Human Reproduction, Vol.35, No.12, pp. 2735–2745, 2020
Advance Access Publication on November 30, 2020 doi:10.1093/humrep/deaa243

human reproduction

ORIGINAL ARTICLE *Infertility*

Standardizing definitions and reporting guidelines for the infertility core outcome set: an international consensus development study† ‡

J.M.N. Duffy[1,2,*], S. Bhattacharya[3], S. Bhattacharya[3], M. Bofill[4], B. Collura[5], C. Curtis[6,7], J.L.H. Evers[8], L.C. Giudice[9,10], R.G. Farquharson[11], S. Franik[12], M. Hickey[13], M.L. Hull[14], V. Jordan[4], Y. Khalaf[15], R.S. Legro[16], S. Lensen[13], D. Mavrelos[17], B.W. Mol[18], C. Niederberger[19], E.H.Y. Ng[20,21], L. Puscasiu[22], S. Repping[23,24], I. Sarris[1], M. Showell[25], A. Strandell[26], A. Vail[27], M. van Wely[23], M. Vercoe[25], N.L. Vuong[28], A.Y. Wang[29], R. Wang[18], J. Wilkinson[27], M.A. Youssef[30], C.M. Farquhar[4,25], and the Core Outcome Measure for Infertility Trials (COMMIT) initiative[§]

[1]King's Fertility, Fetal Medicine Research Institute, London, UK [2]Institute for Women's Health, University College London, London, UK [3]School of Medicine, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, UK [4]Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand [5]RESOLVE, The National Infertility Association, VA, USA [6]Fertility New Zealand, Auckland, New Zealand [7]School of Psychology, University of Waikato, Hamilton, New Zealand [8]Maastricht University Medical Centre, Maastricht, The Netherlands [9]Center for Research, Innovation and Training in Reproduction and Infertility, Center for Reproductive Sciences, University of California, San Francisco, CA, USA [10]International Federation of Fertility Societies, Philadelphia, PA, USA [11]Department of Obstetrics and Gynaecology, Liverpool Women's NHS Foundation Trust, Liverpool, UK [12]Department of Obstetrics and Gynaecology, Münster University Hospital, Münster, Germany [13]Department of Obstetrics and Gynaecology, University of Melbourne, Victoria, Australia [14]Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia [15]Department of Women and Children's Health, King's College London, Guy's Hospital, London [16]Department of Obstetrics and Gynaecology, Penn State College of Medicine, PA, USA [17]Reproductive Medicine Unit, University College Hospital, London, UK [18]Department of Obstetrics and Gynaecology, Monash University, Melbourne, Australia [19]Department of Urology, University of Illinois at Chicago College of Medicine, Chicago, IL, USA [20]Department of Obstetrics and Gynaecology, The University of Hong Kong, Hong Kong [21]Shenzhen Key Laboratory of Fertility Regulation, The University of Hong Kong-Shenzhen Hospital, China [22]Pharmacy, Sciences and Technology, University of Medicine, Targu Mures, Romania [23]Amsterdam University Medical Centers, Amsterdam, The Netherlands [24]National Health Care Institute, Diemen, The Netherlands [25]Cochrane Gynaecology and Fertility Group, University of Auckland, Auckland, New Zealand [26]Department of Obstetrics and Gynecology, Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden [27]Centre for Biostatistics, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK [28]Department of Obstetrics and Gynaecology, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam [29]Faculty of Health, University of Technology, Sydney, Broadway, Australia [30]Department of Obstetrics & Gynaecology, Faculty of Medicine, Cairo University, Cairo, Egypt

*Correspondence address. King's Fertility, Fetal Medicine Research Institute, 16-20 Windsor Walk, Denmark Hill, London SE5 8BB, UK. E-mail: james.duffy3@nhs.net

Submitted on May 13, 2020; resubmitted on July 8, 2020; editorial decision on July 22, 2020

STUDY QUESTION: Can consensus definitions for the core outcome set for infertility be identified in order to recommend a standardized approach to reporting?

SUMMARY ANSWER: Consensus definitions for individual core outcomes, contextual statements and a standardized reporting table have been developed.

WHAT IS KNOWN ALREADY: Different definitions exist for individual core outcomes for infertility. This variation increases the opportunities for researchers to engage with selective outcome reporting, which undermines secondary research and compromises clinical practice guideline development.

†This article has not been externally peer reviewed.

‡This article has been published simultaneously in *Fertility and Sterility*.

§Members of the COMMIT initiative are listed in the Appendix.

STUDY DESIGN, SIZE, DURATION: Potential definitions were identified by a systematic review of definition development initiatives and clinical practice guidelines and by reviewing Cochrane Gynaecology and Fertility Group guidelines. These definitions were discussed in a face-to-face consensus development meeting, which agreed consensus definitions. A standardized approach to reporting was also developed as part of the process.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Healthcare professionals, researchers and people with fertility problems were brought together in an open and transparent process using formal consensus development methods.

MAIN RESULTS AND THE ROLE OF CHANCE: Forty-four potential definitions were inventoried across four definition development initiatives, including the Harbin Consensus Conference Workshop Group and International Committee for Monitoring Assisted Reproductive Technologies, 12 clinical practice guidelines and Cochrane Gynaecology and Fertility Group guidelines. Twenty-seven participants, from 11 countries, contributed to the consensus development meeting. Consensus definitions were successfully developed for all core outcomes. Specific recommendations were made to improve reporting.

LIMITATIONS, REASONS FOR CAUTION: We used consensus development methods, which have inherent limitations. There was limited representation from low- and middle-income countries.

WIDER IMPLICATIONS OF THE FINDINGS: A minimum data set should assist researchers in populating protocols, case report forms and other data collection tools. The generic reporting table should provide clear guidance to researchers and improve the reporting of their results within journal publications and conference presentations. Research funding bodies, the Standard Protocol Items: Recommendations for Interventional Trials statement, and over 80 specialty journals have committed to implementing this core outcome set.

STUDY FUNDING/COMPETING INTEREST(S): This research was funded by the Catalyst Fund, Royal Society of New Zealand, Auckland Medical Research Fund and Maurice and Phyllis Paykel Trust. Siladitya Bhattacharya reports being the Editor-in-Chief of *Human Reproduction Open* and an editor of the Cochrane Gynaecology and Fertility Group. J.L.H.E. reports being the Editor Emeritus of *Human Reproduction*. R.S.L. reports consultancy fees from Abbvie, Bayer, Ferring, Fractyl, Insud Pharma and Kindex and research sponsorship from Guerbet and Hass Avocado Board. B.W.M. reports consultancy fees from Guerbet, iGenomix, Merck, Merck KGaA and ObsEva. C.N. reports being the Editor-in-Chief of *Fertility and Sterility* and Section Editor of the *Journal of Urology*, research sponsorship from Ferring, and a financial interest in NexHand. E.H.Y.N. reports research sponsorship from Merck. A.S. reports consultancy fees from Guerbet. J.W. reports being a statistical editor for the Cochrane Gynaecology and Fertility Group. A.V. reports that he is a Statistical Editor of the Cochrane Gynaecology & Fertility Review Group and of the journal *Reproduction*. His employing institution has received payment from Human Fertilisation and Embryology Authority for his advice on review of research evidence to inform their 'traffic light' system for infertility treatment 'add-ons'. N.L.V. reports consultancy and conference fees from Ferring, Merck and Merck Sharp and Dohme. The remaining authors declare no competing interests in relation to the work presented. All authors have completed the disclosure form.

