes	1	2
Previous livebirth	37 (26%)	35 (24%)
Duration of subfertility (months)	28.1 (15.0)	26.6 (17.2)
Cycle pattern before treatment†		
Amenorrhoea	15 (11%)	25 (17%)
Oligomenorrhoea	111 (77%)	103 (72%)
Unknown	18 (12%)	17 (11%)
TMC ($\times 10^6$)	70 (76)	96 (121)
Polycystic ovaries on ultrasound‡	91 (64%)	89 (62%)
Life birth in M-ovin study	75 (52%)	60 (42%)

Data are mean (SD) or n (%). BMI =body mass index. TMC= total motile sperm count. †Amenorrhoea: absence of menstrual bleeding for >6 months. Oligomenorrhoea: irregular menstrual bleedings with intervals of >35 days but ≤6 months. ‡Defined as the presence of 12 or more follicles in each ovary measuring 2–9 mm in diameter.

REFERENCE LIST

1. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod* 2018;33: 1602-1618.
2. Costello M, Garad R, Hart R, Homer H, Johnson L, Jordan C, Mocanu E, Qiao J, Rombauts L, Teede HJ *et al.* A Review of First Line Infertility Treatments and Supporting Evidence in Women with Polycystic Ovary Syndrome. *Med Sci (Basel)* 2019;7.
3. Costello MF, Garad RM, Hart R, Homer H, Johnson L, Jordan C, Mocanu E, Qiao J, Rombauts L, Teede HJ *et al.* A Review of Second- and Third-line Infertility Treatments and Supporting Evidence in Women with Polycystic Ovary Syndrome. *Med Sci* 2019;7.
4. Nahuis MJ, Weiss NS, van der Veen F, Mol BW, Hompes PG, Oosterhuis J, Lambalk NB, Smeenk JM, Koks CA, van Golde RJ *et al.* The M-OVIN study: does switching treatment to FSH and / or IUI lead to higher pregnancy rates in a subset of women with world health organization type II anovulation not conceiving after six ovulatory cycles with clomiphene citrate - a randomised controlled trial. *BMC Womens Health* 2013;13: 42.
5. Weiss NS, Nahuis MJ, Bordewijk E, Oosterhuis JE, Smeenk JM, Hoek A, Broekmans FJ, Fleischer K, de Bruin JP, Kaaijk EM *et al.* Gonadotrophins versus clomifene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomifene failure (M-OVIN): a randomised, two-by-two factorial trial. *Lancet* 2018;391: 758-765.
6. Bordewijk EM, Weiss NS, Nahuis MJ, Bayram N, van Hooff MHA, Boks DES, Perquin DAM, Janssen CAH, van Golde RJT, Lambalk CB *et al.* Gonadotrophins versus clomiphene citrate with or without IUI in women with normogonadotropic anovulation and clomiphene failure: a cost-effectiveness analysis. *Hum Reprod* 2019;34: 276-284.
7. Bordewijk EM, Weiss NS, Nahuis MJ, Kwee J, Lambeek AF, van Unnik GA, Vrouenraets FPJ, Cohlen BJ, van de Laar-van Asseldonk TAM, Lambalk CB, Goddijn M, Hompes PG, van der Veen F, Mol BWJ, van Wely M; M-ovin study group. Gonadotrophins or clomiphene citrate in women with normogonadotropic anovulation and CC failure: does the endometrium matter? Hum Reprod. 2020;35(6):1319-1324.
8. Duffy JMN, AlAhwany H, Bhattacharya S, Collura B, Curtis C, Evers JLH, Farquharson RG, Franik S, Giudice LC, Khalaf Y *et al.* Developing a core outcome set for future infertility research: an international consensus development study†‡. *Human Reproduction* 2020;35: 2725-2734.
9. Eijkemans MJ, Imani B, Mulders AG, Habbema JD, Fauser BC. High singleton live birth rate following classical ovulation induction in normogonadotrophic anovulatory infertility (WHO 2). *Hum Reprod* 2003;18: 2357-2362.
10. Elting MW, Korsen TJ, Rekers-Mombarg LT, Schoemaker J. Women with polycystic ovary syndrome gain regular menstrual cycles when ageing. *Hum Reprod* 2000;15: 24-28.

Chapter 6

First-line ovulation induction for polycystic ovary syndrome: an individual participant data meta-analysis.

Wang R, Li W, Bordewijk EM, Legro RS, Zhang H, Wu X, Gao J, Morin-Papunen L, Homburg R, König TE, Moll E, Kar S, Huang W, Johnson NP, Amer SA, Vegetti W, Palomba S, Falbo A, Özmen Ü, Nazik H, Williams CD, Federica G, Lord J, Sahin Y, Bhattacharya S, Norman RJ, van Wely M, Mol BW; Reproductive Medicine Network+; International Ovulation Induction IPDMA Collaboration.

Human Reproduction Update. 2019 Nov

ABSTRACT

Background

Polycystic ovary syndrome (PCOS) is the most frequent cause of anovulatory infertility. In women with PCOS, effective ovulation induction serves as an important first-line treatment for anovulatory infertility. Individual participant data (IPD) meta-analysis is considered as the gold standard for evidence synthesis which provides accurate assessments of outcomes from primary randomised controlled trials (RCTs) and allows additional analyses for time-to-event outcomes. It also facilitates treatment-covariate interaction analyses and therefore offers an opportunity for personalised medicine.

Objective and rationale

We aimed to evaluate the effectiveness of different ovulation induction agents, in particular letrozole alone and clomiphene citrate (CC) plus metformin, as compared to CC alone, as the first-line choice for ovulation induction in women with PCOS and infertility, and to explore interactions between treatment and participant-level baseline characteristics.

Search methods

We searched electronic databases including MEDLINE, EMBASE and Cochrane Central Register of Controlled Trials up to 20 December 2018. We included RCTs comparing the following interventions with each other or placebo/no treatment in women with PCOS and infertility: CC, metformin, CC plus metformin, letrozole, gonadotrophin and tamoxifen. We excluded studies on treatment-resistant women. The primary outcome was live birth. We contacted the investigators of eligible RCTs to share the IPD and performed IPD meta-analyses. We assessed the risk of bias by using the Cochrane risk of bias tool for RCTs.

Outcomes

IPD of 20 RCTs including 3962 women with PCOS were obtained. Six RCTs compared letrozole and CC in 1284 women. Compared with CC, letrozole improved live birth rates (3 RCTs, 1043 women, risk ratio [RR] 1.43, 95% confidence interval [CI] 1.17-1.75, moderate-certainty evidence) and clinical pregnancy rates (6 RCTs, 1284 women, RR 1.45, 95% CI 1.23-1.70, moderate-certainty evidence) and reduced time-to-pregnancy (6 RCTs, 1235 women, hazard ratio [HR] 1.72, 95% CI 1.38-2.15, moderate-certainty evidence). Meta-analyses of effect modifications showed a positive interaction between baseline serum total testosterone levels and treatment effects on live birth (interaction RR 1.29, 95% CI 1.01-1.65). Eight RCTs compared CC plus metformin to CC alone in 1039 women. Compared with CC alone, CC plus metformin might improve clinical pregnancy rates (8 RCTs, 1039 women, RR 1.18, 95% CI 1.00-1.39, low-certainty

evidence) and might reduce time-to-pregnancy (7 RCTs, 898 women, HR 1.25, 95% CI 1.00-1.57, low-certainty evidence), but there was insufficient evidence of a difference on live birth rates (5 RCTs, 907 women, RR 1.08, 95% CI 0.87-1.35, low-certainty evidence). Meta-analyses of effect modifications showed a positive interaction between baseline insulin levels and treatment effects on live birth in the comparison between CC plus metformin and CC (interaction RR 1.03, 95% CI 1.01-1.06).

Wider implications

In women with PCOS, letrozole improves live birth and clinical pregnancy rates and reduces time-to-pregnancy compared to CC and therefore can be recommended as the preferred first-line treatment for women with PCOS and infertility. CC plus metformin may increase clinical pregnancy and may reduce time-to-pregnancy compared to CC alone, while there is insufficient evidence of a difference on live birth. Treatment effects of letrozole are influenced by baseline serum levels of total testosterone, while those of CC plus metformin are affected by baseline serum levels of insulin. These interactions between treatments and biomarkers on hyperandrogenaemia and insulin resistance provide further insights into a personalised approach for the management of anovulatory infertility related to PCOS.

INTRODUCTION

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder of reproductive age women, and the prevalence among different geographic regions ranges from 5 to 21%, depending on the criteria used.[1] PCOS is a heterogeneous syndrome comprising of at least two of the following clinical characteristics according to the Rotterdam diagnostic criteria: oligo-/anovulation, clinical and/or biochemical hyperandrogenism or polycystic ovaries morphology based on ultrasound assessment.[2]

Anovulatory infertility is usually one of the key features that women with PCOS are confronted with. Simple and effective infertility treatments as the first-line choice are therefore important. Our previous network meta-analysis compared available first-line treatment options for women with PCOS with infertility and found that letrozole and combined clomiphene citrate (CC)–metformin were superior to other ovulation induction medications in terms of clinical pregnancy and that letrozole resulted in more live births than other interventions, including CC.[3] These findings are in agreement with the evidence summarised in the international evidence based guideline for the assessment and management of PCOS.[4]

As women with PCOS represent a heterogeneous population according to the diagnostic criteria, it is important to identify which individuals benefit most from a particular treatment so that clinicians can provide personalised care.[5] However, primary RCTs are usually underpowered to detect subgroup effects.[6] Subgroup analyses in meta-analyses of aggregate data are at risk of ecological bias due to the ignorance of within-study interactions or are even impossible to perform due to heterogeneous reporting of subgroup data in the primary trials.[6]

Moreover, time-to-pregnancy is also an important patient-centred outcome, but it has never been reported in previous meta-analyses on PCOS. This is likely due to the unavailability of the data in the publication as well as the methodological challenges on data extraction and synthesis. In addition, the primary trials are not always of high quality in terms of analyses and reports,[7] which can directly affect the data extraction, analysis and risk of bias assessment process in subsequent meta-analyses.

These deficiencies in aggregate data meta-analyses can potentially be overcome by using individual participant data (IPD). IPD meta-analysis has been described as the gold standard in evidence synthesis, by engaging investigators of the primary trials to provide the raw data of the primary trials.[8] Such strategy facilitates derivation of the information beyond the primary publication, standardisation of inclusion criteria, outcomes and analyses across trials and investigations of subgroup effects and time-to-event outcomes.[6,8]

We therefore performed an IPD meta-analysis to evaluate the effectiveness of different ovulation induction agents, in particular letrozole alone and CC plus metformin, as compared to CC alone, as the first-line choice for ovulation induction in women with PCOS and infertility, and to explore interactions between treatment and participant-level baseline characteristics.

METHODS

Registration and literature search

This IPD meta-analysis was conducted based on a registered protocol (PROSPERO CRD42017059251) and reported according to the Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data (PRISMA-IPD) statement.[9]

We updated the searches in MEDLINE, EMBASE and Cochrane Central Register of Controlled Trials in September 2017, based on our previous search strategies for a network meta-analysis on treatment strategies for World Health Organization (WHO) II anovulation.[3] In brief, the search terms included both index terms as well as free words on PCOS, anovulation and ovulation induction. After completing data requesting process, we further updated the search on 20 December 2018 to identify the latest studies. We also searched the WHO International Clinical Trials Registry Platform (WHO ICTRP) and US National Institutes of Health (clinicaltrials.gov) and ISRCTN registry to identify ongoing trials. In addition, we reviewed the references lists of relevant papers and corresponded with trialists in PCOS to identify potential eligible trials that we might have missed.

Eligibility criteria

We included RCTs comparing the following interventions with each other or placebo/no treatment: CC, metformin, CC and metformin combined, letrozole, gonadotrophins and tamoxifen in women with WHO II anovulation, including PCOS. We excluded trials reporting on treatment-resistant women, trials comparing different doses of the same intervention and quasi-RCTs. We did not apply language restrictions. For crossover trials, we only included the data in the first phase.

The primary outcome was live birth. The secondary outcomes were clinical pregnancy, ovulation, miscarriage, multiple pregnancy and time to pregnancy.
Study selection and data collection

Study selection and data collection

Two members of the review team (from R.W., W.L. and E.M.B.) independently assessed the titles and abstracts to exclude irrelevant studies and subsequently reviewed the full-text

articles to evaluate their eligibility. Disagreements were resolved by discussion with a third author (B.W.M., M.v.W. or R.J.N.).

We contacted investigators of eligible RCTs to share the de-identified IPD and established the International Ovulation Induction IPDMA Collaboration. We sent at least two more reminders when we did not receive responses.

We obtained de-identified IPD including baseline characteristics including age, body mass index (BMI), ethnicity, type of infertility (primary/secondary), treatment history (treatment-naïve or not), fasting glucose, fasting insulin, total testosterone, sex hormone-binding globulin (SHBG), ovarian volume and the Ferriman–Gallwey score for hirsutism. We also obtained data on allocated treatments, number of ovulation induction cycles, ovulation and fertility outcomes including live birth, clinical pregnancy, miscarriage and multiple pregnancy.

We checked data for consistency by comparing the analyses from obtained IPD with the original publications. We discussed any inconsistencies or obvious errors with investigators of primary RCTs and solved discrepancies by consensus.

Risk of bias assessment

Two members of the review team independently evaluated the risk of bias in each included RCT, using the domain-based evaluation tool described in the Cochrane Handbook for Systematic Reviews of Interventions.[10] We assessed the following domains as low risk of bias, unclear or high risk of bias: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome data, selective reporting (reporting bias) and other sources of bias. When the risk of bias for a domain was unclear, investigators of these RCTs were asked to provide additional information to resolve the uncertainty.

We assessed the overall certainty of the evidence across RCTs by using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, including the risk of bias, consistency of effect, imprecision, indirectness and publication bias.

Data synthesis

We conducted all analyses based on an intention-to-treat principle using women randomised per allocated group as the unit of all analyses. We performed two-stage random-effects IPD meta-analyses for letrozole versus CC alone and CC with metformin versus CC alone. For dichotomous outcomes, we calculated RRs and 95% confidence intervals (CIs) and presented statistical heterogeneity by using I^2 statistic.[10] For time-to-

event outcomes, we used the number of treatment cycles as an approximate estimate for time and visualised the summary time-to-event in simple non-stratified Kaplan–Meier curves. We also estimated hazard ratios (HRs) in Cox proportional hazards regression models for discrete time and pooled HRs and 95% CI, by using the generic inverse variance method.[11]

Subgroup effects were estimated for the primary outcome by treatment–covariate interaction terms within trials and subsequent meta-analyses of interactions, as interactions using within-trial information alone without considering between-trial interactions are recommended as the standard practice to avoid ecological bias.[12] We explored the treatment–covariate interactions of the following pre-specified baseline covariates: age, BMI, ethnicity, primary/secondary infertility, treatment history, hirsutism score, insulin resistance (serum glucose and insulin level), hyperandrogenaemia status (testosterone, SHBG, free androgen index) and ovarian volume. We also added the analysis of homeostatic model assessment for insulin resistance (HOMA-IR) as requested during the peer review process. For dichotomous covariates with statistically significant interaction, we further performed stratified analyses to illustrate the treatment effects in different strata of the subgroups. Continuous variables were analysed as such without categorisation. For continuous covariates with statistically significant interaction, we further presented a weighted mean curve and pointwise CI based on treatment–covariate interactions estimated in relevant studies. Due to the potential type I error, the results of subgroup analyses were all considered exploratory.

To evaluate the IPD availability bias, we performed a network meta-analysis of RCTs with IPD in a random-effects multivariate meta-analysis model[13,14] on live birth and clinical pregnancy and then compared the results with a network meta-analysis of all eligible RCTs. If these results were consistent, we considered the included RCTs with IPD representative of all the eligible RCTs.

We performed a sensitivity analysis on studies with low risk of bias in allocation concealment as planned. As the majority of eligible studies focused only on treatment-naïve women with PCOS, these studies did not contribute to within-study interaction for treatment history and were not included in the treatment-covariate analysis. We performed a post hoc sensitivity analysis by including only treatment-naïve women to demonstrate the robustness of the results.

We conducted all the analyses in Stata software version 15.1 (Stata Corp, College Station, TX, USA).

RESULTS

Characteristics of included studies

The final updated search yielded 709 non-duplicated studies (Fig. 1). After screening the titles and abstracts, 636 irrelevant studies were excluded. Finally, a total of 62 studies (61 publications, 9356 women) fulfilled the inclusion criteria and were included. These studies were published in English ($n = 58$), French ($n = 1$) (Boudhraa *et al.*, 2010), Italian ($n = 1$) (Santonocito *et al.*, 2009), Turkish ($n = 1$) (Aygen *et al.*, 2007) and Persian ($n = 1$) (Lorzadeh *et al.*, 2011).

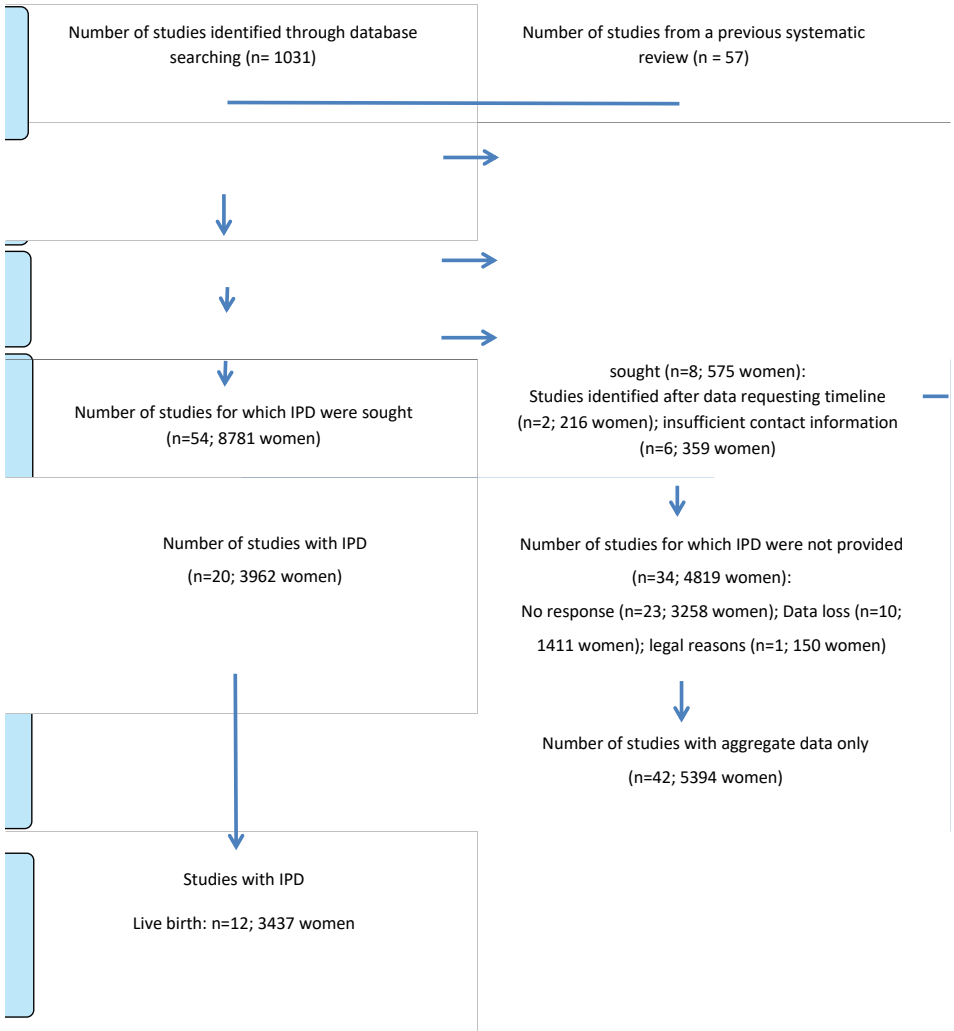


Figure 1. PRISMA-IPD flow diagram

IPD was not sought from eight studies (575 women), due to insufficient contact information ($n = 6$; 359 women) (Beigi, 2006; Boudhraa *et al*., 2010; Cudmore and Tupper, 1966; El-Biely and Habba, 2001; Garcia *et al*., 1985; Johnson *et al*., 1966) or because the studies were identified after our data requesting timeline ($n = 2$; 216 women) (Fatima *et al*., 2018; Topçu *et al*., 2017). For the remaining 54 studies (8781 women), the primary investigators were contacted to share IPD of the primary studies. IPD from 34 studies (4819 women) were not available, due to no response ($n = 23$; 3258 women) (Abuelghar *et al*., 2013; Atay *e al*., 2006; Ayaz *et al*., 2013; Banerjee Ray *et al*., 2012; Basirat *et al*., 2012; Boostanfar *et al*., 2001; Chen *et al*., 2016; Dasari and Pranahita, 2009; Dehbashi *et al*., 2009; Hossein-Rashidi *et al*., 2016; Jahan, 2015; Karimzadeh *et al*., 2007; Karimzadeh and Javedani, 2010; Lopez *et al*., 2004; Lorzadeh *et al*., 2011; Maged *et al*., 2015; Robinson *et al*., 2003; Roy *et al*., 2012; Selim and Borg, 2012; Seyedoshohadaei *et al*., 2012; Sharief and Nafee, 2015; Sheikh-El-Arab Elsedeek and Elmaghraby, 2011; Zeinalzadeh *et al*., 2010), data loss ($n = 10$; 1411 women) (Aygen *et al*., 2007; Badawy *et al*., 2009; Badawy and Gibreal, 2011; Fleming *et al*., 2002; Keikha and Shahraki, 2011; Khorram *et al*., 2006; Mobusher, 2014; Santonocito *et al*., 2009; Tang *et al*., 2006; Zain *et al*., 2009) or legal reasons ($n = 1$; 150 women) (Moussa *et al*., 2016). These studies are listed in Supplementary Table S1.

IPD were available for at least one outcome from 20 studies (3962 women Table 1), including three from the USA (Legro *et al*., 2007; Legro *et al*., 2014; Williams *et al*., 2009), three from Italy (Leanza *et al*., 2014; Palomba *et al*., 2005; Vegetti *et al*., 1999), three from Turkey (Bayar *et al*., 2006; Nazik and Kumtepe, 2012; Sahin *et al*., 2004), two from the UK (Amer *et al*., 2017; Lord *et al*., 2006), two from China (Liu *et al*., 2017; Wu *et al*., 2017), two from India (Kar, 2012; Kar and Sanchita, 2015), two studies (in one publication) from New Zealand (Johnson *et al*., 2010), one from the Netherlands (Moll *et al*., 2006), one from Finland (Morin-Papunen *et al*., 2012) and one from multiple countries (the Netherlands, UK, Malta, Belgium, Argentina and Colombia) (Homburg *et al*., 2012). These RCTs were published in English between 1999 and 2017, with 11 (55%) published after 2010.

Participants in all 20 RCTs were women with PCOS. In one RCT, participants were diagnosed with PCOS by fulfilling at least three of the following: PCO morphology, oligo/amenorrhoea, hirsutism, hyperandrogenaemia and elevated serum LH/FSH ratio (Sahin *et al*., 2004), while in the remaining 19 RCTs, the participants were women with PCOS based on the Rotterdam criteria (Bayar *et al*., 2006; Kar, 2012; Leanza *et al*., 2014; Liu *et al*., 2017; Nazik and Kumtepe, 2012) or different phenotypes, including Phenotype B (ovulatory dysfunction + androgen excess) (Amer *et al*., 2017; Homburg *et al*., 2012; Johnson *et al*., 2010; Kar and Sanchita, 2015; Legro *et al*., 2007; Legro *et al*., 2014; Lord *et al*., 2006; Morin-Papunen *et al*., 2012; Palomba *et al*., 2005; Williams *et al*., 2009; Wu *et al*., 2017) or Phenotype D (ovulatory dysfunction + PCO) (Moll *et al*., 2006; Vegetti *et al*., 1999).

For RCTs involving two stages of different interventions, including cross-over studies, we only included the data in the first stage. We included the IPD comparing letrozole versus CC before crossing over (Amer *et al*., 2017) and included the IPD comparing metformin versus placebo within the first 3 months before starting other ovulation induction agents (Morin-Papunen *et al*., 2012). In one RCT (Nazik and Kumtepe, 2012), switching between intervention and the control after the first cycle was allowed during the trial and the analysis in the primary publication was on a per-cycle basis, and therefore, we only included the IPD of the first cycle.

In summary, four RCTs compared three interventions (CC plus metformin or CC alone versus metformin (Johnson *et al*., 2010; Kar and Sanchita, 2015; Legro *et al*., 2007) or CC with metformin or letrozole versus CC (Liu *et al*., 2017)) and the remaining 16 compared two interventions. The most common comparisons were CC with metformin versus CC alone (8 RCTs) (Johnson *et al*., 2010; Kar and Sanchita 2015; Leanza *et al*., 2014; Legro *et al*., 2007; Liu *et al*., 2017; Moll *et al*., 2006; Sahin *et al*., 2004; Williams *et al*., 2009) and letrozole versus CC alone (6 RCTs) (Amer *et al*., 2017; Bayar *et al*., 2006; Kar, 2012; Legro *et al*., 2014; Liu *et al*., 2017; Nazik and Kumtepe, 2012).

Quality of evidence of individual studies

The details of risks of bias assessments within individual studies are presented in Figure 2. All RCTs ($n = 20$) reported adequate methods of random sequence generation. Sixteen RCTs (80%) reported adequate methods of allocation concealment while the other four used an open allocation schedule without concealment (Kar, 2012; Kar and Sanchita, 2015; Liu *et al*., 2017; Nazik and Kumtepe, 2012). Fourteen RCTs (70%) blinded the participants and personnel during the trial while six RCTs applied an open label design (Homburg *et al*., 2012; Kar, 2012; Kar and Sanchita, 2015; Liu *et al*., 2017; Nazik and Kumtepe, 2012; Vegetti *et al*., 1999). Given that all outcomes of interest were objective outcomes, it is unlikely that the non-blinded design will affect the outcome measurement and therefore detection bias was rated at low risk for all the included studies. One RCT (5%) had high risk of attrition bias, with 22% overall missing outcome data and 31% missing outcome data in the metformin group (Kar and Sanchita, 2015). One RCT (5%) was at another risk of bias due to allowing imbalanced CC in both groups.

Table 1. Characteristics of included studies

Study	Comparisons	Sample size	Age (mean)	BMI (mean)	Treatment-naïve(%)	Outcomes
(Amer *et al*., 2017)	Letrozole vs CC	159	28.2 ± 4.3	27.5 ± 4.8	100%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Bayar *et al*., 2006)	Letrozole vs CC	80 (74)	31.4 ± 4.0	NA	100%	Clinical pregnancy, multiple pregnancy, time to pregnancy
(Homburg *et al*., 2012)	FSH vs CC	302	29.5 ± 3.9	25.4 ± 5.6	100%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Johnson *et al*., 2010)A	Metformin vs placebo	65	29.6 ± 4.2	37.8 ± 3.5	69%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Johnson *et al*., 2010)B	CC + metformin vs CC vs metformin	106	28.7 ± 4.4	26.5 ± 3.7	78%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Kar 2012)	Letrozole vs CC	103	NA	25.9 ± 3.4	100%	Clinical pregnancy, time to pregnancy, ovulation
(Kar and Sanchita 2015)	CC + metformin vs CC vs metformin	105 (81)	25.6 ± 3.3	26.1 ± 4.3	100%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Leanza *et al*., 2014)	CC + metformin vs CC	56	31.1 ± 2.0	29.5 ± 1.4	100%	Clinical pregnancy, miscarriage, ovulation
(Legro *et al*., 2007)	CC + metformin vs CC vs metformin	626	28.1 ± 4.0	35.2 ± 8.7	45%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Legro *et al*., 2014)	Letrozole vs CC	750	28.9 ± 4.3	35.1 ± 9.3	45%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Liu *et al*., 2017)	CC + metformin vs letrozole vs CC	203	27.0 ± 3.0	21.5 ± 2.9	100%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Lord *et al*., 2006)	Metformin vs placebo	44	29.1 ± 4.9	34.8 ± 7.0	unknown	Clinical pregnancy, ovulation
(Moll *et al*., 2006)	CC + metformin vs CC	225	28.4 ± 3.8	28.1 ± 6.9	100%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Morin-Papunen *et al*., 2012)	Metformin vs placebo	320	28.2 ± 4.0	27.2 ± 6.3	69%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Nazik and Kumtepe, 2012)	Letrozole vs CC	64	26.8 ± 5.6	25.1 ± 4.3	100%	Clinical pregnancy, time to pregnancy, ovulation
(Palomba *et al*., 2005)	CC vs metformin	100	26.2 ± 4.4	26.7 ± 2.3	100%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation
(Sahin *et al*., 2004)	CC + metformin vs CC	21	25.1 ± 3.3	28.2 ± 3.7	100%	Clinical pregnancy, time to pregnancy, ovulation
(Vegetti *et al*., 1999)	Tamoxifen vs CC	95 (108)	30.9 ± 3.1	22.7 ± 4.2	100%	Clinical pregnancy, time to pregnancy, ovulation
(Williams *et al*., 2009)	CC + metformin vs CC	59 (55)	NA	NA	100%	Clinical pregnancy, time to pregnancy, ovulation
(Wu *et al*., 2017)	CC vs placebo	500	27.9 ± 3.3	24.5 ± 4.2	70%	Live birth, clinical pregnancy, time to pregnancy, miscarriage, multiple pregnancy, ovulation

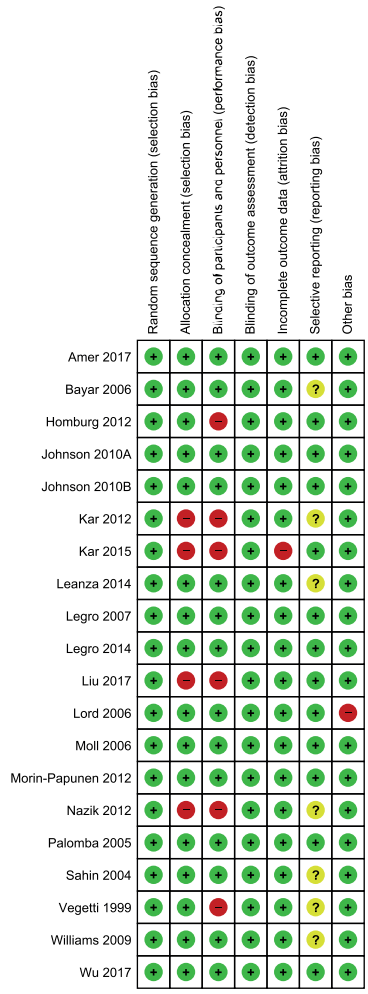


Figure 2. Risk of bias assessments of individual RCTs

Meta-analyses of letrozole versus CC

Live birth

IPD were available in six RCTs comparing letrozole and CC, including 1284 women with PCOS. The forest plot of IPD meta-analysis on live birth is presented in Figure 3a. Compared with CC, letrozole increased live birth rates (3 RCTs, 1043 women, RR 1.43, 95% CI 1.17–1.75, $I^2 = 0$, moderate certainty of evidence). Sensitivity analysis on studies with low risk of bias at allocation concealment and on treatment-naïve women was consistent with the main findings (2 RCTs, 909 women, RR 1.42, 95% CI 1.14–1.76, $I^2 = 0$; 3 RCTs, 627 women, RR 1.41, 95% CI 1.11–1.79, $I^2 = 0$) (Supplementary Table S2).

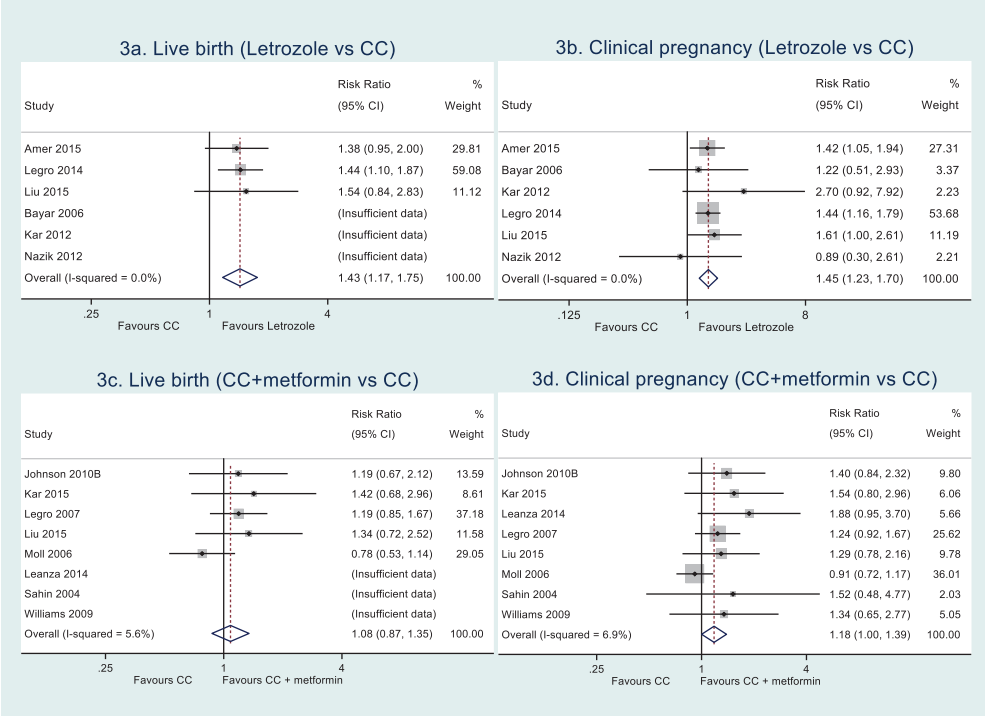


Figure 3. Meta-analyses of letrozole versus CC and CC plus metformin versus CC on live birth and clinical pregnancy

Secondary outcomes

Compared with CC alone, letrozole improved clinical pregnancy (6 RCTs, 1284 women, RR 1.45, 95% CI 1.23–1.70, $I^2 = 0$, moderate certainty of evidence, Fig. 3b) and ovulation rates (5 RCTs, 1210 women, RR 1.13, 95% CI 1.07–1.20, $I^2 = 0$, moderate certainty of evidence, Table 2). There was insufficient evidence of a difference between letrozole and CC alone in terms of multiple pregnancy or miscarriage (Table 2).

The summary Kaplan–Meier curve for time to pregnancy is presented in Fig. 4a. Subsequent pooled analysis of HRs showed that compared to CC, letrozole reduced time-to-pregnancy (6 RCTs, 1235 women, HR 1.72, 95% CI 1.38–2.15, $I^2 = 0$, moderate certainty of evidence).

Table 2. Meta-analyses and GRADE assessments of all outcomes

Comparison	Outcome	Number of RCTs	Number of participants	Risk ratio (RR)	95% confidence interval (CI)	*I2*	Overall certainty of evidence (GRADE)
Letrozole vs CC	Live birth	3	1043	1.43	1.17–1.75	0	Moderate[a]
	Clinical pregnancy	6	1284	1.45	1.23–1.70	0	Moderate[a]
	Multiple pregnancy	2	909	1.45	0.17–12.45	50.9%	Very low[a,b,c]
	Miscarriage	3	1043	1.50	0.95–2.38	0	Low[a,c]
	Ovulation	5	1210	1.13	1.07–1.20	0	Moderate[a]
CC + metformin vs CC	Live birth	5	907	1.08	0.87–1.35	5.6%	Low[a,c]
	Clinical pregnancy	8	1039	1.18	1.00–1.39	6.9%	Low[a,c]
	Multiple pregnancy	4	771	0.76	0.24–2.42	0	Low[a,c]
	Miscarriage	6	963	1.33	0.79–2.26	0	Low[a,c]
	Ovulation	7	968	1.02	0.93–1.12	35.2%	Low[a,c]

[a]Downgraded by one level due to concerns on risk of bias.
[b]Downgraded by one level due to inconsistency.
[c]Downgraded by one level due to imprecision.

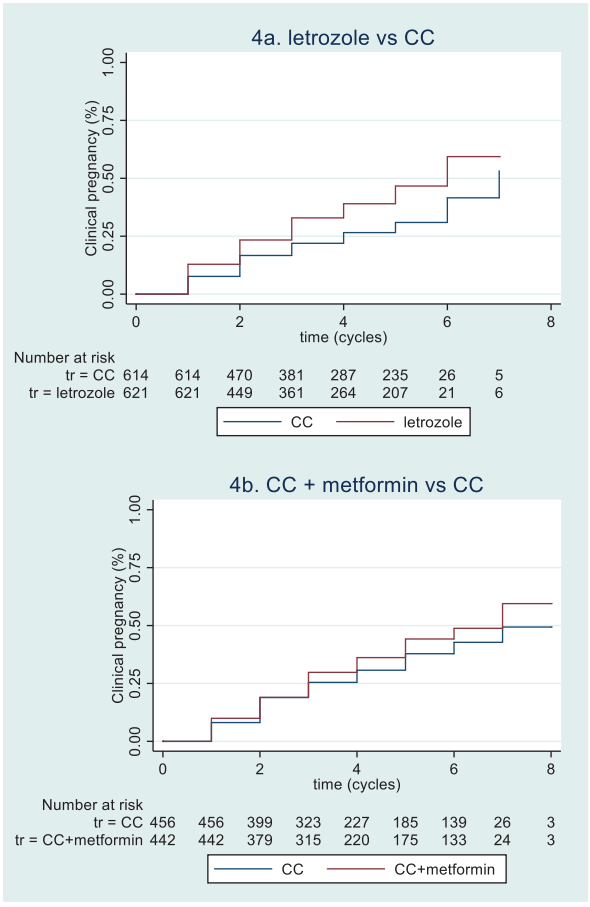


Figure 4. Summary of Kaplan–Meier curves for time-to-event outcomes

(a, b) illustrates the non-stratified summary Kaplan–Meier curves for time-to-pregnancy in the comparisons of letrozole versus CC and CC plus metformin versus CC, respectively. Participants with pregnancy before the first treatment cycles were not included in the 'Numbers at risk' table below, and data were not stratified by trial in this Kaplan–Meier curve. The figures were intended to visualise time-to-event outcomes, but not to show statistical significance.

Treatment–covariate interactions

A meta-analyses of effect modifications showed a positive interaction between baseline serum total testosterone levels and treatment effects on live birth in the comparison between letrozole and CC (interaction RR 1.29, 95% CI 1.01–1.65, 3 RCTs, 1039 women, Fig. 5a). This suggests that women with a higher baseline serum total testosterone level have a larger treatment effect of letrozole versus CC on live birth, compared to women with a lower baseline serum total testosterone level. Such an interaction was consistent across studies ($I^2 = 0$). To directly illustrate the association between baseline serum total testosterone level and relative treatment effects, this interaction is also presented

in a weighted mean curve with 95% CI (Fig. 5b). Meta-analyses did not find any other treatment–covariate interactions (Table 3).

