Accepted: 30 December 2022

DOI: 10.1111/1471-0528.17411

BJOG An International Journal of Obstetrics and Gynaecology

COMMENTARY

Assessing the usefulness of randomised trials in obstetrics and gynaecology

Janneke van 't Hooft[1,2] | **Charlotte E. van Dijk**[2] | **Cathrine Axfors**[1,3] | **Zarko Alfirevic**[4] | **Martijn A. Oudijk**[5] | **Khalid S. Khan**[6] | **Ben W. J. Mol**[7,8] | **Patrick M. Bossuyt**[9] | **John P. A. Ioannidis**[1,10,11,12,13]

[1]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Palo Alto, California, USA

[2]Department of Obstetrics and Gynaecology, Amsterdam Reproduction and Development Institute, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[3]Department for Women's and Children's Health, Uppsala University, Uppsala, Sweden

[4]Center for Women's Health Research, Liverpool Women's Hospital, Liverpool, UK

[5]Department of Obstetrics and Gynaecology, Amsterdam Reproduction and Development Institute, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands

[6]Department of Preventive Medicine and Public Health, University of Granada, and Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Granada, Spain

[7]Department of Obstetrics and Gynecology, Monash University, Melbourne, Victoria, Australia

[8]Aberdeen Centre for Women's Health Research, School of Medicine, University of Aberdeen, Aberdeen, UK

[9]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam AUMC, University of Amsterdam, Amsterdam, The Netherlands

[10]Department of Medicine, Stanford University, Palo Alto, California, USA

[11]Department of Epidemiology and Population Health, Stanford University, Palo Alto, California, USA

[12]Department of Biomedical Data Science, Stanford University, Palo Alto, California, USA

[13]Department of Statistics, Stanford University, Palo Alto, California, USA

Correspondence
Janneke van 't Hooft, Department of Obstetrics and Gynecology, Amsterdam Reproduction and Development Institute, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands.
Email: j.vanthooft1@amsterdamumc.nl

Funding information
ZonMw, Grant/Award Number: 40-45200-98-306

1 | INTRODUCTION

Healthcare professionals will be familiar with how little of what they read in medical journals is of direct value (useful) to their practice.[1] Randomised controlled trials (RCTs) are traditionally placed at the top of the hierarchy of evidence validity for healthcare interventions. However, even the majority of RCTs do not seem useful in that they may not lead to any tangible improvements in clinical decision making. Increasing the usefulness of RCTs can benefit society, providing better solutions for patient problems and reducing waste in medical research budgets.[2,3] In this commentary we discuss the usefulness of RCTs for healthcare interventions and propose a tool for structured assessment. We describe some examples of where RCTs have demonstrated usefulness in the field of women's health and some where they may have been less useful.

2 | WHY ANOTHER TOOL?

Many collaborative initiatives have developed checklists to address issues like bias and transparency in RCTs.[4,5] These tools, in their various versions, can help make judgements about the validity and reporting quality an RCT, but they fail to caputure whether the trial itself is

This article includes Author Insights, a video abstract available at: https://vimeo.com/785911255

useful or wasteful. A 2016 article highlighted the need to consider the usefulness of a study in the conception, design and planning phases.[6] Will a particular study, upon completion, have the potential to improve outcomes? This question matters for clinical researchers, their employers, funding bodies, ethics committees, patient organisations and prospective participants. For completed studies an assessment of usefulness before publication is of interest for journal editors and peer reviewers, even if the improvement of fundamental flaws in the design will be difficult, if not impossible, at this stage. For studies already published, assessments of usefulness matter for clinicians, patients and other stakeholders involved in healthcare provision and research, e.g. an assessment of usefulness of RCTs in a specific domain in medicine may offer feedback on where to improve future study design. The assessment of usefulness should be comprehensive (covering multiple aspects of usefulness), simple enough so that it can be routinely applied and unambiguous to avoid interpretations that are too subjective.[6] Here, we show how the proposed usefulness features (eight criteria with 13 items; Figure 1) can be operationalised for assessing RCTs evaluating healthcare interventions in obstetrics and gynaecology.

3 | PROBLEM BASE: IS THERE A HEALTH PROBLEM THAT IS BIG OR IMPORTANT ENOUGH TO FIX?