TRIAL REGISTRATION NUMBER: Core Outcome Measures in Effectiveness Trials Initiative: 1023.

Key words: female infertility / infertility / male infertility / effectiveness / safety / outcomes

Introduction

Randomized controlled trials (RCTs) evaluating potential treatments for infertility have reported many different outcomes (Wilkinson *et al.*, 2019a). Such variation contributes to challenges in comparing, contrasting and combining individual trials, limiting the usefulness of research to inform clinical practice (Duffy *et al.*, 2019b). The development, dissemination and implementation of a minimum data set, known as a core outcome set, will help to standardize outcome selection, collection and reporting across future infertility research.

A core outcome set for infertility (Fig. 1) has been developed (Duffy *et al.*, 2020a). However, there are inconsistencies in how individual core outcomes are currently defined by fertility trials. For example, definitions of live birth include a viable fetus after 24 weeks of gestation, pregnancy continuation beyond 28 weeks of gestation and delivery of a living baby (Wilkinson *et al.*, 2016). Such variation makes it possible for researchers to selectively report their results based on statistical significance. For example, researchers can undertake multiple statistical analyses at different gestational thresholds for live birth and selectively report the most favorable result.

There are unique challenges when reporting the results of infertility research because of the multistage nature of the treatment, particularly in the context of IVF (Wilkinson *et al.*, 2016). Multiple clinical and procedural events can occur during treatment. These events can be reported in subgroups containing only those patients who reach a certain milestone, for example, oocyte retrieval, embryo transfer and

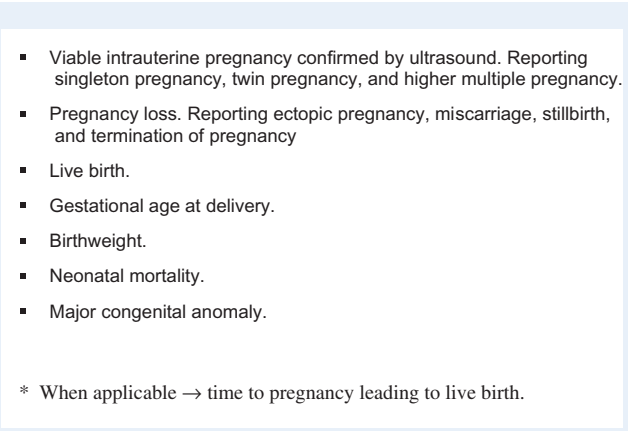

Figure 1. A core outcome set for future infertility research.

implantation. When reporting individual core outcomes, there could be many denominators available. This enables researchers to undertake multiple analyses using different denominators and selectively report results.

The variation in definitions and poor reporting practices makes comparing and combining individual RCTs challenging. When these practices are common, it is likely the benefits of fertility treatments are being overestimated and the harms of treatments are being underestimated (Duffy *et al.*, 2019b). This undermines secondary research, including individual patient data (IPD) meta-analysis and network meta-analysis, and compromises clinical practice guideline development. Standardizing definitions and improving reporting for individual core outcomes creates an opportunity to develop additional consistency in future infertility trials and ensure that secondary research can be undertaken prospectively, efficiently and harmoniously (Duffy *et al.*, 2017a).

No guidelines have established recommendations regarding the development of consensus definitions and reporting guidelines for individual core outcomes (Williamson *et al.*, 2017). Outside the context of core outcome set development, the Harbin Consensus Conference Workshop has developed a standardized definition for live birth (The Harbin Consensus Conference Workshop Group, 2014) and the International Committee for Monitoring Assisted Reproductive Technologies (ICMART) has standardized definitions related to infertility and ART (Zegers-Hochschild *et al.*, 2017).

Motivated by the desire to maximize the potential of infertility research to inform clinical practice, an international collaboration coordinated by the Cochrane Gynaecology and Fertility Group has brought healthcare professionals, researchers and people with infertility together to standardize definitions for the core outcome set for infertility.

Materials and methods

The study was prospectively registered with the Core Outcome Measures in Effectiveness Trials (COMET) initiative, registration number 1023. An international steering group, including healthcare professionals, researchers and people with fertility problems, was established to provide a perspective to inform key methodological decisions.

The important work of the Harbin Consensus Conference Working Group and ICMART is complementary to this study.

A protocol describing the study's consensus methods has been published (Duffy *et al.*, 2018). The protocol was developed with reference to the COMET initiative handbook (Williamson *et al.*, 2017). The protocol was also informed by a systematic review of registered, progressing and completed core outcome sets relevant to women's and newborn health (Duffy *et al.*, 2017b) and the experiences of steering group members involved in other core outcome set development studies (Hirsch *et al.*, 2016; Khalil *et al.*, 2017; Webbe *et al.*, 2017; Whitehouse *et al.*, 2017; Khalil *et al.*, 2019).

Potential definitions for individual core outcomes were extracted from definition development initiatives and national and international clinical practice guidelines and by reviewing the Cochrane Gynaecology and Fertility Group's standardized guidance for infertility reviews. A systematic review was undertaken searching the COMET initiative register to identify definition development initiatives relevant to infertility research, from inception to October 2018. Clinical practice guidelines relevant to infertility were identified by searching bibliographical databases, including Embase, MEDLINE and PubMed, from inception to October 2018. The Cochrane Gynaecology and Fertility Group provided access to their editorial policy, which describes their standardized approach to the selection of outcomes and definitions across Cochrane reviews evaluating potential fertility treatments. Using a pilot-tested and standardized data extraction form, definitions were extracted verbatim from all sources. An inventory was developed by organizing potential definitions within an organizational framework (Supplementary Fig. S1). Steering group members with expertise in statistics and research methodology prepared discussion points related to the analysis and reporting of the core outcome set. The inventory and discussion points were discussed during a face-to-face consensus development meeting held in Auckland, New Zealand.

The consensus development conference is a formal consensus development method developed by the US National Institutes of Health and has been used to reach consensus for definitions, clinical practice recommendations and professional competencies (Ferguson, 1996). The consensus method was developed to include aspects of judicial decision-making, scientific conferences and the town hall meeting. Participants hear evidence on which they will later deliberate and are able to ask questions as the evidence is presented. The chairperson is responsible for directing the discussion. The group discussion follows an informal format.

