Human Reproduction, Vol.35, No.12, pp. 2725–2734, 2020
Advance Access Publication on November 30, 2020 doi:10.1093/humrep/deaa241

human reproduction

ORIGINAL ARTICLE *Infertility*

Developing a core outcome set for future infertility research: an international consensus development study† ‡

J.M.N. Duffy[1,2,*], H. AlAhwany[3], S. Bhattacharya[4], B. Collura[5], C. Curtis[6,7], J.L.H. Evers[8], R.G. Farquharson[9], S. Franik[10], L.C. Giudice[11,12], Y. Khalaf[13], J.M.L. Knijnenburg[14], B. Leeners[15], R.S. Legro[16], S. Lensen[17], J.C. Vazquez-Niebla[18], D. Mavrelos[19], B.W.J. Mol[20], C. Niederberger[21], E.H.Y. Ng[22,23], A.S. Otter[24], L. Puscasiu[25], S. Rautakallio-Hokkanen[26], S. Repping[27], I. Sarris[1], J.L. Simpson[28], A. Strandell[29], C. Strawbridge[30], H.L. Torrance[31], A. Vail[32], M. van Wely[27], M.A. Vercoe[33], N.L. Vuong[34], A.Y. Wang[35], R. Wang[20], J. Wilkinson[32], M.A. Youssef[36], C.M. Farquhar[33], 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, University of Nottingham, Derby, UK [4]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, UK [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]Department of Obstetrics and Gynaecology, Liverpool Women's NHS Foundation Trust, Liverpool, UK [10]Department of Obstetrics and Gynaecology, Münster University Hospital, Münster, Germany [11]Center for Research, Innovation and Training in Reproduction and Infertility, Center for Reproductive Sciences, University of California, San Francisco, CA, USA [12]International Federation of Fertility Societies, Philadelphia, PA, USA [13]Department of Women and Children's Health, King's College London, Guy's Hospital, London, UK [14]Freya Dutch Infertility Association, Gorinchem, The Netherlands [15]Department of Reproductive Endocrinology, University Hospital Zurich, Zurich, Switzerland [16]Department of Obstetrics and Gynaecology, Penn State College of Medicine, PA, USA [17]Department of Obstetrics and Gynaecology, University of Melbourne, VIC, Australia [18]Cochrane Iberoamerica, Biomedical Research Institute Sant Pau, Barcelona, Spain [19]Reproductive Medicine Unit, University College Hospital, London, UK [20]Department of Obstetrics and Gynaecology, Monash University, Melbourne, Australia [21]Department of Urology, University of Illinois at Chicago College of Medicine, Chicago, IL, USA [22]Department of Obstetrics and Gynaecology, The University of Hong Kong, Hong Kong [23]Shenzhen Key Laboratory of Fertility Regulation, The University of Hong Kong-Shenzhen Hospital, China [24]Osakidetza OSI, Bilbao, Basurto, Spain [25]University of Medicine, Pharmacy, Sciences and Technology, Targu Mures, Romania [26]Fertility Europe, Evere, Belgium [27]Center for Reproductive Medicine, Amsterdam Reproduction and Development Institute, Amsterdam University Medical Centres, Amsterdam, The Netherlands [28]Department of Human and Molecular Genetics, Florida International University, FL, USA [29]Department of Obstetrics and Gynecology, Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden [30]Fertility Network UK, London, UK [31]Department of Reproductive Medicine, University Medical Centre Utrecht, Utrecht, the Netherlands [32]Centre for Biostatistics, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK [33]Cochrane Gynaecology and Fertility Group, University of Auckland, Auckland, New Zealand [34]Department of Obstetrics and Gynaecology, University of Medicine and Pharmacy in Ho Chi Minh City, Ho Chi Minh City, Vietnam [35]Faculty of Health, University of Technology, Sydney, Broadway, Australia [36]Department of Obstetrics and Gynaecology, Faculty of Medicine, Cairo University, Cairo, Egypt

*Correspondence address. King's Fertility, Fetal Medicine Research Institute, London, UK. E-mail: james.duffy3@nhs.net

Submitted on May 12, 2020; resubmitted on July 8, 2020; editorial decision on July 22, 2020

STUDY QUESTION: Can a core outcome set to standardize outcome selection, collection and reporting across future infertility research be developed?

SUMMARY ANSWER: A minimum data set, known as a core outcome set, has been developed for randomized controlled trials (RCTs) and systematic reviews evaluating potential treatments for infertility.

†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.

WHAT IS KNOWN ALREADY: Complex issues, including a failure to consider the perspectives of people with fertility problems when selecting outcomes, variations in outcome definitions and the selective reporting of outcomes on the basis of statistical analysis, make the results of infertility research difficult to interpret.