There is a weak or modest correlation between the volume of research done and the burden of various diseases.[7] For example, preterm birth is without doubt an important health problem, given the 10% incidence worldwide.[8] The method of induction of labour might have less impact on an individual woman and her child, but given the fact that around one in three pregnancies are induced, it is also important. This does not mean that rare diseases should be discarded. Answering questions about disease with low prevalence is still valuable if there is a large health impact. Eradicating or markedly decreasing the impact of an uncommon condition, e.g. neural tube defect via folic acid supplementation, may yield at least as much or more usefulness than achieving minor benefits for a more common problem. In general, the size of the health problem should be determined by the product of the individual health impact, the prevalence and the financial cost of that health problem.

4 | CONTEXT PLACEMENT: HAS PRIOR EVIDENCE BEEN SYSTEMATICALLY ASSESSED TO INFORM (THE NEED FOR) NEW STUDIES?

Useful clinical research adds relevant information to what we already know, i.e. findings identified by a systematic assessment of the existing evidence. Each new study should therefore be preceded by a systematic review. A new RCT report should not only present the data but also integrate the data with existing evidence in an updated meta-analysis.[9,10] The STOPPIT trial (Progesterone for the prevention of preterm birth in twin pregnancy), published in 2009, and also older studies like the ACT trial (Calcium supplementation in nulliparous women for the prevention of pregnancy-induced hypertension, preeclampsia and preterm birth), published in 1999, offer examples of such updated meta-analyses.[11,12]

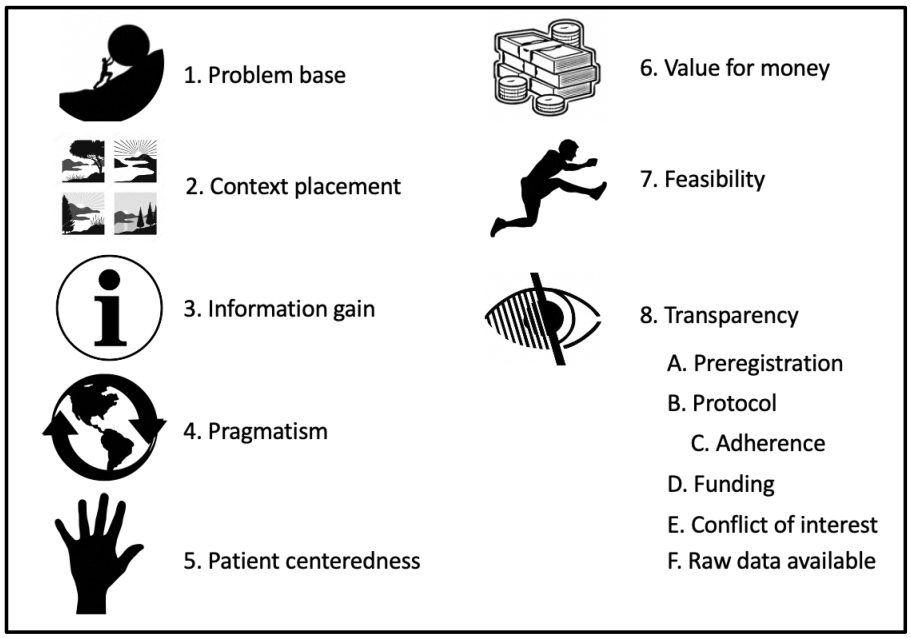

FIGURE 1 Usefulness features, including eight criteria with 13 items.

5 | INFORMATION GAIN: IS THE PROPOSED STUDY LARGE AND LONG ENOUGH TO BE SUFFICIENTLY INFORMATIVE?

Useful clinical research provides sufficient precision in the findings with respect to critical outcomes to inform practice. If a clinical study is underpowered or primarily evaluates surrogate outcomes, the information gain for changing practice is minimal. Furthermore, if a minimally important difference for a patient to accept a treatment is known, this will help each trial to target their sample size appropriately.