Healthcare professionals, researchers and people with fertility problems who had participated in the Delphi survey, which informed the development of the core outcome set for infertility, were invited to participate (Duffy *et al.*, 2020a). The study aimed to recruit between 10 and 15 participants, as this number has yielded sufficient results and assured validity in other studies (Murphy *et al.*, 1998).

Before starting the meeting, participants provided demographic details. The group discussion followed an informal format with the chairperson providing direction. Each core outcome was discussed in turn. Potential definitions were displayed within the definition hierarchy. Each participant was asked to contribute their opinions. Participants were encouraged to suggest other potential definitions or reformulate individual definitions to improve clarity or comprehension. Although the group was encouraged to reach consensus, members were able to express minority or alternative views when consensus could not be achieved. Participants were encouraged to agree contextual statements to highlight important methodological issues which would need to be considered when reporting individual core outcomes. Participants also developed consensus guidance regarding statistical analysis and a reporting table.

Results

Potential definitions were inventoried across four definition development initiatives, including the Brighton Collaboration (Chen *et al.*, 2016), Harbin Consensus Conference Workshop Group, ICMART and World Health Organization (WHO) (World Health Organization, 2018), 12 clinical practice guidelines (American Urological Association, 2010a,b,c; Jarvi *et al.*, 2010; Kroon *et al.*, 2011; Koch *et al.*, 2012; National Institute for Health and Care Excellence, 2013; Loh *et al.*, 2014; Carranza-Mamane *et al.*, 2015; Practice Committee of the

American Society for Reproductive Medicine; 2017a,b; Jungwirth et al., 2018), and the standardized methods advocated by the Cochrane Gynaecology and Fertility Group for the preparation of systematic reviews evaluating potential fertility treatments.

Forty-four potential definitions were discussed during the consensus development meeting. Twenty-seven participants, including 14 healthcare professionals, 7 researchers and 6 people with fertility problems, from 11 countries, participated in the consensus development meeting (Table I).

Live birth

When considering live birth, participants noted the Improving the reporting of clinical trials of infertility treatments (IMPRINT) statement recommended a gestational age threshold of 20 completed weeks. This statement was specifically developed to improve outcome reporting in infertility trials by modifying the Consolidated Standards of Reporting Trials (CONSORT) statement (Moher et al., 2010). Given this context, participants agreed the IMPRINT gestational threshold should be recommended to ensure consistency across comparable initiatives standardizing outcome reporting in RCT (Table II).

When considering the reporting of live birth, participants recommended twin and higher multiple births should be reported as a single live birth event (Table III). This will ensure treatments which increase twin and higher multiple births are not favored. The participants agreed that the summary effect size estimate and 95% CI should be calculated for live birth events only, and recommended the number of participants randomized as the most appropriate denominator. In addition to reporting live birth events, singleton, twin and higher multiple births should be reported narratively. When calculating the corresponding percentages for live birth events and singleton, twin and higher multiple births, the number of participants randomized is the recommended denominator.

Carefully selecting an appropriate denominator will avoid common issues associated with the analyses of data arising from infertility trials, particularly for studies related to ART. These issues are discussed in detail within the discussion.

Table I Participant characteristics.

	Participants n = 27
Stakeholder group, n	
Health professionals	14
Researchers	7
People with fertility problems	6
Gender, n	
Male	12
Female	15
Age (years), n	
Under 29	1
30–39	6
40–49	3
50–59	9
Over 60	5
Prefer not to say	3
Geographical location, n	
Africa	0
Asia	3
Australia and New Zealand	9
Europe	12
North America	3
South America	0

Viable intrauterine pregnancy confirmed by ultrasound

Participants agreed a consensus definition, which included visualization of a heartbeat. Participants discussed the reporting of twin and higher multiple pregnancies and recommended they should be reported as a single pregnancy event. The effect size estimate and 95% CI should be calculated for pregnancy events only. Participants concluded that it was also important for singleton, twin and higher multiple pregnancy to be routinely reported. When calculating the corresponding percentages for pregnancy events and singleton, twin and higher multiple pregnancies, the number of participants randomized is the denominator which should be used.

Participants discussed the importance of embedding RCT within routine clinical practice and were reluctant to insist upon mandatory urinary or serum beta-hCG testing or ultrasonographic examinations in addition to routine care. The variation in routine ultrasonographic examination between countries was discussed, for example, routine ultrasound scans are performed between 6 and 8 weeks in the USA, between 11 and 13 weeks in the UK and following 16 weeks in the Netherlands. Following the discussion, a contextual statement was recommended to ensure researchers consistently reported the gestation at which the ultrasonographic examination diagnosing viable intrauterine pregnancy was performed.

Pregnancy loss

Ectopic pregnancy

Following discussion, consensus was reached to adopt the ICMART definition of ectopic pregnancy.

Miscarriage

Participants discussed the WHO's definition for miscarriage and observed this definition was the most widely used within an international context. The definition includes a gestational age threshold of 20 completed weeks. They observed such a threshold would correlate well with the IMPRINT statement's definition of live birth, which was previously adopted. Participants unanimously agreed to modify the ICMART definition of late fetal loss to include an estimated gestational age threshold of 20 completed weeks. Within the context of this core outcome set, participants recommended miscarriage should only be reported after a viable pregnancy has been confirmed by ultrasound.

Stillbirth

Participants discussed the variety of contextual factors including local cultural influences, legislative framework and national and international reporting requirements, which would influence the different gestational age thresholds incorporated in different definitions of stillbirth. They

Table II Standardized definitions for the core outcome set for infertility.