STUDY DESIGN, SIZE, DURATION: A three-round Delphi survey (372 participants from 41 countries) and consensus development workshop (30 participants from 27 countries).

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

MAIN RESULTS AND THE ROLE OF CHANCE: The core outcome set consists of: viable intrauterine pregnancy confirmed by ultrasound (accounting for singleton, twin and higher multiple pregnancy); pregnancy loss (accounting for ectopic pregnancy, miscarriage, stillbirth and termination of pregnancy); live birth; gestational age at delivery; birthweight; neonatal mortality; and major congenital anomaly. Time to pregnancy leading to live birth should be reported when applicable.

LIMITATIONS, REASONS FOR CAUTION: We used consensus development methods which have inherent limitations, including the representativeness of the participant sample, Delphi survey attrition and an arbitrary consensus threshold.

WIDER IMPLICATIONS OF THE FINDINGS: Embedding the core outcome set within RCTs and systematic reviews should ensure the comprehensive selection, collection and reporting of core outcomes. Research funding bodies, the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statement, and over 80 specialty journals, including the Cochrane Gynaecology and Fertility Group, *Fertility and Sterility* and *Human Reproduction*, 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. The funder had no role in the design and conduct of the study, the collection, management, analysis or interpretation of data, or manuscript preparation. B.W.J.M. is supported by a National Health and Medical Research Council Practitioner Fellowship (GNT1082548). S.B. was supported by University of Auckland Foundation Seelye Travelling Fellowship. S.B. 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*. J.M.L.K. reports research sponsorship from Ferring and Theramex. 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.J.M. reports consultancy fees from Guerbet, iGenomix, Merck, Merck KGaA and ObsEva. C.N. reports being the Co Editor-in-Chief of *Fertility and Sterility* and Section Editor of the *Journal of Urology*, research sponsorship from Ferring, and retains a financial interest in NexHand. A.S. reports consultancy fees from Guerbet. E.H.Y.N. reports research sponsorship from Merck. 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: Consensus development study / core outcome sets / modified Delphi method / modified Nominal Group Technique / outcome measures; outcomes

Introduction

Randomized controlled trials (RCTs) evaluating potential fertility treatments should select, collect and report outcomes that are relevant to people with infertility and reflect the realities of clinical practice (Duffy *et al.*, 2017a). Unfortunately, many infertility trials fall short of this requirement (Wilkinson *et al.*, 2019a). Complex issues, including a failure to take into account the perspectives of people with infertility when designing RCT, variations in outcomes and selective reporting of outcomes, make research evidence difficult to interpret, undermining the translation of research into clinical practice (Duffy *et al.*, 2019a).

Historically, there has been a limited emphasis upon the engagement of people with fertility problems in the design of research, which may have inadvertently led to the selection of outcomes based upon the preferences of researchers. A systematic review has characterized outcome reporting across infertility trials and demonstrates the wide variation in reporting, for example, the majority of infertility trials have not reported live birth, major congenital anomalies and adverse events (Dapuzzo *et al.*, 2011). Even when relevant outcomes are reported, different definitions can limit interpretation. For example, live birth has been inconsistently defined, using different definitions, including 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 provides sufficient flexibility for researchers to selectively report favorable results based on statistical significance. Selective reporting of outcomes based on statistical significance, commonly referred to as result cherry picking, is thought to be widespread across infertility research and can result in the overestimation of treatment efficacy and underestimation of harm (Duffy *et al.*, 2019a). Without consistent outcome selection, collection and reporting, evidence synthesis can be challenging and can make comparisons and combining these data within a meta-analysis impossible (Braakhekke *et al.*, 2014).

These problems can be addressed by the development of a core outcome set for RCT and systematic reviews evaluating potential treatments for infertility. A core outcome set represents a minimum collection of particularly important outcomes and outcome measures which have been developed using formal consensus methods engaging health care professionals, researchers and people with fertility problems (Duffy *et al.*, 2017a). Core outcomes should be routinely utilized by researchers, collected in a standardized manner and reported

consistently in the final publication (Core Outcomes in Women's and Newborn Health Initiative, 2014).

Motivated by the desire to increase the utility of future infertility research, an international collaboration embedded within the Cochrane Gynaecology and Fertility Group, has brought health care professionals, researchers and people with fertility problems together to develop a core outcome set for future infertility research.

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 health care professionals, researchers and people with fertility problems, was established. The steering group was convened during the development of the study protocol, before the launch of the Delphi survey and before the consensus development meeting, to obtain advice regarding the participant sample, data collection and data analysis.