In this case, 'negative trials', if they are sufficiently powered, may be as useful as 'positive' trials: they can support patients and clinicians to reconsider interventions and empower policymakers to de-implement them. A gynaecologic example is the LUNA trial (laparoscopic uterosacral nerve ablation), randomising almost 500 women to the intervention or not, followed by individual participant data (IPD) meta-analysis of over 850 women, evaluating pain scores over a sufficient time horizon in patients suffering with chronic pelvic pain. Both studies convincingly conclude that LUNA does not improve the outcomes for these patients.[13,14] This practice has therefore been abandoned today.[15]

An example of minimal information gain in obstetrics can be found in the group of individual studies evaluating progesterone in women with a short cervix. Many of these 'trials' individually fail to show any benefit. Two trials did not reach their target sample size[16,17]; four trials did not mention a sample size and power calculation and are probably underpowered.[18–21] When combined in an IPD meta-analysis, there is convincing evidence of the benefit of progesterone in these women.[22]

To some extent, existing previous research decides how informative a trial can be. When previous information is scarce, a small trial may function as a pilot for a more informative, larger trial, even when it is not expected to be sufficiently informative on its own.

It is important to avoid the use of surrogate outcomes as they can lack clinical insight.[6] Examples of surrogate outcomes for fertility treatments are the use of embryo quality or ovarian response for in vitro fertilisation (IVF) and the use of ovulation frequency in the treatment for polycystic ovary syndrome (PCOS).[23] Those outcomes cannot be equated with the true outcome of interest, which is a healthy liveborn child. Similar examples are seen in obstetrics for induction of labour trials. A large majority of these trials measure 'delivery within 24 hours' as a primary outcome, whereas the key goal of induction of labour is to achieve vaginal birth (otherwise caesarean section could be performed immediately) of a healthy newborn.[24]

Composite outcomes can increase the power of the study, but can be double-edged swords. For instance, a potential drawback is that they can include opposite effects in the components, possibly leading to erroneous effect estimates.[25] Composite outcomes work best when an intervention that is anticipated to reduce a morbidity measure is also expected to improve survival, but this correlation may not always be present. In the SUPPORT trial,[26] a study investigating target ranges of oxygen saturation (a lower range of 85%–89% vs a higher range of 91%–95%) in extremely preterm infants, the results showed no evidence of a difference in the composite outcome. However, the lower target range of oxygenation resulted in an increase in mortality and a substantial decrease in severe retinopathy of prematurity among survivors. For these cases a composite 'survival without disability' may have some value.[27] Another example of a universally accepted composite outcome is the measure of quality-adjusted life years (QALY).

6 | PRAGMATISM: DOES THE RESEARCH REFLECT REAL LIFE? IF IT DEVIATES, DOES THIS MATTER?

To explain pragmatism, the definitions of *efficacy* and *effectiveness* are relevant. Efficacy denotes both the beneficial and harmful effects of an intervention when it is applied under ideal circumstances, whereas effectiveness denotes the beneficial and harmful effects when applied under the usual circumstances that apply in health care (i.e. pragmatism). Both efficacy trials (traditionally phase 2) and effectiveness/pragmatic trials (traditionally phase 3) are important in the development and licensing of drugs or other new interventions. However, most phase 3 trials are still designed in a highly controlled and monitored setting, making them less useful for clinical practice, as they still explore mostly efficacy rather than effectiveness. If efficacy has been demonstrated, the introduction of this intervention in a real-life setting may be influenced by other factors, modifying the actual effect of the intervention. This could be the case in studies evaluating surgical interventions (e.g. cervical cerclage) in which the experience and learning curve of surgeons may differ between the well-controlled explanatory trial and the pragmatic trial.

Pragmatic trials are designed to evaluate effectiveness in the real world. Their features are described by the PRECIS (pragmatic explanatory continuum indicator summary) group.[28] In general, pragmatic studies are multi-centred studies, unblinded, not placebo controlled and not addressing a new intervention or new indication. The inclusion criteria (often very limited) are designed for generalisability, with intention-to-treat analysis of critical outcomes.[29]

Today, many trials have a long list of inclusion and exclusion criteria, potentially jeopardising not only recruitment but also the generalisability of the results. For example, in many IVF trials inclusion is often limited to participants in their first IVF attempt, below 35 years of age and with a lower BMI, which represents probably less than half of the population requesting IVF. In general, every candidate for whom randomisation to both treatments is ethical and relevant should be eligible for a pragmatic study.

7 | PATIENT CENTREDNESS: DOES THE RESEARCH REFLECT TOP PATIENT PRIORITIES?