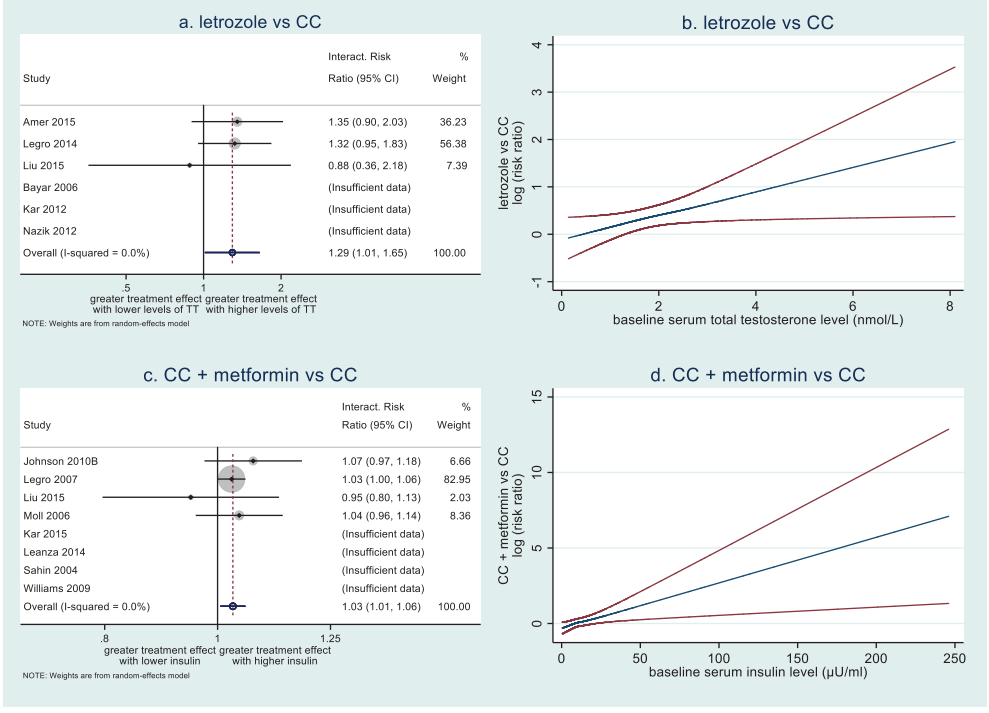


Figure 5. Forest plots and weighted mean curves for treatment-covariate interactions

(a) Forest plot of interactions between baseline serum total testosterone (TT) level and effect of letrozole versus CC on live birth. (b) Weighted mean curve with pointwise 95% CI of interactions between baseline serum total testosterone level and relative effect of letrozole versus CC on live birth. (c) Forest plot of interactions between baseline serum insulin level and effect of CC plus metformin versus CC on live birth. (d) Weighted mean curve with pointwise 95% CIs of interactions between baseline serum insulin level and effect of CC plus metformin versus CC on live birth. (a, c) Circles are used to depict the interaction effects within individual trials as well as the overall interaction effect. The sizes of the circles are in proportion to the inverse of the variance of the estimates. (b, d) Blue line represents for the weighted mean effect of covariate on log risk ratios in the comparison between letrozole and CC. Red lines represent for pointwise 95% CI of interactions.

Table 3. Meta-analyses of treatment-covariate interactions on live birth

Comparison	Baseline covariate	Number of RCTs	Number of participants	Interaction RR	Interaction 95% CI	Interaction *I2*
Letrozole vs CC	Age	3	1043	0.98	0.93–1.05	24.9%
	BMI	3	1043	0.98	0.90–1.05	65.2%
	Ethnicity (non-Caucasian vs Caucasian)	2	909	1.42	0.80–2.45	0
	Treatment history (yes vs no)	1	750	1.07	0.63–1.82	/
	Type of infertility (secondary vs primary)	3	1043	0.83	0.43–1.60	52%
	Total testosterone (nmol/L)	3	1039	1.29	1.01–1.65	0
	SHBG (nmol/L)	2	907	1.00	0.99–1.02	69.7%
	Free androgen index	2	907	1.02	0.91–1.15	79.2%
	Fasting glucose (mmol/L)	3	1002	1.27	0.93–1.73	0
	Fasting insulin (μU/mL)	3	977	1.01	1.00–1.02	0
	HOMA-IR	3	975	1.04	0.98–1.09	0
	Ferriman–Gallwey score for hirsutism	2	884	1.03	0.99–1.06	0
	Ovarian volume (ml)	3	837	1.01	0.95–1.07	33.9%
CC + metformin vs CC	Age	5	895	1.06	0.98–1.15	43.7%
	BMI	5	885	1.02	0.98–1.07	25.1%
	Ethnicity (non-Caucasian vs Caucasian)	3	705	0.91	0.21–3.90	66.8%
	Treatment history (yes vs no)	1	418	0.90	0.46–1.78	/
	Type of infertility (secondary vs primary)	3	622	0.91	0.50–1.65	0
	Total testosterone (nmol/L)	4	824	1.02	0.95–1.08	0
	SHBG (nmol/L)	2	550	1.00	0.99–1.01	0
	Free androgen index	2	546	1.04	0.98–1.09	50.2%
	Fasting glucose (mmol/L)	4	812	1.01	0.74–1.37	0
	Fasting insulin (μU/mL)	4	741	1.03	1.01–1.06	0
	HOMA-IR	4	736	1.14	1.03–1.25	0
	Ferriman–Gallwey score for hirsutism	3	705	0.91	0.21–3.9	66.8%
	Ovarian volume (mL)	2	495	0.99	0.95–1.04	0

Meta-analyses of CC plus metformin versus CC

Live birth

IPD were available in eight RCTs comparing CC with metformin and CC alone, including 1039 women with PCOS. The forest plot of IPD meta-analysis on live birth is presented in Figure 3c. Compared with CC alone, there was insufficient evidence of a difference between CC with metformin and CC alone on live birth (5 RCTs, 907 women, RR 1.08, 95% CI 0.87–1.35, $I^2 = 5.6\%$, low certainty of evidence). Sensitivity analyses on studies with low risk of bias at allocation concealment and on treatment-naïve women showed very small treatment effects with wide CIs (3 RCTs, 714 women, RR 1.02, 95% CI 0.76–1.37, $I^2 = 33.2\%$; 5 RCTs, 662 women, RR 1.06, 95% CI 0.83–1.34, $I^2 = 3.9\%$) (Supplementary Table SII).

Secondary outcomes

Compared with CC alone, CC with metformin might improve clinical pregnancy (8 RCTs, 1039 women, RR 1.18, 95% CI 1.00–1.39, $I^2 = 6.9\%$, low certainty of evidence, Fig. 3b). There was insufficient evidence of a difference between CC with metformin and CC alone on ovulation, multiple pregnancy or miscarriage (Table II).

The summary Kaplan–Meier curve is presented in Figure 4b. Pooled analysis of HRs showed that compared to CC alone, CC with metformin might reduce time-to-pregnancy (7 RCTs, 898 women, HR 1.25, 95% CI 1.00–1.57, $I^2 = 0$, low certainty of evidence).

Treatment–covariate interactions

Meta-analyses of effect modifications showed a positive interaction between baseline insulin levels and treatment effects on live birth in the comparison between CC with metformin and CC alone (interaction RR 1.03, 95% CI 1.01–1.06, 4 RCTs, 741 women, Fig. 5c). Such an interaction was consistent across studies ($I^2 = 0$). This suggests that women with a higher baseline serum insulin level have larger treatment effects of CC with metformin versus CC alone on live birth, compared to women with a lower baseline serum insulin level. Such an interaction was also presented in a weighted mean curve with 95% CI (Fig. 5d). Additional meta-analysis of interactions for HOMA-IR was performed as requested during the peer review process, and it also showed a positive interaction between baseline HOMA-IR and treatment effects on live birth in the comparison between CC with metformin and CC alone (interaction RR 1.14, 95% CI 1.03–1.25, 4 RCTs, 736 women, $I^2 = 0$, Table 3). Meta-analyses did not find any other treatment-covariate interactions (Table 3).

IPD availability bias

With regards to IPD availability bias, network meta-analyses of 20 RCTs with IPD showed similar results to network meta-analyses of all eligible RCTs on both live birth and clinical pregnancy (Supplementary Table S3). Therefore, the participants in RCTs with IPD were

representative of all the eligible participants with PCOS. The transitivity assumption of network meta-analyses was considered valid as the interventions of interest and placebo/no treatment were jointly randomisable.

DISCUSSION

Summary of evidence

This IPD meta-analysis showed that in women with PCOS, letrozole increased live birth rates compared to CC alone and the overall certainty of evidence was moderate. Such treatment benefits of letrozole compared to CC alone were more predominant in women with higher baseline serum levels of total testosterone. There was insufficient evidence of a difference between CC plus metformin and CC alone in live birth rates, and the overall certainty of evidence was low, mainly due to risk of bias and imprecision. The potential benefit of CC in combination with metformin compared to CC alone was more pronounced in women with higher baseline serum insulin or HOMA-IR levels. We did not find other treatment–covariate interactions on live birth for other pre-specified covariates including age, BMI, ethnicity, primary/secondary infertility, treatment history, Ferriman–Gallwey score for hirsutism, SHBG, free androgen index, fasting glucose levels or ovarian volume.

Strengths and limitations

Establishing the International Ovulation Induction IPDMA Collaboration facilitated a platform for key trialists in PCOS to collaborate and share the IPD of the primary trials. It provided us the opportunity to collect unpublished information of the primary trials including the details of randomisation and allocation concealment, treatment history, subgroup data and time-to-pregnancy. Such information allowed us to assess the quality of included trials precisely, to investigate treatment–covariate interactions and to take account of the time in the analyses. The findings of this IPD meta-analysis provide the best available up-to-date evidence.

Moreover, we applied a comprehensive search strategy without language restrictions and updated the search after completing data requesting in case we missed the most recent RCTs. Of the newly identified RCTs, one compared CC plus metformin vs CC in 128 women but did not report live birth,[15] while the other one compared tamoxifen vs CC in 88 women.[16] Although we did not seek IPD from two RCTs identified after the data requesting deadline, adding IPD of these two studies is unlikely to change the main findings.

In addition, the investigation of subgroup effects includes within-study interaction only according to current statistical practice for IPD meta-analyses[12] and therefore is free from ecological bias. For continuous covariates, without categorisation of the data, the

statistical power was not compromised. Further illustration of interactions in weighted mean curve makes the interactions easier to interpret.

Nevertheless, this IPD meta-analysis has a few limitations. First, we were not able to access the IPD of all eligible studies. IPD were available for 32% (20/62) of the included trials, comprising 42% (3962/9356) of the eligible women with PCOS, and the proportions of IPD availability was higher for studies reporting live birth (44% trials including 65% eligible women, Supplementary Table S3). This seems to be partly due to the long history of research on ovulation induction, with the first trial published in 1966. We were however able to access IPD of the highest-quality trials published within the last 15 years, and we did not detect evidence of availability bias. Second, most of the planned subgroup analyses were based on two to three of the included studies and therefore may still be underpowered due to the unavailability of data on relevant covariates and/or live birth. Some primary trials only included a relatively homogeneous ethnicity group, and therefore, IPD in such trials could not contribute to the analysis of treatment–ethnicity interaction as no within-trial interaction was available. Third, as treatment-resistant women were excluded from this IPD meta-analysis, the findings can be applied in clinical practice on the choice of first-line treatment only. Last, we planned a one-stage IPD meta-analysis in the protocol but decided to use a two-stage approach before the final analysis. A two-stage approach allows graphical presentations for both overall treatment effects and treatment–covariate interactions, which is important for clinical interpretation, while it is not obvious how best to present graphically the results of a one-stage model.[12] In addition, the two-stage approach automatically avoids ecological bias by accounting for within-trial interactions only.[12] Given the relatively large number of participants, low heterogeneity and overall good to moderate quality of included studies, we would expect both approaches to give very similar results.

Interpretations and clinical implications

The overall effects of letrozole and CC plus metformin vs CC on live birth and clinical pregnancy in this IPD meta-analysis were in agreement with existing systematic reviews[3,17,18] as well as the most recent international evidence-based guideline recommendations.[4] Based on the findings of this IPD meta-analysis, letrozole can be recommended as the first-line ovulation induction medication in women with PCOS and infertility, provided off-label use is allowed and women are fully informed. Compared to CC alone, CC plus metformin may increase clinical pregnancy rates but the evidence on live birth was insufficient. Sensitivity analysis showed that the treatment effects on live birth seemed very small. The discrepancies between clinical pregnancy and live birth were likely due to the bias arising from low quality of studies which did not report live birth. Further evidence is needed to address this question.

Subgroup analyses showed that women with higher baseline serum levels of total testosterone may benefit more from letrozole compared to CC and women with higher baseline serum levels of insulin may benefit more from CC plus metformin compared to CC alone. Such positive interactions were consistent across trials and supported from a biological perspective. Letrozole has been introduced as an ovulation induction agent since 2001, and it inhibits aromatase, therefore increasing gonadotropin secretion by release of the hypothalamic/pituitary axis from estrogenic negative feedback and resulting in stimulation of ovarian follicle development.[19] According to the recent 'two triangles hypothesis' for folliculogenesis in PCOS, pre-antral follicle growth is excessive due to intrinsic androgen excess that renders granulosa cells hypersensitive to FSH, with consequently excessive AMH expression.[20] Therefore, hyperandrogenaemia may improve the response to letrozole by enhancing the sensitivity of FSH receptors. However, such an interaction was not observed in other biomarkers of hyperandrogenaemia or hirsutism. This is likely due to the fact that the severity of hirsutism does not correlate well with the magnitude of androgen excess, as hirsutism is an expression of hyperandrogenism on hair follicles mediated through different pathways from those affecting the ovaries and follicles.[21] Metformin is an insulin-sensitising agent that decreases gluconeogenesis and lipogenesis and enhances peripheral glucose uptake and therefore increases insulin sensitivity.[22] The addition of metformin may further improve insulin resistance in women with higher fasting insulin or HOMA-IR levels and therefore improve pregnancy outcomes. We acknowledge that insulin levels are affected by many factors, ranging from physical activity and pre-test duration of fasting to sample handling and assay variability.[23] Therefore, the international evidence-based guideline does not recommend clinical measurement of insulin resistance at present due to the lack of accuracy.[4] In addition, SHBG has been proposed as a measure of insulin resistance,[23] but the findings in our IPD meta-analysis did not support treatment-by-SHBG interactions. Our work provides preliminary evidence that there may be a role for assessing insulin resistance in PCOS and infertility and supports the need to assess insulin resistance in infertility studies.

We did not find ethnicity differences on treatment effects. This could be partly due to self-reported ethnicity without objective or DNA validation in all trials. We also did not find other treatment–covariate interactions on live birth for other pre-specified covariates including age, BMI, primary/secondary infertility, treatment history, Ferriman–Gallwey score for hirsutism, SHBG, free androgen index, fasting glucose levels or ovarian volume. Although analyses of subgroup effects were pre-specified in the protocol, these results should still be considered exploratory due to multiplicity.

Time is an important measurement for infertility outcomes, especially in the assessment of the effectiveness of multi-cycle treatments. However, time-to-event outcomes have seldom been reported in meta-analyses of infertility trials as fertility outcomes are usually

considered as dichotomous outcomes and Kaplan–Meier curves are rarely presented. Our IPD meta-analysis used number of cycles as a measure of time and evaluated time-to-pregnancy by estimating HRs and presenting summary Kaplan–Meier curves. Time-to-event analysis takes time and censored participants into account and provides more accurate estimates of treatment effect. Our analyses on time-to-pregnancy were inconsistent with those of clinical pregnancy.
Research implications

Research implications

IPD meta-analyses are useful to inform the design, conduct, analysis and interpretation of trials.[24] Given the consistent treatment benefits of letrozole across different fertility outcomes, future trials investigating new interventions for PCOS should choose letrozole as the reference arm. New trials are encouraged to incorporate treatment selection markers in their design to guide treatment decision,[25] and the impact of these, including age, BMI and other biomarkers, needs to be confirmed in future trials. More specifically, biomarkers for hyperandrogenaemia and insulin resistance could be applied in trials that evaluate metformin. Due to the limited accuracy for measuring existing insulin resistance biomarkers, optimal methods to assess insulin resistance in future trials should also be considered.

Developing and implementing a core outcome set for infertility[26] and PCOS should be recommended to ensure outcomes are reported and collected consistently across future trials on infertility and PCOS to reduce research waste.

CONCLUSIONS

Our IPD meta-analysis shows that in women with PCOS, letrozole improves live birth and clinical pregnancy rates and reduces time-to-pregnancy compared to CC alone. CC plus metformin may improve clinical pregnancy rates and may reduce time-to-pregnancy compared to CC alone, but there is insufficient evidence of a difference on live birth.

Treatment effects of letrozole are influenced by baseline serum levels of total testosterone while those of CC plus metformin are affected by baseline serum levels of insulin. These interactions between treatments and biomarkers on hyperandrogenaemia and insulin resistance provide further insights into a personalised approach towards the clinical management of anovulatory infertility related to PCOS and therefore should be confirmed in future studies.

Acknowledgements

We would like to thank Mr M. Draper from Barr Smith Library, University of Adelaide, for his assistance in developing the search strategies and Dr M. H. Zafarmand from University of Amsterdam for assisting with the translation. We would like to acknowledge all the investigators and participants of the primary trials. The investigators of individual trials are listed in Supplementary Table SIV. We would like to acknowledge the assistance of NICHD, the Reproductive Medicine Network (RMN) and the Protocol Subcommittee, in making the database for PPCOS I and II available. +The authors of the Reproductive Medicine Network are R.S.L., R.G. Brzyski, M.P. Diamond, C. Coutifaris, W.D. Schlaff, P. Casson, G.M. Christman, H. Huang, Q. Yan, R. Alvero, D.J. Haisenleder, K.T. Barnhart, G.W. Bates, R. Usadi, S. Lucidi, V. Baker, J.C. Trussell, S.A. Krawetz, P. Snyder, D. Ohl, N. Santoro, H.X. Barnhart, B.R. Carr, S.A. Carson, M.P. Steinkampf, P.G. McGovern, N.A. Cataldo, G.G. Gosman, J.E. Nestler, L.C. Giudice, P.C. Leppert, E.R. Myers, E. Eisenberg and H. Zhang. The details of their affiliations and NIH Grants are listed in Supplementary Table SV.

Funding

An Australian government research training programme scholarship (to R.W.); Australian National Health and Medical Research Council-funded Centre for Research Excellence in Polycystic Ovary Syndrome (APP1078444).

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflict of interest

R.S.L. reports consultancy fees from AbbVie, Bayer, Fractyl and Ogeda and research sponsorship from Ferring. N.P.J. has received conference expenses from Bayer Pharma, Merck Serono and Merck, Sharp and Dohme (MSD) and a research funding from AbbVie and Myovant Sciences and is a consultant to Vifor Pharma, Guerbet and Myovant Sciences. W.V. has received conference expenses from Ferring and Merck Serono, and his department has received research funding from Ferring and Merck Serono. S.B. is Editor in Chief of Human Reproduction Open and receives an honorarium and support for travel to conferences from Oxford University Press for his role. R.J.N. has received grant funding from Ferring and conference support from Merck. B.W.M. is supported by a NHMRC Practitioner Fellowship (GNT1082548) and reports consultancy for ObsEva, Merck and Guerbet. The other authors have no conflict of interest to declare.

SUPPLEMENTAL DATA

Supplementary Table 1. List of included studies without IPD and reasons

List of included studies without IPD	Reasons
Aygen 2007; Badawy 2009; Badawy 2011; Fleming 2002; Keikha 2011; Khorram 2006; Mobusher 2014; Santonocito 2009; Tang 2006*; Zain 2009	Data loss (n = 10)
Moussa 2016	Legal reasons (n = 1)
Abuelghar 2013; Atay 2006; Ayaz 2013; Banerjee Ray 2012; Basirat 2012; Boostanfar 2001; Chen 2016; Dasari 2009; Dehbashi 2009; Hossein-Rashidi 2016; Jahan 2015; Karimzadeh 2007; Karimzadeh 2010; Lopez 2004; Lorzadeh 2011; Maged 2015; Robinson 2003; Roy 2012; Selim 2012; Seyedoshohadaei 2012; Sharief 2015; Sheikh-El-Arab Elsedeek 2011; Zeinalzadeh 2010	No response (n = 23)
Beigi 2006; Boudhraa 2010; Cudmore 1966; El-Biely 2001; Garcia 1985; Johnson 1966	IPD not sought due to insufficient contact information (n = 6)
Fatima 2018; Topçu 2017	IPD not sought as studies were identified after the data requesting timeline (n = 2)

*Note: Although IPD of baseline and other outcomes in this study were provided, IPD of outcomes of interest for this IPD meta-analysis were not available.

Supplementary Table 2. Sensitivity analyses for live birth

Comparison	Sensitivity analyses	Number of RCTs	Number of participants	Risk Ratio (RR)	95% confidence interval (CI)	I^2
Letrozole vs CC	RCTs with low risk of bias at allocation concealment	2	909	1.42	1.14-1.76	0
	Treatment naïve women with PCOS	3	627	1.41	1.11-1.79	0
CC+metformin vs CC	RCTs with low risk of bias at allocation concealment	3	714	1.02	0.76-1.37	33.2%
	Treatment naïve women with PCOS	5	662	1.06	0.83-1.34	3.9%

Supplementary Table 3. IPD availability bias

Comparison (vs. CC)	Network meta-analyses of RCTs with and without IPD RR (95% CI)	Network meta-analyses of RCTs with IPD RR (95% CI)	Network meta-analyses of RCTs without IPD RR (95% CI)
Live birth	27 RCTs	12 RCTs	15 RCTs
	5257 women	3437 women	1820 women
Placebo	0.58 (0.31-1.07)	0.56 (0.26-1.20)	NA
Metformin	0.90 (0.64-1.28)	0.87 (0.51-1.47)	0.95 (0.57-1.58)
CC + Metformin	1.27 (0.91-1.78)	1.18 (0.73-1.90)	1.70 (0.88-3.30)
Letrozole	1.46 (1.09-1.95)	1.42 (0.79-2.55)	1.47 (1.07-2.02)
Tamoxifen	1.16 (0.61-2.18)	NA	1.12 (0.65-1.94)
Gonadotrophins	1.31 (0.73-2.34)	1.22 (0.45-3.34)	1.45 (0.65-3.22)
Clinical pregnancy	62 RCTs	20 RCTs	42 RCTs
	9356 women	3962 women	5394 women
Placebo	0.49 (0.33-0.71)	0.61 (0.37-1.01)	0.30 (0.16-0.57)
Metformin	1.06 (0.83-1.34)	0.94 (0.67-1.34)	1.13 (0.80-1.59)
CC + Metformin	1.46 (1.21-1.76)	1.34 (1.02-1.76)	1.62 (1.23-2.13)
Letrozole	1.37 (1.16-1.61)	1.48 (1.07-2.05)	1.30 (1.08-1.58)
Tamoxifen	0.91 (0.66-1.25)	0.72 (0.26-1.95)	0.91 (0.65-1.26)
Gonadotrophins	1.34 (0.87-2.08)	1.22 (0.64-2.31)	1.57 (0.81-3.06)

This table shows the results of network meta-analyses of RCTs with IPD and network meta-analyses of all eligible RCTs on live birth and clinical pregnancy. The results are presented in the comparisons of different interventions versus CC for live birth and clinical pregnancy, respectively.

Supplementary Table 4. List of investigators of the primary RCTs

Primary RCTs	Investigators
CLET trial (Amer 2017)	S.A. Amer, J. Smith, A. Mahran, and P. Fox, A. Fakis
Bayar 2006	Ülkü Bayar, Mustafa Basaran, Sibel Kiran, Ayhan Coskun and Sener Gezer
COFFI trial (Homburg 2012)	R. Homburg, M.L. Hendriks, T.E. König, R.A. Anderson, A.H. Balen, M. Brincat, T. Child, M. Davies, T. D'Hooghe, A. Martinez, M. Rajkhowa, R. Rueda-Saenz, P. Hompes and C.B. Lambalk
PCOSMIC trial* (Johnson 2010)	N.P. Johnson, A.W. Stewart, J. Falkiner, C.M. Farquhar, S. Milsom, V.-P. Singh, Q.L. Okonkwo, K.L. Buckingham, REACT-NZ (REproductionAnd Collaborative Trials in New Zealand)
Kar 2012	Sujata Kar
Kar 2015	Sujata Kar and Smriti Sanchita
Leanza 2014	V Leanza, L Coco, F Grasso, G Leanza, G Zarbo, and M Palumbo.
PPCOS I trial (Legro 2007)	Richard S. Legro, Huiman X. Barnhart, William D. Schlaff, Bruce R. Carr, Michael P. Diamond, Sandra A. Carson, Michael P. Steinkampf, Christos Coutifaris, Peter G. McGovern, Nicholas A. Cataldo, Gabriella G. Gosman, John E. Nestler, Linda C. Giudice, Phyllis C. Leppert, and Evan R. Myers, for the Cooperative Multicenter Reproductive Medicine Network
PPCOS II trial (Legro 2014)	Richard S. Legro, Robert G. Brzyski, Michael P. Diamond, Christos Coutifaris, William D. Schlaff, Peter Casson, Gregory M. Christman, Hao Huang, Qingshang Yan, Ruben Alvero, Daniel J. Haisenleder, Kurt T. Barnhart, G. Wright Bates, Rebecca Usadi, Scott Lucidi, Valerie Baker, J.C. Trussell, Stephen A. Krawetz, Peter Snyder, Dana Ohl, Nanette Santoro, Esther Eisenberg, and Heping Zhang, for the NICHD Reproductive Medicine Network
Liu 2017	Chang Liu, Guimei Feng, Wei Huang, Qiuyi Wang, Shiyuan Yang, Jing Tan, Jing Fu and Dong Liu
Lord 2006	J Lord, R Thomas, B Fox, U Acharya and T Wilkin
Moll 2006	Etelka Moll, Patrick M MBossuyt, Johanna C Korevaar, Cornelis B Lambalk, and Fulco van der Veen,
Morin-Papunen 2012	Laure Morin-Papunen, Anni S. Rantala, Leila Unkila-Kallio, AilaTiitinen, Maritta-Hippeläinen, Antti Perheentupa, Helena Tinkanen, RistoBloigu, Katri Puukka, AimoRuokonen and Juha S. Tapanainen
Nazik 2012	Hakan Nazik and YakupKumtepe
Palomba 2005	Stefano Palomba, Francesco Orio, Jr., Angela Falbo, Francesco Manguso, Tiziana Russo, Teresa Cascella, Achille Tolino, Enrico Carmina, Annamaria Colao and Fulvio Zullo
Sahin 2004	Yılmaz Şahin, Ünal Yirmibeş, Fahrettin Keleştimur and Ercan Aygen
Vegetti 1999	W. Vegetti, A. Riccaboni, M. Columbo, E. Baroni, D. Diaferia, G. Ragni and P.G. Crosignani.
Williams 2009	C. D. Williams, L. M. Pastore, W. B.Shelly, A. P.Bailey, D. C.Baras and B. G.Bateman,
PCOSAct trial (Wu 2017)	Xiao-Ke Wu, ElisabetStener-Victorin, Hong-Ying Kuang, Hong-Li Ma, Jing-Shu Gao, Liang-Zhen Xie, Li-Hui Hou, Zhen-Xing Hu, Xiao-Guang Shao, Jun Ge, Jin-Feng Zhang, Hui-Ying Xue, Xiao-Feng Xu, Rui-Ning Liang, Hong-Xia Ma, Hong-Wei Yang, Wei-Li Li, Dong-Mei Huang, Yun Sun, Cui-Fang Hao, Shao-Min Du, Zheng-Wang Yang, Xin Wang, Ying Yan, Xiu-Hua Chen, Ping Fu, Cai-Fei Ding, Ya-Qin Gao, Zhong-Ming Zhou, Chi Chiu Wang, Tai-Xiang Wu, Jian-Ping Liu, Ernest H. Y. Ng, Richard S. Legro and Heping Zhang, for the PCOSAct Study Group

*Note: This publication included two studies, both of which were included.

Supplementary Table 5. Eunice Kennedy Shriver National Institutes of Child Health and Human Development, Reproductive Medicine Network

Names	Affiliations	Support by NIH Grants
Richard S. Legro, M.D.	Pennsylvania State University College of Medicine, Hershey, PA	U10 HD27049, U10 HD38992, U10 HD055925, U10 HD39005, U10 HD38998, U10 HD055936, U10 HD055942, U10 HD055944, U54 HD29834, UL1 TR000127, U01 HD38997, U10 HD27011, U10 HD33172, U10 HD38988, U10 HD38999, MO1RR00056, MO11RR10732, C06 RR016-499
Robert G. Brzyski, M.D., Ph.D.	University of Texas Health Science Center at San Antonio, San Antonio	
Michael P. Diamond, M.D.	Georgia Regents University, Augusta; Wayne State Univeristy, Detroit	
Christos Coutifaris, M.D., Ph.D.	University of Pennsylvania School of Medicine, Philadelphia	
William D. Schlaff, M.D.	University of Colorado, Aurora	
Peter Casson, M.D.	University of Vermont, Burlington	
Gregory M. Christman, M.D.	University of Michigan, Ann Arbor	
Hao Huang, M.D., M.P.H.	Yale University School of Public Health, New Haven, CT	
Qingshang Yan, Ph.D.	Yale University School of Public Health, New Haven, CT	
Ruben Alvero, M.D.	University of Colorado, Aurora	
Daniel J. Haisenleder, Ph.D.	Ligand Core Lab, Univ of Virginia Center for Research in Reproduction, Charlottesville	
Kurt T. Barnhart, M.D.	University of Pennsylvania School of Medicine, Philadelphia	
G. Wright Bates, M.D.	University of Alabama at Birmingham, Birmingham	
Rebecca Usadi, M.D.	Carolinas Medical Center, Charlotte, NC	
Scott Lucidi, M.D.	Virginia Commonwealth University, Richmond	
Valerie Baker, M.D.	Stanford University Medical Center, Stanford, CA	
J.C. Trussell, M.D.	State University of New York Upstate Medical University, Onondaga	
Stephen A. Krawetz, Ph.D.	Wayne State Univeristy, Detroit	
Peter Snyder, M.D.	University of Pennsylvania School of Medicine, Philadelphia	
Dana Ohl, M.D.	University of Michigan, Ann Arbor	
Nanette Santoro, M.D.	University of Colorado, Aurora	
Huiman X. Barnhart, Ph.D.	Duke University Medical Center, Durham, NC	
Bruce R. Carr, M.D.	University of Texas Southwestern Medical Center, Dallas	
Michael P. Diamond, M.D.	Wayne State Univeristy, Detroit	
Sandra A. Carson, M.D.	Baylor College of Medicine, Houston	
Michael P. Steinkampf, M.D.	University of Alabama, Birmingham	
Peter G. McGovern, M.D.	University of Medicine and Dentistry of New Jersey, Newark	
Nicholas A. Cataldo, M.D.	Stanford University, Stanford, CA	
Gabriella G. Gosman, M.D.	University of Pittsburgh, Pittsburgh	
John E. Nestler, M.D.	Virginia Commonwealth University School of Medicine, Richmond	
Linda C. Giudice, M.D., Ph.D.	University of California at San Francisco, San Francisco	
Phyllis C. Leppert, M.D., Ph.D.	National Institute of Child Health and Human Development, Bethesda, MD	
Evan R. Myers, M.D., M.P.H.	Duke University Medical Center, Durham	
Esther Eisenberg, M.D., M.P.H.	Fertility and Infertility Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, MD	
Heping Zhang, Ph.D.	Yale University School of Public Health, New Haven, CT	

REFERENCES TO STUDIES INCLUDED IN THIS REVIEW

1. Abuelghar WM, Elkady OS, Khamees AA. Clomiphene citrate alone, in combination with metformin or in combination with pioglitazone as first line therapy in induction of ovulation in infertile women with polycystic ovary syndrome, a randomized controlled trial. *Middle East Fertil Soc J* 2013;18:135–141.
2. Amer SA, Smith J, Mahran A, Fox P, Fakis A. Double-blind ran-domized controlled trial of letrozole versus clomiphene citrate in subfertile women with polycystic ovarian syndrome. *Hum Reprod* 2017;32:1631–1638.
3. Atay V, Cam C, Muhcu M, Cam M, Karateke A. Comparison of letrozole and clomiphene citrate in women with polycystic ovaries undergoing ovarian stimulation. *J Int Med Res* 2006;34 73–76.
4. Ayaz A, Alwan Y, Farooq MU. Metformin-clomiphene citrate vs. clomiphene citrate alone: polycystic ovarian syndrome. *J Hum Reprod Sci* 2013;6:15–18.
5. Aygen EM, Güzel Z, Özgün T, Atakul T, Şahin Y. The use of letrozole for ovulation induction in infertile women with polycystic ovarian syndrome [in Turkish]. *Erciyes Tip Dergisi* 2007;29:195–200.
6. Badawy A, Abdel Aal I, Abulatta M. Clomiphene citrate or letro- zole for ovulation induction in women with polycystic ovarian syndrome: a prospective randomized trial. *Fertil Steril* 2009;92: 849–852.
7. Badawy A, Gibreal A. Clomiphene citrate versus tamoxifen for ovula- tion induction in women with PCOS: a prospective randomized trial. *Eur J Obstet Gynecol Reprod Biol* 2011;159:151–154.
8. Banerjee Ray P, Ray A, Chakraborti PS. Comparison of efficacy of letrozole and clomiphene citrate in ovulation induction in Indian women with polycystic ovarian syndrome. *Arch Gynecol Obstet* 2012;285:873–877.
9. Basirat Z, Kashifard M, Amiri MG. Enhanced ovarian folliclular devel- opment by metformin does not correlate with pregnancy rate: a randomized trial. *Int J Fertil Steril* 2012;6:31–36.
10. Bayar U, Basaran M, Kiran S, Coskun A, Gezer S. Use of an aromatase inhibitor in patients with polycystic ovary syndrome: a prospective randomized trial. *Fertil Steril* 2006;86:1447–1451.
11. Boostanfar R, Jain JK, Mishell DR Jr, Paulson RJ. A prospective ran- domized trial comparing clomiphene citrate with tamoxifen citrate for ovulation induction. *Fertil Steril* 2001;75:1024–1026.
12. Chen Z, Zhang M, Qiao Y, Yang J. Effects of letrozole in combina- tion with low-dose intramuscular injection of human menopausal gonadotropin on ovulation and pregnancy of 156 patients with polycystic ovary syndrome. *Pak J Med Sci* 2016;32:1434–1438.
13. Dasari P, Pranahita GK. The efficacy of metformin and clomiphene citrate combination compared with clomiphene citrate alone for ovulation induction in infertile patients with PCOS. *J Hum Reprod Sci* 2009;2:18–22.
14. Dehbashi S, Dehbashi S, Kazerooni T, Robati M, Alborzi S, Parsanezhad ME, Shadman A. Comparison of the effects of letrozole and clomiphene citrate on ovulation and pregnancy rate in patients with polycystic ovary syndrome. *Iran J Med Sci* 2009;34:23–28.

15. Fleming R, Hopkinson ZE, Wallace AM, Greer IA, Sattar N. Ovarian function and metabolic factors in women with oligomenorrhea treated with metformin in a randomized double blind placebo-controlled trial. *J Clin Endocrinol Metab* 2002;87:569–574.
16. Homburg R, Hendriks ML, Konig TE, Anderson RA, Balen AH, Brincat M, Child T, Davies M, D'Hooghe T, Martinez A *et al.* Clomifene citrate or low-dose FSH for the first-line treatment of infertile • women with anovulation associated with polycystic ovary syn-drome: a prospective randomized multinational study. *Hum Reprod* 2012;27:468–473.
17. Hossein-Rashidi B, Khandzad B, Shahrokh-Tehraninejad E, Bagheri M, Gorginzadeh M. Recombinant FSH compared to clomiphene citrate as the first-line in ovulation induction in polycystic ovary syndrome using newly designed pens: a randomized controlled trial. *J Family Reprod Health* 2016;10:42–48.
18. Jahan S. Comparative study of efficacy among metformin, clomiphene citrate and aromatase inhibitor (letrozole) as the first-line medication for ovulation induction, achievement of pregnancy and live birth in Asian women with polycystic ovarian syndrome: a prospective trial. *Int J Gynecol Obstet* 2015;131:E503.
19. Johnson NP, Stewart AW, Falkiner J, Farquhar CM, Milsom S, Singh VP, Okonkwo QL, Buckingham KL, React-Nz am-cftg. PCOSMIC: a multi-centre randomized trial in women with polycystic ovary syndrome evaluating metformin for infertility with clomiphene. *Hum Reprod* 2010;25: 1675–1683.
20. Kar S. Clomiphene citrate or letrozole as first-line ovulation induction drug in infertile PCOS women: a prospective randomized trial. *J Hum Reprod Sci* 2012;5:262–265.
21. Kar S, Sanchita S. Clomiphene citrate, metformin or a combination of both as the first line ovulation induction drug for Asian Indian women with polycystic ovarian syndrome: a randomized controlled trial. *J Hum Reprod Sci* 2015;8:197–201.
22. Karimzadeh MA, Eftekhar M, Taheripanah R, Tayebi N, Sakhavat L, Zare F. The effect of administration of metformin on lipid profile changes and insulin resistance in patients with polycystic ovary syn- drome. *Middle East Fertil Soc J* 2007;12:174–178.
23. Karimzadeh MA, Javedani M. An assessment of lifestyle modification versus medical treatment with clomiphene citrate, metformin, and clomiphene citrate-metformin in patients with polycystic ovary syn- drome. *Fertil Steril* 2010;94:216–220.
24. Keikha F, Shahraki MB. Induction ovulation in polycystic ovary patient with clomiphene citrate and letrozole. *Iran J Reprod Med* 2011;9:46.
25. Khorram O, Helliwell JP, Katz S, Bonpane CM, Jaramillo L. Two weeks of metformin improves clomiphene citrate-induced ovulation and metabolic profiles in women with polycystic ovary syndrome. *Fertil Steril* 2006;85:1448–1451.
26. Leanza V, Coco L, Grasso F, Leanza G, Zarbo G, Palumbo M. Ovulation induction with clomiphene citrate and metformin in women with polycystic ovary syndrome. *Minerva Ginecol* 2014;66:299–301.
27. Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, Coutifaris C, McGovern PG, Cataldo NA *et al.* Clomiphene, metformin, or both for infertility in the polycystic

ovary syndrome. *N Engl J Med* 2007;356:551–566.