The core outcome set was developed in a three-stage process using consensus science methods advocated by the COMET initiative (Williamson *et al.*, 2017). A protocol describing the methods has previously been published (Duffy *et al.*, 2018). The protocol was 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 (Duffy *et al.*, 2016; Hirsch *et al.*, 2016a,b; Khalil *et al.*, 2017, 2019; Webbe *et al.*, 2017; Whitehouse *et al.*, 2017).

The important work of the Harbin Consensus Working Group (Harbin Consensus Conference Workshop Group, 2014) and International Committee for Monitoring Assisted Reproductive Technologies (Zegers-Hochschild *et al.*, 2017) is complementary to this study.

A comprehensive inventory of outcomes was developed by extracting outcomes from systematic reviews that had already quantified outcome reporting across infertility trials (Dapuzzo *et al.*, 2011; Braakhekke *et al.*, 2014; Wilkinson *et al.*, 2016). Lay definitions were developed for individual outcomes. The outcome inventory and lay definitions were entered into a modified Delphi method (Murphy *et al.*, 1998).

The study aimed to recruit key stakeholders including health care professionals, researchers and people with fertility problems. Healthcare professionals and researchers were recruited through the British Fertility Society, Core Outcomes in Women's and Newborn Health initiative, Cochrane Gynaecology and Fertility Group, International Federation of Fertility Societies, the International Federation of Gynecology and Obstetrics Committee for Reproductive Medicine, Endocrinology and Infertility, Reproductive Medicine Clinical Study Group and Royal College of Obstetricians and Gynaecologists. People with fertility problems were recruited through Fertility Europe, Fertility Network UK, Fertility New Zealand and RESOLVE: The National Infertility Association. Recruitment was supported by an active social media campaign. The Delphi method does not depend on statistical power. Working from its underlying principles, group error should decrease and the decision quality increase as the number of participants increases. Between 10 and 15 participants have been demonstrated to yield sufficient results and assure validity (Murphy *et al.*, 1998). Anticipating a 20% attrition rate, we aimed to recruit 18 participants for each stakeholder group.

The modified Delphi method was delivered through sequential online surveys using Delphi survey software (Delphi Manager, University of Liverpool, Liverpool, UK). Potential participants received an explanatory video abstract, a plain language summary and Delphi survey instructions. In round one, participants scored individual outcomes on a nine-point Likert scale. Participants were able to select an 'unable to score' category if they considered themselves not to have sufficient expertise or experience to score an individual outcome. Before completing the survey, participants were able to suggest additional outcomes. After the round one survey had closed, the scores for each outcome were aggregated across individual stakeholder groups. The percentage of participants scoring each outcome at every possible response from one to nine was calculated and tabulated for individual stakeholder groups: healthcare professionals, researchers and people with fertility problems. Additional outcomes were considered by the steering group and novel outcomes were entered into the round two survey.

In round two, participants were asked to reflect on their own scores and on the scores of other participants, before rescoring each outcome. Before completing the survey, participants were able to score additional outcomes suggested by participants in the round one survey. After the round two surveys had closed, the percentage of participants scoring each outcome at every possible response from one to nine was calculated and tabulated for individual stakeholder groups. An *a priori* consensus definition, a median score of eight in each stakeholder group, was applied to identify consensus outcomes.

The round two Delphi survey results were reviewed by the steering group to consider whether a further Delphi survey round was required. The steering group members concluded it was unlikely a further Delphi survey round would identify additional consensus outcomes. However, as there is uncertainty regarding the use of the modified Delphi method in core outcome set development, the steering group recommended proceeding with a third Delphi survey round, to ensure that no further consensus outcomes would have been identified (Williamson *et al.*, 2017).

Following the round two survey, a face-to-face consensus development meeting was arranged. A modified Nominal Group Technique was used to further prioritize consensus outcomes. Healthcare professionals, researchers and people with fertility problems who had completed all three rounds of the Delphi survey were invited to participate. The modified Nominal Group Technique does not depend on statistical power. In consultation with the steering group, we aimed to recruit between 10 and 15 participants, as this number has yielded sufficient results and assured validity in other settings (Murphy *et al.*, 1998).