Useful clinical research should be patient centred. This approach is designed to benefit patients through preserving health and improving well-being, and not to focus on the needs of physicians, investigators or sponsors. Including patients in research prioritisation processes and in the development of core outcomes can help address research questions that are important to patients, using outcomes that are relevant to them. Nevertheless, one should be aware that biases can also affect such prioritisation exercises.

In the case of preterm birth studies, a research priority list of 15 research questions related to preterm birth has already been developed through the James Lind Alliance.[30] This nonprofit organisation brings patients, caretakers and clinicians together to identify and prioritise unanswered questions or uncertainties in the evidence that they agree are the most important. Also, women experiencing preterm birth have been involved in the development of a 'core outcome set' for preterm birth prevention studies, listing 13 crucial outcomes that should be listed in all preterm birth studies.[31] The essential step is now the implementation of the prioritised research questions and core outcome sets in future preterm birth trials. To date there are more than 75 'core outcome sets' in development or developed in the field of obstetrics and gynaecology, 23 of which are published (for the list of published core outcome sets, see Appendix S1; www.comet-initiative.org).

8 | VALUE FOR MONEY: IS THE RESEARCH WORTH THE MONEY?

Useful clinical research considers the value of the information under investigation at the design and planing phase of the study. How much can we learn from the study and does this offset the cost of performing it? Especially in an era of limited resources, a value-of-information analysis or budget impact calculation is useful before starting the trial.[32,33] A value-of-information analysis is a method to provide insights on the expected benefits from clinical research by characterising the uncertainty of the effects of interventions on health outcomes.[32] This information is then used to inform decisions about the design and priority of those studies. A budget impact calculation estimates a difference in healthcare costs before and after implementing the research findings (in case they were found to be effective).[33] Both value-of-information and budget impact analyses can be performed when designing a trial, and therefore crucially differ from cost-effectiveness analysis (performed after knowing study results). Funders like ZonMw in the Netherlands and the National Institute for Health and Care Research (NIHR) in the UK ask for budget impact analysis in the grant applications. Presenting this information to ethics committees and also in the protocol of a final article can help to raise awareness about the resources invested to target a specific health problem.

9 | FEASIBILITY: CAN THIS RESEARCH BE DONE?

A useful trial adds meaningfully to existing knowledge when, among other factors, it has sufficient power and reaches its calculated sample size. Unwarranted optimism among investigators and funders can lead to a trial that is unfeasible being stopped earlier than intended, but this can be hard to predict. An example of this is the PLUTO trial (Percutaneous vesicoamniotic shunting versus conservative management for fetal lower urinary tract obstruction).[34] In this trial, seven of the 21 participating centres recruited 31 women over a period of 4 years, instead of the planned sample size of 150 women. Among the reasons given, the authors mentioned the higher than expected proportion of parents choosing termination of pregnancy, a lower prevalence of disease than reported in the scientific literature and a high proportion of parents and clinicians choosing to enter a registry, rather than be randomly assigned to a trial group.[34]

This phenomenon is already described by Louis C. Lasagna in 1979, with Lasagna's law stating that 'the incidence of patients availability sharply decreases when a clinical trial begins', with empirical evidence demonstrating that this can range from one-tenth to one-third of what was originally estimated.[35] Performing pilot feasibility studies beforehand, whether or not randomised, can help address this overestimation and function as a stepping stone for a larger clinically relevant trial.

In the last decade we have seen many strategies to improve recruitment speed in the field of women's health. Maybe the most crucial one is the formation of different research networks and consortia, e.g. the Maternal Fetal Medicine Units Network (Eunice Kennedy Shriver National Institute of Child Health and Human Development, NICHD) and the Obstetric Collaborative Research Network in the USA, Maternal–Fetal Medicine Research in Canada, the NIHR Comprehensive Local Research Network (CCRN), the Interdisciplinary Maternal Perinatal Australasian Clinical Trial in Australia and New Zealand (IMPACT), the Groupe de Recherche and Obstétrique et Gynécologie (GROG) in France, the Dutch Consortium for Healthcare Evaluation and Research in Obstetrics and Gynaecology (NVOG consortium), the Perinatal Italian Network (PIN) and the Hong Kong Maternal Fetal Medicine Network.