Viable intrauterine pregnancy confirmed by ultrasound	A pregnancy diagnosed by ultrasonographic examination of at least one fetus with a discernible heartbeat.	• Researchers should report at which gestation the ultrasound examination was performed. • Pregnancies are counted as pregnancy events, for example, a twin pregnancy is counted as one pregnancy event. • Effect size estimates and 95% confidence interval should be reported for pregnancy events. The denominator should be per participant randomized. • Singleton, twin and higher multiple pregnancy should be reported separately.
Pregnancy loss		• When considering twin and higher multiple pregnancies, pregnancy loss should be explicitly accounted for.
Ectopic pregnancy	A pregnancy outside the uterine cavity, diagnosed by ultrasound, surgical visualization or histopathology.	
Miscarriage	The spontaneous loss of an intrauterine pregnancy prior to 20 completed weeks of gestational age.	• Miscarriage should be reported after a viable pregnancy has been confirmed by ultrasound.
Stillbirth	The death of a fetus prior to the complete expulsion or extraction from its mother after 20 completed weeks of gestational age. The death is determined by the fact that, after such separation, the fetus does not breathe or show any other evidence of life, such as heartbeat, umbilical cord pulsation or definite movement of voluntary muscles.	• When considering stillbirth involving twins and higher multiple births they should be reported as a single event.
Termination of pregnancy	Intentional loss of an intrauterine pregnancy, through intervention by medical, surgical or unspecified means.	• Selective embryo or fetal reduction should be reported.
Live birth	The complete expulsion or extraction from a woman of a product of fertilization, after 20 completed weeks of gestational age; which, after such separation, breathes or shows any other evidence of life, such as heart beat, umbilical cord pulsation or definite movement of voluntary muscles, irrespective of whether the umbilical cord has been cut or the placenta is attached. A birth weight of 350 g or more can be used if gestational age is unknown.	• Live births are counted as birth events, for example, twin live birth is counted as one live birth event. • Effect size estimates and 95% confidence interval should be reported for live birth events. The denominator should be per participant randomized. • Singletons, twin and higher multiple births should be reported separately.
Gestational age at birth	The age of a fetus is calculated by the best obstetric estimate determined by assessments which may include early ultrasound, and the date of the last menstrual period, and/or perinatal details. In the case of assisted reproductive techniques, it is calculated by adding 14 days to the number of completed weeks since fertilization.	• The gestational age of both live births and stillbirths should be reported. • Gestational age at birth should be reported as a median and interquartile range. Reporting the mean and standard deviation in addition would support future meta-analysis.
Birthweight	Birth weight should be collected within 24 h of birth and assessed using a calibrated electronic scale with 10-g resolution.	• The birthweight of singletons, twins and higher multiples should be reported separately. • Birthweight for each newborn infant of the multiple birth set should be reported. • Birthweight should not be adjusted for gestational age. • The birthweight of stillbirths should be reported.
Neonatal mortality	Death of a live born baby within 28 days of birth. This can be sub-divided into early neonatal mortality, if death occurs in the first 7 days after birth and late neonatal, if death occurs between 8 and 28 days after birth.	• Mortality related to preterm infants should be collected up to 28 days beyond their estimated due date. • If a member of a multiple birth set dies in the neonatal period this should be explicitly reported.
Major congenital anomaly	Structural, functional and genetic anomalies, that occur during pregnancy, and identified antenatally, at birth, or later in life, and require surgical repair of a defect, or are visually evident, or are life-threatening, or cause death.	• Major congenital anomalies should be classified using a standardized taxonomy. • Major congenital anomaly should be reported as an infant with at least one major congenital anomaly detected. • If a major congenital anomaly is identified in a member of a multiple set this should be explicitly reported.

Table III Generic reporting table.

	Experimental N	Control N	Effect size estimate (95% CI)*
Live birth event, no. (%)†			
Singleton, no. (%)			
Twin, no. (%)			
Higher multiples, no. (%)			
Viable pregnancy confirmed by ultrasound, no. (%)†			
Singleton pregnancy, no. (%)			
Twin pregnancy, no. (%)			
Higher multiple pregnancy, no. (%)			
Pregnancy loss‡			
Ectopic pregnancy, no.			
Miscarriage, no.			
Stillbirth, no.			
Termination of pregnancy, no.			
Gestational age at delivery (weeks of gestation), median (IQR)§			
Birthweight			
Singleton, g. (mean, SD)			
Twin, g. (mean, SD)‖			
Higher multiples, g (mean, SD)‖			
Neonatal mortality, no.¶			
Major congenital anomaly, no.#			

*Effect size estimates and 95% CI should only be reported for live birth event and viable pregnancy confirmed by ultrasound. The remaining data should be summarized narratively.
†For live birth event and viable pregnancy confirmed by ultrasound the number of participants randomized should be used as the denominator.
‡When considering twin and higher multiple pregnancies, pregnancy loss should be explicitly accounted for within the table footnote.
§For gestational age at delivery reporting the mean and SD within the table footnote would support future meta-analysis.
‖The birthweight for each newborn infant of the multiple birth set should be reported.
¶If a member of a multiple birth set dies in the neonatal period this should be explicitly stated within the table footnote.
#Reported as an infant with at least one major congenital anomaly detected. If a major congenital anomaly is identified in a member of a multiple set this should be explicitly stated within the table footnote.
g, grams; N, number of randomized participants; No, number of events; IQR, interquartile range.

highlighted the importance of accounting for all pregnancy losses and the gestational age threshold for stillbirth would need to consider the threshold already agreed for miscarriage. Participants unanimously agreed to modify the ICMART definition to include a gestational age threshold of 20 completed weeks with an appropriate adjustment for birthweight. When considering stillbirth involving twins and higher multiple pregnancies, participants recommended they should be reported as a single event.

Termination of pregnancy

Following discussion, consensus was reached to adopt the ICMART definition of termination of pregnancy. Participants noted the importance of reporting selective embryo or fetal reduction.

Participants discussed the reporting of pregnancy loss and recommended ectopic pregnancies, miscarriages, stillbirths and terminations of pregnancy should be reported numerically. Percentages and effect size estimates should not be reported. When considering twin and higher multiple pregnancies, participants recommended pregnancy losses should be accounted for within the footnotes of the reporting table and summarized narratively within the study report.

Gestational age at delivery

Following discussion, consensus was reached to adopt the ICMART definition of gestational age. Participants recommended that gestational age at delivery should be reported for both live births and stillbirths. Participants agreed gestational age at delivery should be reported as the median and interquartile range. An effect size estimate should not be reported. Participants recommended that researchers should be encouraged to report the mean and SD within the reporting table footnote to support future meta-analysis.

Birthweight

Participants noted the measurement of birthweight as being well-characterized. Participants noted best practice recommendations, which recommend collecting birthweight within 24 h of birth and using a calibrated electronic scale with 10-g resolution. If there is limited availability of correctly calibrated electronic scales, the type of scale and its calibration should be clearly reported. Participants recommended birthweight should not be adjusted for gestational age. Participants agreed birthweight, reported as a mean and SD, should be

recorded separately for singleton, twin and higher multiple infants. The birthweight for each infant of a multiple birth set should be reported.

Neonatal mortality

Participants noted the consistent use of the WHO definition for neonatal mortality across definition development initiatives, including ICMART, international and national guidelines and Cochrane systematic reviews. A contextual statement was agreed to ensure researchers report any mortality of preterm infants up to 28 days beyond their estimated due date. Participants agreed neonatal mortality should be reported numerically. Percentages and effect size estimates should not be reported. If a member of a multiple birth set dies in the neonatal period this should be stated within the reporting table footnote and summarized narratively within the study report.

Major congenital anomaly

Participants discussed how congenital anomalies varied in severity, with severe anomalies impacting upon an infant's health, development and survival. Participants reached a view that future RCT should consistently report major congenital anomalies. Participants unanimously agreed to modify the ICMART definition to include criteria to ensure only major congenital anomalies are reported. Participants stated the importance of classifying congenital anomalies using a standardized taxonomy (DeSilva *et al.*, 2016). Participants agreed major congenital anomalies should be reported as an infant with at least one major congenital anomaly detected. If a major congenital anomaly is identified in a member of a multiple set this should be stated within the reporting table footnote and summarized narratively within the study report. Percentages and effect size estimates should not be reported.

Time to pregnancy leading to live birth

Detailed guidance regarding the collection, analysis and reporting of time to pregnancy leading to live birth was approved by the meeting participants and has been provided as Supplementary Data File S1.