28. Legro RS, Brzyski RG, Diamond MP, Coutifaris C, Schlaff WD, Casson P, Christman GM, Huang H, Yan Q, Alvero R *et al.* Letrozole versus clomiphene for infertility in the polycystic ovary syndrome. *N Engl J Med* 2014;371:119–129.
29. Liu C, Feng G, Huang W, Wang Q, Yang S, Tan J, Fu J, Liu D. Compar- ison of clomiphene citrate and letrozole for ovulation induction in women with polycystic ovary syndrome: a prospective randomized trial. *Gynecol Endocrinol* 2017;1–5.
30. Lopez E, Gunby J, Daya S, Parrilla JJ, Abad L, Balasch J. Ovulation induction in women with polycystic ovary syndrome: randomized trial of clomiphene citrate versus low-dose recombinant FSH as first line therapy. *Reprod Biomed Online* 2004;9:382–390.
31. Lord J, Thomas R, Fox B, Acharya U, Wilkin T. The effect of metformin on fat distribution and the metabolic syndrome in women with polycystic ovary syndrome–a randomised, double-blind, placebo- controlled trial. *BJOG* 2006;113:817–824.
32. Lorzadeh N, Kazemirad S, Mohammadi Z. Comparison of effects letrozole and clomiphene citrate for ovulation induction in women with polycystic ovary syndrome [in Persian]. *Iran J Obstet Gynecol Infertil* 2011;14.
33. Maged AM, Elsawah H, Abdelhafez A, Bakry A, Mostafa WAI. The adjuvant effect of metformin and N-acetylcysteine to clomiphene citrate in induction of ovulation in patients with polycystic ovary syndrome. *Gynecol Endocrinol* 2015.
34. Mobusher I. Comparison of the efficacy of letrozole and clomiphene citrate for ovulation induction in infertile women with polycystic ovary syndrome. *Pak J Med Health Sci* 2014;8:905–908.
35. Moll E, Bossuyt PM, Korevaar JC, Lambalk CB, van der Veen F. Effect of clomifene citrate plus metformin and clomifene citrate plus placebo on induction of ovulation in women with newly diagnosed polycystic ovary syndrome: randomised double blind clinical trial. *BMJ* 2006;332:1485.
36. Morin-Papunen L, Rantala AS, Unkila-Kallio L, Tiitinen A, Hippelainen M, Perheentupa A, Tinkanen H, Bloigu R, Puukka K, Ruokonen A *et al.* Metformin improves pregnancy and live-birth rates in women with polycystic ovary syndrome (PCOS): a multicenter, double-blind, placebo-controlled randomized trial. *J Clin Endocrinol Metab* 2012;97:1492–1500.
37. Moussa AA, Torky H, Dief O, Elwahed AA, Senna HA. The effect of clomiphene citrate versus tamoxifen versus letrozol on endometrial thickness and blood flow in ovulation induction in women with polycystic ovaries. *Acta Medica Int* 2016;3:88–92.
38. Nazik H, Kumtepe Y. Comparison of efficacy of letrozole and clomiphene citrate in ovulation induction for women with polycystic ovarian syndrome. *HealthMED* 2012;6:879–883.
39. Palomba S, Orio F Jr, Falbo A, Manguso F, Russo T, Cascella T, Tolino A, Carmina E, Colao A, Zullo F. Prospective parallel randomized, double-blind, double-dummy controlled clinical trial comparing clomiphene citrate and metformin as the first-line treatment for ovulation induction in nonobese anovulatory women with polycystic ovary syndrome. *J Clinical Endocrinol Metab* 2005;90: 4068–4074.

40. Robinson R, Swezey M, Propst A, Bates G. Metformin added to clomiphene citrate does not improve pregnancy rates in hyperandrogenic, chronic anovulatory women: a randomized trial. *Fertil Steril* 2003;80:S273–S274 Abstract no: P.
41. Roy K, Baruah J, Singla S, Sharma J, Singh N, Jain S, Goyal M. A prospective randomized trial comparing the efficacy of letrozole and clomiphene citrate in induction of ovulation in polycystic ovarian syndrome. *J Hum Reprod Sci* 2012;5:20–25.
42. Sahin Y, Yirmibes U, Kelestimur F, Aygen E. The effects of metformin on insulin resistance, clomiphene-induced ovulation and pregnancy rates in women with polycystic ovary syndrome. *Eur J Obstet Gynecol Reprod Biol* 2004;113:214–220.
43. Santonocito V, Rapisarda V, Abruzzo SRM, Pollicino R, Coco L, Zarbo G. Comparison between clomiphene citrate and metformin for induction of ovulatory cycles in infertile nonobese women with polycystic ovary syndrome [in Italian]. *Giornale Italiano di Ostetricia e Ginecologia* 2009;31:455–460.
44. Selim MF, Borg TF. Letrozole and clomiphene citrate effect on endometrial and subendometrial vascularity in treating infertility in women with polycystic ovary syndrome. *J Gynecol Surg* 2012;28: 405–410.
45. Seyedoshohadaei F, Zandvakily F, Shahgeibi S. Comparison of the effectiveness of clomiphene citrate, tamoxifen and letrozole in ovu- lation induction in infertility due to isolated unovulation. *Iran J Reprod Med* 2012;10:531–536.
46. Sharief M, Nafee NR. Comparison of letrazole and clomiphene citrate in women with polycystic ovaries undergoing ovarian stimulation. *J Pak Med Assoc* 2015;65:1149–1152.
47. Sheikh-El-Arab Elsedeek M, Elmaghraby HAH. Predictors and char- acteristics of letrozole induced ovulation in comparison with clomiphene induced ovulation in anovulatory PCOS women. *Middle East Fertil Soc J* 2011;16:125–130.
48. 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.
49. Vegetti W, Riccaboni A, Columbo M, Baroni E, Diaferia D, Ragni . G, Crosignani PG. Randomized study of induction of ovulation by two different molecules with antioestrogenic effects, in patients with chronic anovulation disorders. 1999, pp. S234–S235.
50. Williams CD, Pastore LM, Shelly WB, Bailey AP, Baras DC, Bateman BG. A randomized, placebo-controlled study of the influence of instant-release metformin on response to clomiphene citrate and time to conception in polycystic ovary syndrome. *Fertil Steril* 2009;92:S105.
51. Wu XK, Stener-Victorin E, Kuang HY, Ma HL, Gao JS, Xie LZ, Hou LH, Hu ZX, Shao XG, Ge J *et al.* Effect of acupuncture and clomiphene in Chinese women with polycystic ovary syndrome: a randomized clinical trial. *JAMA* 2017;317:2502–2514.
52. Zain MM, Jamaluddin R, Ibrahim A, Norman RJ. Comparison of clomiphene citrate, metformin, or the combination of both for first-line ovulation induction, achievement of pregnancy, and live birth in Asian women with polycystic ovary syndrome: a randomized controlled trial. *Fertil Steril* 2009;91:514–521.

53. Zeinalzadeh M, Basirat Z, Esmailpour M. Efficacy of letrozole in ovulation induction compared to that of clomiphene citrate in patients with polycystic ovarian syndrome. *J Reprod Med* 2010;55: 36–40.

REFERENCES TO STUDIES EXCLUDED FROM THIS REVIEW

1. Beigi A. Randomized trial comparing clomiphene citrate and metformin as the first-line treatment for ovulation induction in polycystic ovary syndrome. *Hum Reprod* 2006;21:i129.
2. Boudhraa K, Jellouli MA, Amri M, Farhat M, Torkhani F, Gara MF. Indication of metformin in the management of hormonal dysfunction secondary to polycystic ovarian syndrome: prospective comparative study of 63 cases [in French]. *Tunis Med* 2010;88:335–340.
3. Cudmore DW, Tupper WR. Induction of ovulation with clomiphene citrate. A double-blind study. *Fertil Steril* 1966;17:363–373.
4. El-Biely MM, Habba M. The use of metformin to augment the induction of ovulation in obese infertile patients with polycystic ovary syndrome. *Mid East Fertil Soc J* 2001;6:43–49.
5. Fatima A, Khan SA, Saifuddin Z, Aslam R. Comparison of efficacy of clomiphene citrate alone and with metformin for treatment of infertility in polycystic ovarian syndrome. *Rawal Med J* 2018;43:285–288.
6. Garcia CR, Freeman EW, Rickels K, Wu C, Scholl G, Galle PC, Boxer AS. Behavioral and emotional factors and treatment responses in a study of anovulatory infertile women. *Fertil Steril* 1985;44: 478–483.
7. Johnson JE Jr, Cohen MR, Goldfarb AF, Rakoff AE, Kistner RW, Plotz EJ, Vorys N. The efficacy of clomiphene citrate for induction of ovulation. A controlled study. *Int J Fertil* 1966;11:265–270.
8. Topçu HO, Batiog Batioğlu AS, İslimye M. Tamoxifen versus clomiphene citrate for ovulation induction in women with polycystic ovary syndrome: a prospective randomized trial. *J Reprod Med* 2017; 62:507–512.

ADDITIONAL REFERENCES

1. Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril* 2016;106:6–15.
2. 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.
3. Wang R, Kim BV, van Wely M, Johnson NP, Costello MF, Zhang H, Ng EH, Legro RS, Bhattacharya S, Norman RJ *et al*. Treatment strategies for women with WHO group II anovulation: systematic review and network meta-analysis. *BMJ* 2017;356:j138.
4. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ, International

PN. Recommendations from the international evidence-based guideline for the assess-ment and management of polycystic ovary syndrome. *Hum Reprod* 2018;33:1602–1618.
5. Wang R, Mol BW. The Rotterdam criteria for polycystic ovary syndrome: evidence-based criteria? *Hum Reprod* 2017;32:261–264.
6. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
7. Eshre Capri Workshop Group. Protect us from poor-quality medical research. *Hum Reprod* 2018;33:770–776.
8. Broeze KA, Opmeer BC, van der Veen F, Bossuyt PM, Bhattacharya S, Mol BW. Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. *Hum Reprod Update* 2010;16:561–567.
9. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, Tierney JF, Group P-ID. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA- IPD statement. *JAMA* 2015;313:1657–1665.
10. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011. The Cochrane Collaboration.
11. Fisher DJ. Two-stage individual participant data meta-analysis and generalized forest plots. *Stata J* 2015;15:369–396.
12. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? *BMJ* 2017;356:j573.
13. Riley RD, Jackson D, Salanti G, Burke DL, Price M, Kirkham J, White IR. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* 2017;358:j3932.
14. White IR. Network meta-analysis. *Stata Journal* 2015;15:951–985.
15. Fatima A, Khan SA, Saifuddin Z, Aslam R. Comparison of efficacy of clomiphene citrate alone and with metformin for treatment of infertility in polycystic ovarian syndrome. *Rawal Med J* 2018;43:285–288.
16. Topçu HO, Batioğlu AS, İslimye M. Tamoxifen versus clomiphene citrate for ovulation induction in women with polycystic ovary syndrome: a prospective randomized trial. *J Reprod Med* 2017; 62:507–512.
17. Franik S, Eltrop SM, Kremer JA, Kiesel L, Farquhar C. Aromatase inhibitors (letrozole) for subfertile women with polycystic ovary syndrome. *Cochrane Database Syst Rev* 2018;5:CD010287.
18. Morley LC, Tang T, Yasmin E, Norman RJ, Balen AH. Insulin-sensitising drugs (metformin, rosiglitazone, pioglitazone, D-chiro-inositol) for women with polycystic ovary syndrome, oligo amenorrhoea and subfertility. *Cochrane Database Syst Rev* 2017;11:CD003053.
19. Mitwally MF, Casper RF. Use of an aromatase inhibitor for induction of ovulation in patients with an inadequate response to clomiphene citrate. *Fertil Steril* 2001;75:305–309.
20. Dewailly D, Robin G, Peigne M, Decanter C, Pigny P, Catteau-Jonard S. Interactions between androgens, FSH, anti-Mullerian hormone and estradiol during folliculogenesis in the human normal and polycystic ovary. *Hum Reprod Update* 2016;22:709–724.

21. Escobar-Morreale HF, Carmina E, Dewailly D, Gambineri A, Keles- timur F, Moghetti P, Pugeat M, Qiao J, Wijeyaratne CN, Witchel SF *et al.* Epidemiology, diagnosis and management of hirsutism: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome Society. *Hum Reprod Update* 2012;18:146–170.
22. Naderpoor N, Shorakae S, de Courten B, Misso ML, Moran LJ, Teede HJ. Metformin and lifestyle modification in polycystic ovary syndrome: systematic review and meta-analysis. *Hum Reprod Update* 2015;21:560–574.
23. Cassar S, Misso ML, Hopkins WG, Shaw CS, Teede HJ, Stepto NK. Insulin resistance in polycystic ovary syndrome: a systematic review and meta-analysis of euglycaemic-hyperinsulinaemic clamp studies. *Hum Reprod* 2016;31:2619–2631.
24. Tierney JF, Pignon JP, Gueffyier F, Clarke M, Askie L, Vale CL, Burdett S, Cochrane IPDM-aMG. How individual participant data meta analyses have influenced trial design, conduct, and analysis. *J Clin Epidemiol* 2015;68:1325–1335.
25. Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. *Ann Intern Med* 2011;154:253–259.
26. Duffy JMN, Bhattacharya S, Curtis C, Evers JLH, Farquharson RG, Franik S, Khalaf Y, Legro RS, Lensen S, Mol BW *et al.* A protocol developing, disseminating and implementing a core outcome set for infertility. *Hum Reprod Open 2018* 2018;hoy007.

Chapter 7

To share or not to share data: how valid are trials evaluating first-line ovulation induction for polycystic ovary syndrome?

Bordewijk EM, Wang R, van Wely M, Costello MF, Norman RJ, Teede H, Gurrin LC, Mol BW, Li W.

Human Reproduction Update. 2020 Nov

ABSTRACT

Background

In our recent individual participant data (IPD) meta-analysis evaluating the effectiveness of first-line ovulation induction for polycystic ovary syndrome (PCOS), IPD were only available from 20 studies of 53 randomized controlled trials (RCTs). We noticed that the summary effect sizes of meta-analyses of RCTs without IPD sharing were different from those of RCTs with IPD sharing. Granting access to IPD for secondary analysis has implications for promoting fair and transparent conduct of RCTs. It is, however, still common for authors to choose to withhold IPD, limiting the impact of and confidence in the results of RCTs and systematic reviews based on aggregate data.

Objective and rationale

We performed a meta-epidemiologic study to elucidate if RCTs without IPD sharing have lower quality and more methodological issues than those with IPD sharing in an IPD meta-analysis evaluating first-line ovulation induction for PCOS.

Search methods

We included RCTs identified for the IPD meta-analysis. We dichotomized RCTs according to whether they provided IPD (shared group) or not (non-shared group) in the IPD meta-analysis. We restricted RCTs to full-text published trials written in English. We assessed and compared RCTs in the shared and non-shared groups on the following criteria: Risk of Bias (RoB 2.0), GRADE approach, adequacy of trial registration; description of statistical methods and reproducibility of univariable statistical analysis; excessive similarity or difference in baseline characteristics that is not compatible with chance; and other miscellaneous methodological issues.

Outcomes

In total, 45 trials (8697 women) were included in this study. IPD were available from 17 RCTs and 28 trials were categorized as the non-shared IPD group. Pooled risk rates obtained from the shared and non-shared groups were different. Overall low risk of bias was associated with 13/17 (76%) of shared RCTs versus 7/28 (25%) of non-shared RCTs. For RCTs that started recruitment after 1 July 2005, adequate trial registration was found in 3/9 (33%) of shared IPD RCTs versus 0/16 (0%) in non-shared RCTs. In total, 7/17 (41%) of shared RCTs and 19/28 (68%) of non-shared RCTs had issues with the statistical methods described. The median (range) of inconsistency rate per study, between reported and reproduced analyses for baseline variables, was 0% (0-92%) (6 RCTs applicable) in the shared group and 54% (0-100%) (13 RCTs applicable) in the non-shared group. The median (range) of inconsistency rate of univariable statistical results for the outcome(s) per study was 0% (0-63%) (14 RCTs applicable) in the shared group

and 44% (0-100%) (24 RCTs applicable) in the non-shared group. The distributions of simulation-generated P-values from comparisons of baseline continuous variables between intervention and control arms suggested that RCTs in the shared group are likely to be consistent with properly conducted randomization (P = 0.163), whereas this was not the case for the RCTs in the non-shared group (P = 4.535 × 10-8).

Wider implications

IPD meta-analysis on evaluating first-line ovulation induction for PCOS preserves validity and generates more accurate estimates of risk than meta-analyses using aggregate data, which enables more transparent assessments of benefits and risks. The availability of IPD and the willingness to share these data may be a good indicator of quality, methodological soundness and integrity of RCTs when they are being considered for inclusion in systematic reviews and meta-analyses.

INTRODUCTION

Polycystic ovary syndrome (PCOS) is a common endocrine disorder that may result in anovulatory infertility. Effective first-line ovulation induction is vital for the treatment of infertility in women troubled by this disease. Several meta-analyses on this topic using aggregate data from randomized controlled trials (RCTs) have been performed[1,2] and evidence-based guidelines have been released based on multiple meta-analyses and systematic reviews.[3]

In our recent individual participant data (IPD) meta-analysis evaluating the effectiveness of first-line ovulation induction for PCOS,[4] IPD were available from 20 RCTs, whereas IPD from 34 RCTs were not available, due to no response (n = 23), data loss (n = 10) or legal reasons (n = 1). While performing the IPD meta-analysis,[4] we noticed the summary effect sizes of meta-analyses of RCTs not providing IPD were different from those of RCTs that provided IPD. This raised concern about the comparability of RCTs with and without IPD sharing.

The lack of availability of IPD from individual trials is a common issue that troubles IPD meta-analyses,[5] and may limit the validity and precision of the results of evidence synthesis. The concerns expressed and reasons given by teams of investigators that refuse or are unable to share IPD include the heterogeneity and complexity of the contributing RCTs, administrative difficulties, lack of agreement on the purpose of sharing, the need for confidentiality and secure storage of data, data quality issues, ethical or ownership restrictions, failure to retain trial data and other personal considerations.[6,7] The willingness to share IPD may depend on the study's type of funding, sample size, risk of bias and the magnitude of the estimated treatment effect.[8] Other trials may not have retained data sufficiently well for subsequent use due to poor storage policies or investigators moving away from the centre that retains the data. The attitude of researchers' institutions towards data sharing is also important.

Little is known about the comparability between shared and non-shared IPD RCTs with regard to quality and integrity. RCTs with IPD sharing are usually performed better compared to non-shared RCTs on a quantitative 'risk of bias' assessment.[9] There has not been a head-to-head comparison between shared and non-shared RCTs contributing data to the same set of IPD meta-analyses in terms of methodological issues, which potentially endangers the robustness of synthesized evidence.

In this meta-epidemiological study, our aim was to try to determine if there is a difference in the frequency and magnitude of methodological and integrity issues between shared and non-shared RCTs, especially those that failed to respond in data enquiry.

Such a difference, if it exists, may assist in understanding whether the risk of publishing disproportionately large or small estimates of effect size is associated with the provision of data for inclusion in IPD meta-analysis.

METHODS

This meta-epidemiological study was prospectively registered at Center for Open Science 'OSF REGISTRIES' open registries network (10.17605/OSF.IO/RCJAV) before data extraction. The conduct of this study followed a pre-defined protocol.

Inclusion of RCTs

We included RCTs that were identified systematically in the IPD meta-analysis evaluating the effectiveness of first-line ovulation induction for PCOS.[4] We distinguished between RCTs where IPD were shared (shared group) and RCTs where IPD were not shared (non-shared group). According to the stated reason for not being able to share the minimum IPD to contribute to the IPD meta-analysis, the non-shared group was subdivided into no response following enquiry and data loss or legal reasons subgroups. We restricted the study to full-text published trials written in English. Trials that do not have full-text were excluded because they could not undergo the assessments of methodological issues used in this study.

Data extraction

We extracted information regarding author, year of publication, country/region, number of study centres, number of participants, definitions of treatment and control protocols, study start and end dates, ethics approval, registration (status and date), date of submission to journal, statement of ethics approval, statistical methods planned, used and reported, trial withdrawn (and reason if given), summary statistics of baseline characteristics and univariable analyses. We used the results of the GRADE (Grading of Recommendations, Assessment, Development and Evaluations) approach for the IPD meta-analysis[4] in this study.

Data synthesis

We performed random effects meta-analyses using aggregate data for the pre-defined comparisons letrozole versus clomiphene citrate (CC) and CC plus metformin versus CC in the shared and non-shared groups, respectively. For live birth and clinical pregnancy, we calculated risk ratios (RRs) and 95% CIs and presented statistical heterogeneity by using I^2 statistic.[10] We conducted all analyses on the basis of intention-to-treat, with women allocated to the treatment group to which they were randomized at the outset as the unit of analysis.[4]

Quality assessments

We conducted the following analyses in the shared and non-shared groups separately and compared the results with statistical methods whenever appropriate.

Assessment of risk of bias of individual studies.

Two review authors (E.M.B. and R.W.) independently assessed risk of bias for each eligible study by using the new Cochrane 'Risk of Bias 2' assessment tool[11] which included five domains: bias arising from the randomization process; bias due to deviations from intended interventions; bias due to missing outcome data; bias in measurement of the outcome; and bias in selection of the reported result. Disagreements were resolved by discussion with a third review author (W.L.).

We also produced a figure to visualize the chronological pattern of overall risk of bias. The intention of this is to investigate the hypothesis that the quality of trials that are published later is better, potentially because they have the opportunity to build upon data and results from earlier published trials.

Overall certainty of evidence: the GRADE approach.

We used the GRADE approach to assess the overall certainty of evidence in two pre-defined treatment comparisons (letrozole vs CC and CC plus metformin vs CC) for the outcomes live birth and clinical pregnancy.[12] As this assessment was performed separately in the shared and non-shared groups, we did not downgrade confidence in evidence due to publication bias.

Trial registration.

We applied this assessment to RCTs that initiated recruitment after 1 July 2005 as the requirement of compulsory prospective registration for any clinical trials starting enrolment after 1 July 2005 was not launched until 2004.[13] We obtained the trial registration number(s) from publication(s), or from searching of the World Health Organization (WHO) and International Standard Randomized Controlled Trial Number (ISRCTN) registries for trials conducted by the same research team and/or organization. Trial registration status was categorized into one of four groups, with registrations occurring: prior to commencing recruitment or within 6 months of initiating recruitment (adequate); after 6 months of initiating recruitment but before the completion of recruitment (late); trials registered after completing recruitment (retrospective); or not registered at all (absent). Where the recruitment period was not clear, the start and end dates of the study captured in the registration database were used. Late, retrospective and absent registration were considered as 'inadequate registration' in the analysis. We compared the distribution of registration status (adequate vs inadequate) between the shared and non-shared groups using Pearson's χ^2 test or Fisher's exact test for contingency tables.

Assessment of statistical method.

We reviewed the statistical methods described and used in the included trials. We considered the following issues with statistical methods: an absent or extremely brief description of the statistical method that hindered us from reproducing univariable analysis; any inconsistency between the specified method and analyses presented in the results section (including tables and figures); use of statistical methods that were not pre-specified; and inappropriate choice of statistical methods given the data type.[14] Studies without any of the above-mentioned issues were considered as having adequate reporting of statistical methods. A composite of these issues was compared between the shared and non-shared groups using Pearson's χ^2 test or Fisher's exact test. Odds ratios (ORs) as a measure of association and 95% CIs were computed.

Reproducibility of statistical analysis.

We attempted to reproduce the results of any univariable statistical analyses with summary statistics (aggregate data) provided in publications. The methods to reproduce these analyses have been elaborated elsewhere.[14] We calculated the inconsistency rate (number of irreproducible analyses/total number of univariable analyses) for baseline characteristics and outcomes in each RCT by comparing our computed results with those published. We compared inconsistency rates of trials between the shared and non-shared groups using the Mann–Whitney test. We plotted the cumulative distribution of the ratio [calculated *P*-value ÷ reported *P*-value] for irreproducible *P*-values separating baseline characteristics and outcomes in the shared and non-shared groups in the hope of inferring the reasons for irreproducibility. If inconsistencies were due to random errors or other unintentional reasons in different RCTs then, when the ratios are combined, we expect that the frequency of ratios <1 and >1 should be roughly equal and the value of the cumulative curve of the ratios should be close to 1 on the *x*-axis (the value of the ratio) when cumulative proportion on the *y*-axis reaches 0.5 (the point on the curve where there are an equal number of ratios above and below the corresponding value on the *x*-axis). If irreproducible *P*-values were intentionally manipulated, the numbers of the ratios <1 and >1 would be disproportionate and the observed cumulative distribution would violate what is expected under the null hypothesis of *P*-value ratios distributed symmetrically about 1.

Probability of random sampling for baseline characteristics.

We used Monte Carlo simulations (computational algorithms that use random sampling to generate numerical results) to generate a *P*-value for differences between treatment and control groups for each baseline continuous variable.[15] If randomization had been done correctly and data recorded accurately in all or most RCTs, then the set of simulation-generated *P*-values from baseline variables (especially those that are continuously valued) should not deviate much from a uniform [0,1] distribution, that is, they should be a

sample of randomly drawn values between 0 and 1. More details of this method and the assumptions required can be found in previous papers.[4,15,16] We used statistical conversions to estimate mean and SD if they were not displayed explicitly in the publications.[17] We used the Kolmogorov–Smirnov test, against a reference distribution of uniform [0,1], for the simulation-generated *P*-values of baseline variables, to check for the effectiveness of randomization across RCTs in the shared and non-shared groups. The smaller the *P*-value from the Kolmogorov–Smirnov test, the less likely it is that the data presented were generated from trials where the randomization process for treatment group allocation was performed properly. We also compared directly the distribution of simulation-generated *P*-values between the shared and non-shared groups.

Other miscellaneous issues.

We looked for additional concerns that present possible issues for the quality of the research, such as the absence of a statement confirming ethics approval in the original publications, unfeasible recruitment or study timelines, and reporting of any details about the participants who withdrew. Due to the possible complexity of these issues, we reported these results qualitatively.

We also performed subgroup analyses for the assessments of statistical method, statistical reproducibility and randomization in the non-shared group according to the reason for not being able to share IPD. All statistical analyses were performed using Review Manager (v5.3), Stata (v16.0) or the R statistical software (v3.5.1).

RESULTS

Inclusion of RCTs

There were 53 RCTs identified in the IPD meta-analysis.[4] Eight RCTs did not fulfil our inclusion criteria and were excluded; three RCTs were not published in English (Persian, Turkish and Italian one each) (Aygen *et al.*, 2007; Santonocito *et al.*, 2009; Lorzadeh *et al.*, 2011) and for five RCTs no full text was available such that methodological quality could not be assessed (Vegetti *et al.*, 1999; Robinson *et al.*, 2003; Williams *et al.*, 2009; Keikha and Shahraki, 2011; Jahan, 2015).

In total, 45 trials (8697 women) were included in this study. IPD were available from 17 RCTs (shared IPD group) (Sahin *et al.*, 2004; Palomba *et al.*, 2005; Bayar *et al.*, 2006; Lord *et al.*, 2006; Moll *et al.*, 2006; Legro *et al.*, 2007; Johnson *et al.*, 2010; Homburg *et al.*, 2012; Kar, 2012; Morin-Papunen *et al.*, 2012; Nazik and Kumtepe, 2012; Leanza *et al.*, 2014; Legro *et al.*, 2014; Kar and Sanchita, 2015; Amer *et al.*, 2017; Liu *et al.*, 2017; Wu *et al.*, 2017). Twenty-eight trials were categorized as the non-shared IPD group due to no response (n = 20)

(Boostanfar *et al.*, 2001; Lopez *et al.*, 2004; Atay *et al.*, 2006; Karimzadeh *et al.*, 2007; Dasari and Pranahita, 2009; Dehbashi *et al.*, 2009; Karimzadeh and Javedani, 2010; Zeinalzadeh *et al.*, 2010; Sheikh-El-Arab Elsedeek and Elmaghraby, 2011; Banerjee Ray *et al.*, 2012; Basirat *et al.*, 2012; Roy *et al.*, 2012; Selim and Borg, 2012; Seyedoshohadaei *et al.*, 2012; Abuelghar *et al.*, 2013; Ayaz *et al.*, 2013; Maged *et al.*, 2015; Sharief and Nafee, 2015; Chen *et al.*, 2016; Hossein-Rashidi *et al.*, 2016), data loss (n = 7) (Fleming *et al.*, 2002; Khorram *et al.*, 2006; Tang *et al.*, 2006; Badawy *et al.*, 2009; Zain *et al.*, 2009; Badawy and Gibreal, 2011; Mobusher, 2014) or legal reasons (n = 1) (Moussa *et al.*, 2016).

Characteristics of RCTs

The median number of trial participants included in the shared group was 159 (range 21–1000), with a median average number of randomized participants per site per month of 2.0 (range 0.3–16.7) (Supplementary Table S1). The median number of trial participants included in the non-shared group was 118 (range 24–438), with a median average number of randomizations per site per month of 6.3 (range 1.0–14.3) (Supplementary Table S2). In total, 11/17 (65%) RCTs in the shared group and 25/28 (89%) RCTs in the non-shared were single-centre trials.

Table 1 shows the results of meta-analyses in the shared and non-shared groups. For the comparison of letrozole versus CC, the pooled risk rates were higher in the shared group (RR 1.43, 95% CI 1.17–1.75 for live birth and RR 1.45, 95% CI 1.23–1.70 for clinical pregnancy) compared with the non-shared group (RR 1.09, 95% CI 0.68–1.74 for live birth and RR 1.29, 95% CI 1.04–1.60 for clinical pregnancy). For the comparison of CC plus metformin versus CC, the pooled risk rates were lower in the shared group (RR 1.08, 95% CI 0.87–1.35 for live birth and RR 1.18, 95% CI 1.00–1.39 for clinical pregnancy) than the non-shared group (RR 1.17, 95% CI 0.43–3.17 for live birth and RR 1.34, 95% CI 1.03–1.76 for clinical pregnancy). The CIs were narrower in the shared group compared with the non-shared group, enabling more precise estimates in the shared group.

Table 1. Meta-analyses and GRADE assessments of randomized controlled trials evaluating first-line ovulation induction for polycystic ovary syndrome by sharing or not sharing individual participant data

Comparison	Outcome	Group	No. of trials	No. of women	RR	95% CI	I^2*	GRADE
Letrozole vs CC	Live birth	Shared	3	1043	1.43*	1.17–1.75*	0%*	Moderate[a,*]
		Non-shared	2	200	1.09	0.68–1.74	13%	Low[a,b]
	Clinical pregnancy	Shared	6	1284	1.45*	1.23–1.70*	0%*	Moderate[a,*]
		Non-shared	13	1933	1.29	1.04–1.60	48%	Very low[a,b,c]
CC plus metformin vs CC	Live birth	Shared	5	907	1.08*	0.87–1.35*	6%*	Low[a,b,*]
		Non-shared	1	82	1.17	0.43–3.17	NA	Very low[d]
	Clinical pregnancy	Shared	7	1039	1.18*	1.00–1.39*	7%*	Low[a,b,*]
		Non-shared	8	853	1.34	1.03–1.76	0%	Low[a,b]

CC, clomiphene citrate; GRADE, Grading of Recommendations, Assessment, Development and Evaluations; I^2, statistics of heterogeneity; NA, not applicable; No., number; RR, risk ratio.
* Extracted from Wang *et al.* (2019).
[a] Downgraded by one level due to some concerns on risk of bias.
[b] Downgraded by one level due to imprecision.
[c] Downgraded by one level due to inconsistency.
[d] Downgraded by three levels due to major concerns on risk of bias and serious imprecision.

Quality of evidence of shared and non-shared IPD RCTs

Risk of bias and the GRADE assessment.

Overall, RCTs in the shared group performed better than those in the non-shared group on the assessment of the risk of bias (Fig. 1, Supplementary Table S3). Overall low risk of bias was associated with 13/17 (76%) of shared RCTs versus 7/28 (25%) of non-shared RCTs. RCTs in the non-shared group had higher risks of bias due to deviations from intended interventions and bias arising from the randomization process than RCTs in the shared group. We found no evidence that trials published later have better performance on overall risk of bias (Supplementary Fig. S1).

For the comparison of letrozole versus CC, the non-shared group was associated with lower overall certainty of evidence compared to the shared group using the GRADE approach (Table 1). When comparing CC plus metformin versus CC for live birth, the shared group had low overall certainty of evidence and the non-shared group had very low overall certainty.

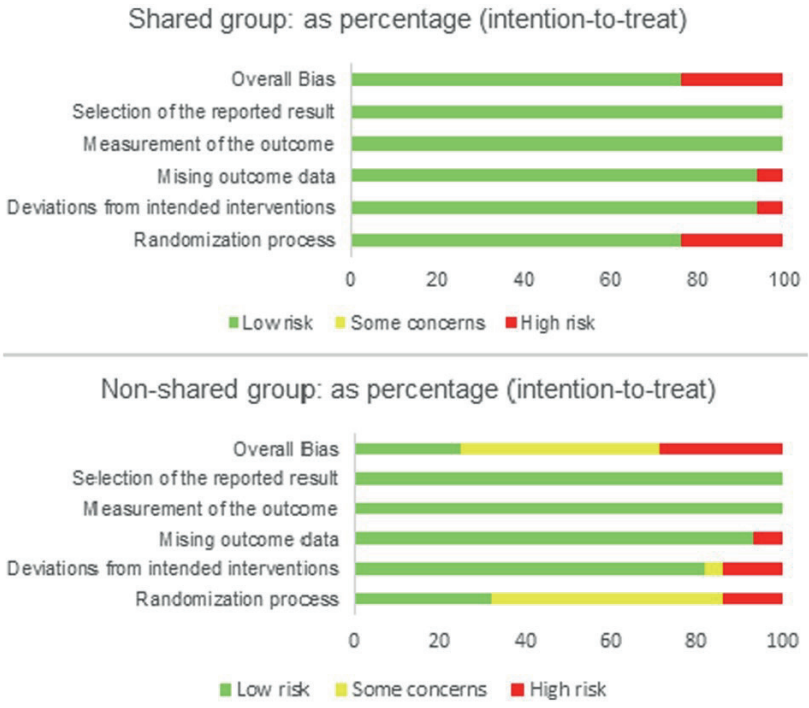


Figure 1. Risk of bias graph by sharing or not sharing individual participant data: review of authors' judgements about each risk of bias item presented as percentages across all included studies

Assessment of trial registration.

There were 25/45 RCTs that started recruitment after 1 July 2005. Adequate trial registration was found in 3/9 (33%) of RCTs in the shared group versus 0/16 (0%) in the non-shared group ($P = 0.037$). In the shared group, 2/9 (22%) trials were registered late or retrospectively and 4/9 (44%) were not registered at all. In the non-shared group, late or retrospective registration was found in 2/16 (13%) of the trials and the majority (14/16, 88%) of the trials had no registration (Table 2).

Table 2. Registration, dates of recruitment and submission to journal for randomized controlled trials evaluating first-line ovulation induction for polycystic ovary syndrome.

Study	Registration[a]	Recruitment (month-year)		Date registered (date-month-year)[b]	Article submission (date-month-year)[c]
		Start	End		
RCTs without IPD sharing					
Abuelghar *et al.* (2013)	Absent	May-2012	January-2013	NA	20-March-2013
Atay *et al.* (2006)	NA	NA	NA	NA	26-July-2005
Ayaz *et al.* (2013)	Absent	February-2008	December-2008	NA	26-June-2012
Badawy *et al.* (2009)	NA	January-2004	September-2006	NA	25-October-2006
Badawy and Gibreal (2011)	Absent	December-2005	December-2009	NA	8-February-2011
Basirat *et al.* (2012)	Retrospective	2007	2009	24-August-2010	29-September-2010
Boostanfar *et al.* (2001)	NA	August-1997	November-1999	NA	27-August-2000
Chen *et al.* (2016)	Absent	January-2013	January-2015	NA	30-August-2016
Dasari and Pranahita (2009)	NA	August-2003	August-2005	NA	6-October-2008
Dehbashi *et al.* (2009)	NA	February-2004	November-2006	NA	20-September-2008
Fleming *et al.* (2002)	NA	NA	NA	NA	16-August-2001
Hossein-Rashidi *et al.* (2016)	Absent	March-2013	June-2014	NA	September-2015
Karimzadeh *et al.* (2007)	Absent	August-2005	September-2006	NA	27-January-2007
Karimzadeh and Javedani (2010)	Absent	NA	NA	NA	22-October-2008
Khorram *et al.* (2006)	NA	NA	NA	NA	28-June-2005
Lopez *et al.* (2004)	NA	April-2000	December-2001	NA	2-July-2004
Maged *et al.* (2015)	Absent	September-2012	March-2014	NA	29-January-2015
Mobusher (2014)	Absent	July-2012	June-2013	NA	NA
Moussa *et al.* (2016)	Absent	August-2014	January-2015	NA	18-February-2016
Banerjee Ray *et al.* (2012)	Absent	January-2008	December-2009	NA	31-May-2011
Roy *et al.* (2012)	NA	January-2005	January-2010	NA	5-June-2011
Selim and Borg (2012)	Absent	November-2008	September-2011	NA	NA
Seyedoshohadaei *et al.* (2012)	Retrospective	November-2007	September-2009	6-June-2011	6-July-2011

Table 2. Continued

Sharief and Nafee (2015)	Absent	January-2012	April-2013	NA	NA
Sheikh-El-Arab Elsedeek and Elmaghraby (2011)	Absent	NA	NA	NA	7-October-2010
Tang *et al.* (2006)	NA	1999	2003	12-September-2003	25-April-2005
Zain *et al.* (2009)	Absent	September-2005	December-2006	NA	5-October-2007
Zeinalzadeh et al. (2010)	Absent	2006	2007	NA	NA
RCTs with IPD sharing					
Amer *et al.* (2017)	Late	April-2007	June-2014	23-May-2007	8-January-2017
Bayar *et al.* (2006)	NA	2004	2005	NA	28-October-2005
Homburg *et al.* (2012)	Late	August-2005	March-2009	14-February-2006	17-February-2011
Johnson et al. (2010)	NA	August-2003	February-2007	21-November-2008 (Protocol published in 2005)	29-September-2009
Kar (2012)	Absent	July-2010	July-2011	NA	1-April-2012
Kar and Sanchita (2015)	Absent	November-2011	December-2013	NA	18-April-2015
Leanza *et al.* (2014)	Absent	NA	NA	NA	17-April-2013
Legro *et al.* (2007)	NA	November-2002	December-2004	10-September-2003	NA
Legro *et al.* (2014)	Adequate	February-2009	January-2012	21-July-2008	NA
Liu *et al.* (2017)	Adequate	April-2012	March-2014	12-December-2011	10-June-2016
Lord *et al.* (2006)	NA	NA	NA	NA	NA
Moll *et al.* (2006)	NA	June-2001	June-2006	27-January-2006	NA
Morin-Papunen *et al.* (2012)	NA	January-2003	December-2009	9-October-2009	8-November-2011
Nazik and Kumtepe (2012)	Absent	December-2005	March-2007	NA	NA
Palomba *et al.* (2005)	NA	April-2003	September-2003	NA	19-January-2005
Sahin *et al.* (2004)	NA	NA	NA	NA	29-October-2002
Wu *et al.* (2017)	Adequate	July-2012	November-2014	4-April-2012	NA

IPD, individual participant data; RCT, randomized controlled trial.

a Registrations that happened 6 months after the initiation of recruitment dates but before the completion of recruitment were considered as 'late'. Attempts that registered after the completion of recruitment were defined as 'retrospective' registration.

b As described in corresponding clinical trial registry.

c As stated in the published article.