The NICHD, for example, collaborates with centres that have proven to recruit patients successfully, based on their track record and based on the demographic that increased the rate of eligible patients. They also pay research staff to screen and enrol patients. This is a common strategy for many consortia. The French consortium GROG organises four full-day meetings a year, well attended by all level-3 maternity centres in France. These meetings provide an opportunity to improve study proposals/protocols and to find

centres that would like to participate in case a study is selected for funding. In the Netherlands, regional and national protocols are adjusted to the default of 'no treatment' in case there is a real equipoise.[36]

10 | TRANSPARENCY: ARE METHODS, DATA AND ANALYSES VERIFIABLE AND UNBIASED IN TERMS OF TRANSPARENCY?

Useful clinical research should be transparent in all possible ways throughout its life cycle.

1. *Preregistration*. Defined as any registration in an official registry database (e.g. www.clinicaltrials.gov) prior to the date of recruitment of the first patient.
2. *Protocol and data analysis plan made publicly available prior to trial commencement and prior to analysis of the data*. This can be done in an official journal, but also in other (free-of-charge) platforms like the Open Science Framework registry (www.osf.io). When submitted with manuscripts, these should be supplied in unredacted form.
3. *Protocol adherence or modification statements*. Trials are not always completed as planned. In case authors decide not to adhere to the original protocol, e.g. when useful modifications are advised by an independent trial steering committee (adaptive design exists for exactly this reason), the changes should be justified and protocol updates subjected to ethics committee reapproval. Authors should provide modification statements in their manuscript.
4. *Funding and conflict of interest disclosures*. There should be transparency in funding sources for the study (private/industry, public, not-for profit) and any conflicts of interests.
5. *Freely available raw data*. Sharing raw data, with statistical code and output, ideally throughout the lifecycle of the clinical trial, is set to become the new norm as the International Committee of Medical Journal Editors (ICMJE) has recommended this and it is slowly permeating into journal instructions for authors. This transparency item is not only relevant in facilitating individual participant data meta-analyses but will also help to identify studies with serious concerns about trustworthiness. An important issue that requires urgent attention, as falsified and fabricated studies have the potential to cause serious harm.[37]

The association of transparency issues and study findings was also illustrated in an example on progesterone for preterm birth prevention by Thornton and his team.[38] This study identified 93 RCTs and 29 systematic reviews and found a remarkable difference in the reported effectiveness of progesterone when evaluating the subset of trials reporting a preregistered primary outcome only ($n = 22$), compared with the totality of trials and reviews. This example illustrates that transparency in research is probably a necessary condition for more useful RCTs.

11 | CONCLUSION

Overall, this multidimensional assessment of RCT usefulness may help map the strengths and weaknesses of a large field. Applying these usefulness criteria in the planning and designing phase of an RCT evaluating healthcare interventions can help researchers and other stakeholders to focus their research question on a design with the most useful strategy. Usefulness assessment can help grant assessors and ethics committees to provide structural feedback in a phase when research ideas can still be changed and improved. Subsequently, it can help peer reviewers, journal editors, ethics committees, systematic reviewers, guideline developers and policymakers and, most importantly, practicing physicians and patients. This tool has no intention to restrict individual researchers' space to innovate and investigate novel approaches. In the end, a day without randomisation is a day without progress.

AUTHOR CONTRIBUTIONS

All authors contributed substantially to the design, content and review of the manuscript of this commentary.

ACKNOWLEDGEMENTS

None.

FUNDING INFORMATION

The study was funded by a grant from the Netherlands Organization for Health Research and Development (ZonMw Rubicon grant #452182306). The funder had no involvement in any phase of this study.

CONFLICT OF INTEREST STATEMENT

Meta-Research Innovation Center at Stanford (METRICS), Stanford University, is supported by a grant from the Laura and John Arnold Foundation. JvtH is supported by a postdoctoral grant from the Netherlands Organization for Health Research and Development (Rubicon grant 452182306). CA is supported by postdoctoral grants from the Knut and Alice Wallenberg Foundation (KAW 2019.0561), Uppsala University, and the Sweden–America Foundation. BWJM is supported by a National Health and Medical Research Council (NHMRC) investigator grant (GNT1176437). BWJM reports consultancy for Guerbet, has been a member of the ObsEva advisory board and holds stock options for ObsEva. The work of JPAI has been supported by an unrestricted gift from Sue and Bob O'Donnell. All the funders mentioned above had no role in the study design, data collection and analysis, decision to publish, or preparation of the article. Completed disclosure of interests form available to view online as supporting information.