Discussion

Definition development initiatives, clinical practice guidelines and Cochrane reviews have defined individual core outcomes in different ways. Through formal consensus methods, 14 healthcare professionals, 7 researchers and 6 people with fertility problems, from 11 countries, have successfully developed consensus definitions for all core outcomes. Specific recommendations have been made to improve the reporting of core outcomes.

This study has used formal consensus methods to develop consensus definitions for the core outcome set for infertility. The consensus development conference is a formal consensus development method developed by the US National Institutes of Health and has been used to reach consensus on a variety of topics in many different countries including, Canada, UK and Sweden. The study has engaged a range of different stakeholders, including healthcare professionals, researchers and people with fertility problems, from different countries. Such diversity should secure the generalizability of the results and increase its credibility with other researchers. The study has developed clear and concise recommendations to enable future researchers to collect core outcomes in a standardized approach and report their results in a clear and transparent manner.

This study is not without limitations. There is significant uncertainty regarding the optimal methods for core outcome set development (Duffy and McManus, 2016; Williamson *et al.*, 2017; Duffy *et al.*, 2019a). The COMET initiative has made no formal recommendations regarding the development of consensus definitions. They advocate the use of formal consensus development method in other aspects of core outcome set development, which informed the methodological choices we made in this study. Different formal consensus methods, including the modified Delphi method and modified Nominal Group Technique, could have been used. Further methodological research is required to evaluate the most appropriate consensus methods for studies similar to ours. Consideration should be given to the representativeness of the steering group and consensus meeting participants. Many consensus meeting participants were from European countries ($n = 12$; 44%) and there was limited representation from low- and middle-income countries, which could have impacted upon the development of consensus definitions. Further research should be undertaken to evaluate virtual or blended formats to improve representation while preserving limited resources.

Analyses of data arising from infertility trials, particularly for studies related to ART, are frequently undermined by the use of an inappropriate denominator (Wilkinson *et al.*, 2016). Two main issues exist. The first is the use of a post-randomization denominator, for example, when live birth rates are calculated per embryo transferred, rather than per woman randomized. Analyses conducted on this basis do not reflect the randomized comparisons, as the groups being compared may differ with respect to their characteristics, and therefore, also with respect to their outcomes (Hirji and Fagerland, 2009). The second issue relates to analyses that commit a unit of analysis error (Vail and Gardener, 2003). This error occurs when proportions are calculated using an inappropriate denominator, for example, the number of oocytes or number of embryos. Unit of analysis errors commonly occurs when researchers calculate the pregnancy rate by dividing the number of gestational sacs on ultrasound by the number of embryos transferred. As the outcomes of a couple's embryos are correlated, this approach is incorrect as standard statistical tests assume that the tested observations are independent.

To avoid these important issues, it is good practice to calculate viable pregnancy confirmed by ultrasound and live birth events using the number of participants randomized as the denominator. This approach is explicitly stated within the core outcome set recommendations. Sophisticated statistical analysis methods capable of accommodating post-randomization comparisons and clustered data do exist. They could be reported in addition to the core outcome set if researchers had access to the necessarily statistical expertise.

This study has developed the generic building blocks for future infertility research. A minimum data set affords the opportunity for researchers to easily populate protocols, case report forms and other data collection tools with core outcomes and consensus definitions. The generic reporting table should assist researchers to clearly report their results within journal publications and conference presentations. Implementing a standardized approach should reduce poor reporting practices, for example, incomplete reporting, selective reporting based on statistical significance and inappropriate use of denominators

(Wilkinson et al., 2016). It is anticipated that research studies with limited access to methodological and statistical advice will benefit the most.

Systematic implementation of this core outcome set should ensure the core outcomes are consistently defined by individual trials. Symmetrical application of standardized definitions in all trial arms is known to reduce measurement bias, including observer and verification bias (Mansournia et al., 2017). Blinding outcome assessors to the treatment allocation would further reduce bias (Sterne et al., 2016). Outcome assessors should also undertake comprehensive training. Other strategies can help to ensure consensus definitions are applied correctly and, in a manner, which is unlikely to vary, including standardized data collection tools, internal validation studies and independent adjudication panels. A freely available electronic case report form and data repository are currently being planned to standardize the collection of the core outcome set within future infertility trials (COMMIT-Collection).

The Core Outcomes in Women's and Newborn Health (CROWN) initiative, supported by over 80 specialty journals, including *the Cochrane Gynaecology and Fertility Group*, *Fertility and Sterility* and *Human Reproduction*, have resolved to implement the core outcome set for infertility (Core Outcomes in Women's and Newborn Health Initiative, 2014). In the future participating journals will request researchers to report the definitions for individual core outcomes within published trial reports. When the consensus definition has not been used, the researchers will be asked to report this observation and its implications for their findings. Reporting will be facilitated by the recommendations made within this study.

The need to combine the results of individual trials evaluating fertility treatments should be anticipated by researchers (Wilkinson et al., 2019b). Implementing the core outcome set, including consensus definitions, should be considered good practice and could make a significant contribution in improving the coordination, development and delivery of fertility research within regional, national and international settings (Devall et al., 2020). Standardization will facilitate pairwise meta-analysis and more sophisticated secondary research, including IPD and network meta-analysis (Duffy et al., 2019b). These approaches could provide unique insights into the effectiveness and safety of fertility treatments.

The consensus definitions developed as part of this study could be incorporated into other core outcome sets to promote additional harmony across women's health. Other core outcome sets have been developed for endometriosis, hyperemesis gravidarum and preterm birth, which share common core outcomes including live birth, neonatal mortality and major congenital anomalies (van 't Hooft et al., 2016; Jansen et al., 2020; Duffy et al., 2020b). Core outcome set developers should be encouraged to use the consensus definitions developed as part of this study.

Standardized consensus definitions are not meant to limit regional, national and international requirements to collect and report collect core outcomes using specific definitions, including live birth, stillbirth and congenital anomalies. For example, researchers undertaking research in the UK may wish to define stillbirth as occurring after 22 weeks of gestation, in line with national recommendations (Da Silva et al., 2016). Researchers wishing to collect data using other definitions in the context of their own RCT would continue to be able to do so. Selective reporting should be avoided by presenting findings for both the consensus definition and any other definition used. Researchers would need to carefully consider how these data would be collected to fulfill different definitions and reporting obligations.

The ultimate objective of an infertility trial is a healthy baby who develops normally. There are significant challenges in developing an objective consensus definition regarding what constitutes a 'healthy baby' as contextual factors, including local practices, cultural influences and legal implications, are important considerations. Consensus was reached to define live birth based on a 20-week gestational age threshold, reflecting IMPRINT recommendations and WHO guidelines. The current limit of viability is considered to be 22 weeks of gestation; however, the threshold is constantly challenged as perinatal and neonatal medicine advances. This context was also considered and a clear threshold has been decided through a robust consensus process to facilitate clear reporting across future infertility research (The Harbin Consensus Conference Workshop Group, 2014).