Assessment of statistical methods.

In total, 7/17 (41%) of RCTs in the shared group and 19/28 (68%) of RCTs in the non-shared group had issues with the statistical methods described (OR 0.33, 95% CI 0.10–1.16, $P=0.079$). In the subgroup of no response, there were 14/20 (70%) trials that had issues with the statistical methods described (OR 3.33, 95% CI 0.71–16.15, $P=0.078$ vs the shared group). In the subgroup of data loss or legal reasons, 5/8 (62.5%) RCTs had issues regarding statistical methods description (OR 2.38, 95% CI 0.32–20.03, $P=0.411$ vs the shared group).

Reproducibility of statistical analysis.

The median (range) of inconsistency rate per study, between reported and reproduced analysis for baseline variables, was 0% (0–92%) (6 RCTs applicable) in the shared group and 54% (0–100%) (13 RCT applicable) in the non-shared group ($P=0.061$). The subgroup analyses showed a median (range) of inconsistency rate of 65% (0–100%) (10 RCTs applicable, $P=0.118$ vs the shared group) in the no response subgroup and 54% (40–100%) (3 RCTs applicable, $P=0.059$ vs the shared group) in the data loss or legal reason subgroup.

The median (range) of inconsistency rate per study, between reported and reproduced analysis for the outcome(s), was 0% (0–63%) (14 RCTs applicable) in the shared group and 44% (0–100%) (24 RCTs applicable) in the non-shared group ($P=0.003$). The subgroup analyses showed a median (range) of inconsistency rate of 39% (0–100%) (16 RCTs applicable, $P=0.005$ vs the shared group) in the no response subgroup and 55% (0–100%) (eight RCTs applicable, $P=0.032$ vs the shared group) in the data loss or legal reason subgroup.

Regarding the shared group, there were equal proportions of ratios of calculated *P*-value/reported *P*-value that are <1 and >1 for both irreproducible baseline characteristics and outcomes (Fig. 2A). For the non-shared group, we found more ratios that are <1 than >1 for both baseline characteristics (73.6% vs 26.4%) and outcomes (59.7% vs 40.3%) (Fig. 2B). In the subgroup analysis, we found similar patterns to what is shown in the total non-shared group (Fig. 2C and D).

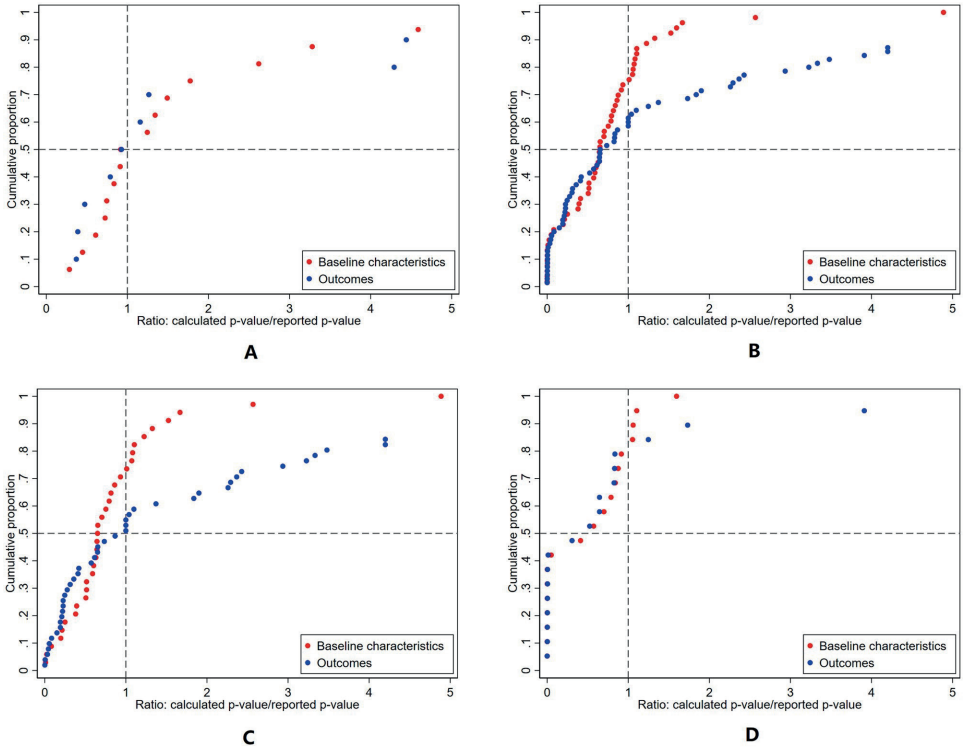


Figure 2. Cumulative distribution of the ratio of calculated P-value/reported P-value for irreproducible analyses of baseline characteristics and outcomes.

Given that the analyses were irreproducible because of random errors and were not manipulated, the numbers of the ratio 1 should be roughly equal and the cumulative curve approximates 1 on the x-axis when cumulative proportion reaches 0.5. (A) The distribution matches the expected observation in the shared group. (B) More ratios <1 were observed for baseline characteristics and outcomes in the non-shared group. (C) More ratios <1 were observed for baseline characteristics in the no response subgroup. (D) More ratios <1 were observed for baseline characteristics and outcomes in the data loss or legal reason subgroup.

Probability of random sampling for baseline characteristics.

In the shared group, the distribution of simulation-generated *P*-values from all baseline variables was consistent with the expected uniform distribution (Kolmogorov–Smirnov test *P*-value = 0.1626) (Fig. 3a, blue dots). In the non-shared group, the distribution of simulation-generated *P*-values from all baseline variables was substantially different to the expected uniform distribution (Kolmogorov–Smirnov test *P*-value = 4.535×10^{-8}) (Fig. 3A, red dots). A direct comparison of the distribution of *P*-values from the shared group and non-shared group also showed these to be very different ($P = 2.227 \times 10^{-4}$).

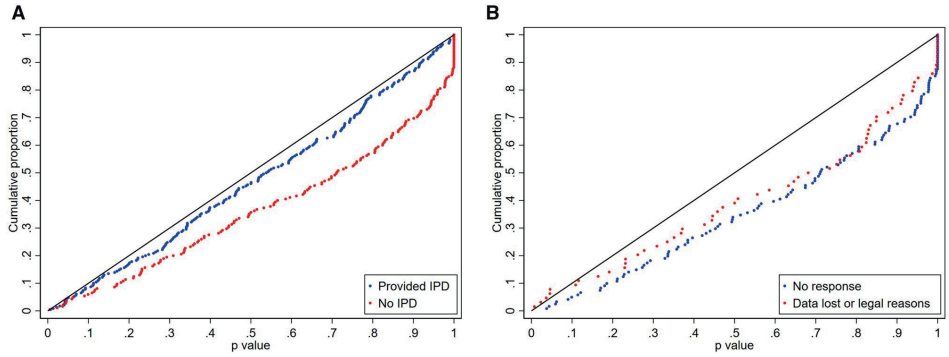


Figure 3. Cumulative distribution of Monte Carlo simulations generated P-values for baseline characteristics from trials evaluating first-line ovulation induction for polycystic ovary syndrome.

The null hypothesis is that the baseline characteristics in intervention and controls groups in these randomized controlled trials are the results of a properly conducted randomization process. (A) The distribution was consistent with the null hypothesis (diagonal line) for shared individual participant data (IPD) studies (blue dots, P = 0.1626), but inconsistent with the null hypothesis for non-shared IPD studies (red dots, P = 4.535 × 10−8). (B) The distribution was inconsistent with the null hypothesis for no response (blue dots, P = 2.004 × 10−6) and of IPD data loss or legal reasons (red dots, P = 0.001597).

Subgroup analyses of trials in the non-shared group with no response ($P = 2.004 \times 10^{-6}$) (Fig. 3B, blue dots) and of IPD data loss or legal reasons ($P = 0.002$) (Fig. 3B, red dots) showed that they were unlikely to follow the expected uniform distribution.

Miscellaneous issues.

Two RCTs in the shared-group (Leanza *et al.*, 2014; Kar and Sanchita, 2015) and two RCTs in the non-shared group (Seyedoshohadaei *et al.*, 2012; Mobusher, 2014) did not declare in the primary publications whether or not approval to conduct the study was obtained from an institutional review board or human research ethics committee. All other RCTs declared that they had received approval from a local ethics committee.

In the shared group, there were 13/17 (76%) trials that reported participants withdrawn, in which 8/13 (62%) performed intention-to-treat analysis. In the non-shared group, only 10/28 (36%) trials reported participants withdrawn and 2/10 (20%) trials applied intention-to-treat analysis.

DISCUSSION

In this article, we assessed methodological issues of RCTs studying first-line ovulation induction for PCOS by the availability of IPD. Effect estimates generated from trials that shared IPD were different from those using data of trials not sharing IPD. We found that for all pre-specified quality criteria the shared group performed better than the non-shared group. Compared to trials in the shared group, trials in the non-shared group had poorer performance on the risk of bias assessment and had lower certainty of evidence based on the GRADE approach. We found a higher proportion of trials with unsatisfactory trial registration, problematic statistical method reporting and irreproducible analyses in the non-shared group. While the randomization of participants to treatment and control groups was robust in the shared group, an analysis of baseline characteristics in the non-shared group indicated that many of these results were extremely unlikely to be the results of a properly conducted randomization. These findings did not vary according to the reason for not sharing IPD.

We used novel methods to assess the methodological issues of these RCTs. There are several limitations of this meta-epidemiological study. First, papers for which the full text was not available were excluded. However, studies available only as abstracts are generally of poorer quality than those published as full text.[18] Second, we used aggregated data from the primary publications to assess statistical methods for RCTs of the shared IPD group since using IPD for methodological assessments was not part of the data sharing agreement for the IPD meta-analysis. Third, the assumed null hypothesis about the distribution of simulation-generated *P*-values under proper randomization is not precisely correct if baseline variables are heavily correlated or more complex randomization techniques, such as stratification, are used or there is significant publication bias. The extent of the problem caused by the violation of the null hypothesis remains the subject of current research.[19,20] The issue of complex randomization methods, if it has a significant impact on the assumed null hypothesis, only affects the analysis of the shared group because there are five trials in the shared group that used more complex randomization methods whereas no trials in the non-shared group applied these techniques. Lastly, the sample sizes for some comparisons were small, limiting the power to interpret further.

Several systematic reviews have provided evidence regarding first-line ovulation induction for PCOS. A Cochrane aggregate data meta-analyses,[1] a network meta-analysis,[21] and our IPD meta-analysis[4] all concluded that letrozole increases live birth rates compared to CC. Similarly, aggregate data meta-analyses and IPD meta-analysis unanimously concluded that there was uncertainty or insufficient evidence of a difference between CC plus metformin and CC alone on live birth.[2,4] While the directions of these conclusions are consistent, the precision of risk estimates in the aggregate data meta-analyses is likely

to be troubled by the inclusion of trials that could not provide IPD as these trials were found to have lower quality and more methodological issues. In this study, all pooled risk estimates generated from the shared and non-shared groups are different in terms of the magnitude of effect sizes and the CIs of the non-shared group estimates are much wider. Thus, the IPD meta-analysis had higher validity and more accurate risk estimates than meta-analyses using aggregate data on this topic, which enables a more transparent and precise estimate of the effectiveness and safety of treatment options. We found that trials in the non-shared group had smaller effect estimates than those in the shared group for the comparison of letrozole versus CC while the opposite was true for the comparison of CC plus metformin versus CC alone. It was hypothesized that trials not contributing to IPD meta-analysis tend to overestimate effect sizes. However, a recent meta-epidemiological study including 31 IPD meta-analyses showed that trials not contributing IPD could yield either larger or smaller effect estimates than trials that shared IPD and there is no consistent pattern of data availability bias.[22] Findings in our study support this conclusion.

We found that trials with IPD sharing were overall of lower risk of bias than non-shared ones. This finding is in line with the statement of the International Weight Management in Pregnancy (i-WIP) Collaborative Group (2017),[9] which found RCTs that agreed to share IPD usually outperformed non-shared RCTs on a quantitative the 'risk of bias' assessment. This also agrees with the finding from a previous study that willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results.[23] The Risk of Bias tool and the GRADE approach are cornerstones supporting the validity of systematic reviews and meta-analyses of randomized trials. In theory, there is a potential risk that some authors may choose to withhold weaknesses of the trials in the manuscripts to improve the chance of them being published. Data integrity issues are another potential reason for unwillingness to share data and data 'unavailability'. While low-quality trials do not necessarily need to be excluded from meta-analyses, trials that have concerns on integrity or methodology, such as multiple irreproducible results and evidence of non-randomization, are meant to be handled carefully, examined critically and viewed cautiously before inclusion to ensure the robustness of synthesized evidence. We can ill afford to include trials with compromised data integrity in meta-analyses because policy-makers, insurance companies, doctors and patients using these meta-analyses could be misled and harms are irreparable.[24] At present, there is little consideration about data reproducibility and integrity of trials when performing meta-analyses using aggregate data, and guidelines on how to handle these critical issues are still lacking.

The poor performance in the non-shared group regarding all pre-specified criteria challenges the legitimacy of the traditional approach to include all relevant trials in systematic reviews and meta-analyses. While this inclusive approach minimizes the risk of publication bias, methodological and integrity issues of included trials cannot be rectified

by rigorous evidence synthesis. The availability of IPD and the willingness to share these data may be an indicator of quality, methodological soundness and integrity of trials when they are being considered for inclusion in systematic reviews. Meanwhile, some trials that shared IPD also had issues regarding quality assessments in this study, suggesting the limitation of using this indicator alone when realizing that IPD meta-analysis also cannot solve most of the issues of bias that arise from the poor conduct of trials. Rigorous assessments of quality and integrity of trials before inclusion is important to all forms of systematic reviews and meta-analyses.[25] Although assessing data availability bias is an intrinsic part of IPD meta-analysis, not all IPD meta-analyses assessed the potential impact of including aggregate data from studies not contributing IPD and almost no IPD meta-analyses analysed the underlying reasons for the differences between results with or without aggregate data.[26] Findings in this study highlight the importance of assessing data availability bias and the reasons beneath.

Increasingly, researchers are agreeing to share IPD from RCTs to create databases of studies for secondary analysis. Apart from conducting IPD meta-analyses and, more generally, addressing questions requiring evidence synthesis, combining IPD across multiple studies allows for harmonization of variables, standardization of data analysis, uniform presentation of results across the contributing RCTs, the investigation of subgroup effects and the generation (and potential testing) of new hypotheses.[6,27] Granting open access to the unit record data of RCTs has implications for promoting transparency and reproducibility of research. The results of this study provide another rationale for increasing the obligation of investigators to share RCT data, especially publicly funded trials, for IPD meta-analysis, and to evaluate more carefully the methodological features of RCTs that fail to share data. As trial data technically belongs to the institutions rather than investigators, more efforts should be made to stress the obligations of institutions to share trial data.

In conclusion, the results of this study suggest RCTs without IPD sharing have lower quality and substantially more methodological issues compared with RCTs that shared IPD on evaluating first-line ovulation induction for PCOS. IPD meta-analysis should be encouraged in all areas of obstetrics and gynaecology and, indeed, all areas of medicine when arguing for the reproducibility of research, which is a necessary precursor to the generation of valid scientific evidence.

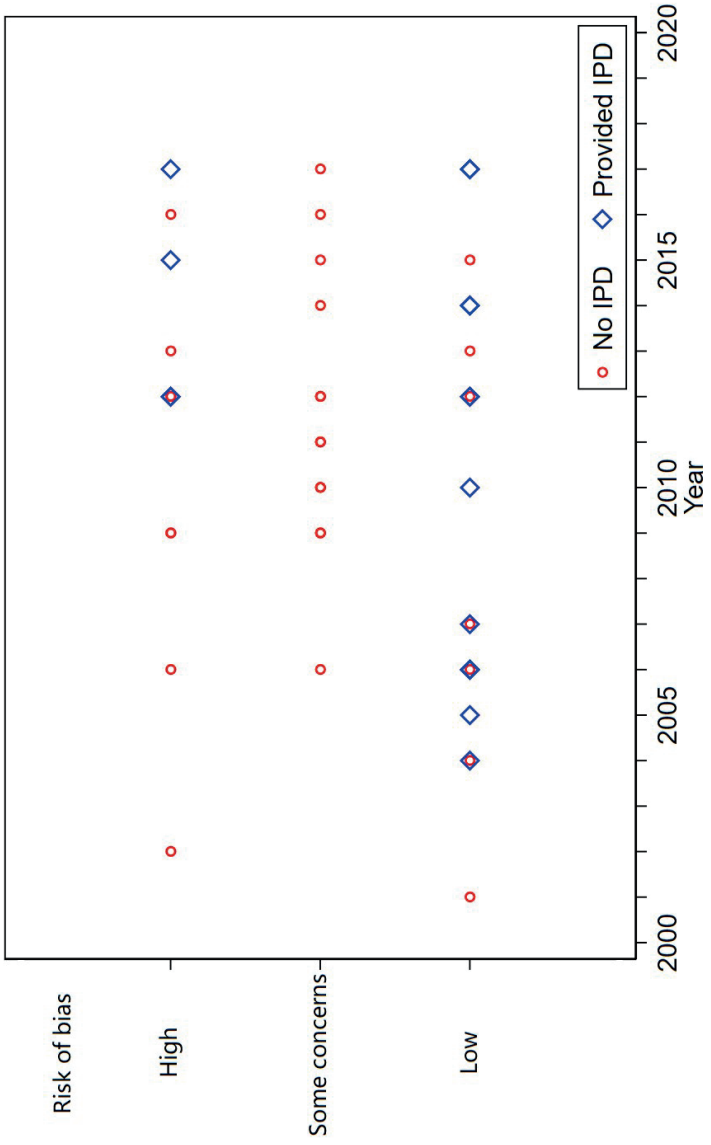
Funding

No external funding was sought to support this work.

Conflict of interest

B.W.M. reports grants from NHMRC, personal fees from ObsEva, personal fees from Merck Merck KGaA, personal fees from Guerbet, personal fees from iGenomix, outside the submitted work. M.F.C. reports shares in Virtus Health and past sponsorship from Merck Serono for scientific conference presentations. R.J.N. report a minor ownership in a fertility company Fertility SA and speaking fees for Merck for work outside of this area. Other authors have no conflicts of interest.

SUPPLEMENTARY DATA



Supplementary Figure 1. Distribution of risk of bias of trials according to the order of publication year by data availability. No evidence that trials published later have better performance on overall risk of bias.

Supplementary Table S1 Characteristics of randomized controlled trials that shared individual participant data evaluating first-line ovulation induction for polycystic ovary syndrome.

Study	Country	No. of patients	No. of sites	No. of randomisation per month[a]	No. of arms	Comparison	RR (95% CI)		Conclusion in the primary paper
							Clinical pregnancy	Live birth	
Amer *et al.*, 2017	UK	159	1	2	2	CC, letrozole	1.4 (1.1, 2.0)	1.4 (0.95, 2.0)	Benefit letrozole
Bayar *et al.*, 2006	Turkey	80	1	7	2	CC, letrozole	1.2 (0.5, 2.9)	NA	No difference
Homburg *et al.*, 2012	Netherlands	302	10	1	2	CC, FSH	1.2 (0.95-1.6)	1.2 (0.9-1.6)	Benefit FSH
Johnson *et al.*, 2010	New Zealand	171	4	1	4	Lifestyle, metformin, CC, CC plus metformin	CC, CC plus metformin 1.4 (0.8, 2.3) Placebo, metformin 1.4 (0.5-4.1) CC, metformin 1.0 (0.6-1.8)	CC, CC plus metformin 1.2 (0.7, 2.1) Placebo, metfromin 2.6 (0.5-12.3) CC, metformin 0.8 (0.4-1.6)	No difference
Kar *et al.*, 2012	India	103	1	8	2	CC, letrozole	2.7 (0.9, 7.9)	NA	Benefit letrozole
Kar *et al.*, 2015	India	105	1	4	3	CC, metformin, CC plus metformin	CC, CC plus metformin 1.5 (0.8, 3.0) CC, metformin 1.7 (0.9-3.2)	CC, CC plus metformin 1.4 (0.7, 3.0) CC, metformin 1.3 (0.6-2.8)	Benefit CC plus metformin
Leanza *et al.*, 2014	Italy	60	1	NA	2	CC plus metformin, CC	1.9 (0.95, 3.7)	NA	Benefit CC plus metformin
Legro *et al.*, 2007	USA	626	12	2	3	CC, metformin, CC plus metformin	CC, CC plus metformin 1.2 (0.9, 1.7) CC, metformin 0.4 (0.2-0.6)	CC, CC plus metformin 1.2 (0.9, 1.7) CC, metformin 0.3 (0.2-0.6)	No difference between CC and combination
Legro *et al.*, 2014	USA	750	11	2	2	CC, letrozole	1.4 (1.2, 1.8)	1.4 (1.1, 1.9)	Benefit letrozole

Liu *et al.*, 2017	China	268	1	11	4	CC, CC plus metformin, letrozole, letrozole plus metformin	1.3 (0.8, 2.2)	1.3 (0.7, 2.5)	Benefit letrozole
Lord *et al.*, 2006	UK	40	1	NA	2	Metformin, placebo	1.5 (0.3-8.1)	NA	No difference
Moll *et al.*, 2006	Netherlands	228	20	0.3	2	CC plus metformin, CC plus placebo	0.9 (0.7, 1.2)	0.8 (0.5, 1.1)	No difference
Morin-Papunen *et al.*, 2012	Finland	320	5	1	2	Metformin, placebo	1.4 (1.1-1.8)	1.5 (1.1-2.0)	Benefit metformin
Nazik *et al.*, 2012	Turkey	98	1	6	2	CC, letrozole	0.9 (0.3, 2.6)	NA	No difference
Palomba *et al.*, 2005	Italy	100	1	17	2	Metformin plus placebo, CC plus placebo	0.5 (0.3-0.8)	NA	Benefit metformin
Sahin *et al.*, 2004	Turkey	21	1	NA	2	CC plus metformin, CC	1.5 (0.5, 4.8)	NA	No difference
Wu *et al.*, 2017	China	1000	21	2	4	Acupuncture plus CC, control acupuncture plus CC, acupuncture plus placebo, control acupuncture plus placebo			No difference

NA= not applicable; RR= risk ratio; No. = number; CC = clomiphene citrate

[a] Calculated with the number of randomization and duration of study

Supplementary Table S2 Characteristics of randomized controlled trials that did not share individual participant data evaluating first-line ovulation induction for polycystic ovary syndrome.

Study	Country	No. of patients	No. of sites	No. of randomisation per month[a]	No. of arms	Comparison	RR (95% CI)		Conclusion in paper
							Clinical pregnancy	Live birth	
Abuelghar *et al.*, 2013	Egypt	106	1	12	3	CC, CC plus metformin, CC plus pioglitazone	1.4 (0.3-7.9)	NA	No difference
Atay *et al.*, 2006	Turkey	106	1	NA	2	CC, letrozole	2.4 (0.9-6.4)	NA	Benefit letrozole
Ayaz *et al.*, 2013	Saudi Arabia	42	1	4	2	CC plus metformin, CC	2.2 (1.0-4.6)	NA	Benefit CC plus metformin
Badawy *et al.*, 2009	Egypt	438	NA	NA	2	CC, letrozole	0.9 (0.7-1.1)	NA	No difference
Badawy *et al.*, 2011	Egypt	371	2	4	2	CC, tamoxifen	0.6 (0.4-0.97)		Benefit CC
Basirat *et al.*, 2012	Iran	334	2	7	2	CC plus metformin, CC	1.2 (0.8-1.7)	NA	No difference
Boostanfar *et al.*, 2001	USA	86	1	3	2	CC, tamoxifen	1.6 (0.6-4.1)	NA	No difference
Chen *et al.*, 2016	China	156	1	6	3	CC, letrozole, letrozole plus HMG	0.9 (0.5-1.7)	NA	Benefit letrozole plus hmg
Dasari *et al.*, 2009	India	24	1	1	2	CC plus metformin, CC	3.0 (0.6-14.5)	NA	Benefit CC plus metformin
Dehbashi *et al.*, 2009	Iran	100	1	3	2	CC, letrozole	1.9 (0.8-4.3)	1.7 (0.7-4.2)	Benefit letrozole
Fleming *et al.*, 2002	UK	94	1	NA	2	Metformin, placebo	1.7 (0.4-6.9)	NA	Benefit metformin
Hossein-Rashidi *et al.*, 2016	Iran	104	1	7	2	CC, FSH	1.2 (0.4-3.7)	1.0 (0.3-3.3)	No difference
Karimzadeh *et al.*, 2007	Iran	200	1	14	2	Metformin, placebo	3.6 (2.0-6.7)	NA	Benefit metformin
Karimzadeh *et al.*, 2010	Iran	343	1	NA	4	Lifestyle, metformin, CC, CC plus metformin	CC, CC plus metformin 1.2 (0.6-2.6) CC, metformin 1.2 (0.6-2.5)	NA	No difference
Khorram *et al.*, 2006	USA	31	1	NA	2	CC plus metformin, CC	10.4 (0.6-172.6)	NA	Not described
Lopez *et al.*, 2004	Spain	76	1	4	2	CC, FSH	1.8 (0.9-3.5)	1.8 (0.8-4.5)	No difference

Maged *et al.*, 2015	Egypt	120	1	6	3	CC, CC plus NAC, CC plus metformin	1.0 (0.3-3.7)	NA	Benefit CC plus NAC
Mobusher *et al.*, 2014	Pakistan	100	1	8	2	CC, letrozole	2.50 (0.8-7.5)	NA	Benefit letrozole
Moussa *et al.*, 2016	Egypt	150	2	13	3	CC, letrozole, tamoxifen	CC, letrozole 1.8 (0.8-3.8) CC, tamoxifen 0.9 (0.7-1.1)	NA	No difference
Banerjee Ray *et al.*, 2012	India	147	1	6	2	CC, letrozole	1.6 (0.9-3.0)	NA	Benefit letrozole
Roy *et al.*, 2012	Inda	204	1	3	2	CC, letrozole	1.7 (1.1-2.5)	NA	Benefit letrozole
Selim *et al.*, 2012	Egypt	220	1	6	2	CC, letrozole	1.5 (0.9-2.4)	NA	Benefit letrozole
Seyedoshohadaei *et al.*, 2012	Iran	150	1	7	3	CC, tamoxifen, letrozole	CC, letrozole 0.8 (0.6-1.1) CC, tamoxifen 0.6 (0.4-0.9)	CC, letrozole 0.95 (0.6-1.5) CC, tamoxifen 0.8 (0.5-1.3)	Benefit CC
Sharief *et al.*, 2015	Iraq	75	1	5	2	CC, letrozole	1.6 (0.7-3.8)	NA	No difference
Sheikh-El-Arab Elsedeek *et al.*, 2011	Egypt	124	1	NA	2	CC, letrozole	1.3 (0.7-2.2)	NA	No difference
Tang *et al.*, 2006	UK	143	1	3	2	Metformin, placebo	3.2 (0.7-15.4)	NA	No difference
Zain *et al.*, 2009	Malaysia	115	1	7	3	Metformin, CC, CC plus metformin	CC, CC plus metformin 1.3 (0.5-3.5) CC, metformin 0.5 (0.1-1.9)	CC, CC plus metformin 1.2 (0.4-3.2) CC, metformin 0.5 (0.1-1.9)	No difference
Zeinalzadeh *et al.*, 2010	Iran	107	1	9	2	CC, letrozole	1.4 (0.6-3.3)	NA	No difference

NAC: N-acetylcysteine

[a] Calculated with the number of randomization and duration of study

Supplementary Table S3 Risk of bias summary: review of authors' judgements about each risk of bias item for each included study.

Study	IPD or No IPD	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result	Overall Bias
Amer *et al.*, 2017	IPD	Low	Low	Low	Low	Low	Low
Bayar *et al.*, 2006	IPD	Low	Low	Low	Low	Low	Low
Homburg *et al.*, 2012	IPD	Low	Low	Low	Low	Low	Low
Johnson *et al.*, 2010	IPD	Low	Low	Low	Low	Low	Low
Kar *et al.*, 2012	IPD	High	Low	Low	Low	Low	High
Kar *et al.*, 2015	IPD	High	High	High	Low	Low	High
Leanza*et al.*, 2014	IPD	Low	Low	Low	Low	Low	Low
Legro*et al.*, 2007	IPD	Low	Low	Low	Low	Low	Low
Legro *et al.*, 2014	IPD	Low	Low	Low	Low	Low	Low
Liu *et al.*, 2017	IPD	High	Low	Low	Low	Low	High
Lord *et al.*, 2006	IPD	Low	Low	Low	Low	Low	Low
Moll *et al.*, 2006	IPD	Low	Low	Low	Low	Low	Low
Morin-Papunen *et al.*, 2012	IPD	Low	Low	Low	Low	Low	Low
Nazik *et al.*, 2012	IPD	High	Low	Low	Low	Low	High
Palomba *et al.*, 2005	IPD	Low	Low	Low	Low	Low	Low
Sahin *et al.*, 2004	IPD	Low	Low	Low	Low	Low	Low
Wu *et al.*, 2017	IPD	Low	Low	Low	Low	Low	Low
Badawy *et al.*, 2011	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Badawy *et al.*, 2009	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Fleming *et al.*, 2002	No IPD	Low	Low	High	Low	Low	High
Khorram *et al.*, 2006	No IPD	High	Low	Low	Low	Low	High
Mobusher *et al.*, 2014	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Tang *et al.*, 2006	No IPD	Low	Low	Low	Low	Low	Low

Zain *et al.*, 2009	No IPD	High	Low	Low	Low	Low	High
Moussa *et al.*, 2016	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Abueelghar *et al.*, 2013	No IPD	Low	High	Low	Low	Low	High
Atay *et al.*, 2006	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Ayaz *et al.*, 2013	No IPD	Low	Low	Low	Low	Low	Low
Basirat *et al.*, 2012	No IPD	High	High	Low	Low	Low	High
Boonstanfar *et al.*, 2001	No IPD	Low	Low	Low	Low	Low	Low
Chen *et al.*, 2017	No IPD	Some concerns	Some concerns	Low	Low	Low	Some concerns
Dasari *et al.*, 2009	No IPD	Some concerns	High	Low	Low	Low	High
Dehbasi *et al.*, 2009	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Hossein-Rashidi *et al.*, 2016	No IPD	Some concerns	High	High	Low	Low	High
Karimzadeh *et al.*, 2007	No IPD	Low	Low	Low	Low	Low	Low
Karimzadeh *et al.*, 2010	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Lopez *et al.*, 2004	No IPD	Low	Low	Low	Low	Low	Low
Maged *et al.*, 2015	No IPD	Low	Low	Low	Low	Low	Low
Ray *et al.*, 2012	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Roy *et al.*, 2012	No IPD	Low	Low	Low	Low	Low	Low
Selim *et al.*, 2012	No IPD	High	Low	Low	Low	Low	High
Seyedoshohadaei *et al.*, 2012	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Sharief *et al.*, 2015	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Sheikh-El-Arab Elsedeek *et al.*, 2011	No IPD	Some concerns	Low	Low	Low	Low	Some concerns
Zeinalzadeh *et al.*, 2010	No IPD	Some concerns	Low	Low	Low	Low	Some concerns

IPD: individual participant data

REFERENCES TO STUDIES INCLUDED IN THIS REVIEW

1. Abuelghar WM, Elkady OS, Khamees AA. Clomiphene citrate alone, in combination with metformin or in combination with pioglitazone as first line therapy in induction of ovulation in infertile women with polycystic ovary syndrome, a randomized controlled trial. Middle East Fertil Soc J 2013;18:135–141.
2. Amer SA, Smith J, Mahran A, Fox P, Fakis A. Double-blind randomized controlled trial of letrozole versus clomiphene citrate in subfertile women with polycystic ovarian syndrome. Hum Reprod 2017;32:1631–1638.
3. Atay V, Cam C, Muhcu M, Cam M, Karateke A. Comparison of letrozole and clomiphene citrate in women with polycystic ovaries undergoing ovarian stimulation. J Int Med Res 2006;34:73–76.
4. Ayaz A, Alwan Y, Farooq MU. Metformin-clomiphene citrate vs. clomiphene citrate alone: polycystic ovarian syndrome. J Hum Reprod Sci 2013;6:15–18.
5. Badawy A, Abdel Aal I, Abulatta M. Clomiphene citrate or letrozole for ovulation induction in women with polycystic ovarian syndrome: a prospective randomized trial. Fertil Steril 2009;92: 849–852.
6. Badawy A, Gibreal A. Clomiphene citrate versus tamoxifen for ovulation induction in women with PCOS: a prospective randomized trial. Eur J Obstet Gynecol Reprod Biol 2011;159:151–154.
7. Banerjee Ray P, Ray A, Chakraborti PS. Comparison of efficacy of letrozole and clomiphene citrate in ovulation induction in Indian women with polycystic ovarian syndrome. Archiv Gynecol Obstet 2012;285:873–877.
8. Basirat Z, Kashifard M, Amiri MG. Enhanced ovarian folliclular development by Metformin does not correlate with pregnancy rate: a randomized trial. Int J Fertil Steril 2012;6:31–36.
9. Bayar U, Basaran M, Kiran S, Coskun A, Gezer S. Use of an aromatase inhibitor in patients with polycystic ovary syndrome: a prospective randomized trial. Fertil Steri 2006;86:1447–1451.
10. Boostanfar R, Jain JK, Mishell DR Jr, Paulson RJ. A prospective randomized trial comparing clomiphene citrate with tamoxifen citrate for ovulation induction. Fertil Steril 2001;75:1024–1026.
11. Chen Z, Zhang M, Qiao Y, Yang J. Effects of letrozole in combination with low-dose intramuscular injection of human menopausal gonadotropin on ovulation and pregnancy of 156 patients with polycystic ovary syndrome. Pak J Med Sci 2016;32:1434–1438.
12. Dasari P, Pranahita GK. The efficacy of metformin and clomiphene citrate combination compared with clomiphene citrate alone for ovulation induction in infertile patients with PCOS. J Hum Reprod Sci 2009;2:18–22.
13. Dehbashi S, Dehbashi S, Kazerooni T, Robati M, Alborzi S, Parsanezhad ME, Shadman A. Comparison of the effects of letrozole and clomiphene citrate on ovulation and pregnancy rate in patients with polycystic ovary syndrome. Iran J Med Sci 2009;34: 23–28.
14. Fleming R, Hopkinson ZE, Wallace AM, Greer IA, Sattar N. Ovarian function and metabolic factors in women with oligomenorrhea treated with metformin in a randomized double blind placebo controlled trial. J Clin Endocrinol Metab 2002;87:569–574.

15. Homburg R, Hendriks ML, Konig TE, Anderson RA, Balen AH, Brincat M, Child T, Davies M, D'Hooghe T, Martinez A et al. Clomifene citrate or low-dose FSH for the first-line treatment of infertile women with anovulation associated with polycystic ovary syndrome: a prospective randomized multinational study. Hum Reprod 2012;27:468–473. ·
16. Hossein-Rashidi B, Khandzad B, Shahrokh-Tehraninejad E, Bagheri M, Gorginzadeh M. Recombinant FSH compared to clomiphene citrate as the first-line in ovulation induction in polycystic ovary syndrome using newly designed pens: a randomized controlled trial. J Family Reprod Health 2016;10:42–48.
17. Johnson NP, Stewart AW, Falkiner J, Farquhar CM, Milsom S, Singh VP, Okonkwo QL, Buckingham KL; REACT-NZ (REproduction And Collaborative Trials in New Zealand), a multi-centre fertility trials group. PCOSMIC: a multi-centre randomized trial in women with PolyCystic Ovary Syndrome evaluating Metformin for Infertility with Clomiphene. Hum Reprod 2010;25:1675–1683.
18. Kar S. Clomiphene citrate or letrozole as first-line ovulation induction drug in infertile PCOS women: a prospective randomized trial. J Hum Reprod Sci 2012;5:262–265.
19. Kar S, Sanchita S. Clomiphene citrate, metformin or a combination of both as the first line ovulation induction drug for Asian Indian women with polycystic ovarian syndrome: a randomized controlled trial. J Hum Reprod Sci 2015;8:197–201.
20. Karimzadeh MA, Eftekhar M, Taheripanah R, Tayebi N, Sakhavat L, Zare F. The effect of administration of metformin on lipid profile changes and insulin resistance in patients with polycystic ovary syn- drome. Middle East Fertil Soc J 2007;12:174–178.
21. Karimzadeh MA, Javedani M. An assessment of lifestyle modification versus medical treatment with clomiphene citrate, metformin, and clomiphene citrate-metformin in patients with polycystic ovary syn- drome. Fertil Steril 2010;94:216–220.
22. Khorram O, Helliwell JP, Katz S, Bonpane CM, Jaramillo L. Two weeks of metformin improves clomiphene citrate-induced ovula- tion and metabolic profiles in women with polycystic ovary syn- drome. Fertil Steril 2006;85:1448–1451.
23. Leanza V, Coco L, Grasso F, Leanza G, Zarbo G, Palumbo M. Ovulation induction with clomiphene citrate and metformin in women with polycystic ovary syndrome. Minerva Ginecol 2014;66: 299–301.
24. Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, Coutifaris C, McGovern PG, Cataldo NA et al. Clomiphene, metformin, or both for infertility in the polycystic ovary syndrome. N Engl J Med 2007;356:551–566.
25. Legro RS, Brzyski RG, Diamond MP, Coutifaris C, Schlaff WD, Casson P, Christman GM, Huang H, Yan Q, Alvero R et al. Letrozole versus clomiphene for infertility in the polycystic ovary syndrome. N Engl J Med 2014;371:119–129.
26. Liu C, Feng G, Huang W, Wang Q, Yang S, Tan J, Fu J, Liu D. Comparison of clomiphene citrate and letrozole for ovulation in- duction in women with polycystic ovary syndrome: a prospective randomized trial. Gynecol Endocrinol 2017:33:872–876.
27. Lopez E, Gunby J, Daya S, Parrilla JJ, Abad L, Balasch J. Ovulation in- duction in women

with polycystic ovary syndrome: randomized trial of clomiphene citrate versus low-dose recombinant FSH as first line therapy. Reprod Biomed Online 2004;9:382–390.