The modified Nominal Group Technique provides an opportunity to generate ideas, which are discussed, and ranked by a group of experts (Murphy *et al.*, 1998). At the start of the meeting, the results of the Delphi survey were reviewed. All potential core outcomes reaching the standardized consensus definition were entered into the process. Participants were able to enter other potential core outcomes which had not reached the standardized consensus definition, upon request. Each participant was asked to contribute their opinions. Following the initial discussion, outcomes were divided into three initial categories: outcomes to be considered for inclusion in the final core outcome set; outcomes where no consensus existed; and outcomes

which should not be considered for inclusion in the final core outcome set. Participants were invited to discuss the ordering of the outcomes within each category. The discussion focused upon ranking the outcomes being considered for inclusion in the final core outcome set and the outcomes where no consensus existed. During the discussion, the outcomes could be moved between the categories. Finally, the core outcome set was agreed.

Results

An outcome inventory, which included 101 outcomes, was developed (Supplementary Table SI). These outcomes were thematically ordered into 23 thematic domains, including early pregnancy outcomes, patient-reported outcomes and adverse events immediately following treatment. Outcome domains, outcomes and lay definitions were entered into the modified Delphi method.

When considering the Delphi survey, round one was completed by 261 healthcare professionals, 57 researchers and 54 people with fertility problems, from 41 countries (Table I). Round two was completed by 275 participants and round three was completed by 227 participants. One hundred and one outcomes were entered into the Delphi survey (Fig. 1). In response to the outcomes suggested by participants, the steering group added 32 additional outcomes to round two, including cumulative live birth, experimental intervention feasibility and cost effectiveness. Therefore, 133 outcomes were scored during round two. Following round two, 28 outcomes reached the consensus threshold. No additional consensus outcomes were identified following the completion of the round three survey.

Fifteen healthcare professionals, six researchers and nine people with fertility problems, including four men with fertility problems, from 27 countries, participated in the consensus development meeting. Twenty-eight consensus outcomes were entered into the modified Nominal Group Technique. Participants entered an additional eight no consensus outcomes into the process. These outcomes had been highly scored by people with infertility (median score nine), however, had not met the consensus threshold because of lower scores in other stakeholder groups. Participants prioritized outcomes for inclusion in the core outcome set for infertility (Fig. 2).

Discussion

Using formal consensus science methods, health care professionals, researchers and people with fertility problems have developed a core outcome set which should be used to standardize outcome selection,

Table I Participant characteristics.

	Modified Delphi method				Modified nominal group technique
	Round 1 n = 372	Round 2 n = 275	Round 3 n = 227	Withdrawals n = 145	n = 30
Stakeholder group, n					
Health professionals	261	203	176	85	15
Researchers	57	44	38	19	6
People with infertility	54	28	13	41	9
Gender, n					
Male	124	94	76	48	15
Female	244	178	148	96	13
Not stated	4	3	3	1	2
Age (years), n					
Under 29	75	64	62	13	3
30 to 39	116	81	66	50	6
40 to 49	76	54	39	37	5
50 to 59	7	54	44	34	8
Over 60	22	18	12	10	6
Prefer not to say	5	4	4	1	0
Geographical location, n					
Africa	13	5	5	8	1
Asia	118	99	94	24	2
Australia and New Zealand	42	34	29	13	3
Europe	134	92	70	64	17
North America	37	26	18	19	4
South America	15	9	5	10	2
Prefer not to say	13	10	6	7	1