The core outcome set should be reported by all future RCTs evaluating potential fertility treatments. This context is important when considering the consensus definition developed for pregnancy and miscarriage. Routine urinary or serum beta-hCG testing is a common feature of IVF research, however, is less likely when evaluating other interventions. To take this into account, the consensus definition for pregnancy and miscarriage includes ultrasound, which is a common component of antenatal care. An extension to the core outcome set specifically for IVF research (COMMIT-IVF) is currently being developed and includes pregnancy confirmed by urinary or serum beta-hCG testing and early miscarriage.

The development of consensus definitions has provided additional focus upon the language researchers commonly use when reporting infertility research. People with fertility problems and the patient organizations involved in this study have routinely commented upon terminology. It has been often perceived as lacking a patient-centric approach including terms such as missed spontaneous abortion, induced abortion and fetal loss. Researchers should recognize the language used to report fertility research is important and holds significance to people with fertility problems. The standardization of terminology within this core outcome set has been developed to ensure precision and with consideration to good practice guidelines in partnership with people with fertility problems and the patient organizations.

The COMMIT initiative has committed to undertaking further research to assess the uptake and implementation of the core outcome set for infertility (COMMIT-Implementation). Assessing the uptake of the core outcome set, including the use of consensus definitions, will be undertaken by examining registry records, published protocols and RCT. Further research is planned to examine and understand the reasons why researchers do, and do not, implement the core outcome set for infertility. By identifying perceived barriers to implementation, strategies will be developed to promote implementation of the core outcome set across future infertility research.

In conclusion, ensuring that core outcomes are consistently defined across RCT evaluating potential fertility treatments will secure evidence which is more accessible and facilitate the translation of research into clinical practice. Standardized reporting should help limit poor reporting practices. Future researchers should benefit from core outcomes and consensus definitions, which can be included in protocols, case report forms and other data collection tools. The generic

reporting table should assist researchers in clearly reporting their results in journal publications and conference presentations.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Acknowledgments

We would like to thank the consensus development meeting participants and colleagues at the Cochrane Gynaecology and Fertility Group, University of Auckland, New Zealand.

Authors' roles

Study concept and design: J.M.N.D., S.B., S.B., M.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., M.H., M.L.H., V.J., Y.K., R.S.L., S.L., D.M., B.W.M., C.N., E.H.Y.N., L.P., S.R., I.S., M.S., A.S., A.V., M.v.W., M.V., N.L.V., A.Y.W., R.W., J.W., M.A.Y. and C.M.F. Acquisition of data: J.M.N.D., S.B., S.B., M.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., M.H., M.L.H., V.J., Y.K., R.S.L., S.L., D.M., B.W.M., C.N., E.H.Y.N., L.P., S.R., I.S., M.S., A.S., A.V., M.v.W., M.V., N.L.V., A.Y.W., R.W., J.W., M.A.Y. and C.M.F. Analysis and interpretation of data: J.M.N.D., S.B., S.B., M.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., M.H., M.L.H., V.J., Y.K., R.S.L., S.L., D.M., B.W.M., C.N., E.H.Y.N., L.P., S.R., I.S., M.S., A.S., A.V., M.v.W., M.V., N.L.V., A.Y.W., R.W., J.W., M.A.Y. and C.M.F.

Drafting of manuscript: J.M.N.D., S.L., A.V., J.W. and C.M.F. Critical revision of the manuscript for important intellectual content: S.B., S.B., M.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., M.H., M.L.H., V.J., Y.K., R.S.L., D.M., B.W.M., C.N., E.H.Y.N., L.P., S.R., I.S., M.S., A.S., M.v.W., M.V., N.L.V., A.Y.W., R.W. and M.A.Y. Statistical analysis: J.M.N.D., A.V. and J.W. Study supervision: C.M.F.

Funding

This research was funded by the Catalyst Fund, Royal Society of New Zealand, Auckland Medical Research Fund and Maurice and Phyllis Paykel Trust. The funder had no role in the design and conduct of the study, the collection, management, analysis or interpretation of data or manuscript preparation. Siladitya Bhattacharya was supported by the University of Auckland Foundation Seelye Travelling Fellowship. B.W.M. is supported by a National Health and Medical Research Council Practitioner Fellowship (GNT1082548).

Conflict of interest

Siladitya Bhattacharya reports being the Editor-in-Chief of *Human Reproduction Open* and an editor of the Cochrane Gynaecology and Fertility Group. J.L.H.E. reports being the Editor Emeritus of *Human Reproduction*. R.S.L. reports consultancy fees from Abbvie, Bayer, Ferring, Fractyl, Insud Pharma and Kindex and research sponsorship from Guerbet and Hass Avocado Board. B.W.M. reports consultancy fees from Guerbet, iGenomix, Merck, Merck KGaA and ObsEva. C.N. reports being the Editor-in-Chief of *Fertility and Sterility* and Section Editor of the *Journal of Urology*, research sponsorship from Ferring, and a financial interest in NexHand. E.H.Y.N. reports research sponsorship from Merck. A.S. reports consultancy fees from Guerbet. J.W. reports being a statistical editor for the Cochrane Gynaecology and Fertility Group. A.V. reports that he is a Statistical Editor of the *Cochrane Gynaecology & Fertility Review Group* and of the journal *Reproduction*. His employing institution has received payment from Human Fertilisation and Embryology Authority for his advice on review of research evidence to inform their 'traffic light' system for infertility treatment 'add-ons'. N.L.V. reports consultancy and conference fees from Ferring, Merck and Merck Sharp and Dohme. The remaining authors declare no competing interests in relation to the work presented. All authors have completed the disclosure form.