28. Lord J, Thomas R, Fox B, Acharya U, Wilkin T. The effect of metfor- min on fat distribution and the metabolic syndrome in women with polycystic ovary syndrome - a randomised, double-blind, placebo-controlled trial. BJOG 2006;113:817–824.
29. Maged AM, Elsawah H, Abdelhafez A, Bakry A, MostafaWAI. The adjuvant effect of metformin and N-acetylcysteine to clomiphene citrate in induction of ovulation in patients with Polycystic Ovary Syndrome. Gynecol Endocrinol 2015;356:635–638.
30. Mobusher I. Comparison of the efficacy of letrozole and clomiphene citrate for ovulation induction in infertile women with polycystic ovary syndrome. Pak J Med Health Sci 2014;8:905–908.
31. Moll E, Bossuyt PM, Korevaar JC, Lambalk CB, van der Veen F. Effect of clomifene citrate plus metformin and clomifene citrate plus placebo on induction of ovulation in women with newly diagnosed polycystic ovary syndrome: randomised double blind clinical trial. BMJ 2006;332:1485.
32. Morin-Papunen L, Rantala AS, Unkila-Kallio L, Tiitinen A, Hippelainen M, Perheentupa A, Tinkanen H, Bloigu R, Puukka K, Ruokonen A et al. Metformin improves pregnancy and live-birth rates in women with polycystic ovary syndrome (PCOS): a multicenter, double-blind, placebo-controlled randomized trial. J Clin Endocrinol Metab 2012;97:1492–1500.
33. Moussa AA, Torky H, Dief O, Elwahed AA, Senna HA. The effect of clomiphene citrate versus tamoxifen versus letrozol on endometrial thickness and blood flow in ovulation induction in women with polycystic ovaries. Acta Med Int 2016;3:88–92.
34. Nazik H, Kumtepe Y. Comparison of efficacy of letrozole and clomiphene citrate in ovulation induction for women with polycystic ovarian syndrome. HealthMED 2012;6:879–883.
35. Palomba S, Orio F Jr, Falbo A, Manguso F, Russo T, Cascella T, Tolino A, Carmina E, Colao A, Zullo F. Prospective parallel randomized, double-blind, double-dummy controlled clinical trial comparing clomiphene citrate and metformin as the first-line treatment for ovulation induction in nonobese anovulatory women with polycystic ovary syndrome. J Clin Endocrinol Metab 2005;90: 4068–4074.
36. Roy K, Baruah J, Singla S, Sharma J, Singh N, Jain S, Goyal M. A prospective randomized trial comparing the efficacy of Letrozole and Clomiphene citrate in induction of ovulation in polycystic ovarian syndrome. J Hum Reprod Sci 2012;5:20–25.
37. Sahin Y, Yirmibes U, Kelestimur F, Aygen E. The effects of metformin on insulin resistance, clomiphene-induced ovulation and pregnancy rates in women with polycystic ovary syndrome. Eur J Obstet Gynecol Reprod Biol 2004;113:214–220.
38. Selim MF, Borg TF. Letrozole and clomiphene citrate effect on endo- metrial and subendometrial vascularity in treating infertility in women with polycystic ovary syndrome. J Gynecol Surg 2012;28: 405–410.
39. Seyedoshohadaei F, Zandvakily F, Shahgeibi S. Comparison of the effectiveness of clomiphene citrate, tamoxifen and letrozole in ovu- lation induction in infertility due to isolated unovulation. Iranian J Reprod Med 2012;10:531–536.

40. Sharief M, Nafee NR. Comparison of letrazole and clomiphene cit- rate in women with polycystic ovaries undergoing ovarian stimula- tion. J Pak Med Assoc 2015;65:1149–1152.
41. Sheikh-El-Arab Elsedeek M, Elmaghraby HAH. Predictors and charac- teristics of letrozole induced ovulation in comparison with clomi- phene induced ovulation in anovulatory PCOS women. Middle East Fertil Soc J 2011;16:125–130.
42. 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.
43. Wu XK, Stener-Victorin E, Kuang HY, Ma HL, Gao JS, Xie LZ, Hou LH, Hu ZX, Shao XG, Ge J et al. Effect of acupuncture and clomiphene in Chinese women with polycystic ovary syndrome: a randomized clinical trial. JAMA 2017;317:2502–2514.
44. Zain MM, Jamaluddin R, Ibrahim A, Norman RJ. Comparison of clo- miphene citrate, metformin, or the combination of both for first- line ovulation induction, achievement of pregnancy, and live birth in Asian women with polycystic ovary syndrome: a randomized controlled trial. Fertil Steril 2009;91:514–521.
45. Zeinalzadeh M, Basirat Z, Esmailpour M. Efficacy of letrozole in ovulation induction compared to that of clomiphene citrate in patients with polycystic ovarian syndrome. J Reprod Med 2010;55: 36–40.

REFERENCES TO STUDIES EXCLUDED FROM THIS REVIEW

1. Aygen EM, Güzel Z, Özgün T, Atakul T, S ahin Y. The use of letrozole for ovulation induction in infertile women with polycystic ovarian syndrome [in Turkish]. Erciyes Tip Dergisi 2007;29: 195–200.
2. Jahan S. Comparative study of efficacy among metformin, clomiphene citrate and aromatase inhibitor (letrozole) as the first-line medication for ovulation induction, achievement of pregnancy and live birth in Asian women with polycystic ovarian syndrome: a prospective trial. Int J Gynaecol Obstet 2015;131:E503.
3. Keikha F, Shahraki MB. Induction ovulation in polycystic ovary patient with clomiphene citrate and letrozole. Iran J Reprod Med 2011; 356:46.
4. Lorzadeh N, Kazemirad S, Mohammadi Z. Comparison of effects letrozole and clomiphene citrate for ovulation induction in women with polycystic ovary syndrome [in Persian]. Iranian J Obstet Gynecol Infertil 2011;356:13–19.
5. Robinson RD, Swezey M, Propst A, Bates GW. Metformin added to clomiphene citrate does not improve pregnancy rates in hyperandrogenic, chronic anovulatory women: a randomized trial. Fertil Steril 2003;80:273–274.
6. Santonocito V, Rapisarda V, Abruzzo SRM, Pollicino R, Coco L, Zarbo G. Comparison between clomiphene citrate and metformin for induction of ovulatory cycles in infertile nonobese women with polycystic ovary syndrome [in Italian]. G Ital Ostet Ginecol 2009;31: 455–460.

7. Vegetti W, Riccaboni A, Colombo M, Baroni E, Diaferia D, Ragni G, Crosignani P. Randomized study of induction of ovulation by two different molecules with antiestrogenic effects, in patients with chronic anovulation disorders. Fertil Steril 1999;72:S234–S235.
8. Williams C, Pastore L, Shelly W, Bailey A, Baras D, Bateman B. A randomized, placebo-controlled study of the influence of instant release metformin on response to clomiphene citrate and time to conception in polycystic ovary syndrome. Fertil Steril 2009;92:S105.

ADDITIONAL REFERENCES

1. Franik S, Eltrop SM, Kremer JA, Kiesel L, Farquhar C. Aromatase inhibitors (letrozole) for subfertile women with polycystic ovary syndrome. Cochrane Database Syst Rev 2018;5:CD010287.
2. Sharpe A, Morley LC, Tang T, Norman RJ, Balen AH. Metformin for ovulation induction (excluding gonadotrophins) in women with polycystic ovary syndrome. Cochrane Database Syst Rev 2019;12: CD013505.
3. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ, International PN. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Hum Reprod 2018;33: 1602–1618.
4. Wang R, Li W, Bordewijk EM, Legro RS, Zhang H, Wu X, Gao J, Morin-Papunen L, Homburg R, Konig TE et al. First-line ovulation induction for polycystic ovary syndrome: an individual participant data meta-analysis. Hum Reprod Update 2019;25:717–732.
5. Bergeris A, Tse T, Zarin DA. Trialists' intent to share individual participant data as disclosed at ClinicalTrials.gov. JAMA 2018;319: 406–408.
6. Mbuagbaw L, Foster G, Cheng J, Thabane L. Challenges to complete and useful data sharing. Trials 2017;18:71.
7. Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. BMJ 2017;357:j1390.
8. Veroniki AA, Ashoor HM, Le SPC, Rios P, Stewart LA, Clarke M, Mavridis D, Straus SE, Tricco AC. Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. J Clin Epidemiol 2019;113:176–188.
9. International Weight Management in Pregnancy (i-WIP) Collaborative Group. Effect of diet and physical activity based interventions in pregnancy on gestational weight gain and pregnancy outcomes: meta-analysis of individual participant data from randomised trials. BMJ 2017;358:j3119.
10. Higgins J, Green S (eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011. The Cochrane Collaboration. https://handbook-5-1.cochrane.org/ (10 February 2020, date last accessed).
11. Sterne J, Savović J, Page M, Elbers R, Blencowe N, Boutron I, Cates C, Cheng H-Y, Corbett M, Eldridge S et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898.

12. Schünemann H, Brozek J, Guyatt G, Oxman A (eds). Handbook for grading the quality of evidence and the strength of recommenda- tions using the GRADE approach. updated October 2013. GRADE Working Group. gdt.guidelinedevelopment.org/app/handbook/ handbook.html 2013 (8 February 2020, date last accessed).
13. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJ et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet 2004;364:911–912.
14. Li W, Suke S, Wertaschnigg D, Lensen S, Wang R, Gurrin L, Mol BW. Randomised controlled trials evaluating endometrial scratching: assessment of methodological issues. Hum Reprod 2019;34: 2372–2380.
15. Carlisle JB, Dexter F, Pandit JJ, Shafer SL, Yentis SM. Calculating the probability of random sampling for continuous variables in submitted or published randomised controlled trials. Anaesthesia 2015; 70:848–858.
16. Carlisle JB, Loadsman JA. Evidence for non-random sampling in randomised, controlled trials by Yuhji Saitoh. Anaesthesia 2017;72: 17–27.
17. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014;14:135.
18. Scherer RW, Meerpohl JJ, Pfeifer N, Schmucker C, Schwarzer G, von Elm E. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev 2018;11:Mr000005.
19. Betensky RA, Chiou SH. Correlation among baseline variables yields non-uniformity of p-values. PLoS One 2017;12:e0184531.
20. Bolland MJ, Gamble GD, Avenell A, Grey A. Rounding, but not randomization method, non-normality, or correlation, affected baseline P-value distributions in randomized trials. J Clin Epidemiol 2019; 110:50–62.
21. Wang R, Kim BV, van Wely M, Johnson NP, Costello MF, Zhang H, Ng EH, Legro RS, Bhattacharya S, Norman RJ *et al*. Treatment strategies for women with WHO group II anovulation: systematic review and network meta-analysis. *BMJ* 2017;356:j138.
22. Tsujimoto Y, Fujii T, Onishi A, Omae K, Luo Y, Imai H, Takahashi S, Itaya T, Pinson C, Nevitt SJ et al. No consistent evidence of data availability bias existed in recent individual participant data meta- analyses: a meta-epidemiological study. J Clin Epidemiol 2020;118: 107–114.e5.
23. Wicherts JM, Bakker M, Molenaar D. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. PLoS One 2011;6:e26828.
24. Li W, van Wely M, Gurrin L, Mol BW. Integrity of randomized controlled trials: challenges and solutions. Fertil Steril 2020;113: 1113–1119.
25. Wang R, van Wely M. Understand low-quality evidence: learn from food chains. Fertil Steril 2020;113:93–94.
26. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selec- tion bias, and unavailable data in meta-analyses using individual par- ticipant data: a database survey. BMJ 2012;344:d7762.
27. Broeze KA, Opmeer BC, van der Veen F, Bossuyt PM, Bhattacharya S, Mol BW. Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. Hum Reprod Update 2010;16:561–567.

Chapter 8

Methods to assess research misconduct in health-related research: A scoping review.

Bordewijk EM, Li W, van Eekelen R, Wang R, Showell M, Mol BW, van Wely M.

Journal of Clinical Epidemiology. 2021 May

ABSTRACT

Objective

To give an overview of the available methods to investigate research misconduct in health-related research.

Study design and setting

In this scoping review, we conducted a literature search in MEDLINE, Embase, The Cochrane CENTRAL Register of Studies Online (CRSO), and The Virtual Health Library portal up to July 2020. We included papers that mentioned and/or described methods for screening or assessing research misconduct in health-related research. We categorized identified methods into the following four groups according to their scopes: overall concern, textual concern, image concern, and data concern.

Results

We included 57 papers reporting on 27 methods: two on overall concern, four on textual concern, three on image concern, and 18 on data concern. Apart from the methods to locate textual plagiarism and image manipulation, all other methods, be it theoretical or empirical, are based on examples, are not standardized, and lack formal validation.

Conclusion

Existing methods cover a wide range of issues regarding research misconduct. Although measures to counteract textual plagiarism are well implemented, tools to investigate other forms of research misconduct are rudimentary and labour-intensive. To cope with the rising challenge of research misconduct, further development of automatic tools and routine validation of these methods is needed.

Trial registration number

Center for Open Science (OSF) (https://osf.io/mq89w).

INTRODUCTION

Science relies on the integrity of findings that are re- ported. It was found that about 2% of scientists admitted to having fabricated, falsified, or modified data or results at least once and on average over 14% of scientists observed these behaviours among their colleagues.[1] Research misconduct may result in a waste of financial and human resources and, more importantly, it might pose an immediate risk to human health.[1]

The US Office of Research Integrity (ORI) defines research misconduct as fabrication (1), falsification (2), or plagiarism (3) in proposing, performing, or reviewing research, or in reporting research results.[2] Fabrication is making up data, results, or recordings, and reporting them. Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the re- search record. Plagiarism is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit. Research misconduct does not include honest error or differences of opinion.[2]

In case of suspected misconduct, according to the Committee on Publication Ethics (COPE) code of conduct,[3] editors have the duty to take action.[4] However, only a third of top-ranking peer-reviewed journals have publicly available definitions of misconduct and less than half describe editorial procedures for handling allegations of misconduct.[5] Admittedly, investigating research misconduct is usually not straightforward, and therefore dealing with possible misconduct is not an easy task. Failure to adequately investigate possible misconduct may perpetuate unreliable research findings in the literature. When researchers who commit fraud go unchecked, they may continue to practice misconduct.[4] Methods that investigate research misconduct accumulate and evolve. However, to date, there is no complete overview of these methods and their applicability. Here, we reviewed the literature for articles that mention, describe, validate, or apply methods for screening or assessing research misconduct in health-related research.

METHODS

The protocol of this scoping review is registered in the Center for Open Science (OSF) on July 14, 2020 (https://osf.io/mq89w). We followed the reporting guidelines for meta-analyses and systematic reviews extension for scoping reviews, as outlined by the PRISMA statement.[6]

Literature search

A comprehensive and systematic literature search was undertaken in MEDLINE, Embase, The Cochrane CENTRAL Register of Studies Online (CRSO), and The Virtual Health Library for reports up to the July 14, 2020 by an information specialist (MS, Appendix 1). To identify any additional studies, we scanned reference lists of appropriate reports and communicated with experts in this field. All references were imported in Covidence (covidence.org). There was no language restriction or date restriction, but we excluded conference abstracts.

In- and exclusion criteria and study selection

Studies that refer to methods to investigate research misconduct, i.e., fabrication, falsification, and/or plagiarism in health-related research, were eligible for this scoping review. We excluded editorials, education plagiarism tools, and studies on data audits, meta-data, peer-review, and p-hacking as these methods are not directed at detecting research misconduct.

Two review authors (EB and MvW) independently screened all records on basis of titles and abstracts. After the eligibility screening, we critically reviewed the full text of the selected studies to assess eligibility. Any discrepancies between the reviewers were solved by consensus.

Data extraction and categorization

We used a data charting XLS sheet developed by EB with the help of MvW. Data were extracted by EB and checked by MvW for the non-statistical papers, and WL or RvE for the statistical papers. We extracted any method provided concerning research misconduct.

From each included study we extracted information regarding author, year of publication, journal, and the method. For each method, if applicable, we recorded the link to the method, description on how to use the method, information needed, whether qualitative or quantitative, validation method, automatic application, and performance if available. We categorised identified methods into the following four groups according to their scopes: overall concern, textual concern, image concern, and data concern. We did not perform a critical assessment because there was insufficient information to support a fair critical appraisal of the identified methods.

RESULTS

Literature search

We identified 6,112 articles (Fig. 1). After removing duplications, we screened 4,956 articles, of which 4,750 irrelevant articles were excluded (proportionate agreement between reviewers was 0.91). After assessing the full text of 206 articles, we excluded another 149 (proportionate agreement between reviewers was 0.83). Therefore, 57 papers were included in this review.

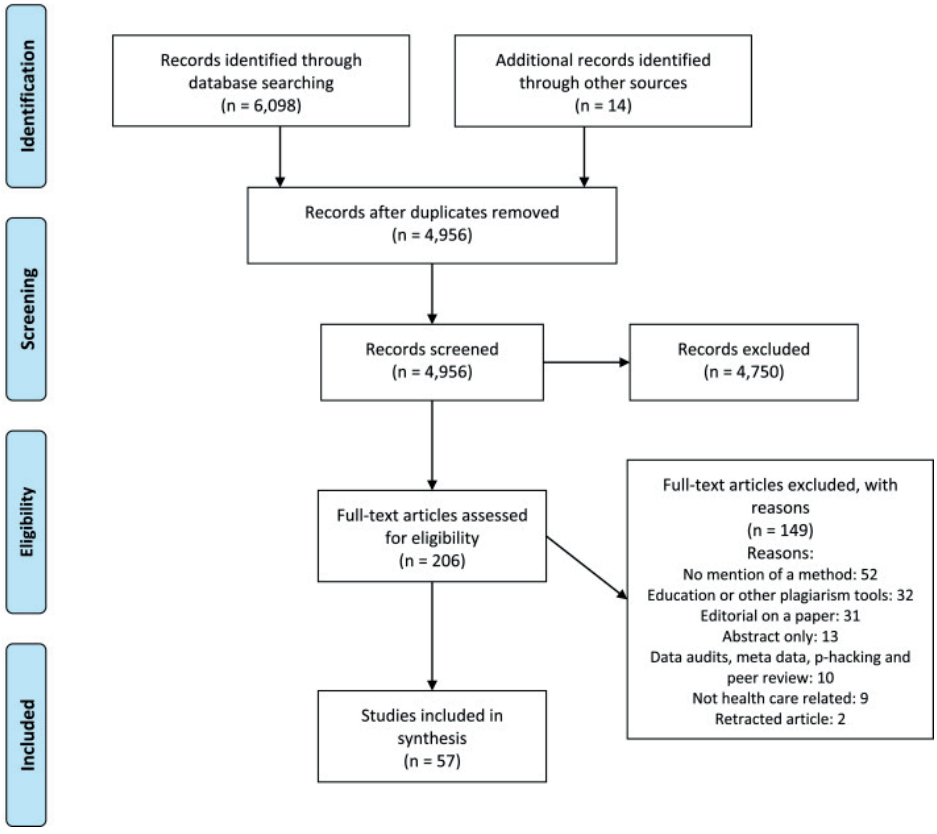


Figure 1. PRISMA 2009 flow diagram.

Included studies

The included papers reported on 27 different methods, two on overall concern, four on textual concern, three on image concern, and 18 on data concern. The characteristics of the articles are in Table 1. The following sections briefly explain the methods and their rationale. Table 2 expresses the applicability of the available methods. Available software

links and programs can be found in Table 3, with further details on how to use the statistical methods described in Appendix 2.

Table 1. Characteristics of the included studies

No. in reference list	Study	Journal/source	Title	Method
8	Grey et al. 2020	Nature	Check for publication integrity before misconduct	Overall concern: Screening (REAPRAISSED checklist)
9	Smith 2005	BMJ	Investigating the previous studies of a fraudulent author.	Overall concern: Investigating all publications of one author
24	Errami et al. 2007	Nucleic Acids Research	eTBLAST: a web server to identify expert reviewers, appropriate journals and similar publications.	Textual concern: Textual plagiarism (Helioblast)
26	Errami et al. 2008	Bioinformatics	Déjà vu—A study of duplicate citations in Medline.	Textual concern: Textual plagiarism (Helioblast)
25	Errami et al. 2010	Bioinformatics	Identifying duplicate content using statistically improbable phrases.	Textual concern: Textual plagiarism (Helioblast)
27	Garner 2012	Nature	How to stop plagiarism.	Textual concern: Textual plagiarism (Helioblast)
28	Higgings et al. 2016	Research integrity and peer review	Plagiarism in submitted manuscripts: incidence, characteristics and optimization of screening-case study in a major specialty medical journal.	Textual concern: Textual plagiarism (iThenticate)
29	Taylor 2017	American Roentgen Ray Society	Plagiarism in Manuscripts Submitted to the AJR: Development of an Optimal Screening Algorithm and Management Pathways	Textual concern: Textual plagiarism (iThenticate)
16	Bordewijk et al. 2020a	European Journal of Obstetrics & Gynecology and Reproductive Biology	Data integrity of 35 randomized controlled trials in women' health.	Overall concern: Investigating all publications of one author Textual concern: Compare baseline characteristics and outcome tables Data concern: Baseline P value distribution for RCTs & Digit preference checks
15	Bordewijk et al. 2020b	Fertility and Sterility Dialog	Data integrity of 10 other randomized controlled trials of an author with a retracted paper.	Overall concern: Investigating all publications of one author Textual concern: Compare baseline characteristics and outcome tables Data concern: Baseline P value distribution for RCTs
22	Baydik et al. 2016	Journal of Korean medical science	How to Act When Research Misconduct Is Not Detected by Software but Revealed by the Author of the Plagiarized Article.	Textual concern Translated plagiarism

Table 1. Continued

No. in reference list	Study	Journal/source	Title	Method
23	Wiwanitkit 2016	Macedonian Journal of Medical Sciences	How to verify and manage the translational plagiarism?	Textual concern Translated plagiarism
14	Spiroski 2016	Open Access Macedonian Journal of Medical Sciences	How to verify plagiarism of the paper written in Macedonian and translated in foreign language?	Overall concern: Investigating all publications of one author Textual concern Translated plagiarism
19	Bohannon 2015	Science	Scientific publishing. Hoax-detecting software spots fake papers.	Textual concern: Automatically generated fake papers (Scidetect)
20	Nguyen et al. 2016	BIR 2016 Workshop	Engineering a Tool to Detect Automatically Generated Papers.	Textual concern: Automatically generated fake papers (Scidetect)
21	Springer et al. 2015	Springer press release	Springer and Université Joseph Fourier release SciDetect to discover fake scientific papers	Textual concern: Automatically generated fake papers (Scidetect)
30	ORI	ORI	https://ori.hhs.gov/forensic-tools	Image concern
33	Parrish et al. 2009	Science and Engineering Ethics	Image manipulation as research misconduct.	Image concern
31	Koppers et al. 2017	Science and engineering ethics	Toward a Systematic Screening Tool for Quality Assurance and Semiautomatic Fraud Detection for Images in the Life Sciences.	Image concern
32	Acuna et al. 2018	bioRxiv	Bioscience-scale automated detection of figure element reuse.	Image concern
58	Hartgerink 2016	Data	688,112 Statistical Results: Content Mining Psychology Articles for Statistical Test Results.	Data concern: Statistics check (Statcheck)
59	van der Zee et al. 2017	BMC Nutrition	Statistical heartburn: an attempt to digest four pizza publications from the Cornell Food and Brand Lab.	Data concern: Statistics check (Statcheck and Test statistics)
57	Epskamp et al. 2015	R-project	Statcheck: Extract statistics from articles and recompute p values. R package version 1.0.1.	Data concern: Statistics check (Statcheck)
61	Anaya 2016	PeerJ Preprints	The GRIMMER test: A method for testing the validity of reported measures of variability	Data concern: Statistics check (Grimmer test)
60	Brown et al. 2017	Social Psychological and Personality Science	The GRIM Test: A Simple Technique Detects Numerous Anomalies in the Reporting of Results in Psychology	Data concern: Statistics check (Grimmer test)
62	Heathers et al. 2018	PeerJ Preprints	Recovering data from summary statistics: Sample Parameter Reconstruction via Iterative TEchniques (SPRITE)	Data concern: Statistics check (SPRITE)

Table 1. Continued

No. in reference list	Study	Journal/source	Title	Method
63	Li et al. 2020	Fertility and sterility	Integrity of randomized controlled trials: challenges and solutions.	Data concern: Statistics check (Test statistics)
17	Dahlberg 2010	Sci Eng Ethics	Scientific Forensics: How the Office of Research Integrity can Assist Institutional Investigations of Research Misconduct During Oversight Review	Overall concern: Investigating all publication of one author Data concern: Statistics check (Test statistics), Benford's law, Digit preference checks & Inliers
52	Al-Marzouki et al. 2005	BMJ	Are these data real? Statistical methods for the detection of data fabrication in clinical trials.	Data concern: Statistics check (Test statistics) & Digit preference checks
18	Carlisle 2020	Anesthesia	False individual patient data and zombie randomized controlled trials submitted to Anesthesia	Overall concern: Investigating all publication of one author Data concern: Statistics check (Test statistics), Digit preference checks, Repeated measurements & Outliers
56	Hüllemann et al. 2017	Anaesthesist	Application of Benford's law: a valuable tool for detecting scientific papers with fabricated data?	Data concern: Benford's law
53	Orita et al. 2012	Expert opinion on drug discovery	Agreement of drug discovery data with Benford's law.	Data concern: Benford's law
54	Hein et al. 2012	Anaesthesist	Scientific fraud in 20 falsified anesthesia papers Detection using financial auditing methods	Data concern: Benford's law
55	Pollach et al. 2016	Medical Hypotheses	The "first digit law" – A hypothesis on its possible impact on medicine and development aid	Data concern: Benford's law
10	Carlisle 2012	Anesthesia	The analysis of 168 randomized controlled trials to test data integrity	Overall concern: Investigating all publications of one author Data concern: Baseline P value distribution for RCTs
44	Carlisle et al. 2015	Anesthesia	Calculating the probability of random sampling for continuous variables in submitted or published randomized controlled trials.	Data concern: Baseline P value distribution for RCTs
11	Bolland 2016	Neurology	Systematic review and statistical analysis of the integrity of 33 randomized controlled trials.	Overall concern: Investigating all publications of one author Data concern: Baseline P value distribution for RCTs
45	Carlisle et al. 2017	Anesthesia	Evidence for non-random sampling in randomized, controlled trials by Yuhji Saitoh.	Data concern: Baseline P value distribution for RCTs

Table 1. Continued

No. in reference list	Study	Journal/source	Title	Method
47	Mascha et al. 2017	Anesthesia and analgesia	An Appraisal of the Carlisle-Stouffer-Fisher Method for Assessing Study Data Integrity and Fraud.	Data concern: Baseline P value distribution for RCTs
48	Kharasch et al. 2017	Anesthesia	Seeking and reporting apparent research misconduct: errors and integrity.	Data concern: Baseline P value distribution for RCTs
49	Bolland et al. 2019a	Journal of clinical epidemiology	Rounding, but not randomization method, non-normality, or correlation, affected baseline P-value distributions in randomized trials.	Data concern: Baseline P value distribution for RCTs
50	Bolland et al. 2019b	Journal of clinical epidemiology	Baseline P value distributions in randomized trials were uniform for continuous but not categorical variables.	Data concern: Baseline P value distribution for RCTs
46	Myles 2019	Anesthesia	Evidence for compromised data integrity in studies of liberal peri-operative inspired oxygen.	Data concern: Baseline P value distribution for RCTs
51	Bolland et al. 2020	Anesthesia	Empirically generated reference proportions for baseline p values from rounded summary statistics.	Data concern: Baseline P value distribution for RCTs
34	Buyse et al. 1999	Statistics in medicine	The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials.	Data concern: Benford's law, Digit preference checks, Plausibility of Correlations between variables, Date checking, Recruitment over time, Repeated measurements, Inliers, Outliers & Centre with possible data fabrication
35	Kirkwood et al. 2013	Clinical Trials	Application of methods for central statistical monitoring in clinical trials.	Data concern: Benford's law, Digit preference checks, Plausibility of Correlations between variables, Date checking, Repeated measurements, Inliers, Outliers & Centre with possible data fabrication
37	van den Bor et al. 2017	Journal of clinical epidemiology	A computationally simple central monitoring procedure, effectively applied to empirical trial data with known fraud.	Data concern: Benford's law, Digit preference checks, Plausibility of Correlations between variables, Date checking, Recruitment over time, Missing data, Outliers & Centre with possible data fabrication:

Table 1. Continued

No. in reference list	Study	Journal/source	Title	Method
43	Hartgerink et al. 2016	PsyArXiv	Detection of Data Fabrication Using Statistical Tools	Data concern: Benford's law, Digit preference checks, Plausibility of Correlations between variables, Standard deviations & Centre with possible data fabrication
36	Taylor et al. 2002	Drug Information Journal	Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data.	Data concern: Digit preference checks, Date checking, Inliers & Centre with possible data fabrication
38	O'Kelly 2004	Pharmaceutical Statistics	Using statistical techniques to detect fraud: A test case.	Data concern: Digit preference checks, Inliers, Outliers & Centre with possible data fabrication
41	Pogue et al. 2013	Clinical trials	Central statistical monitoring: detecting fraud in clinical trials.	Data concern: Digit preference checks, Repeated measurements & Centre with possible data fabrication
42	Knepper et al. 2016	Therapeutic Innovation and Regulatory Science	Statistical Monitoring in Clinical Trials: Best Practices for Detecting Data Anomalies Suggestive of Fabrication or Misconduct.	Data concern: Digit preference checks, Plausibility of correlations between variables, Date checking, Missing data & Centre with possible data fabrication
7	Bailey 1991	Controlled clinical trials	Detecting fabrication of data in a multicenter collaborative animal study.	Overall concern: Investigating all publications of one author Data concern: Plausibility of Correlations between, Inliers, Outliers & Centre with possible data fabrication
40	Wu et al. 2012	Pharmaceutical statistics	Detecting data fabrication in clinical trials from cluster analysis perspective.	Data concern: Plausibility of Correlations between variables
13	Hudes et al. 2017	FASEB journal	Unusual clustering of coefficients of variation in published articles from a medical biochemistry department in India.	Overall concern: Investigating all publications of one author Data concern: Plausibility of Coefficients of variation
12	Simonsohn 2013	Psychological science	Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone.	Overall concern: Investigating all publications of one author Data concern: Plausibility of Standard deviations
39	Venet et al. 2012	Clinical Trials	A statistical approach to central monitoring of data quality in clinical trials	Data concern: Repeated measurements & Centre with possible data fabrication

Table 2. Applicability of the methods to assess research misconduct in health-related research

Method/technique	Application				
	Minimum information requireda	**Type**	**Scope**	**Automated**	**Validated**
Overall concern					
Screening: REAPRAISSED checklist	Manuscript + tables + figures	Mixed	Fabrication / falsification / plagiarism	-	-
Detection of patterns of misconduct in all publications of one author/group	Manuscript + tables + figures	Qualitative	Fabrication / falsification / self-plagiarism	-	-
Textual concern					
Textual plagiarism: Helioblast/ iThenticate	Manuscript	Quantitative	Plagiarism	√	√
Compare baseline and outcome tables	Tables	Mixed	Fabrication / falsification	-	-
Translated plagiarism	Manuscript	Qualitative	Plagiarism	-	-
Automatically generated fake papers: SciDetect	Manuscript	Quantitative	Fabrication	√	-
Image concern					
Image manipulation detection tools	Figures	Quantitative	Fabrication / falsification	√	√
Data concern					
Statcheck	Manuscript	Quantitative	Fabrication / falsification	√	√
Grimmer test	Manuscript + tables	Quantitative	Fabrication / falsification	√	-
SPRITE	Manuscript + tables	Quantitative	Fabrication / falsification	√	-
Recalculate test statistics	Manuscript + tables	Quantitative	Fabrication / falsification	-[b]	-
Benford's law and digit preference checks	Manuscript + tables + figures	Quantitative	Fabrication / falsification	-[b]	-[c]
Baseline P value distribution for RCTs	Manuscript + baseline table	Quantitative	Fabrication / falsification	-[b]	-[c]
Plausibility of Correlations between variables	Manuscript + tables	Quantitative	Fabrication / falsification	-[b]	-
Plausibility of Coefficients of variation	Manuscript + tables	Quantitative	Fabrication / falsification	-	-
Plausibility of Standard deviations	Manuscript + tables	Quantitative	Fabrication / falsification	-[b]	-
Date checking	Manuscript + tables + raw data	Quantitative	Fabrication / falsification	-[b]	-

Table 2. Continued

Method/technique	Application				
	Minimum information requireda	**Type**	**Scope**	**Automated**	**Validated**
Recruitment over time	Manuscript + tables	Quantitative	Fabrication / falsification	-[b]	-
Repeated measurements	Manuscript + tables + raw data	Quantitative	Fabrication / falsification	-[b]	-
Missing data	Manuscript + tables + raw data	Quantitative	Fabrication / falsification	-[b]	-
Inliers and Outliers	Manuscript + tables + raw data	Quantitative	Fabrication / falsification	-[b]	-
Centre with data manipulation (multicenter study)	Manuscript + tables + raw data	Quantitative			
	Fabrication / falsification	-	-		

[a]May be complemented by supplementary materials

[b]R-program available although automated software is absent at the moment

[c]Preliminary attempts for validation exist

Table 3. Methods and the links to find the software or R-programs

Method/technique	Link
Overall integrity	
Screening: REAPRAISSED checklist (M)	http://resource-cms.springernature.com/springer-cms/rest/v1/content/17589730/data/v1 (This checklist is licensed under a Creative Commons licence: CC BY-NC-SA).
Investigate all publications of one author / author group (M)	Manual
Textual integrity	
Textual plagiarism: Helioblast/ iThenticate (S)	Helioblast: https://helioblast.heliotext.com Turnitin: https://www.crossref.org/services/similarity-check/, http://www.ithenticate.com/ & https://www.turnitin.com/
Compare baseline characteristics and outcome tables (M)	Manual
Translated plagiarism (S/M)	Manual or after translation use of textual plagiarism software
Scidetect (S)	https://gricad-gitlab.univ-grenoble-alpes.fr/labbecy/scidetect
Image Integrity	
Image manipulation detection tools (S)	https://ori.hhs.gov/forensic-tools,https://github.com/lkoppers/FraudDetTool
Data Integrity	
Statcheck (S)	https://cran.r-project.org/web/packages/statcheck/index.html.
Grimmer test (S)	http://www.prepubmed.org/grimmer/
Sprite algorithm (S)	http://www.prepubmed.org/sprite/
Recalculate test statistics (R) (M)	https://cran.r-project.org/web/packages/rpsychi/index.html,https://github.com/OmnesRes/pizzapizza
Benford's law (leading digit) and Digit preference (last digit) (R) (M)	http://www.ctc.ucl.ac.uk/Training.aspx, R-program available of the Web Appendix of van den Bor, et al. 20, https://github.com/chartgerink/ddfab
Baseline P value distribution for RCTs (R) (M)	Manual, https://cran.r-project.org/web/packages/simdistr/index.html
Plausibility of Correlations between variables (R) (M)	http://www.ctc.ucl.ac.uk/Training.aspx, R-program available of the Web Appendix of van den Bor, et al. 20
Plausibility of Coefficients of variation (M)	Manual
Plausibility of Standard deviations (M)	Manual, https://github.com/chartgerink/ddfab
Date checking (R) (M)	http://www.ctc.ucl.ac.uk/Training.aspx, R-program available of the Web Appendix of van den Bor, et al. 20
Recruitment over time (R) (M)	R-program available of the Web Appendix of van den Bor, et al. 20
Repeated measurements (R) (M)	http://www.ctc.ucl.ac.uk/Training.aspx
Missing data (R) (M)	R-program available of the Web Appendix of van den Bor, et al. 20
Inliers and Outliers (R) (M)	http://www.ctc.ucl.ac.uk/Training.aspx, R-program available of the Web Appendix of van den Bor, et al. 20
Centre with possible data fabrication	
Mean at each center to overall mean of other centers (R) (M)	http://www.ctc.ucl.ac.uk/Training.aspx
Substantial differences in outcomes/treatment effects (M)	Manual
(S) Software package, (R) R program available, (M) manual	

OVERALL CONCERN

Screening

The "REAPRAISSED checklist" for evaluation of publication integrity is a screening tool to assess whether a paper has characteristics that question its trustworthiness.[8] The checklist facilitates systematic evaluation through 11 categories. It covers ethical oversight and funding, research productivity and investigator workload, validity of randomization, plausibility of results, and duplicate data reporting.

Detection of patterns of misconduct in all publications of one author/group

When a fraudulent research paper is discovered, it is reasonable to assume that there may be similar problems with previous works of the authors involved.[9] Some patterns of research misconduct that are unique to the leading author/group can only be identified when all relevant works are compared, such as copying data of the group's previous works and overlapping publications. Also, comparing conference posters or abstracts, research grants, and protocols of one author or author group can be useful in the detection of research misconduct.[17]

TEXTUAL CONCERN

Methods that detect textual concern are summarised in Appendix 3. Methods for anti-textual plagiarism have been widely implemented.

IMAGE CONCERN

ORI offers Forensic Image Analysis Tools to detect data image manipulation in the field of biomedicine, especially Western Blots.[30]

Koppers, Wormer[31] created a tool that uses mathematical methods to detect suspicious images in large image archives, the R package called "FraudDetTools". The tool can detect manipulation by deleting unwanted data information, duplication by reusing images in different papers or contexts, and manipulation by adding information/data points.

Acuna, Brookes[32] created a tool that analyses potential inappropriate reuse of images. This algorithm detects figure region reuse and is robust to rotation, cropping, resizing, and contrast changes, and estimates which of the reuses have biological meaning.

For all these algorithm-based tools, the final decision should always be made by human experts to avoid false positives.

DATA CONCERN

Methods to check the authenticity of the data are directed at the given statistical results and the original raw data. Some of the methods described in the following sections are sufficiently complicated that to apply them, we refer the readers to the original papers.

Statistics check

Reported statistical results can be reproduced with summary statistics in publications. Inconsistencies may be explained by data fabrication or falsification as well as other possible reasons such as honest error. We found four software packages: Statcheck, the GRIMMER test, SPRITE, and the R package rpsychi.

The free *Statcheck*[57-59] software extracts statistical values reported in the text. For each extracted statistical test result, the reported statistical values are used to recalculate the *P*-value for the reported statistical result. Recalculated *P*-values are checked against the reported *P*-values for inconsistencies. This tool is not able to search tables and can miss tests that are not in APA format. It checks only if the *P*-value is consistent with the test statistic and degrees of freedom. It cannot check if the test statistic or degrees of freedom are correct.[59]

The GRIMMER Test[60,61] (Granularity-Related Inconsistency of Means Mapped to Error Repeats) is built upon the GRIM test.[60] This freely available software allows for testing whether reported measures of variability are mathematically possible. GRIMMER relies upon the statistical phenomenon that variances display a simple repetitive pattern when the data is discrete, i.e., granular. The algorithm created by Anaya[61] can identify whether a reported statistic is consistent with the sample size and granularity. The ability of the test relies upon: (1) the sample size; (2) the granularity of the data; (3) the precision (number of decimals) of the reported statistic; and (4) the size of the standard deviation or standard error (but not the variance). A limitation of the test is that it is at present restricted to a sample size of 99.