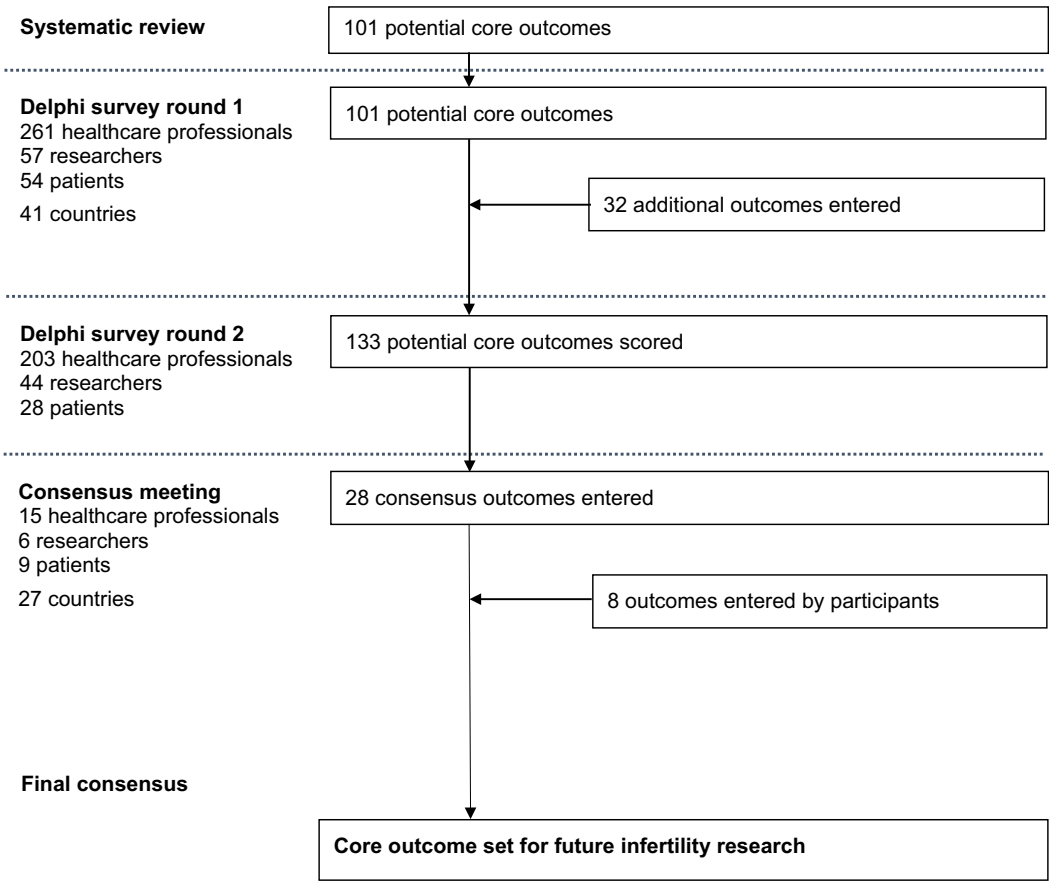

Figure 1. Flow of participants and outcomes.

- Viable intrauterine pregnancy confirmed by ultrasound. Accounting for singleton pregnancy, twin pregnancy, and higher multiple pregnancy.
- Pregnancy loss. Accounting for ectopic pregnancy, miscarriage, stillbirth, and termination of pregnancy.
- Live birth.
- Gestational age at delivery.
- Birth weight.
- Neonatal mortality.
- Major congenital anomaly.

* When applicable → time to pregnancy leading to live birth.

Figure 2. A core outcome set for future infertility research.

collection and reporting across RCT and systematic reviews evaluating potential treatments for infertility.

The COMET initiative has recently published methodological standards for core outcome set development (Kirkham *et al.*, 2017). This study has met these standards. With 372 participants, from 41 countries, participating in the Delphi survey and 30 participants, from 27 countries, participating in the consensus development meeting, the global participation achieved in this study should secure the

generalizability of the results across diverse research settings. The study included people with fertility problems as steering group members and participants. As participants, they shared their views regarding the importance of potential core outcomes during the Delphi survey and participated fully in the consensus development meeting, which prioritized the final core outcome set. This contribution should ensure the final core outcome set holds the necessary reach and relevance to people with fertility problems.

This consensus study is not without limitations. Consideration should be given to the representativeness of the study's participants. When considering the Delphi survey, there was a higher response from participants who lived in Europe (134 participants; 36%). To participate in the Delphi survey, English proficiency, a computer and internet access were required. We appreciate limitations in the representativeness of the sample could have impacted upon the outcomes prioritized.

There is considerable uncertainty regarding core outcome set development methods (Duffy and McManus, 2016; Williamson et al., 2017; Duffy et al., 2019b). The optimal approaches to selecting participants, structuring interactions, and methods of synthesizing individual judgments are unclear (Murphy et al., 1998). Further methodological research is required to inform future core outcome set development (Williamson et al., 2017).

The Delphi survey's overall attrition rate was 38%, which is comparable to other core outcome development studies (Duffy et al., 2017b). Participants who identified as people with fertility problems were more likely to withdraw. It may have been possible to reduce attrition by reducing the length of the survey; for example, limiting the outcomes entered into the Delphi survey, removing outcomes which reached consensus in subsequent survey rounds, or reducing the number of survey rounds. However, attrition needed to be balanced with the requirement to enter a comprehensive long list of potential core outcomes into the Delphi survey and for participants to be able to reflect on and rescore individual outcomes in relation to each other. Further methodological research is required to understand the impact of attrition on the development of consensus within core outcome set development studies.

Many international initiatives, professional societies and colleagues have strongly advocated for the collection and reporting of many of the core outcomes, including live birth, pregnancy loss and adverse events (Barnhart, 2014; Harbin Consensus Conference Workshop Group, 2014). Despite the clear articulation of the importance of these outcomes, poor reporting persists with only one-third of infertility trials reporting live birth (Wilkinson et al., 2016). Why will this time be different? The Core Outcome Measure for Infertility Trials (COMMIT) initiative has developed a strategic plan in consultation with a broad range of stakeholders across the research pipeline to utilize available enablers to secure the routine selection, collection and reporting of core outcomes across future fertility research (Devall et al., 2020).