Appendix

Core Outcome Measure for Infertility Trials (COMMIT) initiative

Professor Ahmed M. Abou-Setta, University of Manitoba, Canada; Dr Juan J. Aguilera, Argentina; Dr Hisham AlAhwany, University of Nottingham, UK; Dr Oluseyi O. A. Atanda, Ladoke Akintola University of Technology Teaching Hospital, Nigeria; Eva M. E. Balkenende, University of Amsterdam, The Netherlands; Dr Kurt T. Barnhart, University of Pennsylvania, USA; Dr Yusuf Beebeejaun, King's Fertility, Fetal Medicine Research Institute, UK; Megan Black, New Zealand; Associate Professor Georgina M. Chambers, University of New South Wales, Australia; Dr Abrar A. Chughtai, University of New South Wales, Australia; Dr Javier A. Crosby, Clinica Las Condes, Chile; Dr Irene Cuevas-Sáiz, Hospital General Universitario de Valencia, Spain; Dr Cate Curtis, University of Waikato, New Zealand; Dr Arianna D'Angelo, Wales Fertility Institute, UK; Danielle D. Dubois, Ottawa Fertility Centre, Canada; Dr Kirsten Duckitt, University of British Columbia, Canada; Dr Carlos Encinas, Geneva Foundation for Medical Education and Research, Bolivia; Dr Marie-Odile Gerval, Chelsea and Westminster Hospital NHS Foundation Trust, UK; Dr Nhu H. Giang, Vietnam; Dr Ahmed Gibreel, Mansoura University, Egypt; Lynda J. Gingel, UK; Dr Elizabeth J. Glanville, Fertility Plus, National Women's Hospital, New Zealand; Dr Demian Glujovsky, CEGYR Medicina Reproductiva, Argentina; Dr Ingrid Granne, University of Oxford, UK; Professor Georg Griesinger, University Hospital of Schleswig-Holstein, Germany; Dr Devashana Gupta Repromed, New Zealand; Associate Professor Zeinab Hamzehgardeshi, Mazandaran University of Medical Sciences, Iran; Dr Martin Hirsch, University College London Hospitals, UK; Dr Marcos Horton, Pregna Reproductive Medicine, Argentina; Dr Shikha Jain, Dreamz IVF, India; Dr Marta Jansa Perez, UK; Dr Claire A. Jones, University of Toronto, Canada; Professor Mohan S. Kamath, Christian Medical College, Vellore, India; José Knijnenburg, Freya, The Netherlands; Dr Elena Kostova, Cochrane Gynaecology and Fertility, The Netherlands; Professor Antonio La Marca, University Hospital of Modena, Italy; Dr Tien Khac Le, Vietnam; Dr Arthur Leader, Ottawa Hospital Research Institute, Canada; Dr Brigitte Leeviers, University

Hospital Zurich, Switzerland; Dr Jian Li Chinese, University of Hong Kong, China; Professor Olabisi M. Loto, Obafemi Awolowo University, Nigeria; Karen L. Marks, UK; Rodrigo M. Martinez-Vazquez, Cochrane Iberoamerica, Spain; Alison R. McTavish, University of Aberdeen, UK; David J. Mills, UK; Dr Raju R. Nair, Mitera Hospital, India; Dr Dung Thi Phuong Nguyen, Vietnam; Dr Anne-Sophie Otter, Osakidetza OSI, Spain; Professor Allan A. Pacey, University of Sheffield, UK; Satu Rautakallio-Hokkanen, Fertility Europe, Belgium; Dr Lynn C. Sadler, Auckland District Health Board, New Zealand; Dr Peggy Sagle, University of Alberta, Canada; Dr Juan-Enrique Schwarze, Clinica Las Condes, Chile; Dr Heather M. Shapiro, University of Toronto, Canada; Dr Joe L. Simpson, Florida International University, USA; Professor Charalampos S. Siristatidis, Greece; Dr Akanksha Sood, St. Mary's Hospital, UK; Catherine Strawbridge, Fertility Network UK, UK; Dr Helen L. Torrance, University Medical Center Utrecht, The Netherlands; Dr Cam Tu Tran, Vietnam; Emma L. Votteler, Bath Fertility Centre, UK; Professor Chi Chiu Wang, The Chinese University of Hong Kong, Hong Kong; Dr Andrew Watson, Tameside Foundation Trust, UK; and Dr Menem Yossry, City Hospital Sunderland, UK.

References

American Urological Association. *The Optimal Evaluation of the Infertile Male: AUA Best Practice Statement*. Linthicum, MD: American Urological Association, 2010a.

American Urological Association. *The Evaluation of the Azoospermic Male*. Linthicum, MD: American Urological Association, 2010b.

American Urological Association. The Management of Obstructive Azoospermia. Linthicum, MD: American Urological Association, 2010c.

Carranza-Mamane B, Havelock J, Hemmings R, Cheung A, Sierra S, Carranza-Mamane B, Case A, Cathie D, Graham J, Havelock J *et al*. The management of uterine fibroids in women with otherwise unexplained infertility. *J Obstet Gynaecol Can* 2015;**37**:277–285.

Chen RT, Moro PL, Bauwens J, Bonhoeffer J. Obstetrical and neonatal case definitions for immunization safety data. *Vaccine* 2016;**34**: 5991–5992.

Core Outcomes in Women's and Newborn Health Initiative. The CROWN Initiative: journal editors invite researchers to develop core outcomes in women's health. *Hum Reprod* 2014;**29**: 1349–1350.

Da Silva FT, Gonik B, Mcmillan M, Keech C, Dellicour S, Bhange S, Tila M, Harper D M, Woods C, Kawai A T *et al*.. Stillbirth: Case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine 2016;**34**: 6057–6068.

DeSilva M, Munoz FM, McMillan M, Kawai AT, Marshall H, Macartney KK, Joshi J, Oneko M, Rose AE, Dolk H *et al*. Congenital anomalies: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. *Vaccine* 2016;**34**:6015–6026.

Devall AJ, Out JH, Mol BWJ, Duffy JMN, Collura B, Dyer S. Coordination and planning of clinical research on a national and global level. *Fertil Steril* 2020;**113**:1100–1106.

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. *Hum Reprod* 2020a;doi: 10.1093/humrep/deaa241.

Duffy JMN, Bhattacharya S, Curtis C, Evers JLH, Farquharson RG, Franik S, Khalaf Y, Legro RS, Lensen S, Mol BW *et al*.; COMMIT: Core Outcomes Measures for Infertility Trials. A protocol developing, disseminating, and implementing a core outcome set for infertility. *Hum Reprod Open* 2018;**2018**:hoy007.

Duffy JMN, Bhattacharya S, Herman M, Mol B, Vail A, Wilkinson J, Farquhar C, the Cochrane Gynaecology and Fertility Group. Reducing research waste in benign gynaecology and fertility research. *BJOG: Int J Obstet Gy* 2017a;**124**:366–369.

Duffy JMN, Hirsch M, Vercoe M, Abbott J, Barker C, Collura B, Drake R, Evers J, Hickey M, Horne AW *et al*.; endo:outcomes – an International Collaboration Harmonising Outcomes and Outcome Measures for Endometriosis Research. A core outcome set for future endometriosis research: an international consensus development study. *BJOG: Int J Obstet Gy* 2020b;**127**: 967–974.

Duffy JMN, Hirsch M, Ziebland S, McManus RJ, Brown M, Gale C, Grobman W, Fitzpatrick R, Karumanchi SA, Lucas N *et al*.; the International Collaboration to Harmonise Outcomes in Pre-eclampsia (iHOPE). Methodological decisions influence the identification of potential core outcomes in studies related to pre-eclampsia: an analysis informing the development of recommendations for future core outcome set developers. *BJOG: Int J Obstet Gy* 2019a;**126**:1482–1490.

Duffy JMN, McManus R. Influence of methodology upon the identification of potential core outcomes: recommendations for core outcome set developers are needed. *BJOG: Int J Obstet Gy* 2016;**123**: 1599.