SPRITE (Sample Parameter Reconstruction via Interative Techniques) is a technique for reconstructing potential discrete data sets using only basic summary information about a sample, namely the mean, the standard deviation, the sample size, and the lower and upper bounds of the range of item values. SPRITE complements the GRIM and GRIMMER tests.[62] SPRITE does not have a sample size limitation. SPRITE also takes into consideration

the range of possible values of the raw data. It can identify cases in which the summary statistics are theoretically possible, but imply a highly skewed or otherwise unusual distribution of individual responses. SPRITE is not a complete mathematical solution, and some degree of interpretation of its output will always be required.[62]

Statistics checks could be performed for trials that performed univariable analyses. Independent *t*-test, one- and two-way ANOVAs can be checked using the means, standard deviations, and the sample size reported in articles.[59] Chi-square tests, Fisher's exact tests, unadjusted odds ratios, and risk ratios can be reproduced using absolute numbers given in crosstabs.[63]
If an original raw dataset is available the statistics can be recalculated and a comparison can be made between the results of these recalculations and the resulting claims in the paper.[17] ORI developed a method that focuses on the insignificant data or numbers of a paper whenever possible. This principle is based on their repeated observation, that when falsifying or fabricating data, an individual will focus on the desired outcome and pay less attention to the other data to make it appear authentic.[17]

Benford's law and digit preference checks

Benford's law is a description of the probability of the digits in naturally occurred numbers. The first digits tend to follow a non-uniform distribution in the natural occurrence which means that the first digits one, two, and three account for more than 60% of the total probability distribution. The last digits, however, are expected to approach uniform distribution.

It is possible to collect all the numbers from published articles and assess the frequencies of the digits by applying Benford's law. There may be some legitimate reasons for unequal distributions of digits, for example, biomarker measurements might be rounded to the last digit of either zero or five because the technology is insufficiently accurate. However, preference for one digit over another might be an indication of digit preference of a fraudulent researcher. Limited validation studies on cases of proven fraud and non-verified controls found that this approach is highly sensitive, but the specificity is unclear.[56] Only limited exploratory studies have revealed that some natural biomedical data obey Benford's law.[53]

Baseline *P*-value distribution for randomized trials

This method only applies to randomised trials. In a trial, the *P*-values in the baseline characteristics table (often "Table 1") are expected to be "uniform" with equal distributions between zero and one. This balance of baseline characteristics due to random allocation is difficult to be perfectly fabricated. Using summary statistics in publications, baseline *P*-values can be obtained through parametric tests or Monte Carlo simulations.[11, 44]

The effectiveness of this method in locating trials with concerns has been demonstrated empirically in several notorious cases of research misconduct.[11, 44-46] However, there is concern about the validity of the expected uniform distribution given several assumptions may be violated such as independence between variables, exclusive use of simple randomisation, no rounding of summary statistics, and no publication bias.[47, 48] Reassuringly, recent simulation studies suggest that although correlation, randomisation method, and non-normality do not have important effects on baseline *P*-value distribution, those calculated from rounded summary statistics are not uniformly distributed.[49] Also, it was found that baseline *P*-value distributions were uniform for continuous but not categorical characteristics.[50] Based on these findings, the true expected (i.e., reference) distributions for baseline *P*-values from rounded summary statistics were established empirically.[51]

Positive findings using this method may be due to one or a combination of the inaccuracy of the method, honest errors regarding data analysis and reporting, chance, or fraud.[47]

Plausibility of correlations between variables, coefficients of variation and standard deviations

Researchers may create false data and use sensible values for a single variable. However, it is difficult to fabricate several variables that together are consistent with real data.[35] By eyeballing the baseline and results sections, unlikely values may come to light. Some variables should be correlated based on knowledge or common sense, the correlation after manipulations of the data may end up too strong or too weak to be plausible.[7, 34, 35, 37, 40, 42, 43]

Similar to correlations, it is difficult to fabricate multiple means and standard deviations for separate variables or groups in a way that they differ enough to be realistic but not so much that it attracts attention. Coefficients of variation indicate variable variation regardless of its unit, defined as dividing the sample standard deviation by the sample mean. Researchers who commit fraud could unconsciously pick coefficients of variation that are too similar for unrelated variables with very different scales.[13]

Fabricated data may tend to have too similar standard deviations to be plausible.[12, 43] When researchers fabricate different means for two or more study arms, they might be reluctant to change the standard deviation. The standard deviation of multiple standard deviations across groups can indicate that they are unrealistically similar.[12]

Date checking and recruitment over time

In presence of raw data, all dates should occur after the first participant being recruited or randomised, and before final events such as death or the end of the study.[35] Also, it

could be checked whether there is a relative irregularity of subject visits taking place during weekends[37] and whether routine measurements were not taken at weekends or holidays,[36] as randomisation or clinic appointments are unlikely to heavily fall on these days. Care must be taken in choosing which dates to check, because dates of death, emergency treatment, or some clinic visits may occur at any time.[35, 36, 42]

Furthermore, the rate by which real participants are recruited might not be perfectly constant over time as studies often have a "start-up" period. Inclusions for fabricated data might be more constant over time.[37] In trials, a comparison of treatment groups by week or month of randomization can reveal periods with unrealistic inclusion.[34]

Repeated measurements and missing data

Some variables are measured repeatedly on the same individuals. An insufficient variability over time may reveal propagation of previous values rather than genuine observations.[34] If data are fabricated the false values may not vary enough compared to real data.[35] Repeated sequences of values of different included individuals can also be found within a whole column by plotting the column values in order.[18]
Also, fabricated data might be "too perfect" in the sense of containing relatively few missing values. Missing data rates can be checked in raw data and missing data rates can be compared between centres in case of a multicentre trial.[37, 42]

Outliers and Inliers: unrealistically large or low variance in data variables

Outliers are observations that appear to be inconsistent with the rest of the data, usually appearing as too large. Here, this method compares the observed value for a single participant to those from all other participants. Outliers at the participant level are more likely to result from errors rather than fraud.[35] On the other hand, researchers who create false data tend to choose values close to the mean of the other observations, as outliers might be noticed by others.[35] Thus, in their tendency to avoid creating outliers, researchers who commit fraud might create odd distributions in which individual values are unusually close to the overall study mean (inliers).[38] Having several participants with a very small difference from the study mean at the same site, the site may require further investigation.[35] ORI uses this principle to compare suspicious datasets with control datasets of a similar topic.[17]

Centre with possible data fabrication

Central statistical monitoring (CSM), using statistical methods to compare data of one centre with data of all the centres combined, focuses on multicentre studies. This principle is based on the assumption that a fraudulent staff member does not have access to any trial data of other centres. Consequently, fabricated observations might be different from true observations.[37]

Most above-mentioned methods that assess data concerns can be used in CSM. Some further comparisons may be helpful. For example, if a particular site has mean values that are very different from the other sites, it might indicate that some participants have been fabricated, or those recruited are so different from other centres that this may require further investigation.[34, 35, 39, 41] Also, fraudulent researchers might wish to demonstrate positive findings. Hartgerink 2019[43] compared genuine and fabricated summary statistics and found that the fabricated effects were in general larger than the genuine ones.

DISCUSSION

This scoping review describes numerous methods to assess research misconduct. The methods to detect textual plagiarism have been regularly implemented as detection tools. Most other methods have not been adequately validated nor structurally implemented. Some methods are based on eyeballing and experience. There is a need for automation to facilitate the detection of potential misconduct.

The strength of this scoping review is that it brings together all literature-reported methods that detect research misconduct in health-related research. We sorted the collection of methods and summarised their applicability to build a quick reference guide for readers. However, it is possible that some unpublished methods were missed in this literature-based effort, especially in-house methods that belong to publishers. These methods are usually commercial products for certain aims and may not have good generalizability. We could not obtain enough information to assess them via the public domain. We are also aware of some methods that pass from mouth-to-mouth, such as implausible productivity of researchers, implausibly high recruitment rates given the stringent eligibility criteria and the capacity of the recruiting centres, and inability to identify the claimed Institutional Review Board. Another limitation is that it was not possible to make a comparison of the available methods because they focus on different dimensions and behaviours of research misconduct. Only a few tests were validated and there was almost no information on how reliable the results are, let alone systematic critical appraisal of the identified methods. Limitations of few methods have been preliminarily discussed, for example, limitations of the increasingly used baseline *P*-value distribution in randomized trials have been touched upon.[64, 65] But for most methods identified in this review, there is no reference to their strengths and weaknesses. Even the validated tests have limitations, as there are still discussions on setting thresholds for plagiarism. These underpin the necessity to use multiple methods for any investigation.

We advise using multiple methods to detect potential research misconduct because a single method is usually insufficient. At this moment there is no one particular method

that we recommend using alone. The main research gap is that we need to know what minimal set of tests are required to optimize detection of misconduct; this includes the necessity of validation of available methods and determining their diagnostic capacity. Second, it always helps to ask for the raw datasets and apply statistical checks. These attempts are usually hampered by the poor accessibility and stewardship of research data. As an obligation of publication, a unified requirement to submit research data to appropriate data repositories along with meta-data like data dictionaries may be part of the solution. Third, it is important to check research governance including protocols, ethics approval, and documentation of study medication as this will contribute to either trust or distrust of the research. Last, we advise automating these methods as much as possible. Automation of "ready" methods would promote wide use. Automation of methods in development would encourage validation and testing. We also encourage new methods to be automated in advance to expedite the process of validation and application.

A thorough investigation of suspected research misconduct is currently a difficult, time-consuming, and labour-intensive process. The scientific community needs to develop better detection tools that are validated. Subsequently, these tools can be automated for routine assessments and tested by the community to proactively defend the integrity of research before publication.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary data

The supplementary data of this study can be downloaded at https://www.sciencedirect.com/science/article/pii/S0895435621001554?via%3Dihub#sec0030.

REFERENCE LIST

1. Fanelli D. How many scientists fabricate and falsify research? A sys- tematic review and meta-analysis of survey data. PLoS One 2009;4 e5738.
2. Office of Research Integrity. Definition of Research Misconduct. https://ori.hhs.gov/definition-research-misconduct
3. Committee on Publication Ethics (COPE). Code of conduct,https://publicationethics.org/files/Code%20of%20Conduct.pdf
4. Wager E. Coping with scientific misconduct. Bmj 2011;343 d6586.
5. Bosch X, Hernández C, Pericas JM, Doti P, Marušić A. Miscon- duct policies in high-impact biomedical journals. PloS one 2012;7 e51928-e.
6. Tricco A, Lillie E, Zarin W, O'Brien K, Colquhoun H, Levac D. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018;169(7):467–73.
7. Bailey KR. Detecting fabrication of data in a multicenter collabora tive animal study. Control Clin Trials 1991;12:741–52.
8. Grey A, Bolland MJ, Avenell A, Klein AA, Gunsalus C. Check for publication integrity before misconduct. Nature Publishing Group; 2020.
9. Smith R. Investigating the previous studies of a fraudulent author. Br Med J. 2005;331:288–91.
10. Carlisle JB. The analysis of 168 randomised controlled trials to test data integrity. Anaesthesia 2012;67:521–37.
11. Bolland MJ, Avenell A, Gamble GD, Grey A. Systematic review and statistical analysis of the integrity of 33 randomized controlled trials. Neurology 2016;87:2391–402.
12. Simonsohn U. Just post it: the lesson from two cases of fabricated data detected by statistics alone. Psychol Sci 2013;24.
13. Hudes ML, McCann J, Ames B. Unusual clustering of coefficients of variation in published articles from a medical biochemistry de- partment in India. FASEB J 2009;23(3):689–703.
14. Spiroski M. How to verify plagiarism of the paper written in Mace- donian and translated in foreign language? Open Access Maced J Med Sci 2016;4:1–4.
15. Bordewijk EM, Wang R, van Wely M, Li W, Mol BW. Data integrity of 10 other randomized controlled trials of an author with a retracted paper. Fertil Steril 2020. https://www.fertstertdialog.com/posts/ data- integrity- of- 10- other- randomized- controlled- trials- of- an- author- with-a-retracted-paper. [Accessed 05 January 2021].
16. Bordewijk EM, Wang R, Askie LM, Gurrin LC, Thornton JG, van Wely M. Data integrity of 35 randomised controlled trials in women' health. Eur J Obstet Gynecol Reprod Biol 2020;249:72–83.
17. Dahlberg JE, Davidian NM. Scientific forensics: how the office of research integrity can assist institutional investigations of research misconduct during oversight review. Sci Eng Ethics 2010;16:713–35.
18. Carlisle JB. False individual patient data and zombie ran- domised controlled trials submitted to Anaesthesia. Anaesthesia 2020;76(4):472–9.
19. Bohannon J. Scientific publishing. Hoax-detecting software spots fake papers. Science 2015;348:18–19.

20. Nguyen M, Labbé C. Engineering a tool to detect automatically generated papers. BIR@ECIR; 2016.
21. Springer and Université Joseph, Fourier release SciDetect to dis- cover fake scientific papers. 2020, https://www.springer.com/gp/ about-springer/media/press-releases/corporate/ scidetect/541662015. [Accessed 05 January 2021].
22. Baydik OD, Gasparyan AY. How to act when research misconduct is not detected by software but revealed by the author of the plagiarized article. J Korean Med Sci 2016;31:1508–10.
23. Wiwanitkit V. How to verify and manage the translational plagia- rism? Maced J Med Sci 2016;4:533.
24. Errami M, Wren JD, Hicks JM, Garner HR. eTBLAST: a web server to identify expert reviewers, appropriate journals and similar publi- cations. Nucleic Acids Res 2007;35:W12–WW5.
25. Errami M, Sun Z, George AC, Long TC, Skinner MA, Wren JD. Identifying duplicate content using statistically improbable phrases. Bioinformatics 2010;26:1453–7.
26. Errami M, Hicks JM, Fisher W, Trusty D, Wren JD, Long TC. Déjà vu—A study of duplicate citations in Medline. Bioinformatics 2007;24:243–9.
27. How to stop plagiarism. Nature 2012;481:21–3.
28. Higgins JR, Lin F-C, Evans JP. Plagiarism in submitted manuscripts: incidence, characteristics and optimization of screening-case study in a major specialty medical journal. Res Integr Peer Rev 2016;1:13.
29. Taylor DB. Plagiarism in manuscripts submitted to the AJR: Devel- opment of an optimal screening algorithm and management pathways. AJR Am J Roentgenol 2017;208:712–20.
30. The office of research integrity (ORI). Forensic Tools, 2020. https://ori.hhs.gov/forensic-tools. [Accessed 05 January 2021].
31. Koppers L, Wormer H, Ickstadt K. Towards a systematic screen- ing tool for quality assurance and semiautomatic fraud detection for images in the life sciences. Sci Eng Ethics 2017;23:1113–28.
32. Acuna DE, Brookes PS, Kording KP. Bioscience-scale automated detection of figure element reuse. BioRxiv 2018:269415.
33. Parrish D, Noonan B. Image manipulation as research misconduct. Sci Eng Ethics 2009;15:161–7.
34. Buyse M, George SL, Evans S, Geller NL, Ranstam J, Scherrer B. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. Stat Med 1999;18:3435–51.
35. Kirkwood AA, Cox T, Hackshaw A. Application of methods for central statistical monitoring in clinical trials. Clin Trials 2013;10:783–806.
36. Taylor RN, McEntegart DJ, Stillman EC. Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. Drug Inf J 2002;36:115–25.
37. van den Bor RM, Vaessen PWJ, Oosterman BJ, Zuithoff NPA, Grobbee DE, Roes KCB. A computationally simple central mon- itoring procedure, effectively applied to empirical trial data with known fraud. J Clin Epidemiol 2017;87:59–69.
38. O'Kelly M. Using statistical techniques to detect fraud: A test case. Pharm Stat 2004;3:237–46.
39. Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, Genevois-Marlin E. A statistical approach to central monitoring of data quality in clinical trials. Clin Trials 2012;9:705–13.

40. Wu X, Carlsson M. Detecting data fabrication in clinical trials from cluster analysis perspective. Pharm Stat 2011;10:257–64.
41. Pogue JM, Devereaux PJ, Thorlund K, Yusuf S. Central sta- tistical monitoring: detecting fraud in clinical trials. Clin Trials 2013;10:225–35.
42. Knepper D, Lindblad AS, Sharma G, Gensler GR, Manukyan Z, Matthews AG. Statistical monitoring in clinical trials: best practices for detecting data anomalies suggestive of fabrication or misconduct. Ther Innov Regul Sci 2016;50:144–54.
43. Hartgerink CHJ, Voelkel JG, Wicherts JM, van Assen MALM. De- tection of data fabrication using statistical tools. PsyArXiv 2019. doi:10.31234/osf.io/jkws4.
44. Carlisle JB, Dexter F, Pandit JJ, Shafer SL, Yentis SM. Calculat- ing the probability of random sampling for continuous variables in submitted or published randomised controlled trials. Anaesthe- sia 2015;70:848–58.
45. Carlisle JB, Loadsman JA. Evidence for non-random sampling in randomised, controlled trials by Yuhji Saitoh. Anaesthesia 2017;72:17–27.
46. Myles PS, Carlisle JB, Scarr B. Evidence for compromised data integrity in studies of liberal peri-operative inspired oxygen. Anaes- thesia 2019;74:573–84.
47. Mascha EJ, Vetter TR, Pittet J-F. An Appraisal of the Carlisle-S- touffer-Fisher Method for Assessing Study Data Integrity and Fraud. Anesth Analg 2017;125:1381–5.
48. Kharasch ED, Houle TT. Seeking and reporting apparent research misconduct: errors and integrity. Anaesthesia 2018;73:125–6.
49. Bolland MJ, Gamble GD, Avenell A, Grey A. Rounding, but not randomization method, non-normality, or correlation, affected base- line P-value distributions in randomized trials. J Clin Epidemiol 2019;110:50–62.
50. Bolland MJ, Gamble GD, Avenell A, Grey A, Lumley T. Baseline P value distributions in randomized trials were uniform for continuous but not categorical variables. J Clin Epidemiol 2019;112:67–76.
51. Bolland MJ, Gamble GD, Grey A, Avenell A. Empirically generated reference proportions for baseline p values from rounded summary statistics. Anaesthesia 2020.
52. Al-Marzouki S, Evans S, Marshall T, Roberts I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. BMJ. 2005;331:267-70.
53. Orita M, Hagiwara Y, Moritomo A, Tsunoyama K, Watanabe T, Ohno K. Agreement of drug discovery data with Benford's law. Expert Opin Drug Discov 2013;8:1–5.
54. Hein J, Zobrist R, Konrad C, Schuepfer G. Scientific fraud in 20 fal- sified anesthesia papers : detection using financial auditing methods. Der Anaesthesist 2012;61:543–9.
55. Pollach G, Brunkhorst F, Mipando M, Namboya F, Mndolo S, Luiz T. The "first digit law" - A hypothesis on its possible impact on medicine and development aid. Med Hypotheses 2016;97:102–6.
56. Hullemann S, Schupfer G, Mauch J. Application of Benford's law: a valuable tool for detecting scientific papers with fabricated data?: A case study using proven falsified articles against a comparison group. Anaesthesist 2017;66:795–802.

57. Epskamp S, Nuijten MB. statcheck: Extract statistics from articles and recompute p values. R package version 1.0.1. http://CRAN. R-project.org/package=statcheck. 2015.
58. Hartgerink C. 688,112 Statistical results: content mining psychology articles for statistical test results. Data 2016;1:14.
59. van der Zee T, Anaya J, Brown NJL. Statistical heartburn: an attempt to digest four pizza publications from the Cornell Food and Brand Lab. BMC Nutr 2017;3:54.
60. Brown NJL, Heathers JAJ. The GRIM test:a simple technique de- tects numerous anomalies in the reporting of results in psychology. Soc Psychol Personal Sci 2017;8:363–9.
61. Anaya J. The GRIMMER test: A method for testing the validity of reported measures of variability. PeerJ Preprints 2016;4 e2400v1.
62. Heathers J, Anaya J, van der Zee T, Brown N. Recovering data from summary statistics: Sample parameter reconstruction via itera- tive techniques. (SPRITE) Peer J Preprints 2018;6:e26968v1.
63. Li W, van Wely M, Gurrin L, Mol BW. Integrity of ran- domized controlled trials: challenges and solutions. Fertil Steril 2020;113:1113–19.
64. Betensky RA, Chiou SH. Correlation among baseline variables yields non-uniformity of p-values. PLoS One 2017;12:e0184531.
65. Bland M. Do baseline P-values follow a uniform distribution in ran- domised trials? PLoS One 2013;8:e76010.

Chapter 9

Summary, implications for practice and future research

Infertility affects about one in seven of all couples that are trying to conceive.[1,2] In 20 to 25% of these couples, the woman suffers from ovulation disorders.[2] The classification of the World Health Organization (WHO) distinguishes three classes of ovulation disorders. This thesis focused on WHO type II ovulation disorder, which concerns the majority, around 85%, of women with ovulation disorders. WHO type II ovulation disorder results from absent or irregular ovulation due to hypothalamic-pituitary-ovarian dysfunction and is also known as normogonadotropic anovulation. The majority of women with WHO type II have polycystic ovary syndrome (PCOS).[2-4]

If women with normogonadotropic anovulation and wish to conceive, strategies to induce ovulation include treatment with clomiphene citrate (CC), letrozole, gonadotrophins, and CC with metformin.[5] Also, intrauterine insemination (IUI) can be added to replace intercourse[2]. CC and gonadotrophins are both well established and effective treatment options and have been used for many years[2,6]. Currently, letrozole is known to be more effective,[5, 7] but the use of letrozole is in many countries off-label for this indication.

Randomized controlled trials (RCTs) are scientific investigations that have the least potential for bias when evaluating the effects of interventions. Findings of RCTs are considered to be the top level of evidence for clinical practice. Systematic reviews and meta-analyses combine the results of multiple RCTs and have an even higher reliability when adequately done. Individual participant data (IPD) can combine evidence of multiple comparisons and they have the power to evaluate effectiveness and safety of interventions within subgroups. As women with PCOS represent a heterogeneous population according to the diagnostic criteria, it is important to identify which individuals benefit most from a particular treatment so that clinicians can provide personalized care.[8] Also, the availability of IPD and the willingness to share data may be an indicator of quality, methodological soundness and integrity of trials when they are being considered for inclusion in systematic reviews. Integrity problems in scientific research often occurs. About 2% of scientists admitted to having fabricated, falsified, or modified data or results at least once and on average over 14% of scientists observed these behaviors among their colleagues.[9]

In this thesis we evaluated the effectiveness, safety, costs and long-term results of the different treatment options in women with normogonadotropic anovulation. Secondly, we evaluated the quality of shared and non-shared RCT's and give an overview of the available methods to investigate research misconduct in health-related research.

Chapter 1 gives a general introduction and describes the background and outline of this thesis.

In **chapter 2** we performed a multicentre RCT, in which we included women with normogonadotropic anovulation not pregnant after six ovulatory cycles of CC. We

compared the effectiveness of gonadotrophins to continued treatment with CC, both with or without IUI. The primary outcome was conception leading to live birth within eight months. Between December 8th 2008 and December 16th 2015, 666 women were randomized to receive gonadotrophins plus IUI (N=166), gonadotrophins plus intercourse (N=165), CC plus IUI (N=163), or CC plus intercourse (N=172). We made two comparisons, one in which gonadotrophins was compared to CC and one in which IUI was compared to intercourse. Women allocated to gonadotrophins had more live births than women allocated to CC (167 of 327 women (51.5%) versus 138 of 334 (41.3%), RR 1.24, 95% CI 1.05 -1.46). Addition of IUI did not increase live births compared to intercourse (161 of 327 women (49.2%) vs. 144 of 334 (43.1%), RR 1.14, 95% CI 0.97-1.35). Multiple pregnancy rates for the two comparisons were low and not different.

The results of this RCT demonstrate that for women with normogonadotropic anovulation and CC failure, a switch of treatment to gonadotrophins increased the chance of livebirth over treatment with CC. There was no evidence that addition of IUI increased livebirth over intercourse.

In **chapter 3** we presented a cost-effectiveness analysis that was performed alongside the RCT of chapter 2. We calculated the direct medical costs of ovulation induction with gonadotrophins versus CC and of IUI versus intercourse. We included costs of medication, cycle monitoring, interventions, and pregnancy leading to live birth. Resource use was collected from the case report forms and unit costs were derived from various sources. We calculated the incremental cost-effectiveness ratios (ICER) for gonadotrophins compared to CC and for IUI compared to intercourse.

Mean direct medical costs per woman receiving gonadotrophins or CC were €4495 versus €3006 (cost difference €1475 (95% CI €1457–€1493)). Live birth rates were 51.5% in women allocated to gonadotrophins and 41.3% in women allocated to CC (RR 1.24, 95% CI 1.05–1.46). The ICER was €15 258 (95% CI €8721 to €63 654) per additional live birth with gonadotrophins. Mean direct medical costs per woman allocated to IUI or intercourse were €4497 versus €3005 (cost difference €1510 (95% CI €1492–€1529)). Live birth rates were 49.2% in women allocated to IUI and 43.1% in women allocated to intercourse (RR 1.14, 95% CI 0.97–1.35). The ICER was €24 361 (95% CI €–11 290 to €85 172) per additional live birth with IUI.

In summary we found that gonadotrophins are more effective, but more expensive than CC. The use of gonadotrophins depends on society's willingness to pay for an additional child. In view of the uncertainty around the cost-effectiveness estimate of IUI, these data are not sufficient to make recommendations on the use of IUI in these women.

In **Chapter 4** we investigated whether endometrial thickness (EMT) could be used as a biomarker to identify women with CC failure who are better off switching to gonadotropins and those who could continue CC. This was a post hoc analysis of the RCT in chapter 2 in which 666 women with CC failure were randomly assigned to receive six cycles with gonadotropins ($n = 331$) or six cycles continuing with CC ($n = 335$), both with IUI or without IUI. EMT was measured mid-cycle before randomization during their sixth ovulatory CC cycle. The EMT was available in 380 women, of whom 190 were allocated to gonadotropins and 190 were allocated to CC. We performed a spline analysis to evaluate the association of EMT with chance to pregnancy leading to a live birth in the next cycles and to determine the best cut-off point. On the basis of the resulting cut-off point, we calculated the live birth rates for gonadotropins versus CC at EMT values below and above this cut-off point.

Mid-cycle EMT in the sixth cycle interacted with treatment effect ($P < 0.01$). Spline analyses showed a cut-off point of 7 mm. There were 162 women (45%) who had an EMT ≤ 7 mm in the sixth ovulatory cycle and 218 women (55%) who had an EMT > 7 mm. Among the women with EMT ≤ 7 mm, gonadotropins resulted in a live birth in 44 of 79 women (56%), while CC resulted in a live birth in 28 of 83 women (34%) (RR 1.57, 95% CI 1.13–2.19). Per additional live birth with gonadotropins, the ICER was €9709 (95% CI €5117 to €25 302). Among the women with EMT > 7 mm, gonadotropins resulted in a live birth in 53 of 111 women (48%) while CC resulted in a live birth in 52 of 107 women (49%) (RR 0.98, 95% CI 0.75–1.29).

This study showed that on basis of a cut-off of 7 mm for EMT, we are able to make a distinction between women who are better off switching to gonadotropins and those who could continue CC after six earlier failed ovulatory CC cycles. Women with six failed ovulatory cycles on CC and an EMT ≤ 7 mm in the sixth cycle are advised to switch to gonadotropins, since it improves live birth rate over continuing treatment with CC at an extra of €9709 to achieve one additional live birth. Women with an EMT > 7 mm are advised to continue treatment with CC, since live birth rates are similar to those with gonadotropins, without the extra costs.

In **Chapter 5** we conducted a follow-up study to investigate the long-term outcomes of switching to gonadotrophins versus continuing treatment with CC. The study population comprised all women who participated in the RCT in chapter 2. The participating women were asked to complete a web-based questionnaire. The primary outcome of this study was cumulative live birth. Secondary outcomes included fertility treatments, clinical pregnancies, multiple pregnancies, miscarriage, stillbirth, and ectopic pregnancy. We retrieved follow-up data for 374 women (57%) of whom 184 had been originally allocated to gonadotrophins and 190 to CC. The median follow-up time was 8,2 years, 154 women

had a live birth (83.7%) in the gonadotrophin group and 150 women had a live birth (78.9%) in the CC group (RR 1.06, 95% CI 0.96 – 1.17). A second live birth occurred in 85 of 184 women (46,2%) in the gonadotrophin group and in 85 of 190 women (44.7%) in the CC group (RR 1.03, 95% CI 0.83 – 1.29). Women allocated to gonadotrophins had a third live birth in 6 of 184 women (3.3%) and women allocated to CC had a third live birth in 14 of 190 women (7.4%). There were respectively 12 and 11 twins in the gonadotrophin and CC groups. The use of fertility treatments in the follow-up period was comparable between both groups. Women with a mid-cycle EMT ≤7 mm in the sixth ovulatory cycle with CC before randomization had a higher first live birth rate with gonadotrophins compared to CC, respectively 95% versus 80%, while in women with an EMT >7mm the live birth rates were 79% in both the gonadotrophin and CC group.

This long-term follow-up study showed that about four in five women with normogonadotropic anovulation and CC failure had a live birth. In terms of pregnancy rates, the long-term differencesbetween switching to gonadotrophins are small compared to continued treatment with CC.

In **Chapter 6** we aimed to evaluate the effectiveness of different ovulation induction agents, in particular letrozole alone and CC plus metformin, as compared to CC alone, as the first-line choice for ovulation induction in women with PCOS and infertility, and to explore interactions between treatment and participant-level baseline characteristics. We searched electronic databases including MEDLINE, EMBASE and Cochrane Central Register of Controlled Trials up to 20 December 2018. We included RCTs comparing the following interventions with each other or placebo/no treatment in women with PCOS and infertility: CC, metformin, CC plus metformin, letrozole, gonadotrophin and tamoxifen. We excluded studies on treatment-resistant women. The primary outcome was live birth. We contacted the investigators of eligible RCTs to share the IPD and performed IPD meta-analyses. We assessed the risk of bias by using the Cochrane risk of bias tool for RCTs.

IPD of 20 RCTs including 3962 women with PCOS were obtained. Six RCTs compared letrozole and CC in 1284 women. Compared with CC, letrozole improved live birth rates (3 RCTs, 1043 women, RR 1.43, 95% CI 1.17-1.75, moderate-certainty evidence). Meta-analyses of effect modifications showed a positive interaction between baseline serum total testosterone levels and treatment effects on live birth (interaction RR 1.29, 95% CI 1.01-1.65). Eight RCTs compared CC plus metformin to CC alone in 1039 women. There was insufficient evidence of a difference on live birth rates (5 RCTs, 907 women, RR 1.08, 95% CI 0.87-1.35, low-certainty evidence). Meta-analyses of effect modifications showed a positive interaction between baseline insulin levels and treatment effects on live birth in the comparison between CC plus metformin and CC (interaction RR 1.03, 95% CI 1.01-1.06).

This IPD meta-analyses showed that in women with PCOS, letrozole improves live birth and clinical pregnancy rates and reduces time-to-pregnancy compared to CC and therefore can be recommended as the preferred first-line treatment for women with PCOS and infertility. CC plus metformin may increase clinical pregnancy and may reduce time-to-pregnancy compared to CC alone, while there is insufficient evidence of a difference on live birth. Treatment effects of letrozole are influenced by baseline serum levels of total testosterone, while those of CC plus metformin are affected by baseline serum levels of insulin. These interactions between treatments and biomarkers on hyperandrogenaemia and insulin resistance provide further insights into a personalized approach for the management of anovulatory infertility related to PCOS.

In **Chapter 7** we performed a meta-epidemiologic study to elucidate if RCTs without IPD sharing have lower quality and more methodological issues than those with IPD sharing in the IPD meta-analysis evaluating first-line ovulation induction for PCOS. We included RCTs identified for the IPD meta-analysis in chapter 6. We dichotomized RCTs according to whether they provided IPD (shared group) or not (non-shared group).

In total, 45 trials (8697 women) were included. IPD was obtained from 17 trials and not available from 28 trials. Pooled risk rates obtained from the shared and non-shared groups were different. Overall low risk of bias was associated with 13/17 (76%) of shared RCTs versus 7/28 (25%) of non-shared RCTs. For RCTs that started recruitment after 1 July 2005, adequate trial registration was found in 3/9 (33%) of shared RCTs versus 0/16 (0%) in non-shared RCTs. In total, 7/17 (41%) of shared RCTs and 19/28 (68%) of non-shared RCTs had issues with the statistical methods described. The median (range) of inconsistency rate per study, between reported and reproduced analyses for baseline variables, was 0% (0-92%) (6 RCTs applicable) in the shared group and 54% (0-100%) (13 RCTs applicable) in the non-shared group. The median (range) of inconsistency rate of univariable statistical results for the outcome(s) per study was 0% (0-63%) (14 RCTs applicable) in the shared group and 44% (0-100%) (24 RCTs applicable) in the non-shared group. The distributions of simulation-generated P-values from comparisons of baseline continuous variables between intervention and control arms suggested that RCTs in the shared group are likely to be consistent with properly conducted randomization ($P = 0.163$), whereas this was not the case for the RCTs in the non-shared group ($P = 4.535 \times 10^{-8}$).

IPD meta-analysis on evaluating first-line ovulation induction for PCOS preserves validity and generates more accurate estimates of risk than meta-analyses using aggregate data, which enables more transparent assessments of benefits and risks. The availability of IPD and the willingness to share these data is a good indicator of quality, methodological soundness and integrity of RCTs when they are being considered for inclusion in systematic reviews and meta-analyses.

In **Chapter 8** we provided an overview of the available methods to investigate research misconduct in health-related research. In this scoping review, we conducted a literature search in MEDLINE, Embase, The Cochrane CENTRAL Register of Studies Online (CRSO), and The Virtual Health Library portal up to July 2020. We included papers that mentioned and/or described methods for screening or assessing research misconduct in health-related research.

We categorized identified methods into the following four groups according to their scopes: overall concern, textual concern, image concern, and data concern.

We included 57 papers reporting on 27 methods: two on overall concern, four on textual concern, three on image concern, and 18 on data concern. Apart from the methods to locate textual plagiarism and image manipulation, all other methods, be it theoretical or empirical, are based on examples, are not standardized, and lack formal validation.

We presented an overview of the available methods to investigate research misconduct in health-related research. We concluded that existing methods cover a wide range of issues regarding research misconduct. Although measures to counteract textual plagiarism are well implemented, tools to investigate other forms of research misconduct are rudimentary and labor-intensive until now. To cope with the rising challenge of research misconduct, further development of automatic tools and routine validation of these methods is needed.

IMPLICATIONS OF PRACTICE

This thesis focused on women with normogonadotropic anovulation. We showed that over a long follow-up period women with normogonadotropic anovulation with CC-failure have a chance of about 80% to conceive at least one live birth. A second live birth occurred in about half of the women. In the follow-up period of 8 months, gonadotrophins had a higher live birth in women with an endometrial thickness <7 mm at higher costs. There was no difference in live birth between gonadotrophins and CC in women with an EMT > 7mm nor was it in the long term follow up period.

In neither M-ovin nor M-ovin follow-up study a difference was found in multiple pregnancy, miscarriage, ectopic pregnancy, and pregnancy complications.

The IPD meta-analyses showed that letrozole improves live birth rates and clinical pregnancy rates and reduces time-to-pregnancy compared to CC. Letrozole is recommended as the preferred first-line treatment for women with normogonadotropic

anovulation and infertility. If letrozole is not available or not a feasible treatment, ovulation induction with CC over 12 cycles is a good alternative. Using CC for a longer period of time reduces costs and is less invasive compared to gonadotrophins. If the EMT is 7 mm or thinner in the 6th CC cycle it could be considered to switch to gonadotrophins as in these women CC resulted in less live births.

Treatment effects of letrozole are influenced by baseline serum levels of total testosterone, while those of CC plus metformin are affected by baseline serum levels of insulin. These interactions between treatments and biomarkers on hyperandrogenaemia and insulin resistance provide further insights into a personalized approach for the management of anovulatory infertility related to PCOS. Together with the couples, counseling by fertility doctors would be in place with this information to form the most personalized fertility treatment.

We showed that IPD meta-analysis preserves validity and generates more accurate estimates of risk than meta-analyses using aggregate data, which enables more transparent assessments of benefits and risks. The availability of IPD and the willingness to share these data may be a good indicator of quality, methodological soundness and integrity of RCTs when they are being considered for inclusion in systematic reviews and meta-analyses. This is why we suggest that future protocols, if available, should be based on the results of IPD meta-analyses.

Regarding research misconduct, existing methods cover a wide range of issues. Although measures to counteract textual plagiarism and image manipulation are well implemented, tools to investigate other forms of research misconduct are rudimentary and labour-intensive. Further development of automatic tools and routine validation of these methods is needed. To detect potential misconduct, we advise using multiple methods because a single method is usually insufficient. At this moment there is no particular method that we recommend using alone. Second, it helps to ask for the raw datasets and apply statistical checks. As an obligation of publication, a unified requirement to submit research data may be part of the solution against research misconduct.[10] Also, it is important to check research governance including protocols, ethics approval, and documentation of study medication as this will contribute to either trust or distrust of the research. Not only should these checks be applied before publication of an article, but also post-publications should be checked for integrity issues. Especially articles of authors known to already having fabricated their articles.[11] Last, we advise automating these methods as much as possible.

IMPLICATIONS FOR FUTURE RESEARCH

When the M-ovin study started, we did not collect data on EMT in the case record forms, but the participating centres performed EMT measurements during treatment according to their local protocol as part of their routine monitoring. We added the EMT in the case record forms after the trial had already included 286 women, and consequently, we only have EMT measurements of 380 women. We believe EMT has clinical implications, but this needs to be confirmed in future studies. Therefore, we advise in future studies evaluating ovulation induction to include measurements of the mid-cycle EMT. Unfortunately, EMT is not included in the international multi-stakeholder core outcome set to be reported in studies on PCOS.[12]

For women with normogonadotropic anovulation it is currently known that letrozole is the most effective first-line treatment in terms of live birth rates.[5, 7] The ESHRE guideline for PCOS does not state how many cycles letrozole should be used before switching to another ovulation induction treatment. Currently, no research has examined the use of letrozole for 12 months as we did in de M-ovin trial with CC.[13] We recommend using a similar study protocol as the M-ovin trial, to study continued ovulation induction with letrozole versus switching to gonadotrophins in women who did not conceive after six cycles of letrozole. Possibly even with a third arm with switching to 6 cycles of CC.

Metformin is an insulin-sensitizing agent that decreases gluconeogenesis and lipogenesis and enhances peripheral glucose uptake and therefore increases insulin sensitivity.[14] The IPD-analyses provides preliminary evidence that there may be a role for assessing insulin resistance in PCOS and infertility, as addition of metformin may improve insulin resistance in women with higher fasting insulin levels and therefore improve pregnancy rates. The international evidence-based guideline does not recommend clinical measurement of insulin resistance at present due to the lack of accuracy.[13] In addition, SHBG has been proposed as a measure of insulin resistance,[15] but the findings in our IPD meta-analysis did not support treatment-by-SHBG interactions. Our work supports the need to assess insulin resistance in future infertility studies.

