Value of clinical research: Usefulness tool development and systematic review of 350 randomised controlled trials in preterm birth

Janneke van t Hooft[1], Charlotte van Dijk[1], Cathrine Axfors[2], Noah Haber[2], Zarko Alfirevic[3], Martijn Oudijk[1], Ben Mol[4], Patrick Bossuyt[1], and John P. A. Ioannidis[2]

[1]Amsterdam UMC Locatie AMC
[2]Stanford University
[3]University of Liverpool
[4]Monash University Central Medical School

June 22, 2021

Abstract

Objective: We developed a research usefulness tool collating published criteria and examined if randomised controlled trials (RCTs) addressing preterm birth were useful. Search Strategy: Cochrane library. Selection Criteria: Published RCTs within 56 preterm birth Cochrane reviews. Data Collection and Analysis: A usefulness tool was developed with eight criteria combining 13 items identified through literature searches and consensus. RCTs were evaluated for compliance with each item by multiple assessors (reviewer agreement 95-98%). Proportions with 95% confidence interval (CI) were calculated and compared for change over time using [?] 2010 as a cut-off, with relative risks (RR). Main Results: Among 350 selected RCTs, only 38 (11%, 95% CI 8-15%) met half of the usefulness criteria. Compared to trials before 2010, recent trials used composite or surrogate (less informative) outcomes more often (13% vs 25%, RR 1.87, 95% CI 1.19-2.93). Only 17 trials reflected real life (pragmatism) in design (5%, 95% CI 3-8%), with no improvements over time. No trials reported involvement of mothers to reflect patients' top priorities in question definition or outcomes selection. Recent trials were more transparent with prospective registration (0.5% vs 28%, RR 58, 95% CI 8-420%), availability of protocol (0.5% vs 15%, RR 32, 95% CI 4-237%) and data sharing statements (2% vs 8%, RR 3, 95% CI 1-10%). Conclusion: Clinical trials in preterm birth lacked many usefulness features, with one tenth of trials meeting half of the items evaluated. Use of informative outcomes, patient centeredness, pragmatism and transparency should be key targets for future research planning.

Title Page

Value of clinical research: Usefulness tool development and systematic review of 350 randomised controlled trials in preterm birth

Janneke van 't Hooft (postdoctoral fellow, 0000-0001-5303-1503)[1,2], Charlotte E. van Dijk (MD, 0000-0001-9417-4180)[2], Cathrine Axfors (postdoctoral fellow, 0000-0002-2706-1730)[1,3], Noah A. Haber (postdoctoral fellow, 0000-0002-5672-1769)[1], Zarko Alfirevic (professor, 0000-0001-9276-518X)[4], Martijn A. Oudijk (0000-0001-8672-4365)[2], Ben WJ Mol (professor, 0000-0001-8337-550X) [5,6], Patrick M. Bossuyt (professor, 0000-0003-4427-0128)[7], John P.A. Ioannidis (professor, 0000-0003-3118-6859)[1,8].

1. Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Palo Alto, CA, USA.
2. Department of Obstetrics and Gynecology, 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, United Kingdom.
5. Department of Obstetrics and Gynecology, Monash University, Melbourne, Victoria, Australia.
6. Aberdeen Centre for Women's Health Research, School of Medicine, University of Aberdeen, Aberdeen, UK.
7. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam AUMC, University of Amsterdam, Amsterdam, The Netherlands.
8. Departments of Medicine, of Epidemiology and Population Health of Biomedical Data Science, and of Statistics, Stanford University, Palo Alto, CA, USA.

Corresponding author: Janneke van 't Hooft, Department of Obstetrics and Gynecology, Amsterdam Reproduction and Development Institute, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, the Netherlands. J.vanthooft@gmail.com.

T:0031-613477818

Running title: Usefulness of clinical research in preterm birth

Abstract

Objective: We developed a research usefulness tool collating published criteria and examined if randomised controlled trials (RCTs) addressing preterm birth were useful.

Search Strategy: Cochrane library.

Selection Criteria: Published RCTs within 56 preterm birth Cochrane reviews.

Data Collection and Analysis: A usefulness tool was developed with eight criteria combining 13 items identified through literature searches and consensus. RCTs were evaluated for compliance with each item by multiple assessors (reviewer agreement 95-98%). Proportions of compliances with 95% confidence interval (CI) were calculated and compared for change over time using [?] 2010 as a cut-off, with relative risks (RR).

Main Results: Among 350 selected RCTs, only 38 (11%, 95% CI 8-15%) met half of the usefulness criteria. Compared to trials before 2010, recent trials used composite or surrogate (less informative) outcomes more often (13% vs 25%, RR 1.87, 95% CI 1.19-2.93). Only 17 trials reflected real life (pragmatism) in design (5%, 95% CI 3-8%), with no improvements over time. No trials reported involvement of mothers to reflect patients' top priorities in question definition or outcomes selection. Recent trials were more transparent with prospective registration (0.5% vs 28%, RR 58, 95% CI 8-420%), availability of protocol (0.5% vs 15%, RR 32, 95% CI 4-237%) and data sharing statements (2% vs 8%, RR 3, 95% CI 1-10%).

Conclusion: Clinical trials in preterm birth lacked many usefulness features, with one tenth of trials meeting half of the items evaluated. Use of informative outcomes, patient centeredness, pragmatism and transparency should be key targets for future research planning.

Funding: ZonMw grand #452182306

Keywords: Randomised Controlled Trials, Preterm Labour: clinical research

Tweetable abstract: Among 350 RCTs in preterm birth many lacked usefulness features like informative outcomes and transparency.