Duffy JMN, Rolph R, Gale C, Hirsch M, Khan KS, Ziebland S, McManus RJ, On behalf of the International Collaboration to Harmonise Outcomes in Pre-eclampsia (iHOPE). Core outcome sets in women's and newborn health: a systematic review. *BJOG: Int J Obstet Gy* 2017b;**124**:1481–1489.

Duffy JMN, Ziebland S, von Dadelszen P, McManus RJ. Tackling poorly selected, collected, and reported outcomes in obstetrics and gynecology research. *Am J Obstet Gynecol* 2019b;**220**: 71.e71–71.e74.

Ferguson JH.The NIH consensus development program.The evolution of guidelines. Int J Technol Assess Health Care 1996;**12**: 460–474.

Hirji KF, Fagerland MW. Outcome based subgroup analysis: a neglected concern. *Trials* 2009;**10**:33.

Hirsch M, Duffy JMN, Barker C, Hummelshoj L, Johnson NP, Mol B, Khan KS, Farquhar C. Protocol for developing, disseminating and implementing a core outcome set for endometriosis. *BMJ Open* 2016;**6**:e013998.

Jansen LAW, Koot MH, van'T Hooft J, Dean CR, Duffy JMN, Ganzevoort W, Gauw N, Goes BY, Rodenburg J, Roseboom TJ *et al*. A core outcome set for hyperemesis gravidarum research: an international consensus study. *BJOG: Int J Obstet Gy* 2020;**127**: 983–992.

Jarvi K, Lo K, Fischer A, Grantmyre J, Zini A, Chow V, Mak VC. CUA Guideline the workup of azoospermic males. *Can Urol Assoc J* 2010;**4**:163–167.

Jungwirth A, Diemer T, Kopa Z, Krausz C, Minhas S, Tournaye H. *EAU Guidelines on Male Infertility*. The Netherlands: European Association of Urology, 2018, Arnhem, The Netherlands.

Khalil A, Duffy JMN, Perry H, Ganzevoort W, Reed K, Baschat AA, Deprest J, Gratacos E, Hecher K, Lewi L *et al.*; On behalf of the International Collaboration to Harmonise Outcomes for Selective Fetal Growth Restriction (CHOOSE-FGR). Study protocol: developing, disseminating, and implementing a core outcome set for selective fetal growth restriction in monochorionic twin pregnancies. *Trials* 2019;**20**:35.

Khalil A, Perry H, Duffy J, Reed K, Baschat A, Deprest J, Hecher K, Lewi L, Lopriore E, Oepkes D; On behalf of the International Collaboration to Harmonise Outcomes for Twin–Twin Transfusion Syndrome (CHOOSE). Twin-Twin Transfusion Syndrome: study protocol for developing, disseminating, and implementing a core outcome set. *Trials* 2017;**18**:325.

Koch J, Rowan K, Rombauts L, Yazdani A, Chapman M, Johnson N. Endometriosis and infertility—a consensus statement from ACCEPT (Australasian CREI Consensus Expert Panel on Trial evidence). *Aust N Z J Obstet Gynaecol* 2012;**52**:513–522.

Kroon B, Johnson N, Chapman M, Yazdani A, Hart R, on behalf of the Australasian CREI Consensus Expert Panel on Trial evidence (ACCEPT) group. Fibroids in infertility–consensus statement from ACCEPT (Australasian CREI Consensus Expert Panel on Trial evidence. *Aust N Z J Obstet Gynaecol* 2011;**51**:289–295.

Loh SF, Agarwal R, Chan JK, Chia SJ, Cho LW, Lim LH, Lau MS, Loh SK, Hendricks MS, Nair S *et al.* Academy of Medicine-Ministry of Health Clinical Practice Guidelines: assessment and management of infertility at primary healthcare level. *Singapore Med J* 2014;**55**:58–65.

Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology* 2017;**28**:54–59.

Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c869.

Murphy M, Sanderson C, Black N, Askham J, Lamping D, Marteau T, McKee C. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;**2**:1–88.

National Institute for Health and Care Excellence. *Fertility Problems: Assessment and Treatment*. London, UK: National Institute for Health and Care Excellence, 2013.

Practice Committee of the American Society for Reproductive Medicine. Role of metformin for ovulation induction in infertile patients with polycystic ovary syndrome (PCOS): a guideline. *Fertil Steril* 2017a;**108**:426–441.

Practice Committee of the American Society for Reproductive Medicine. Removal of myomas in asymptomatic patients to improve fertility and/or reduce miscarriage rate: a guideline. *Fertil Steril* 2017b;**108**:416–425.

Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;**355**:i4919.

The Harbin Consensus Conference Workshop Group. Improving the reporting of clinical trials of infertility treatments (IMPRINT): modifying the CONSORT statement. *Hum Reprod* 2014;**29**: 2075–2082.

Vail A, Gardener E. Common statistical errors in the design and analysis of subfertility trials. *Hum Reprod* 2003;**18**:1000–1004.

van 't Hooft J, Duffy JMN, Daly M, Williamson PR, Meher S, Thom E, Saade GR, Alfirevic Z, Mol BW, Khan KS; Global Obstetrics Network (GONet). A core outcome set for evaluation of interventions to prevent preterm birth. *Obstet Gynecol* 2016;**127**: 49–58.

Webbe J, Brunton G, Ali S, Duffy JMN, Modi N, Gale C. Developing, implementing and disseminating a core outcome set for neonatal medicine. *BMJ Paediatr Open* 2017;**1**:e000048.

Whitehouse KC, Kim CR, Ganatra B, Duffy JMN, Blum J, Brahmi D, Creinin MD, DePiñeres T, Gemzell-Danielsson K, Grossman D *et al.* Standardizing abortion research outcomes (STAR): a protocol for developing, disseminating and implementing a core outcome set for medical and surgical abortion. *Contraception* 2017;**95**: 437–441.

Wilkinson J, Bhattacharya S, Duffy JMN, Kamath MS, Marjoribanks J, Repping S, Vail A, Wely M, Farquhar CM. Reproductive medicine: still more ART than science? *BJOG: Int J Obstet Gy* 2019a;**126**: 138–141.

Wilkinson J, Brison DR, Duffy JMN, Farquhar CM, Lensen S, Mastenbroek S, van Wely M, Vail A. Don't abandon RCTs in IVF. We don't even understand them. *Hum Reprod* 2019b;**34**: 2093–2098.

Wilkinson J, Roberts SA, Showell M, Brison DR, Vail A. No common denominator: a review of outcome measures in IVF RCTs. *Hum Reprod* 2016;**31**:2714–2722.

Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, Clarke M, Gargon E, Gorst S, Harman N *et al.* The COMET handbook: version 1.0. *Trials* 2017;**18**:280.

World Health Organization. *International Classification of Diseases-11 for Mortality and Morbidity Statistics*. Geneva, Switzerland: World Health Organization, 2018.

Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, Rienzi L, Sunde A, Schmidt L, Cooke ID *et al.* The international glossary on infertility and fertility care, 2017. *Fertil Steril* 2017;**108**:393–406.