IPD meta-analyses are useful to inform the design, conduct, analysis and interpretation of trials.[16] Given the consistent treatment benefits of letrozole across different fertility outcomes, future trials investigating new interventions for PCOS should choose letrozole as the reference arm. New trials are encouraged to incorporate treatment selection markers in their design to guide treatment decision,[17] and the impact of these, including age, BMI and other biomarkers, needs to be confirmed in future trials. More specifically, biomarkers for hyperandrogenaemia and as mentioned above insulin resistance could be applied in trials that evaluate metformin.

The core outcome set for infertility and PCOS includes these biomarkers and this core outcome set[12] should be used to ensure outcomes are reported and collected consistently across future trials on infertility and PCOS to reduce research waste.

Increasingly, researchers are agreeing to share IPD from RCTs to create databases of studies for secondary analysis. Apart from conducting IPD meta-analyses and, more generally, addressing questions requiring evidence synthesis, combining IPD across multiple studies allows for harmonization of variables, standardization of data analysis, uniform presentation of results across the contributing RCTs, the investigation of subgroup effects and the generation of new hypotheses.[18, 19] We encourage authors and institutions to share their trial data.

We can't afford to include trials with compromised data integrity in meta-analyses because policy-makers, doctors and patients using these meta-analyses could be misled and harms are irreparable.[20] At present, there is little consideration about data reproducibility and integrity of trials when performing meta-analyses using aggregate data, and guidelines on how to handle these critical issues are lacking.

The poor performance in the non-shared group regarding all pre-specified criteria challenges the legitimacy of the traditional approach to include all relevant trials in systematic reviews and meta-analyses. While this inclusive approach minimizes the risk of publication bias, methodological and integrity issues of included trials cannot be rectified by rigorous evidence synthesis. Therefore, we advise to perform more IPD meta-analyses and to primarily base guidelines and protocols on IPD meta-analyses. Although assessing data availability bias is an intrinsic part of IPD meta-analysis, only a minority analyzed the underlying reasons for the differences between results with or without aggregate data.[19] The findings in our study highlight the importance of assessing data availability bias and the reasons beneath. We advise, after performing an IPD meta-analyses, to assess the studies and compare the shared trials with the non-shared trials in terms of quality, methodological soundness and integrity.

A thorough investigation of suspected research misconduct is currently a difficult, time-consuming, and labour-intensive process. The scientific community needs to develop better detection tools that are validated. Subsequently, these tools can be automated for routine assessments and tested by the community to proactively defend the integrity of research before publication. Automation of "ready" methods would promote wide use. Automation of methods in development would encourage validation and testing. We also encourage new methods to be automated in advance to expedite the process of validation and application. Currently our group is developing and validating a data integrity checklist. The main research gap is that we need to know what minimal set of tests are required to optimize detection of misconduct; this includes the necessity of validation of available methods and determining their diagnostic capacity.

REFERENCE LIST

1. World Health Organization (WHO). International Classification of Diseases, 11th Revision (ICD-11) Geneva: WHO 2018.
2. NICE. Fertility problems: assessment and treatment. London: National Institute for Health and Care Excellence (NICE); 2017 Sep. PMID: 32134604.
3. National Collaborating Centre for Ws, Children's H. National Institute for Health and Clinical Excellence: Guidance. Fertility: Assessment and Treatment for People with Fertility Problems. London: Royal College of Obstetricians & Gynaecologists National Collaborating Centre for Women's and Children's Health.; 2013.
4. Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril* 2016;106:6–15.
5. Wang R, Kim BV, van Wely M, et al. Treatment strategies for women with WHO group II anovulation: systematic review and network meta-analysis. *BMJ* 2017; 356: j138.
6. Brown J, Farquhar C. Clomiphene and other antioestrogens for ovulation induction in polycystic ovarian syndrome. *Cochrane Database Syst Rev* 2016; 12: CD002249.
7. Franik S, Eltrop SM, Kremer JA, Kiesel L, Farquhar C. Aromatase inhibitors (letrozole) for subfertile women with polycystic ovary syndrome. The Cochrane database of systematic reviews. 2018;5:Cd010287.
8. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
9. Fanelli D. How many scientists fabricate and falsify research? A sys- tematic review and meta-analysis of survey data. PLoS One 2009;4 e5738.
10. Carlisle JB. False individual patient data and zombie randomised controlled trials submitted to Anaesthesia. Anaesthesia. 2021 Apr;76(4):472-479. doi: 10.1111/anae.15263. Epub 2020 Oct 11. PMID: 33040331.
11. Bordewijk EM, Wang R, van Wely M, et al. Data integrity of 10 other randomized controlled trials of an author with a retracted paper. Fertility and Sterility Dialog; 2020 June. https://www.fertstertdialog.com/posts/data-integrity-of-10-oth- er-randomized-controlled-trials-of-an-author-with-a-retracted-paper.
12. Al Wattar BH, Teede H, Garad R, Franks S, Balen A, Bhide P, Piltonen T, Romualdi D, Laven J, Thondan M, Bueno-Cavanillas A, Moss N, Andrews C, Hawkes R, Mol BW, Khan KS, Thangaratinam S. Harmonising research outcomes for polycystic ovary syndrome: an international multi-stakeholder core outcome set. Hum Reprod. 2020 Feb 29;35(2):404-412. doi: 10.1093/humrep/dez272. PMID: 32020203.
13. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Human reproduction (Oxford, England). 2018;33(9):1602-18.
14. Naderpoor N, Shorakae S, de Courten B, Misso ML, Moran LJ, Teede HJ. Metformin and lifestyle modification in polycystic ovary syndrome: systematic review and meta-analysis.

Hum Reprod Update 2015;21:560–574.

15. Cassar S, Misso ML, Hopkins WG, Shaw CS, Teede HJ, Stepto NK. Insulin resistance in polycystic ovary syndrome: a systematic review and meta-analysis of euglycaemic-hyperinsulinaemic clamp studies. *Hum Reprod* 2016;31:2619–2631.
16. Tierney JF, Pignon JP, Gueffyier F, Clarke M, Askie L, Vale CL, Burdett S, Cochrane IPDM-aMG. How individual participant data meta analyses have influenced trial design, conduct, and analysis. *J Clin Epidemiol* 2015;68:1325–1335.
17. Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. *Ann Intern Med* 2011;154:253–259.
18. Mbuagbaw L, Foster G, Cheng J, Thabane L. Challenges to complete and useful data sharing. Trials 2017;18:71.
19. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selec-tion bias, and unavailable data in meta-analyses using individual par- ticipant data: a database survey. BMJ 2012;344:d7762. 20.Li W, van Wely M, Gurrin L, Mol BW. Integrity of randomized controlled trials: challenges and solutions. Fertil Steril 2020;113: 1113–1119.

Chapter 10

Nederlandse samenvatting, implicaties voor de klinische praktijk en toekomstig onderzoek

Verminderde vruchtbaarheid komt voor bij ongeveer een op de zeven paren die proberen zwanger te worden.[1,2] Bij 20 tot 25% van deze paren heeft de vrouw een ovulatiestoornis.[2] De classificatie van de Wereldgezondheidsorganisatie (WHO) onderscheidt drie typen ovulatiestoornissen. Dit proefschrift richt zich op de WHO type II ovulatiestoornis, dat de meerderheid, ongeveer 85%, van de vrouwen met ovulatiestoornissen betreft. WHO type II ovulatiestoornis is het gevolg van afwezige of onregelmatige ovulatie als gevolg van hypothalamus-hypofyse-ovarium disfunctie en is ook bekend als normogonadotrope anovulatie. De meerderheid van de vrouwen met WHO type II heeft polycysteus-ovariumsyndroom (PCOS).

Bij vrouwen met een normogonadotrope anovulatiestoornis die zwanger willen worden, is ovulatie inductie mogelijk middels behandeling met clomifeencitraat (CC), letrozol, gonadotrofines en CC gecombineerd met metformine.[5] Daarnaast kan intra-uteriene inseminatie (IUI) worden ingezet om coïtus te vervangen.[2] Ovulatie-inductie met CC en gonadotrofinen zijn erkende, effectieve behandelopties die al vele jaren worden ingezet.[2,6] Van letrozol is bekend dat het de meest effectieve behandeling is,[5, 7] maar het gebruik van letrozol is in veel landen niet geregistreerd voor deze indicatie.

Gerandomiseerde gecontroleerde onderzoeken (RCT's) zijn wetenschappelijke onderzoeken die de minste kans op vertekening hebben bij het evalueren van de effecten van interventies. Bevindingen van RCT's worden beschouwd als het hoogste niveau van bewijs voor de klinische praktijk. Systematische reviews en meta-analyses combineren de resultaten van meerdere RCT's en hebben een nog hogere betrouwbaarheid als ze adequaat worden uitgevoerd. Door individuele patiëntengegevens uit de databases (IPD) te gebruiken, kunnen we meerdere vergelijkingen combineren en is het mogelijk om de effectiviteit en veiligheid van interventies binnen subgroepen te evalueren. Aangezien vrouwen met PCOS volgens de diagnostische criteria een heterogene populatie vertegenwoordigen, is het belangrijk om vast te stellen welke subgroep het meest baat heeft bij een bepaalde behandeling, zodat clinici zorg op maat kunnen bieden.[8] Ook kunnen de beschikbaarheid van IPD en de bereidheid om gegevens te delen een indicator zijn van kwaliteit, methodologische correctheid en integriteit van onderzoeken. Problemen met integriteit in wetenschappelijke artikelen komen regelmatig voor. Ongeveer 2% van de wetenschappers gaf toe minstens één keer gegevens of resultaten te hebben verzonnen, vervalst of gewijzigd en gemiddeld meer dan 14% van de wetenschappers observeerde dit gedrag bij hun collega's.[9]

In dit proefschrift evalueerden we de effectiviteit, veiligheid, kosten en lange termijn resultaten van de verschillende behandelopties bij vrouwen met normogonadotrope anovulatie. Ten tweede hebben we de kwaliteit van RCT's waarvan de data wel of niet

gedeeld werden vergeleken en een overzicht gegeven van de beschikbare methoden om integriteitsproblemen in medische artikelen te onderzoeken.

Hoofdstuk 1 geeft een algemene introductie en beschrijft de achtergrond en opzet van dit proefschrift.

In **hoofdstuk 2** hebben we een multicenter RCT uitgevoerd, waarin we vrouwen includeerden met normogonadotrope anovulatie, die niet zwanger waren na zes ovulatoire cycli met CC. We vergeleken de effectiviteit van gonadotrofinen met voortgezette behandeling met CC, zowel met als zonder IUI. De primaire uitkomstmaat was conceptie die binnen acht maanden tot een levend geboren kind leidde. Tussen 8 december 2008 en 16 december 2015 werden 666 vrouwen gerandomiseerd om gonadotrofinen plus IUI (N=166), gonadotrofinen plus coïtus (N=165), CC plus IUI (N=163) of CC plus coïtus (N=172) te krijgen. We maakten twee vergelijkingen, één waarin gonadotrofinen werden vergeleken met CC en één waarin IUI werd vergeleken met coïtus. Vrouwen die gerandomiseerd waren voor behandeling met gonadotrofinen hadden vaker een levend geboren kind dan vrouwen toegewezen aan CC (167 van 327 vrouwen (51,5%) versus 138 van 334 (41,3%), RR 1,24, 95% CI 1,05 -1,46). Toevoeging van IUI verhoogde het aantal levend geboren kinderen niet in vergelijking met coïtus (161 van 327 vrouwen (49,2%) versus 144 van 334 (43,1%), RR 1,14, 95% CI 0,97-1,35). Het aantal meerlingzwangerschappen voor de twee vergelijkingen was laag en niet verschillend.

De resultaten van deze RCT laten zien dat voor vrouwen met normogonadotrope anovulatie en CC-falen, een switch naar gonadotrofinen de kans op een levend geboren kind vergroot ten opzichte van behandeling met CC. Er was geen bewijs dat toevoeging van IUI het percentage verhoogde.

In **hoofdstuk 3** presenteerden we een kosteneffectiviteitsanalyse die werd uitgevoerd naast de RCT van hoofdstuk 2. We berekenden de directe medische kosten van ovulatie-inductie met gonadotrofinen versus CC en van IUI versus coïtus. We hebben hierbij de kosten van medicatie, cyclusmonitoring, interventies en zwangerschap tot levend geboren kind meegenomen. Details omtrent het gebruik van interventies werd verkregen uit de patiëntformulieren. Voor de kosten per behandeleenheid zijn eerder vastgestelde kostprijzen gebruikt en voor bevallingskosten schattingen uit de literatuur. We berekenden de incrementele kosteneffectiviteitsratio's (ICER) voor gonadotrofinen in vergelijking met CC en voor IUI in vergelijking met geslachtsgemeenschap.

De gemiddelde directe medische kosten per vrouw die gonadotrofinen of CC kregen, waren €4495 versus €3006 (kostenverschil €1475 (95% CI €1457–€1493)). Het aantal levend geboren kinderen was 51,5% bij vrouwen toegewezen aan gonadotrofinen en

41,3% bij vrouwen toegewezen aan CC (RR 1,24, 95% CI 1,05-1,46). De ICER bedroeg € 15 258 (95% CI € 8721 tot € 63 654) per extra levend geboren kind met gonadotrofinen. De gemiddelde directe medische kosten per vrouw toegewezen aan IUI of coïtus waren € 4497 versus € 3005 (kostenverschil € 1510 (95% CI € 1492–€ 1529). Het aantal levend geboren kinderen was 49,2% bij vrouwen toegewezen aan IUI en 43,1% bij vrouwen toegewezen aan coïtus (RR 1,14, 95% CI 0,97-1,35). De ICER bedroeg €24 361 (95% CI € 11 290 tot € 85 172) per extra levend geboren kind met IUI.

Samenvattend vonden we dat gonadotrofinen effectiever zijn, maar duurder dan CC. Het gebruik van gonadotrofinen hangt af van de bereidheid van de samenleving om voor een extra kind te betalen. Gezien de onzekerheid rond de schatting van de kosteneffectiviteit van IUI, zijn deze gegevens niet voldoende om aanbevelingen te doen over het gebruik van IUI bij deze vrouwen.

In **hoofdstuk 4** hebben we onderzocht of endometriumdikte (EMD) kan worden gebruikt als biomarker om vrouwen met normogonadotrope anovulatie en CC-falen te identificeren die beter af zijn om over te schakelen op gonadotrofinen en degenen die CC kunnen voortzetten. Dit was een post-hoc analyse van de RCT in hoofdstuk 2 waarin 666 vrouwen willekeurig werden toegewezen aan zes cycli met gonadotrofinen (n = 331) of doorgingen met zes cycli met CC (n = 335), zowel met als zonder IUI. De EMD werd halverwege de cyclus gemeten vóór randomisatie tijdens hun zesde ovulatoire CC-cyclus. De EMD was beschikbaar bij 380 vrouwen, van wie 190 werden toegewezen aan gonadotrofinen en 190 werden toegewezen aan CC. We hebben een spline-analyse uitgevoerd om de associatie van EMD met kans op zwangerschap die leidt tot een levend geboren kind in de volgende cycli te evalueren en om het beste afkappunt te bepalen. Op basis van het resulterende afkappunt berekenden we het percentage levend geboren kinderen voor gonadotrofinen versus CC bij EMD-waarden onder en boven dit afkappunt.

Midcyclische EMD in de zesde cyclus interacteerde met het behandeleffect ($P < 0.01$). Spline-analyses toonden een afkappunt van 7 mm. Er waren 162 vrouwen (45%) met een EMD ≤ 7 mm in de zesde ovulatoire cyclus en 218 vrouwen (55%) met een EMD > 7 mm. Bij de vrouwen met EMD ≤ 7 mm resulteerden gonadotrofinen bij 44 van de 79 vrouwen (56%) in een zwangerschap die leidt tot een levend geboren kind, terwijl CC resulteerde in een levend geboren kind bij 28 van de 83 vrouwen (34%) (RR 1,57, 95% CI 1,13– 2.19). De ICER per extra levendgeborene na gonadotrofinen was de ICER €9709 (95% CI € 5117 tot € 25.302). Bij de vrouwen met EMD > 7 mm resulteerden gonadotrofines in een levend geboren kind bij 53 van de 111 vrouwen (48%), terwijl CC resulteerde in een levend geboren kind bij 52 van de 107 vrouwen (49%) (RR 0,98, 95% CI 0,75-1,29).

Deze studie toonde aan dat we op basis van een cut-off van 7 mm voor EMD een onderscheid kunnen maken tussen vrouwen die beter kunnen overstappen op gonadotrofinen en degenen die CC konden voortzetten na zes mislukte ovulatoire CC-cycli. Vrouwen met zes mislukte ovulatoire cycli op CC en een EMD < 7 mm in de zesde cyclus wordt geadviseerd om over te schakelen op gonadotrofinen, aangezien dit het percentage levend geboren kinderen verbetert ten opzichte van voortgezette behandeling met CC waarbij de kosten €9709 hoger zijn om één extra levendgeborene te bereiken. Vrouwen met een EMD > 7 mm wordt geadviseerd om de behandeling met CC voort te zetten, aangezien de geboortecijfers vergelijkbaar zijn met die met gonadotrofinen, zonder de extra kosten.

In **hoofdstuk 5** hebben we een vervolgstudie uitgevoerd om de lange termijn uitkomsten van het overschakelen op gonadotrofines versus het voortzetten van de behandeling met CC te onderzoeken. De onderzoekspopulatie omvatte alle vrouwen die deelnamen aan de RCT in hoofdstuk 2. De deelnemende vrouwen werd gevraagd een online vragenlijst in te vullen. De primaire uitkomst van deze studie was het cumulatieve aantal levend geboren kinderen. Secundaire uitkomsten waren vruchtbaarheidsbehandelingen, klinische zwangerschappen, meerlingzwangerschappen, miskramen, IUVD's en buitenbaarmoederlijke zwangerschappen. We hebben follow-up gegevens verkregen van 374 vrouwen (57%) van wie 184 oorspronkelijk waren toegewezen aan gonadotrofinen en 190 aan CC. De mediane follow-up duur was 8,2 jaar, 154 vrouwen hadden een zwangerschap die resulteerde in een levend geboren kind (83,7%) in de gonadotrofinegroep ten opzichte van 150 vrouwen (78,9%) in de CC-groep (RR 1,06, 95% CI 0,96 – 1,17). In de gonadotrofinegroep kregen 85 van de 184 vrouwen (46,2%) een tweede kind vergeleken met 85 van de 190 vrouwen (44,7%) in de CC-groep (RR 1,03, 95% CI 0,83 – 1,29). Van de vrouwen toegewezen aan gonadotrofinen kregen 6 van de 184 vrouwen (3,3%) een derde kind en in de CC-groep waren dit 14 van de 190 vrouwen (7,4%). Er waren respectievelijk 12 en 11 tweelingen in de gonadotrofine- en CC-groepen. Het gebruik van vruchtbaarheidsbehandelingen in de follow-up periode was vergelijkbaar tussen beide groepen. Vrouwen met een midcyclische EMD ≤7 mm in de zesde ovulatoire cyclus met CC vóór randomisatie hadden een hoger percentage eerste levend geboren kinderen na gonadotrofines vergeleken met CC, respectievelijk 95% versus 80%, terwijl bij vrouwen met een EMD >7 mm de levendgeborene percentages gelijk waren; 79% in zowel de gonadotrofine- als de CC-groep.
Deze follow-up studie toonde aan dat ongeveer vier op de vijf vrouwen met normogonadotrope anovulatie en CC-falen een levend geboren kind kreeg. De lange termijn verschillen in geboortecijfers tussen voortgezette behandeling met CC en overschakelen naar gonadotrofinen bleken klein.

In **Hoofdstuk 6** wilden we de effectiviteit evalueren van verschillende behandelingen voor ovulatie-inductie, in het bijzonder letrozol en CC in combinatie met metformine,

in vergelijking met CC alleen, als eerstekeuze voor ovulatie-inductie bij vrouwen met PCOS en anovulatie. Daarnaast is bekeken of er interacties waren tussen behandeling en baseline kenmerken op patiëntniveau. Elektronische databases zijn doorzocht, waaronder MEDLINE, EMBASE en Cochrane Central Register of Controlled Trials tot 20 december 2018. We hebben RCT's geïncludeerd waarin de volgende interventies met elkaar werden vergeleken of placebo/geen behandeling bij vrouwen met PCOS en anovulatie: CC, metformine, CC plus metformine, letrozol, gonadotrofine en tamoxifen. Studies over therapieresistente vrouwen zijn uitgesloten. De primaire uitkomstmaat was een levend geboren kind. We namen contact op met de onderzoekers van in aanmerking komende RCT's om de IPD te delen en voerden IPD-meta-analyses uit. We hebben het risico op bias beoordeeld met behulp van de Cochrane risk of bias-tool voor RCT's.

Er werden individuele patiënt data (IPD) verkregen vanuit 20 RCT's, met een totaal van 3962 vrouwen met PCOS. Zes RCT's vergeleken letrozol en CC bij 1284 vrouwen. Vergeleken met CC leidde letrozol tot meer levend geboren kinderen (3 RCT's, 1043 vrouwen, RR 1,43, 95% CI 1,17-1,75, bewijs met matige zekerheid). Meta-analyses van effectmodificaties lieten een positieve interactie zien tussen baseline serum spiegels van totaal testosteron en de kans op een levend geboren kind bij het gebruik van letrozol (interactie RR 1,29, 95% CI 1,01-1,65). Acht RCT's vergeleken CC plus metformine met CC alleen bij 1039 vrouwen en vonden onvoldoende bewijs voor een verschil in levendgeborenen (5 RCT's, 907 vrouwen, RR 1,08, 95% CI 0,87-1,35, bewijs met lage zekerheid). Meta-analyses van effectmodificaties lieten een positieve interactie zien tussen baseline insulinespiegels en behandelingseffecten op levend geboren kinderen in de vergelijking met CC plus metformine en CC (interactie RR 1,03, 95% CI 1,01-1,06).

Deze IPD-meta-analyses toonden aan dat letrozol bij vrouwen met PCOS de kans op een levend geboren kind en klinische zwangerschap verbetert en de tijd tot zwangerschap verkort in vergelijking met CC. Daarom kan letrozol worden aanbevolen als eerste keus voor vrouwen met PCOS en anovulatie. CC plus metformine kan de klinische zwangerschapskans verhogen en de tijd tot zwangerschap verkorten in vergelijking met CC alleen, terwijl er onvoldoende bewijs is voor een verschil in percentage levend geboren kinderen. De effecten van de behandeling van letrozol worden beïnvloed door baseline serumspiegels van totaal testosteron, terwijl die van CC plus metformine worden beïnvloed door baseline serumspiegels van insuline. Deze interacties tussen behandelingen en biomarkers op het gebied van hyperandrogenemie en insulineresistentie bieden meer inzicht in een gepersonaliseerde aanpak voor vrouwen met PCOS en anovulatie.

In **hoofdstuk 7** hebben we een meta-epidemiologische studie (IPD meta-analyse) uitgevoerd om op te helderen of RCT's die hun database niet delen ten behoeve van een IPD, een lagere kwaliteit en meer methodologische problemen hebben dan RCTs die hun

database wel delen. Deze studie betrof de RCT's die eerstelijns ovulatie-inductie voor PCOS onderzochten, zoals geïdentificeerd en geïncludeerd in hoofdstuk 6. We hebben RCT's gedichotomiseerd al naar gelang ze hun data wel (gedeelde groep) of niet (niet-gedeelde groep) aanleverden.

In totaal werden 45 trials (8697 vrouwen) geïncludeerd. Data ten behoeve van de IPD-meta-analyse werd verkregen van 17 RCT's en waren niet beschikbaar van 28 trials. Gepoolde risicopercentages verkregen uit de gedeelde en niet-gedeelde groepen waren verschillend. Over het algemeen was een laag risico op bias geassocieerd met 13/17 (76%) gedeelde RCT's versus 7/28 (25%) niet-gedeelde RCT's. Voor RCT's die na 1 juli 2005 waren gestart met rekruteren, werd adequate registratie gevonden in 3/9 (33%) gedeelde RCT's versus 0/16 (0%) in niet-gedeelde RCT's. In totaal had 7/17 (41%) van de gedeelde RCT's en 19/28 (68%) van de niet-gedeelde RCT's problemen met de beschreven statistische methoden. De mediaan (bereik) van inconsistentie per onderzoek, tussen gerapporteerde en gereproduceerde analyses voor baselinevariabelen, was 0% (0-92%) (6 RCT's van toepassing) in de gedeelde groep en 54% (0-100%) (13 RCT's van toepassing) in de niet-gedeelde groep. De mediaan (bereik) van inconsistentie van univariabele statistische resultaten voor de uitkomst(en) per onderzoek was 0% (0-63%) (14 RCT's van toepassing) in de gedeelde groep en 44% (0-100%) (24 RCT's) van toepassing in de niet-gedeelde groep. De verdelingen van door simulatie gegenereerde P-waarden uit vergelijkingen van continue variabelen bij baseline tussen interventie- en controle-armen suggereerden dat RCT's in de gedeelde groep waarschijnlijk consistent zijn met correct uitgevoerde randomisatie (P = 0,163), terwijl dit niet het geval was voor de RCT's in de niet-gedeelde groep ($P = 4.535 \times 10\text{-}8$).

Het verrichten van een IPD-meta-analyse ten behoeve van de evaluatie welke ovulatie-inductie behandeling bij PCOS het beste is heeft als voordeel het behoud van validiteit en genereert nauwkeurigere risicoschattingen dan gewone meta-analyses die gebruik maken van geaggregeerde gegevens. De beschikbaarheid van IPD en de bereidheid om deze gegevens te delen is een goede indicator voor kwaliteit, methodologische betrouwbaarheid en integriteit van RCT's.

In **hoofdstuk 8** hebben we een overzicht gegeven van de beschikbare methoden om problemen met wetenschappelijke integriteit in medische artikelen te onderzoeken. In deze scoping review hebben we tot juli 2020 literatuuronderzoek gedaan in MEDLINE, Embase, The Cochrane CENTRAL Register of Studies Online (CRSO) en The Virtual Health Library-portal. Artikelen zijn geïncludeerd die methoden voor screening of het beoordelen van wetenschappelijk integriteit in medisch onderzoek beschrijven.

We hebben de geïdentificeerde methoden ingedeeld in de volgende vier groepen: algemeen, tekstueel, afbeelding en gerapporteerde gegevens. Er werden 57 publicaties geïncludeerd die rapporteerden over 27 verschillende methoden. Afgezien van de methoden om tekstplagiaat en manipulatie van afbeelding te beoordelen, zijn alle andere methoden, theoretisch of empirisch, gebaseerd op voorbeelden, niet gestandaardiseerd en ontberen formele validatie.

We presenteerden een overzicht van de beschikbare methoden om problemen met wetenschappelijke integriteit in medische artikelen te onderzoeken. We concludeerden dat met bestaande methoden een breed scala aan problemen met wetenschappelijk integriteit onderzocht kunnen worden. Hoewel maatregelen om tekstplagiaat tegen te gaan goed zijn geïmplementeerd, zijn methoden om andere vormen van potentiële wetenschappelijk fraude te onderzoeken tot nu toe rudimentair en arbeidsintensief. Om het hoofd te bieden aan de toenemende uitdaging van fraude in onderzoek, is verdere ontwikkeling van geautomatiseerde hulpmiddelen en routinematige validatie van deze methoden nodig.

IMPLICATIES VOOR DE PRAKTIJK

Dit proefschrift richt zich op onderzoek bij vrouwen met normogonadotrope anovulatie. We toonden aan dat vrouwen met normogonadotrope anovulatie met CC-falen na een lange follow-up periode een kans van ongeveer 80% hebben om ten minste één levend geboren kind te baren. Een bevalling van een tweede kind vond plaats bij ongeveer de helft van de vrouwen. In de korte follow-up periode van 8 maanden resulteerden gonadotrofinen een hoger percentage levend geboren kinderen bij vrouwen met een endometriumdikte <7 mm tegen hogere kosten. Er was geen verschil in percentage levend geboren kinderen tussen gonadotrofines en CC bij vrouwen met een EMD > 7 mm, en evenmin tijdens de follow-up op lange termijn.

Noch in M-ovin, noch het vervolgonderzoek werd een verschil gevonden in meerlingzwangerschappen, miskramen, buitenbaarmoederlijke zwangerschappen en zwangerschapscomplicaties.

De IPD-meta-analyses toonden aan dat letrozol het aantal levend geboren kinderen en klinische zwangerschappen verbetert en de tijd tot zwangerschap verkort in vergelijking met CC. Letrozol wordt aanbevolen als eerste keuze behandeling bij voorkeur voor vrouwen met normogonadotrope anovulatie. Als letrozol niet beschikbaar is of geen haalbare behandeling is, is ovulatie-inductie met CC gedurende 12 cycli een goed alternatief. Het gebruik van CC voor een langere periode verlaagt de kosten en is minder

invasief in vergelijking met gonadotrofines. Als de EMD 7 mm of dunner is in de 6e CC-cyclus, kan worden overwogen om over te schakelen op gonadotrofines, aangezien CC bij deze vrouwen leidde tot minder levend geboren kinderen.

De effecten van de behandeling van letrozol worden beïnvloed door baseline serumspiegels van totaal testosteron, terwijl die van CC plus metformine worden beïnvloed door baseline serumspiegels van insuline. Deze interacties tussen behandelingen en biomarkers op hyperandrogenemie en insulineresistentie bieden meer inzicht in een gepersonaliseerde aanpak van behandelopties bij vrouwen met PCOS en anovulatie. Met deze informatie kunnen fertiliteitsartsen in samenenspraak met de paren de meest gepersonaliseerde vruchtbaarheidsbehandeling kiezen.

We toonden aan dat IPD-meta-analyse de validiteit verhoogd en nauwkeurigere risicoschattingen genereert dan meta-analyses die gebruik maken van geaggregeerde gegevens, wat een transparantere beoordeling van werkzaamheid en veiligheid van de behandelingen mogelijk maakt. De beschikbaarheid van IPD en de bereidheid om deze gegevens te delen, kunnen een goede indicator zijn voor de kwaliteit, methodologische betrouwbaarheid en integriteit van RCT's wanneer ze worden opgenomen in systematische reviews en meta-analyses. Daarom stellen we voor om toekomstige protocollen, indien beschikbaar, te baseren op de resultaten van IPD-meta-analyses.

Met betrekking tot wetenschappelijke fraude bestrijken de bestaande methoden een breed scala aan problemen. Hoewel maatregelen voor het tegengaan van tekstplagiaat en manipulatie van afbeeldingen goed zijn geïmplementeerd, zijn instrumenten om andere vormen van wetenschappelijk fraude te onderzoeken nog rudimentair en arbeidsintensief. Verdere ontwikkeling van geautomatiseerde tools en validatie van deze methoden is nodig. Om mogelijke integriteitsproblematiek op te sporen, raden we aan om meerdere methoden te gebruiken, omdat één methode meestal niet afdoende is. Op dit moment is er niet één specifieke methode die we aanbevelen. Ten tweede helpt het om de ruwe datasets op te vragen en statistische controles uit te voeren. Als publicatieverplichting kan een uniforme eis om onderzoeksgegevens in te dienen een deel van de oplossing zijn ter voorkoming van wetenschappelijke fraude.[10] Het is ook belangrijk om de toezichtsformulieren op het wetenschappelijke onderzoek te controleren, inclusief de studieprotocollen, ethische goedkeuringsprocedure en documentatie van de onderzoeksmedicatie, aangezien dit zal bijdragen aan vertrouwen van het medisch wetenschappelijk onderzoek. Deze controles moeten niet alleen worden uitgevoerd vóór publicatie van een artikel, maar ook na publicatie moet worden gecontroleerd op integriteitskwesties. In het bijzonder artikelen van auteurs van wie bekend is dat ze in eerdere artikelen hebben gefraudeerd.[11] Tenslotte adviseren we deze methoden zoveel mogelijk te automatiseren.

IMPLICATIES VOOR TOEKOMSTIG ONDERZOEK

Toen de M-ovin-studie begon, verzamelden we geen gegevens over de EMD in de dossierformulieren, maar de deelnemende centra voerden EMD-metingen uit tijdens de behandeling volgens hun lokale protocol als onderdeel van hun routinematige monitoring. We hebben de EMD toegevoegd aan de dossierformulieren nadat er al 286 vrouwen waren geïncludeerd, met als gevolg dat we EMD-metingen van de resterende 380 vrouwen hebben. Naar aanleiding van ons onderzoek denken we dat EMD mogelijk klinische implicaties heeft, maar dit moet in toekomstige studies worden bevestigd. Daarom adviseren we om in toekomstige studies die ovulatie-inductie evalueren om metingen van de midcyclische EMD op te nemen. Helaas is EMD niet opgenomen in de internationale multi-stakeholder kern uitkomsten die zullen worden geadviseerd te gebruiken in onderzoeken naar PCOS.[12]

Voor vrouwen met normogonadotrope anovulatie is op dit moment bekend dat letrozol de meest effectieve eerstelijnsbehandeling is in termen van levend geboren kinderen.[5, 7] De ESHRE-richtlijn voor PCOS geeft niet aan hoeveel cycli letrozol geadviseerd worden voordat wordt overgeschakeld op een andere ovulatie-inductie behandeling. Op dit moment is er geen onderzoek gedaan naar het gebruik van letrozol gedurende 12 maanden, zoals wel onderzocht is in de M-ovin-studie voor CC.[13] We raden aan om een vergelijkbaar onderzoeksprotocol te gebruiken als de M-ovin-studie, om voortgezette ovulatie-inductie met letrozol te bestuderen versus over te schakelen op gonadotrofinen bij vrouwen die na zes cycli letrozol niet zwanger zijn geworden. Eventueel zelfs met een derde arm bij overschakeling naar zes cycli CC.

Metformine is een insuline sensibiliserend middel dat de gluconeogenese en lipogenese vermindert en de perifere glucoseopname verhoogt en daardoor de insulinegevoeligheid verhoogt.[14] De IPD-analyse levert voorlopig bewijs dat er een rol kan zijn voor het beoordelen van insulineresistentie bij PCOS en anovulatie. Door toevoeging van metformine kan de insulineresistentie bij vrouwen met hogere nuchtere insulinespiegels verbeteren en daardoor het zwangerschapspercentage verbeteren. De internationale evidence-based richtlijn beveelt momenteel geen klinische meting van insulineresistentie aan vanwege het gebrek aan nauwkeurigheid.[13] Daarnaast is SHBG voorgesteld als een maatstaf voor insulineresistentie,[15] maar in onze IPD-meta-analyse vond geen verbetering van behandeling door een SHBG-interactie. Ons werk ondersteunt de noodzaak om insulineresistentie te beoordelen in toekomstige vruchtbaarheidsstudies.

IPD-meta-analyses zijn nuttig voor het opzetten, uitvoeren, analyseren en interpreteren van onderzoeken.[16] Gezien de consistente behandelvoordelen van letrozol bij verschillende vruchtbaarheidsresultaten, moeten toekomstige onderzoeken naar nieuwe

interventies voor PCOS letrozol als referentiearm kiezen. Nieuwe onderzoeken worden aangemoedigd om markers voor behandelingsselectie op te nemen in hun ontwerp om uiteindelijk zorg op maat te kunnen geven.[17] Ook het effect van onder andere leeftijd, BMI en andere biomarkers, moet in toekomstige onderzoeken worden bevestigd. Meer specifiek zouden biomarkers voor hyperandrogenemie en zoals hierboven vermeld insulineresistentie kunnen worden toegepast in onderzoeken die metformine evalueren. De kern uitkomsten set voor onderzoeken naar PCOS omvat deze biomarkers en moet worden gebruikt om ervoor te zorgen dat resultaten consistent worden gerapporteerd en verzameld in toekomstige onderzoeken naar PCOS en anovulatie. Zo ook dat verspilling van onderzoeksmiddelen wordt tegengegaan.[12]

Steeds vaker stemmen onderzoekers ermee in IPD van RCT's te delen zodat databases met studies voor secundaire analyse gecreëerd kunnen worden. Naast het uitvoeren van IPD-meta-analyses en -meer in het algemeen- het beantwoorden van vragen die synthese van bewijsmateriaal vereisen, maakt het combineren van IPD mogelijk om tot harmonisatie van variabelen, standaardisatie van data-analyse, uniforme presentatie van resultaten over de bijdragende RCT's, het onderzoek van subgroepeffecten en het genereren van nieuwe hypothesen.[18, 19] We moedigen auteurs en instellingen aan om hun onderzoeksgegevens te delen.

We kunnen het ons niet veroorloven om onderzoeken met integriteitsproblemen in meta-analyses op te nemen, omdat beleidsmakers, artsen en patiënten die deze meta-analyses gebruiken, misleid kunnen worden en onherstelbare schade kan ontstaan. [20] Op dit moment wordt er weinig nagedacht over de reproduceerbaarheid en integriteit van onderzoeken bij het uitvoeren van meta-analyses met behulp van geaggregeerde data, en richtlijnen voor het omgaan met deze problemen ontbreken. De zwakke prestatie van de studies waarvan de data niet was gedeeld, daagt de legitimiteit van de traditionele benadering uit om alle relevante onderzoeken op te nemen in systematische reviews en meta-analyses. Hoewel deze benadering het risico op publicatiebias minimaliseert, kunnen methodologische en integriteitskwesties van geïncludeerde onderzoeken niet ongedaan gemaakt worden bij systematische reviews en meta-analyses. Daarom adviseren wij om meer IPD-meta-analyses uit te voeren en richtlijnen en protocollen primair te baseren op IPD-meta-analyses.

Hoewel het beoordelen van data beschikbaarheidsbias een belangrijk onderdeel is van IPD-meta-analyse, analyseerde slechts een minderheid de onderliggende redenen voor de verschillen tussen resultaten van studies met of zonder gedeelde gegevens.[19] De bevindingen in ons onderzoek benadrukken het belang van het beoordelen van de redenen van al dan niet het beschikbaar stellen van de onderzoeksgegevens voor IPD. Wij adviseren om, na het uitvoeren van een IPD-meta-analyse, de onderzoeken te beoordelen

en de gedeelde onderzoeken te vergelijken met de niet-gedeelde onderzoeken op kwaliteit, methodologische betrouwbaarheid en integriteit.

Grondig onderzoek naar vermoedens van wetenschappelijke fraude is momenteel een moeizaam, tijdrovend en arbeidsintensief proces. De wetenschappelijke gemeenschap moet betere gevalideerde detectietools ontwikkelen. Vervolgens kunnen deze tools worden geautomatiseerd voor routinematige beoordelingen en getest door de gemeenschap om de integriteit van onderzoek proactief te verdedigen vóór publicatie. Automatisering van "ready" methoden zou een breed gebruik bevorderen en validatie en testen stimuleren. We moedigen ook aan om nieuwe methoden vooraf te automatiseren om het validatie- en toepassingsproces te versnellen. Momenteel ontwikkelt en valideert onze groep een checklist om data-integriteitsproblemen in studies op te sporen.