Research funding bodies are increasingly advocating for the use of core outcome sets within the research they fund. It is considered good practice for researchers planning RCT to follow the SPIRIT statement, which outlines the scientific, ethical and administrative elements that should be incorporated in a clinical trial protocol (Chan et al., 2013). This statement specifically recommends the collection of core outcomes.

This study has established a core outcome set for infertility, however different definitions exist for individual core outcomes. The study has recently developed standardized definitions, using formal consensus development methods, for individual core outcomes. This additional harmony across future infertility trials should ensure secondary research can be undertaken prospectively, efficiently and harmoniously (Duffy et al., 2020b). This standardization will be supported by the development of a freely available electronic case report form and data repository, which future researchers will be encouraged to use for data collection (COMMIT-Collection). Several core outcomes, including live birth, birthweight and neonatal mortality, are common to other core outcome sets developed for hyperemesis gravidarum, multiple pregnancy research and neonatal care (Perry et al., 2019; Webbe et al., 2020a; Jansen et al., 2020; Townsend et al., 2020a). Additional consistency could be achieved across our specialty if the consensus definitions developed within this initiative were embedded within these core outcome sets.

The CROWN initiative, supported by over 80 specialty journals, including the Cochrane Gynaecology and Fertility Group, *Fertility and Sterility* and *Human Reproduction*, has resolved to implement this core outcome set (Core Outcomes in Women's and Newborn Health Initiative, 2014). CROWN initiative journals will advise researchers to report the core outcome set for infertility within trial reports and offer conclusions based on these outcomes. Where core outcome sets have not been collected, the researchers will be asked to report this deficiency and its implications for their findings. The COMMIT initiative is currently developing reporting tools and templates to assist researchers to clearly report core outcomes within their manuscripts (COMMIT-Reporting).

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 which 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 occur 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 address these important issues the COMMIT initiative has resolved to reach clear recommendations regarding the selection of the most appropriate denominator (Duffy et al., 2020b).

The Cochrane Gynaecology and Fertility Group have published over 100 systematic reviews evaluating potential treatments for infertility and has committed to implementing the core outcome set for infertility when new and updated reviews are being prepared. Secondary research, including pairwise meta-analyses, individual participant data meta-analyses and network meta-analyses, will be more influential when infertility trials routinely collect and report core outcomes.

The COMMIT initiative has committed to undertaking further research to assess the uptake and implementation of the core outcome set for infertility (COMMIT-Implementation). Objectively demonstrating the uptake of the core outcome set for infertility is important to quantify its contribution to improve the value of future research. Assessing the uptake of the core outcome set will be undertaken by examining registry records, published protocols, RCT and systematic reviews, and undertaking a citation analysis. 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 informed by implementation science will be developed to limit, and hopefully overcome, any perceived barriers.

The core outcome set reported in this study is intended to be used across trials evaluating a broad range of potential fertility treatments, for example, male endocrine stimulation protocols, lifestyle interventions for people with fertility problems, and methods for embryo selection during IVF cycles. Extensions to the current core outcome set are planned or currently in development for different patient populations, including men with fertility problems (COMMIT-Male Infertility), women with endometriosis (Duffy *et al.*, 2020c) and interventions including IVF (COMMIT-IVF). Other extensions are planned to ensure future infertility trials and systematic reviews routinely collect and report harms (COMMIT-Harms). Although quality of life was not selected as a core outcome, the COMMIT initiative has committed to undertaking a systematic review and methodological assessment of measurement instruments capable of measuring quality of life and will make recommendations to inform the design of future infertility trials (COMMIT-QoL).

This comprehensive strategy could make a significant contribution in reducing research waste across future fertility research. This approach has acted as a template for other areas of women's health seeking to tackle research waste, including twin and multiple pregnancy research (Townsend *et al.*, 2020b). The variation in outcome reporting and suspected outcome reporting bias has been characterized across women's and newborn health, including endometriosis, twin-twin transfusion syndrome and neonatal care. This study should inform the development of other core outcome sets seeking to tackle poorly selected, collected and reported outcomes (Hirsch *et al.*, 2016a,b; Perry *et al.*, 2018; Webbe *et al.*, 2020a,b).

Research priority setting presents an opportunity to develop a prioritized research agenda (Graham *et al.*, 2020). A research priority setting study has recently been completed for infertility and identified research priorities related to the prevention, diagnosis and treatment of male, female and unexplained infertility (Duffy *et al.*, 2020a). Undertaking an RCT is the only appropriate method to answer many of these research priorities (Wilkinson *et al.*, 2019b). Therefore, it is important for our specialty to work together to improve the design, delivery and reporting of future trials.