ETHICS APPROVAL

None.

ORCID

Janneke van 't Hooft https://orcid.org/0000-0001-5303-1503
Charlotte E. van Dijk https://orcid.org/0000-0001-9417-4180
Cathrine Axfors https://orcid.org/0000-0002-2706-1730

Zarko Alfirevic https://orcid.org/0000-0001-9276-518X
Martijn A. Oudijk https://orcid.org/0000-0001-8672-4365
Khalid S. Khan https://orcid.org/0000-0001-5084-7312
Ben W. J. Mol https://orcid.org/0000-0001-8337-550X
Patrick M. Bossuyt https://orcid.org/0000-0003-4427-0128
John P. A. Ioannidis https://orcid.org/0000-0003-3118-6859

REFERENCES

1. Hayes B, Hayward R, Lomas J. Bridges between evidence and clinical practice. J Am Med Inform Assoc. 1995;2(6):342–50.
2. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, et al. Biomedical research: increasing value, reducing waste. Lancet. 2014;383(9912):101–4.
3. Hackbarth AD. Eliminating waste in US Health Care. JAMA. 2012;307(14):1513.
4. Pandis N, Chung B, Scherer RW, Elbourne D, Altman DG. CONSORT 2010 statement: extension checklist for reporting within person randomised trials. BMJ. 2017;357:j2835.
5. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:14898.
6. Ioannidis JPA. Why most clinical research is not useful. PLoS Med. 2016;13(6):e1002049.
7. Viergever RF, Terry RF, Karam G. Use of data from registered clinical trials to identify gaps in health research and development. Bull World Health Organ. 2013;91(6):416–25C.
8. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller AB, et al. Born too soon: the global epidemiology of 15 million preterm births. Reprod Health. 2013;10(S1):S2.
9. Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. J R Soc Med. 2007;100:187–90.
10. Heneghan C, Aronson JK. Why reports of clinical trials should include updated meta-analyses. BMJ Evid Based Med. 2020;25(1):1–2.
11. Norman JE, Mackenzie F, Owen P, Mactier H, Hanretty K, Cooper S, et al. Progesterone for the prevention of preterm birth in twin pregnancy (STOPPIT): a randomised, double-blind, placebo-controlled study and meta-analysis. Lancet. 2009;373:2034–40.
12. Crowther CA, Hiller JE, Pridmore B, Bryce R, Duggan P, Hague WM, et al. Calcium supplementation in nulliparous women for the prevention of pregnancy-induced hypertension, preeclampsia and preterm birth: an Australian randomized trial. Aust N Z J Obstet Gynaecol. 1999;39(1):12–8.
13. Daniels J, Gray R, Hills RK, Latthe P, Buckley L, Gupta J, et al. Laparoscopic uterosacral nerve ablation for alleviating chronic pelvic pain. JAMA. 2009;302(9):955–61.
14. Daniels JP, Middleton L, Xiong T, Champaneria R, Johnson NP, Lichten EM, et al. Individual patient data meta-analysis of randomized evidence to assess the effectiveness of laparoscopic uterosacral nerve ablation in chronic pelvic pain. Hum Reprod Update. 2010;16(6):568–76.
15. National Institute for Health and Care Excellence. Laparoscopic uterine nerve ablation (LUNA) for chronic pelvic pain [cited 2007 Oct 30]. Available from: www.nice.org.uk/guidance/ipg234
16. Crowther CA, Ashwood P, McPhee AJ, Flenady V, Tran T, Dodd JM, et al. Vaginal progesterone pessaries for pregnant women with a previous preterm birth to prevent neonatal respiratory distress syndrome (the PROGRESS Study): a multicentre, randomised, placebo-controlled trial. PLoS Med. 2017;14(9):e1002390.
17. van Os M, van der Ven A, Kleinrouweler C, Schuit E, Kazemier B, Verhoeven C, et al. Preventing preterm birth with progesterone in women with a short cervical length from a low-risk population: a multicenter double-blind placebo-controlled randomized trial. Am J Perinatol. 2015;32(10):993–1000.
18. Azargoon A, Ghorbani R, Aslebahar F. Vaginal progesterone on the prevention of preterm birth and neonatal complications in high risk women: a randomized placebo-controlled double-blind study. Int J Reprod Biomed. 2016;14(5):309–16.