Het belangrijkste hiaat in het onderzoek is dat we moeten weten welke minimale reeks tests nodig is om de detectie van integriteitsproblemen te optimaliseren; dit omvat de noodzaak van validatie van beschikbare methoden en het bepalen van hun diagnostische capaciteit.

REFERENTIE LIJST

1. World Health Organization (WHO). International Classification of Diseases, 11th Revision (ICD-11) Geneva: WHO 2018.
2. NICE. Fertility problems: assessment and treatment. London: National Institute for Health and Care Excellence (NICE); 2017 Sep. PMID: 32134604.
3. National Collaborating Centre for Ws, Children's H. National Institute for Health and Clinical Excellence: Guidance. Fertility: Assessment and Treatment for People with Fertility Problems. London: Royal College of Obstetricians & Gynaecologists National Collaborating Centre for Women's and Children's Health.; 2013.
4. Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril* 2016;106:6–15.
5. Wang R, Kim BV, van Wely M, et al. Treatment strategies for women with WHO group II anovulation: systematic review and network meta-analysis. *BMJ* 2017; 356: j138.
6. Brown J, Farquhar C. Clomiphene and other antioestrogens for ovulation induction in polycystic ovarian syndrome. *Cochrane Database Syst Rev* 2016; 12: CD002249.
7. Franik S, Eltrop SM, Kremer JA, Kiesel L, Farquhar C. Aromatase inhibitors (letrozole) for subfertile women with polycystic ovary syndrome. The Cochrane database of systematic reviews. 2018;5:Cd010287.
8. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
9. Fanelli D. How many scientists fabricate and falsify research? A sys- tematic review and meta-analysis of survey data. PLoS One 2009;4 e5738.
10. Carlisle JB. False individual patient data and zombie randomised controlled trials submitted to Anaesthesia. Anaesthesia. 2021 Apr;76(4):472-479. doi: 10.1111/anae.15263. Epub 2020 Oct 11. PMID: 33040331.
11. Bordewijk EM, Wang R, van Wely M, et al. Data integrity of 10 other randomized controlled trials of an author with a retracted paper. Fertility and Sterility Dialog; 2020 June. https://www.fertstertdialog.com/posts/data-integrity-of-10-oth- er-randomized-controlled-trials-of-an-author-with-a-retracted-paper.
12. Al Wattar BH, Teede H, Garad R, Franks S, Balen A, Bhide P, Piltonen T, Romualdi D, Laven J, Thondan M, Bueno-Cavanillas A, Moss N, Andrews C, Hawkes R, Mol BW, Khan KS, Thangaratinam S. Harmonising research outcomes for polycystic ovary syndrome: an international multi-stakeholder core outcome set. Hum Reprod. 2020 Feb 29;35(2):404-412. doi: 10.1093/humrep/dez272. PMID: 32020203.
13. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Human reproduction (Oxford, England). 2018;33(9):1602-18.
14. Naderpoor N, Shorakae S, de Courten B, Misso ML, Moran LJ, Teede HJ. Metformin and lifestyle modification in polycystic ovary syndrome: systematic review and meta-analysis.

Hum Reprod Update 2015;21:560–574.

15. Cassar S, Misso ML, Hopkins WG, Shaw CS, Teede HJ, Stepto NK. Insulin resistance in polycystic ovary syndrome: a systematic review and meta-analysis of euglycaemic-hyperinsulinaemic clamp studies. *Hum Reprod* 2016;31:2619–2631.
16. Tierney JF, Pignon JP, Gueffyier F, Clarke M, Askie L, Vale CL, Burdett S, Cochrane IPDM-aMG. How individual participant data meta analyses have influenced trial design, conduct, and analysis. *J Clin Epidemiol* 2015;68:1325–1335.
17. Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. *Ann Intern Med* 2011;154:253–259.
18. Mbuagbaw L, Foster G, Cheng J, Thabane L. Challenges to complete and useful data sharing. Trials 2017;18:71.
19. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selec- tion bias, and unavailable data in meta-analyses using individual par- ticipant data: a database survey. BMJ 2012;344:d7762.
20. Li W, van Wely M, Gurrin L, Mol BW. Integrity of randomized controlled trials: challenges and solutions. Fertil Steril 2020;113: 1113–1119.

Appendices

List of co-authors and affiliations

Author contributions

PhD portfolio

List of publications

Dankwoord

About the author

LIST OF CO-AUTHORS AND AFFILIATIONS

S.A. Amer	Department of Obstetrics and Gynaecology, University of Nottingham, Royal Derby Hospital, Derby, UK.
N. Bayram	Department of Obstetrics and Gynaecology, Zaans Medical Center, Zaandam, Netherlands
S. Bhattacharya	Aberdeen Centre for Women's Health Research, University of Aberdeen, Aberdeen, UK.
D.E.S. Boks	Department of Obstetrics and Gynaecology, Spaarne Gasthuis, Hoofddorp, Netherlands
P. Bourdrez	Department of Obstetrics and Gynaecology, VieCuri Medical Center, Venlo, Netherlands
F.J.M. Broekmans	Department of Reproductive Medicine and Gynaecology, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands
J.P. de Bruin	Department of Obstetrics and Gynecology, Jeroen Bosch Hospital, 's Hertogenbosch, Netherlands
B.J. Cohlen	Isala Fertility Center, Isala Hospital, Zwolle, Netherlands
R. van Eekelen	Amsterdam UMC, University of Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
M.F. Costello	School of Women's and Children's Health, The University of New South Wales, Sydney, Australia.
A.Falbo	Department of Obstetrics and Gynecology, Azienda USL-IRCCS di Reggio Emilia, Italy.
G. Federica	Department of Surgery Obstetrics and Gynecology, University of Catania, Catania, Italy.
K. Fleischer	Department of Obstetrics and Gynaecology, Radboud University, Nijmegen, Netherlands
J. Gao	Department of Obstetrics and Gynecology, First Affiliated Hospital, Heilongjiang University of Chinese Medicine, Harbin, China.
M.H. Gerards	Department of Obstetrics and Gynaecology, Diakonessenhuis, Utrecht, Netherlands
J. Gianotten	Department of Obstetrics and Gynaecology, Spaarne gasthuis, Haarlem, Netherlands
M. Goddijn	Amsterdam UMC, University of Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
R.J.T. van Golde	Department of Obstetrics and Gynecology, Maastricht University Medical Center, Maastricht, Netherlands
L.C. Gurrin	Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Australia.

D.J. Hendriks	Department of Obstetrics and Gynaecology, Amphia Ziekenhuis Breda, Breda, Netherlands
R. Homburg	Homerton Fertility Centre, Homerton University Hospital, London, UK.
A. Hoek	Center for Reproductive Medicine, University Groningen, University medical Center Groningen, Groningen
P.G.A. Hompes	Amsterdam UMC, Vrije Universiteit Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
M.H.A. van Hooff	Department of Obstetrics and Gynaecology, Franciscus Gasthuis, Rotterdam, Netherlands
W. Huang	Department of Obstetrics and Gynecology, West China Second University Hospital of Sichuan University, Chengdu, China.
T.I. Jannink	Amsterdam UMC, University of Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
C.A.H. Janssen	Department of Obstetrics and Gynaecology, Groene Hart Hospital, Gouda, Netherlands
N.P. Johnson	Robinson Research Institute and Adelaide Medical School, University of Adelaide, North Adelaide, SA, Australia. Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand.
E.M. Kaaijk	Department of Obstetrics and Gynaecology, Onze Lieve Vrouwe Hospital, Amsterdam (East), Netherlands
S. Kar	Department of Obstetrics and Gynaecology, Kar Clinic and Hospital, Bhubaneswar, India.
N.F. Klijn	Center for Reproductive Medicine, Leiden University, Leiden, Netherlands
C.A. Koks	Department of Obstetrics and Gynaecology, Máxima Medical Center, Veldhoven, Netherlands
T.E. König	Department of Obstetrics and Gynecology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
J. Kwee	Department of Obstetrics and Gynaecology, Onze Lieve Vrouwe Hospital, Amsterdam (West), Netherlands
T.A.M. van de Laar-van Asseldonk	Department of Obstetrics and Gynaecology, Elkerliek Hospital, Helmond, Netherlands
C.B. Lambalk	Amsterdam UMC, Vrije Universiteit Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
A.F. Lambeek	Department of Obstetrics and Gynaecology, IJsselland hospital, Capelle a/d IJssel, Netherlands

J.S.E. Laven	Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynaecology, Erasmus Medical Center, Rotterdam, Netherlands
R.S. Legro	Department of Obstetrics and Gynecology, Pennsylvania State University, Hershey, PA, USA.
J. Lord	Department of Obstetrics and Gynaecology, Royal Cornwall Hospital, University of Exeter Medical School, Truro, UK.
W. Li	Department of Obstetrics and Gynaecology, Monash University, Clayton, VIC 3168, Australia.
B.W.J. Mol	Department of Obstetrics and Gynaecology, Monash University, Clayton, Victoria, Australia
E. Moll	Department of Obstetrics and Gynaecology, Onze Lieve Vrouwe Gasthuis (OLVG) West, Amsterdam, the Netherlands.
L. Morin-Papunen	Department of Obstetrics and Gynecology, Medical Research Center, PEDEGO Research Unit, University of Oulu and Oulu University Hospital, Oulu, Finland.
M.J. Nahuis	Department of Obstetrics and Gynaecology, Northwest Hospital Group, Alkmaar, Netherlands
A.W. Nap	Department of Obstetrics and Gynaecology, Rijnstate, Arnhem, Netherlands
H. Nazik	Department of Obstetrics and Gynaecology, Adana City Training and Research Hospital, Adana, Turkey.
R.J. Norman	Robinson Research Institute and Adelaide Medical School, University of Adelaide, North Adelaide, SA, Australia. Fertility SA, Adelaide, SA, Australia.
G.J.E. Oosterhuis	Department of Obstetrics and Gynaecology, St Antonius Hospital, Utrecht, Netherlands
Ü. Özmen	Department of Obstetrics and Gynecology, School of Medicine, Zonguldak Bulent Ecevit University, Kozlu, Zonguldak, Turkey.
S. Palomba	Department of Obstetrics and Gynecology, Grande Ospedale Metropolitano of Reggio Calabria, Reggio Calabria, Italy.
D.A.M. Perquin	Department of Obstetrics and Gynaecology, Medical Center Leeuwarden, Leeuwarden, Netherlands
I.A.J. van Rooij	Department of Obstetrics and Gynaecology, ElisabethTweeSteden Hospital, Tilburg, Netherlands
J.M. van Rijn – van Weert	Department of Obstetrics and Gynaecology, Northwest Hospital Group, Alkmaar, Netherlands
H.G.M. Rijnsaardt-Lukassen	Department of Obstetrics and Gynaecology, Albert Schweitzer Hospital, Dordrecht, Netherlands

Y. Sahin	Department of Obstetrics and Gynecology, Faculty of Medicine, Erciyes University, Kayseri, Turkey.
M. Showell	Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand.
J.M.J. Smeenk	Department of Obstetrics and Gynecology, ElisabethTweeSteden Hospital, Tilburg, Netherlands
A.V. Sluijmer	Department of Obstetrics and Gynaecology, Wilhelmina Hospital, Assen, Netherlands
H. Teede	Monash Centre for Health Research and Implementation, Monash University, Clayton, Australia.
G.A. van Unnik	Department of Obstetrics and Gynaecology, Alrijne Hospital, Leiden / Leiderdorp, Netherlands
F. van der Veen	Amsterdam UMC, University of Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
W. Vegetti	Infertility Unit, Fondazione I.R.C.C.S. Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.
J.H.A. Vollebergh	Department of Obstetrics and Gynaecology, Bernhoven Hospital, Uden, Netherlands
T. de Vries	Amsterdam UMC, University of Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
F.P.J.M. Vrouenraets	Department of Obstetrics and Gynaecology, Zuyderland Medical Center, Heerlen, Netherlands
R. Wang	Department of Obstetrics and Gynaecology, Monash University, Clayton, VIC 3168, Australia.
N. Weiss	Amsterdam UMC, University of Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands Centre for Reproductive Medicine Amsterdam UMC, VU University, De Boelelaan 1117, Amsterdam, Netherlands
M. van Wely	Amsterdam UMC, University of Amsterdam, Center for Reproductive Medicine, Amsterdam, Netherlands
C.D. Williams	Reproductive Medicine and Surgery Center of Virginia, Charlottesville, VA, USA.
X. Wu	Department of Obstetrics and Gynecology, First Affiliated Hospital, Heilongjiang University of Chinese Medicine, Harbin, China.
H. Zhang	Department of Biostatistics, Yale University School of Public Health, New Haven, CO, USA.

AUTHOR CONTRIBUTIONS

Chapter 2

MJN, JEO, PGH, FvdV, BWJM, and MvW designed the trial. NSW and MJN were the trial coordinators. NSW and MvW did the statistical analyses. NSW was in charge of drafting the manuscript. PGH, FvdV, BWJM, and MvW participated in the analysis, manuscript drafting, and supervision of the work. All authors acquired the data from the participating centres, provided critical discussion, and contributed in the preparation of the manuscript.

Chapter 3

N.W. and E.B. were in charge of collecting the data. N.W. took the lead in writing the article. E.B. and M.v.W. performed the analyses. B.M., F.v.V., P.H., M.G., C.B. and M.v.W. helped with interpreting the outcomes of the data and reviewed the article. All authors read, edited and approved the final article.

Chapter 4

P.G.H., F.v.d.V., B.W.J.M. and M.v.W. designed the trial. N.S.W. and M.J.N. were the trials coordinators. E.M.B. and N.S.W. were in charge of collecting the data. E.M.B. and M.v.W. performed the analyses. E.M.B. wrote the manuscript. J.K., A.F.L., G.A.v.U., F.P.J.V., B.J.C. and T.A.M.v.d.L.v.A. recruited and counselled participants of this study as local investigators. N.S.W., C.B.L., M.G., P.G.H., F.v.d.V., B.W.J.M. and M.v.W. helped with interpreting the outcomes of the data and reviewed the manuscript. All authors read, edited and approved the final manuscript.

Chapter 5

TJ and EB took the lead in writing the manuscript. TJ, EB and MvW performed the analyses. TJ and EB were in charge of collecting the data. NW, TV, AH, MG, BWM, and MvW helped with interpreting the outcomes of the data and reviewed the manuscript. All authors read, edited and approved the final manuscript.

Chapter 6

R.W., R.S.L., S.B., R.J.N., M.v.W. and B.W.M. conceptualised and designed the study. R.W., W.L., E.M.B., R.J.N., M.v.W. and B.W.M. collected the data. R.S.L., H.Z., X.W., J.G., L.M.P., R.H., T.E.K., E.M., S.K., W.H., N.P.J., S.A.A., W.V., S.P., A.F., U.O., H.N., C.D.W., G.F., J.L. and Y.S. provided and interpreted data from the included trials. R.W., W.L., E.M.B., M.v.W. and B.W.M. cleaned and analysed the data. R.W. drafted the first manuscript. All authors interpreted the pooled data, critically revised the manuscript for important intellectual content and approved the final version.

Chapter 7

W.L. designed the study, extracted and analysed data, and critically revised the manuscript; E.M.B. assisted in the study design, checked data, extracted and interpreted data, and drafted the manuscript; R.W. designed the study, extracted and interpreted data, and critically revised the manuscript; M.F.C., R.J.N., H.T., L.C.G. and M.v.W. oversaw data analysis and interpretation, and critically revised the manuscript; B.W.M. designed the study, oversaw data interpretation and critically revised the manuscript. All authors approved the version to be published.

Chapter 8

EB designed the study, managed the literature search, extracted data, and drafted the manuscript; MvW designed the study, managed the literature search, checked data, and critically revised the manuscript; WL designed the study, helped to interpret the statistical methods and critically revised the manuscript; RW and RvE helped to interpret the statistical methods and critically revised the manuscript; MS performed the literature search and critically revised the manuscript; BwM designed the study, and critically revised the manuscript.

PHD PORTFOLIO

Name PhD student: Esmée M Bordewijk
PhD period: December 2016 – November 2022
Names of PhD supervisor(s) & co-supervisor(s): Prof. dr. M. Goddijn and dr. M. van Wely

1. PhD training

	Year	ECTS
General courses		
- Good Clinical Practice light cursus, GCP central	2019	0.5
- Practical Biostatistics, AMC Graduate School for Medical Sciences	2020	1.5
- Project management, AMC Graduate School for Medical Sciences	2020	1.5
Seminars, workshops and master classes		
- Pre-congres course ESHRE: The PCO syndrome: from diagnosis to health risk management, Barcelona	2018	0.5
Presentations		
- Oral: Cost-effectiveness of gonadotrophins compared to clomiphene citrate with or without IUI in anovulatory women who had not conceived after 6 cycles with clomiphene citrate. ASRM scientific congress & expo, San Antonio, America	2017	0.5
- Oral: Clomiphene citrate or gonadotrophins in women with WHO type II anovulation and CC failure; a role for EMT? ESHRE annual meeting, Barcelona, Spain	2018	0.5
- Poster: Laparoscopic ovarian drilling for ovulation induction in women with anovulatory polycystic ovary syndrome - a cochrane review ESHRE annual meeting, virtual meeting	2020	0.2
- Poster: Long-term outcomes of using gonadotrophins vs clomiphene citrate with or without intrauterine insemination in women with WHO type II anovulation and clomiphene failure: follow-up study of the M-ovin trial. ESHRE annual meeting, Milan, Italy	2022	0.2
(Inter)national conferences		
- ASRM scientific congress & expo, San Antonio, America	2017	1.0
- ESHRE annual meeting, Barcelona, Spain	2018	1.0
Other		
- Clinical Department of Obstetrics and Gynaecology Researchproject investigating quality, methodological soundness and integrity of RCTs in Women's Health. University of Monash in Melbourne, Australia Supervisor: Prof. Dr. Ben Mol	Jan. 2020 – Apr. 2020	3

2. Teaching

	Year	ECTS
Lecturing		
- Lecture complications postpartum to HBO- and MBO-nurses	2019	1.0
Supervising		
- Scientific intership of T.I. Jannink, medical student	2020	3
- Bachelorthesis of I. de Koning, medical student	2020	1
- Scientific intership of J. Aalberts, medical student	2021	3
Other		
- Trained and supervised six Australian medical students how to notice data integrity issues in scientific papers	2020	3

3. Parameters of Esteem	
	Year
Travel grants	
- Amsterdam Universiteitsfonds	2018
- Van Walree beurs	2018
- Jo Kolk Studiefonds	2020
- Amsterdam Universiteitsfonds	2020

4. Publications	
Peer reviewed	**Year**
1. **Bordewijk EM**, Nahuis M, Costello MF, Van der Veen F, Tso LO, Mol BW, van Wely M. Metformin during ovulation induction with gonadotrophins followed by timed intercourse or intrauterine insemination for subfertility associated with polycystic ovary syndrome. Cochrane Database Syst Rev. 2017 Jan 24;1(1):CD009090.	2017
2. Sandberg EM, **Bordewijk EM**, Klemann D, Driessen SRC, Twijnstra ARH, Jansen FW. Medical malpractice claims in laparoscopic gynecologic surgery: a Dutch overview of 20 years. Surg Endosc. 2017 Dec;31(12):5418-5426.	2017
3. *Weiss NS, Nahuis MJ, **Bordewijk E,** Oosterhuis JE, Smeenk JM, Hoek A, Broekmans FJ, Fleischer K, de Bruin JP, Kaaijk EM, Laven JS, Hendriks DJ, Gerards MH, van Rooij IA, Bourdrez P, Gianotten J, Koks C, Lambalk CB, Hompes PG, van der Veen F, Mol BWJ, van Wely M. Gonadotrophins versus clomifene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomifene failure (M-OVIN): a randomised, two-by-two factorial trial. Lancet. 2018 Feb 24;391(10122):758-765.	2018
4. ***Bordewijk EM**, Weiss NS, Nahuis MJ, Bayram N, van Hooff MHA, Boks DES, Perquin DAM, Janssen CAH, van Golde RJT, Lambalk CB, Goddijn M, Hompes PG, van der Veen F, Mol BWJ, van Wely M; M-ovin study group. Gonadotrophins versus clomiphene citrate with or without IUI in women with normogonadotropic anovulation and clomiphene failure: a cost-effectiveness analysis. Hum Reprod. 2019 Feb 1;34(2):276-284.	2019
5. **Bordewijk EM**, Mol F, van der Veen F, Van Wely M. Required amount of rFSH, HP-hMG and HP-FSH to reach a live birth: a systematic review and meta-analysis. Hum Reprod Open. 2019 Jun 1;2019(3):hoz008.	2019
6. *Wang R, Li W, **Bordewijk EM,** Legro RS, Zhang H, Wu X, Gao J, Morin-Papunen L, Homburg R, König TE, Moll E, Kar S, Huang W, Johnson NP, Amer SA, Vegetti W, Palomba S, Falbo A, Özmen Ü, Nazik H, Williams CD, Federica G, Lord J, Sahin Y, Bhattacharya S, Norman RJ, van Wely M, Mol BW; Reproductive Medicine Network+; International Ovulation Induction IPDMA Collaboration. First-line ovulation induction for polycystic ovary syndrome: an individual participant data meta-analysis. Hum Reprod Update. 2019 Nov 5;25(6):717-732.	2019
7. **Bordewijk EM,** Ng KYB, Rakic L, Mol BWJ, Brown J, Crawford TJ, van Wely M. Laparoscopic ovarian drilling for ovulation induction in women with anovulatory polycystic ovary syndrome. Cochrane Database Syst Rev. 2020 Feb 11;2(2):CD001122.	2020

8.	**Bordewijk EM**, Wang R, Askie LM, Gurrin LC, Thornton JG, van Wely M, Li W, Mol BW. Data integrity of 35 randomised controlled trials in women' health. Eur J Obstet Gynecol Reprod Biol. 2020 Jun;249:72-83.	2020
9.	**Bordewijk EM**, Wang R, van Wely M, et al. Data integrity of 10 other randomized controlled trials of an author with a retracted paper. Fertility and Sterility Dialog; 2020 June. https://www.fertstertdialog.com/posts/data-integrity-of-10-oth-er-randomized-controlled-trials-of-an-author-with-a-retracted-paper.	2020
10.	***Bordewijk EM**, Weiss NS, Nahuis MJ, Kwee J, Lambeek AF, van Unnik GA, Vrouenraets FPJ, Cohlen BJ, van de Laar-van Asseldonk TAM, Lambalk CB, Goddijn M, Hompes PG, van der Veen F, Mol BWJ, van Wely M; M-ovin study group. Gonadotrophins or clomiphene citrate in women with normogonadotropic anovulation and CC failure: does the endometrium matter? Hum Reprod. 2020 Jun 1;35(6):1319-1324.	2020
11.	***Bordewijk EM**, Wang R, van Wely M, Costello MF, Norman RJ, Teede H, Gurrin LC, Mol BW, Li W. To share or not to share data: how valid are trials evaluating first-line ovulation induction for polycystic ovary syndrome? Hum Reprod Update. 2020 Nov 1;26(6):929-941.	2020
12.	***Bordewijk EM,** Li W, van Eekelen R, Wang R, Showell M, Mol BW, van Wely M. Methods to assess research misconduct in health-related research: A scoping review. J Clin Epidemiol. 2021 May 24;136:189-202.	2021
13.	**Bordewijk EM**, Li W, Gurrin LC, Thornton JG, van Wely M, Mol BW. An investigation of seven other publications by the first author of a retracted paper due to doubts about data integrity. Eur J Obstet Gynecol Reprod Biol. 2021 Jun;261:236-241.	2021
14.	Li W, **Bordewijk EM**, Mol BW. Assessing Research Misconduct in Randomized Controlled Trials. Obstet Gynecol. 2021 Sep 1;138(3):338-347.	2021
*Presented in this thesis		
Other		
-	***Bordewijk EM,** Jannink TI, Weiss NS. Long-term outcomes of using gonadotrophins vs clomiphene citrate with or without intrauterine insemination in women with normogonadotropic anovulation and clomiphene failure: follow-up study of a factorial randomised clinical trial. Accepted for publication in Human Reproduction Oct 2022.	2022
-	Mol BW, Lai S, Rahim A, **Bordewijk EM**, Wang R, van Eekelen R, Gurrin LC, Thornton JG, van Wely M, Li W. A checklist to assess Trustworthiness in RAndomised Controlled Trials (TRACT checklist). Submitted to J Clin Epidemiol. Sept 2022.	

DANKWOORD

Dr van Wely, beste Madelon, in 2015 zat ik in jaar 3 van de geneeskunde opleiding en ik moest een project vinden voor mijn bachelorthesis. Ik vond jouw emailadres via de website van het AMC en vroeg of je een onderwerp voor mij had. Dat werd uiteindelijk een heel PhD traject. Ik ben zo blij dat je mij toen die kans hebt gegeven. Bedankt voor de ontelbare meetings gedurende de afgelopen 7 jaar, zowel in het AMC, online, bij jou thuis en op congressen. Bedankt voor al je hulp bij het schrijven van de artikelen. Jij was altijd de eerste die mijn artikelen nakeek en goede feedback gaf. Tevens ook veel dank dat ik studenten via jou mocht begeleiden, dat je mij van alles over statistiek hebt geleerd, dat ik mijn presentaties bij jou kon oefenen en dat je mijn financiering voor een fulltime jaar onderzoek hebt geregeld. Samen werken we super samen en hoop dit te kunnen blijven doen.

Professor Goddijn, beste Mariëtte, ik heb het als heel fijn ervaren hoe je mij mijn gang liet gaan rondom het bedenken en vormgeven van mijn proefschrift. Je gaf me altijd een goed gevoel over mijn artikelen, andere proefschrift onderdelen en gaf daarbij duidelijke en rustige instructies over hoe het misschien net iets beter kon. De afgelopen maanden heb je mij ook ontzettend geholpen met alles regelen rondom het daadwerkelijk promoveren. De samenvatting, discussie, contact met promotiezaken, proefpromotie, contact met de promotiecommissie en noem maar op. Heel hartelijk dank voor al je inzet en onze hele fijne samenwerking.

Geachte leden van de leescommissie; prof dr van der Post, prof dr Painter, dr Mochtar, prof dr Dijkgraaff, prof dr Bouter, prof dr Hoek. Hartelijk dank voor jullie beoordeling van mijn proefschrift en de tijd die jullie op 25 november voor mij nemen. Ik kijk er naar uit om met u van gedachten te wisselen gedurende mijn verdediging.

Professor Mol, beste Ben, nooit gedacht dat toen ik aan je vroeg of ik een tijdje naar Australië kon komen, ik mij zou gaan focussen op data integriteit problemen. En wat voor ontdekkingen hebben we gedaan! Ik heb heel veel respect voor je dat je je hier zo enorm voor inzet. Je hebt mij geleerd kritisch te kijken naar onderzoek en de moed gegeven om hierover te publiceren. Bedankt voor je discussies, de ruimte die je mij gaf om de onderzoeken vorm te geven en de kans om nog langer fulltime onderzoek te doen. Ook bedankt voor je vele voetbalquotes, het advies dat ik absoluut geen anesthesioloog moest gaan worden, en dat ik er geen moment over moest twijfelen om heen en weer te vliegen voor de begrafenis van mijn oma. Ik sluit zeker niet uit dat ik nog eens langskom in Australië, maar ook online zullen wij vast blijven samenwerken.

Dear Wentao and Rui, thank you for our great corporation before and in Melbourne, I really enjoyed working with both of you. The last three studies would not have been in this thesis if it weren't for you. Both of you learned me a lot about statistics, data analysis, reviews, quality of studies and writing. I hope we will do more projects together in the future, so I can learn much more. Also, thanks for introducing the great Chinese 'hot pot' to me which we should definitely introduce to the Dutch cuisine.

Professor van der Veen, beste Fulco, jouw eerste tip aan mij was om het boekje van Strunk and White te kopen en dat is voor mij nog steeds een leidraad bij Engels schrijven. Onder andere taaltechnisch heb je mijn stukken echt een niveau hoger getild. In mijn manuscripten stonden soms in Caps Lock zinnen met uitroeptekens, "dit moet echt anders." Als we dan samen een meeting hadden, kwamen we er altijd uit en stond elke zin net wat beter. Ook bedankt voor de 5 gouden regels van de discussie die ik nog altijd gebruik. Ook gebruik ik nu altijd het lettertype Arial en hou ik mijn PowerPoint zo leeg mogelijk.

Beste Nienke Weiss, in 2015 zocht je een student om de missing data van de M-ovin studie op te vragen bij de 48 ziekenhuizen die hieraan mee hadden gedaan. Had ik even geluk dat ik een tussenjaar had. Hierdoor mocht ik jouw 'opvolgster' worden en is dit proefschrift tot stand gekomen. Dankjewel voor de hele fijne en gezellige samenwerking en de enorme kans die je mij hebt gegeven.

Lieve Tamara, hoe leuk dat ik samen met jou, als oud-huisgenoot, ons laatste artikel kon schrijven. Dankzij jouw doorzettingsvermogen is het gelukt om zoveel vrouwen in de follow-up geïncludeerd te hebben gekregen. Ik had het echt niet zonder jou kunnen volbrengen. Op naar de volgende artikelen samen, want we beginnen pas net.

Beste Rik, dankjewel dat je met mij door de enorme berg statistische studies heen bent gegaan voor het scoping review. Uiteindelijk is het voor mij begrijpelijk geworden dankzij jouw hulp.

Dear Red Wing repropeeps, Myrthe, Miriam, Callista, Ivy, Ieva, Qijing, Joanna, Danal, Jitske, and Jan Willem, thank you for adopting me to the Red Wing during my parttime PhD. When I was working in the AMC I always felt super comfortable with you. Thanks for the coffee breaks and all the fun times.

Beste Monique en Femke, wat was ik blij dat ik samen met jullie naar ASRM kon in San Antonio. Ik was natuurlijk mega zenuwachtig voor mijn eerste presentatie op een groot congres, maar door jullie kwam ik daar makkelijk doorheen. Het voelde echt als twee suikertantes die mij begeleidden (inclusief cocktails voeren en me aan de man helpen).

Bedankt Femke, voor het regelen van het telefoonnummer van die jongen in het vliegtuig. Na één date was al duidelijk dat het hem niet werd, maar we hadden wel een mooi verhaal.

Beste Tessa de Vries, heel erg bedankt voor al je hulp bij de follow up van de M-ovin studie. Ook alvast hartelijk dank voor de hulp die je al hebt gestoken en gaat steken in de ANDES studie. Heel fijn om je bij ons team te hebben.

Beste Hildegard, bedankt voor alles wat je voor mij hebt geregeld voor mijn komst naar Melbourne en de gezelligheid op het kantoor. En de overheerlijke Australische koffietjes die we daar gedronken hebben.

Dear Karim and Mae, thank you for the great time we had in Melbourne. Ofcourse Karim for the great time at work, but most of all during the weekends. Renting the car and driving to Mornington Peninsula and the Great Ocean Road together was really one of the highlights of my time there. Happy to still follow your stories trough Instagram and see your beautiful daughter.

Beste Beatrix, Valesca, Barbara en Madeleine, lieve dames van het secretariaat van het CVV: bedankt voor al het plannen en regelen rondom mijn promotie.
Lieve Lidija, dank voor al je support tijdens onze master geneeskunde. We hebben het echt fantastisch gehad met als hoogtepunt Sri Lanka. Dankjewel dat je mijn paranimf bent voor mijn promotie en dit fantastisch regelt. Ik ben super trots op jou dat je helemaal je eigen pad hebt gevonden en je wordt sowieso de leukste psychiater die er is.

Lieve Roselie, als clubgenoot hebben we vooral heel veel gefeest. Allebei geneeskunde studeren maar elk op een andere universiteit; klagen over coschappen, deadlines en examens. Dankjewel dat je alles tot in de puntjes regelt voor mijn promotie als paranimf. Jij gaat een fantastische dermatoloog worden en het duurt niet lang voordat ook jij je prachtige proefschrift mag verdedigen. Ik ben heel trots op jou.

Beste Tatjana, wat heeft Diakonessenhuis een geluk met jou als opleider. Heel erg bedankt voor het leuke sollicitatiegesprek via zoom in coronatijd waar jullie mij aannamen als anios. Je liet altijd heel duidelijk aan mij merken dat je veel vertrouwen in mij had en dat zorgde er ook voor dat ik het super naar mijn zin had en heb. Echt fantastisch hoe je mij geholpen hebt vertrouwen te krijgen in de sollicitatie. En dat heeft zijn vruchten afgeworpen. Ik zat thuis heel zenuwachtig te zijn toen je belde: "Je bent aangenomen!" En je was net zo blij als ik. Echt super. Heel fijn dat je nu nog steeds mijn opleider bent.

Beste Floor, wat ik niet meer vergeet is dat je bij het sollicitatiegesprek zei: "Nou zullen we het maar gewoon gaan zeggen, want eigenlijk weten we al dat we Esmée gaan aannemen toch?" En ook het moment dat je het nieuws kwam brengen dat jullie mij wilden voordragen voor de opleiding van het Amsterdam UMC, echt super. Dankjewel voor al je vertrouwen in mij. Ook bedankt voor wat je me nu allemaal leert in het Diakonessenhuis, ik ben blij dat ik zeker nog 2 jaar van je mag blijven leren.

Beste Jan, heel veel dank voor je fijne gesprekken tijdens mijn anios periode en nu als mentor in mijn beginnende aios tijd. Als er iemand af en toe door mij heen kan prikken, dan ben jij het wel. Ik waardeer het ontzettend dat je zoveel tijd neemt om mij veel dingen te leren, een goed persoonlijk leerpad uit te stippelen en mij kritisch te laten nadenken. Daarnaast ook natuurlijk heel veel dank voor alle gezelligheid en hilarische verhalen.

Lieve Amber, dankjewel dat je mijn buddy was tijdens mijn anios periode en ik af en toe mijn onzekerheden bij je kon spuien. Je hebt me veel tips gegeven en je was iemand waarop ik helemaal kon terugvallen in mijn beginperiode als arts-assistent. Je gaat het zeker weten fantastisch doen in het Antonius.

Beste gynaecologen van het Diakonessenhuis, bedankt dat jullie mij geleerd hebben een echte dokter te worden en dat ik nog veel meer van jullie mag leren in de aankomende 2 jaar. Jullie zijn een topteam.

Lieve (oud) aiossen en anniossen van het Diak, wat een mega toptijd heb ik met jullie (gehad). Dank voor alle gezelligheid, leerzame momenten, skireis, feestjes en ook heel veel steun.

Lieve klinisch verloskundigen van het Diak, dankzij jullie heb ik geleerd hoe ik een bevalling moet begeleiden in een hele fijne leeromgeving. Bedankt voor alle gezelligheid op het werk en jullie support.

Lieve verpleegkudingen van het Diak, dank voor alle enorme steun die ik van jullie krijg. Zonder jullie zou het op het werk niet zo leuk en leerzaam zijn. Ik vind onze samenwerking op de kamer altijd heel fijn en ben blij dat ik nog een hele tijd met jullie mag werken.

Lieve Rovers, Charlotte, Roselie, Puck, Margreet, Malou, Julia, Jessie, Shinta en Annefloor, dank jullie wel voor alle steun en mega veel gezelligheid van de afgelopen bijna 10 jaar tijdens studie en daarna. Dat ik aan jullie mijn presentatie mocht geven voordat ik naar ESHRE ging, ook al hadden jullie eigenlijk soms geen idee waar het over ging.

Lieve Marthe, Tessa, Lieke, Jip, Tessa, Maike en Fien, dank voor alle gezelligheid en steun sinds de middelbare school. En vooral ook voor al jullie vragen over de gynaecologie (inknippen, scheuren, keizersnedes) en dan na 10 minuten aangeven dat jullie het niet meer willen horen.

Lieve Jan Willem, Jacqueline, Shelitha, Samira en Cameron, bedankt voor al het nakijken van mijn Engelse artikelen, samenvattingen en presentaties. Dat scheelt een heleboel spelfouten in dit proefschrift.

Lieve familie, dank voor jullie superfijne steun die ik al mijn hele leven van jullie krijg. Ik heb enorm veel geluk met jullie allemaal. Dankjewel ook voor de vele felicitaties als ik een artikel had gepubliceerd.

Lieve Yaniek, ik kan iedereen aanraden om Bumble te installeren, want op die manier vind je echt de leukste vriend. Zo fijn hoeveel je mij de afgelopen tijd hebt gesteund met van alles, zodat ik tijd had voor dit proefschrift en mijn opleiding. Geen beter gevoel dan van werk naar ons mooie huis te lopen en weten dat jij daar bent.

Lieve mama, dankjewel dat je mij hebt geleerd om doorzettingsvermogen te hebben. Dat dromen uitkomen als je ervoor gaat. En zelfs bij de grootste tegenslag die je kan hebben weer opstaat, doorgaat en blijft genieten. Als er iemand onuitputtelijk kan werken, met een goed humeur, dan ben jij dat wel. Je bent mijn beste voorbeeld.

Lieve Bas, jij bent echt de allerliefste en beste broer die ik maar kan bedenken. Je steunt me in alles en als ik vraag om me te helpen met wat dan ook, dan ben je er. Ik ben heel trots op je.

Lieve papa, dankjewel dat als je naar mij, Bas of mama keek altijd de trots van je afstraalde. En dat je ontzettend trots was dat ik dokter ging worden. Hierdoor weet ik dat als je hier nog was, je naar mij zou kijken als die hele lieve trotse vader.

ABOUT THE AUTHOR

Esmée Bordewijk was born on the 6th of November 1992 in Rotterdam, the Netherlands. She grew up with one brother (Bas) in Arnhem and lived with her father Guido Bordewijk and mother Alexandra Bordewijk-de Vos. After graduation from secondary school at Arentheem college in Arnhem in 2011, she studied Biomedical sciences for one year in Antwerpen, Belgium.

In 2012 Esmée started medical school at the University of Amsterdam. In the third year of medical school, she became involved in research at the Centre of reproductive medicine in the AMC. During her internships she continued research and presented two oral presentations on international congresses (ASRM 2017 and ESHRE 2018). In September 2019 she obtained her medical degree and continued her PhD program for one full-time year of research in which she visited the University of Monash in Melbourne Australia for 3 months at the clinical department of Obstetrics and Gynaecology. Here underlead of Prof. dr. Ben Mol she gained interested in data integrity issues.

In November 2020, she started working as a resident (not in training) at the Department of Obstetrics and Gynecology of the Diakonessenhuis in Utrecht, supervised by dr. T.E. Vogelvang and dr. F. Vernooij.

She was accepted to become a gyneacologist in training at the Amsterdam UMC in September 2021 and started her training at the Diakonessenhuis in Utrecht in June 2022.

Together with Yaniek Schuring she lives in Utrecht.