In summary, this study used formal consensus methods to develop a core outcome set for future RCT and systematic reviews evaluating potential treatments for infertility. Embedding the core outcome set within future infertility research could make a profound contribution to advancing the usefulness of research to inform clinical practice and enhance the care people with infertility problems receive.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Acknowledgements

We would like to thank the Delphi survey and 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., H.A., S.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., Y.K., R.S.L., J.M.L.K., D.M., B.W.J.M., B.L., S.L., C.N., E.H.Y.N., A.S.O., L.P., S.R.H., S.R., I.S., J.L.S., A.S., C.S., H.L.T., A.V., M.v.W., M.A.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., H.A., S.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., Y.K., R.S.L., J.M.L.K., D.M., B.W.J.M., B.L., S.L., C.N., E.H.Y.N., A.S.O., L.P., S.R.H., S.R., I.S., J.L.S., A.S., C.S., H.L.T., A.V., M.v.W., M.A.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., H.A., S.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., Y.K., R.S.L., J.M.L.K., D.M., B.W.J.M., B.L., S.L., C.N., E.H.Y.N., A.S.O., L.P., S.R.H., S.R., I.S., J.L.S., A.S., C.S., H.L.T., A.V., M.v.W., M.A.V., N.L.V., A.Y.W., R.W., J.W., M.A.Y. and C.M.F. Drafting of the manuscript: J.M.N.D. and C.M.F. Critical revision of the manuscript for important intellectual content: H.A., S.B., B.C., C.C., J.L.H.E., L.C.G., R.G.F., S.F., Y.K., R.S.L., J.M.L.K., D.M., B.W.J.M., B.L., S.L., C.N., E.H.Y.N., A.S.O., L.P., S.R.H., S.R., I.S., J.L.S., A.S., C.S., H.L.T., A.V., M.v.W., M.A.V., N.L.V., A.Y.W., R.W., J.W. and M.A.Y. Statistical analysis: J.M.N.D. 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. B.W.J.M. is supported by a National Health and Medical Research Council Practitioner Fellowship (GNT1082548). S.B. was supported by University of Auckland Foundation Seelye Travelling Fellowship.

Conflict of interest

S.B. 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*. J.M.L.K. reports research sponsorship from Ferring and Theramex. 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.J.M. reports consultancy fees from Guerbet, iGenomix, Merck, Merck KGaA and ObsEva. C.N. reports being the Co Editor-in-Chief of *Fertility and Sterility* and Section Editor of the *Journal of Urology*, research sponsorship from Ferring, and retains a financial interest in NexHand. A.S. reports consultancy fees from Guerbet. E.H.Y.N. reports research sponsorship from Merck. 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 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, Universtity of Pennsylvania, USA; Dr Yusuf Beebeejaun, King's Fertility, Fetal Medicine Research Institute, UK; Dr Sohinee Bhattacharya, University of Aberdeen, UK; Megan Black, New Zealand; Magdalena Bofill, University of Auckland, 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 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; Anita Fincham, Fertility Europe, Belgium; 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; Professor Martha Hickey, University of Melbourne, Australia; Dr Martin Hirsch, University College London Hospitals, UK; Dr Marcos Horton, Pregna Reproductive Medicine, Argentina; Associate Professor M. Louise Hull, University of Adelaide, Australia; Dr Shikha Jain, Dreamz IVF, India; Dr Marta Jansa Perez, UK; Dr Claire A. Jones, University of Toronto, Canada; Dr Vanessa Jordan, University of Auckland, New Zealand; Professor Mohan S. Kamath, Christian Medical College, Vellore, India; 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 Jian Li, Chinese University of Hong Kong, China; Professor Olabisi M. Loto, Obafemi Awolowo University, Nigeria; Karen L. Marks, UK; Alison R. McTavish, University of Aberdeen, UK; David J. Mills, UK; Dr Raju R. Nair, Mitera Hospital, India; Dr Dung Thi Phuong Nguyen, Vietnam; Professor Allan A. Pacey, University of Sheffield, UK; 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; Marian Showell, University of Auckland, New Zealand; Professor Charalampos S. Siristatidis, Greece; Dr Akanksha Sood, St. Mary's Hospital, UK; 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

Barnhart KT. Live birth is the correct outcome for clinical trials evaluating therapy for the infertile couple. *Fertil Steril* 2014;**101**:1205–1208.