19. Ahuja R, Sood A, Pal A, Mittal R. Role of micronized progesterone in prevention of preterm labour in women with previous history of one or more preterm births: a research study at a tertiary care hospital. Int J Reprod Contracept Obstet Gynecol. 2015;4:1176–80.
20. Jabeen S, Akhtar M, Fatima N, Akram M. Role of progesterone for the prevention of preterm labour. Pakistan J Med Health Sci. 2012;6(1):253–5.
21. EL-Gharib MN, EL-Hawary TM. Matched sample comparison of intramuscular versus vaginal micronized progesterone for prevention of preterm birth. J Matern Fetal Neonatal Med. 2013;26(7):716–9.
22. Stewart LA, Simmonds M, Duley L, Llewellyn A, Sharif S, Walker RA, et al. Evaluating Progestogens for Preventing Preterm birth International Collaborative (EPPPIC): meta-analysis of individual participant data from randomised controlled trials. Lancet. 2021;397(10280):1183–94.
23. Legro RS. Surrogate end-points or primary outcomes in clinical trials in women with polycystic ovary syndrome? Hum Reprod. 2004;19(8):1697–704.
24. Swarnamani K, Smits LS, Palmer K, Mol BW, Davies-Tuck M. Is cesarean section the right outcome for induction of labor trials? Impact of sample size and primary outcomes. Ultrasound Obstet Gynecol. 2020;56(5):645–6.
25. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA. 2003;289(19):2554.
26. Target ranges of oxygen saturation in extremely preterm infants. N Engl J Med. 2010;362(21):1959–69.
27. Parekh SA, Field DJ, Johnson S, Juszczak E. Accounting for deaths in neonatal trials: is there a correct approach? Arch Dis Child Fetal Neonatal Ed. 2015;100(3):F193–7.
28. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009;62(5):464–75.
29. Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. Criteria for distinguishing effectiveness from efficacy trials in systematic reviews. In: . Technical review 12 (prepared by the RTI-International–University of North Carolina Evidence-based Practice Center under contract no 290-02-0016) AHRQ publication no 06-0046. Rockville: Agency for Healthcare Research and Quality; 2006.
30. Uhm S, Crowe S, Dowling I, Oliver S. The process and outcomes of setting research priorities about preterm birth – a collaborative partnership. Inf Dent. 2014;10(6):178–81.
31. van 't Hooft J, Duffy JMN, Daly M, Williamson PR, Meher S, Thom E, et al. A core outcome set for evaluation of interventions to prevent preterm birth. Obstet Gynecol. 2016;127(1):49–58.
32. Meltzer DO, Hoomans T, Chung JW, Basu A. Minimal modeling approaches to value of information analysis for health research. Med Decis Making. 2011;31(6):E1–22.
33. Mauskopf JA, Earnshaw S, Mullins CD. Budget impact analysis: review of the state of the art. Expert Rev Pharmacoecon Outcomes Res. 2005;5(1):65–79.
34. Morris RK, Malin GL, Quinlan-Jones E, Middleton LJ, Hemming K, Burke D, et al. Percutaneous vesicoamniotic shunting versus conservative management for fetal lower urinary tract obstruction (PLUTO): a randomised trial. Lancet. 2013;382(9903):1496–506.
35. Knottnerus JA, Tugwell P. Prevention of premature trial discontinuation: how to counter Lasagna's law. J Clin Epidemiol. 2016;80:1–2.
36. Klumper J, Breebaart W, Roos C, Naaktgeboren CA, van der Post J, Bosmans J, et al. Study protocol for a randomised trial for atosiban versus placebo in threatened preterm birth: the APOSTEL 8 study. BMJ Open. 2019;9(11):e029101.

37. Papageorghiou AT. Welcome to 2022. BJOG. 2022;129(1):1–2.
38. Prior M, Hibberd R, Asemota N, Thornton J. Inadvertent P-hacking among trials and systematic reviews of the effect of progestogens in pregnancy? A systematic review and meta-analysis. BJOG. 2017;124(7):1008–15.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.