Braakhekke M, Kamphuis EI, van Rumste MM, Mol F, van der Veen F, Mol BW. How are neonatal and maternal outcomes reported in randomised controlled trials (RCTs) in reproductive medicine? *Hum Reprod* 2014;**29**:1211–1217.

Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin JA *et al.* SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;**158**:200–207.

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.

Dapuzzo L, Seitz FE, Dodson WC, Stetter C, Kunselman AR, Legro RS. Incomplete and inconsistent reporting of maternal and fetal outcomes in infertility treatment trials. *Fertil Steril* 2011;**95**:2527–2530.

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, Adamson GD, Benson E, Bhattacharya S, Bhattacharya S, Bofill M, Brain K, Collura B, Curtis C, Evers JLH *et al.* Priorities for future infertility research: an international consensus development study. *Hum Reprod* 2020a;**35**:2715–2724.

Duffy JMN, Bhattacharya S, Bhattacharya S, Bofill M, Collura B, Curtis C, Evers JLH, Giudice LC, Farquharson RG, Franik S *et al.* Standardizing definitions for the infertility core outcome set: an international consensus development study. *Hum Reprod* 2020b;**35**:2735–2745.

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* 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* 2020c;**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* 2019b;**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* 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* 2017b;**124**:1481–1489.

Duffy JMN, van 't Hooft J, Gale C, Brown M, Grobman W, Fitzpatrick R, Karumanchi SA, Lucas N, Magee L, Mol B *et al.* A protocol for developing, disseminating, and implementing a core outcome set for pre-eclampsia. *Pregnancy Hypertens* 2016;**6**: 274–278.

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* 2019a;**220**: 71.e71–71.e74.

Graham L, Illingworth B, Showell M, Vercoe M, Crosbie E, Gingel L, Farquhar C, Horne A, Prior M, Stephenson J *et al.* Research priority setting in women's health: a systematic review. *BJOG* 2020; **127**:694–700.

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.

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* 2016a;**6**:e013998.

Hirsch M, Duffy JMN, Kusznir JO, Davis CJ, Plana MN, Khan KS, Duffy JMN, Farquhar C, Hirsch M, Johnson N *et al.* Variation in outcome reporting in endometriosis trials: a systematic review. *Am J Obstet Gynecol* 2016b;**214**:452–464.

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* 2020;**127**:983–992.

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 JMN, 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.

Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, Williamson PR. Core Outcome Set-STAndards for development: the COS-STAD recommendations. *PLoS Med* 2017;**14**:e1002447.

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.

Perry H, Duffy JMN, Reed K, Baschat A, Deprest J, Hecher K, Lewi L, Lopriore E, Oepkes D, Khalil A *et al.*; the International Collaboration to Harmonise Outcomes for Twin–Twin Transfusion Syndrome (CHOOSE). Core outcome set for research studies evaluating treatments for twin–twin transfusion syndrome. *Ultrasound Obstet Gynecol* 2019;**54**:255–261.

Perry H, Duffy JMN, Umadia O, Khalil A; the International Collaboration to Harmonise Outcomes for Twin-Twin Transfusion Syndrome (CHOOSE). Outcome reporting across randomized trials and observational studies evaluating treatments for twin–twin transfusion syndrome: systematic review. *Ultrasound Obstet Gynecol* 2018;**52**:577–585.

Townsend R, Duffy JMN, Khalil A. Increasing value and reducing research waste in obstetrics: towards woman-centered research. *Ultrasound Obstet Gynecol* 2020b;**55**:151–156.

Townsend R, Duffy JMN, Sileo F, Perry H, Ganzevoort W, Reed K, Baschat AA, Deprest J, Gratacos E, Hecher K *et al.*; International Collaboration to Harmonise Outcomes for Selective Fetal Growth Restriction (CHOOSE-FGR). Core outcome set for studies investigating management of selective fetal growth restriction in twins. *Ultrasound Obstet Gynecol* 2020a;**55**:652–660.

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

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

Webbe JWH, Ali S, Sakonidou S, Webbe T, Duffy JMN, Brunton G, Modi N, Gale C. Inconsistent outcome reporting in large neonatal trials: a systematic review. *Arch Dis Child Fetal Neonatal Ed* 2020b; **105**:69–75.

Webbe JWH, Duffy JMN, Afonso E, Al-Muzaffar I, Brunton G, Greenough A, Hall NJ, Knight M, Latour JM, Lee-Davey C *et al.* Core outcomes in neonatology: development of a core outcome set for neonatal research. *Arch Dis Child Fetal Neonatal Ed* 2020a; **105**:425–431.

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* 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.

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.