Manuscript

Introduction

Research varies in value for clinical practice, with substantial wastage in the evidence pipeline.(1,2) Most current tools focus on reporting,(3,4) which helps readers in critical appraisal after the research has been conducted and analysed. However, this does not capture whether research is useful or not in the first place. Fundamentally, usefulness needs to be considered at the design phase for the reported outputs to be of potential benefit to patients.

A seminal paper in 2016 (5) highlighted the need to capture usefulness at the question, design and planning phase of the study. This way, it will be possible to know prospectively whether undertaking a particular study will upon completion have the potential to improve outcomes.

A search in Medline and Google Scholar (Nov 2019) using the terms 'Usefulness' and 'Randomised Controlled Trial' provided 15,473 results. After screening the most recent 1500 results, we found only one publication discussing Usefulness in medical research (5) but no other publication discussing the usefulness of the randomised controlled trials (RCTs) themselves, rather the usefulness of the intervention under study. Narrowing the search to 'Usefulness' and 'randomised controlled trials' and 'preterm birth*' provided no results.

Here, we aimed to develop a research usefulness tool collating published criteria and examined if RCTs addressing preterm birth were useful. By quantifying these eight criteria consisting of 13 items, we aimed to get insights into the current status of research regarding its usefulness, the evolution of its usefulness over time, and where potential improvement is needed. Preterm birth (PTB) is a biomedical research topic that has a major global health burden (6,7) and has a fairly mature track record of research being performed over many decades. Reducing PTB is a top health priority and there is a large number of RCTs focusing on PTB prevention published to date. However, these trials have been able to provide only limited guidance on how to deal best with PTB.(8) Mapping their usefulness would be helpful for understanding the status quo and developing future PTB trials.

Methods

Study sample

Our protocol and details on the search strategy can be found in PROSPERO (CRD42019153728). We used the search strategy of a previously published Cochrane umbrella review, systematically assessing all Cochrane reviews evaluating interventions to prevent PTB in pregnancy (Supporting information Box S1), performed on November 2nd2017.(8) We updated this search on November 14th, 2019. Reviews were included if they prespecified or reported PTB as an outcome, with PTB defined as birth before 37 weeks' gestation. The population studied in this project are pregnant women with a singleton or multiple pregnancy without signs of preterm labour or ruptured amniotic membranes and irrespective of risk status for PTB or co-morbidities. Interventions that assessed PTB as an indirect effect of their intervention (e.g. insulin treatment vs metformin in diabetic pregnant women) were included. Interventions to prevent miscarriage were not included.

Progesterone is the most studied drug intervention on prevention of PTB. However, we found no update of the 2013 Cochrane review on this subject in singleton pregnant women,(9) therefore we decided to include the trials included in the most recent Individual Participant Data (IPD) meta-analysis project on progesterone (EPPPIC) as we assumed that this project might be the reason for the lack of update from Cochrane.(10)

We included RCTs regardless of language, provided that full-text articles were available. RCTs that did not explicitly report PTB or gestational age at delivery as an outcome were not included. We focused on RCTs included in Cochrane reviews, since these trials have already been systematically searched and screened by Cochrane reviewers.

One reviewer screened the titles and abstracts of all retrieved reports to exclude any obvious reports of non-eligible trials. A copy of the full article was then obtained for all non-excluded reports.

Data extraction

For each trial, we extracted the journal name, dates of publication and extracted general characteristics reported in the trial (e.g., number of centres involved, geographic region, sample size, power calculations, reporting of primary outcomes, number of women randomised, number of women analysed, trial conclusions, P values reported, DSMB involvement, ethics approval and funding) and details on the study group (i.e., gestational age of the pregnant women), intervention, comparison and PTB outcome.

Three reviewers (JH, LD, CA) independently assessed the full text using a standardized data extraction form developed in a pilot of 50 RCTs. After duplicate extraction of 20 RCTs by reviewer JH and LD and 20 RCTs by reviewer JH and CA the overall inter-agreement was 98% and 95%, with the item 'information gain' scoring the lowest inter-agreement (93%). Any discrepancies were resolved by consensus and discussion with a fourth team member (JI). After a second round of 40 duplicate extractions the inter-agreement was stable at 98% with improvement for the item 'information gain' to 95%. Any ambiguous item during the remaining data extractions was discussed in detail within the group.

Development of the usefulness tool

We search in Medline and Google Scholar (March 2021) using the terms 'Usefulness' and 'Randomised Controlled Trial' to identify items that are important to assess usefulness in clinical research. Of the 15,473 results we identified only one paper highlighting usefulness of clinical research itself (other identified papers commented on usefulness of an intervention studied in the trial). This identified paper on usefulness, elaborated on eight specific criteria. (5) These eight features were then discussed within a steering group consisting of two epidemiologist (PB, JI) and four clinicians working in the field of obstetrics and gynaecology (JH, ZA, MO, BWM). After discussion and consent of all group members, a usefulness criteria appraisal checklist consisting of eight criteria combining 13 items was developed (Box S2).

Assessment of clinical research value using the usefulness tool

In the usefulness criteria checklist, one question has to be answered for each feature, except for the second and last criterion, for which there are two and six separate questions to be answered (Box S2). Details on how we have operationalised the eight features can be seen in in supporting information (Box S3).

Statistical analysis

Most analyses are descriptive and report proportions (with their 95% confidence interval [CI] using Wilson score) percentages, medians (interquartile range, IQR) or means (standard deviation, SD). We performed three prespecified subgroup analyses. First, we compared the eight individual usefulness assessments in studies reported before 2000, between 2000-2009 and between 2010-2019. These time cut-offs were chosen as we expected a clear improvement on at least the transparency features due to a wide range of initiatives that started after the millennium. The trial register clinicaltrials.gov was launched in 2000 and the International Committee of Medical Journal Editors (ICMJE) started demanding registration of trials before submission to ICMJE journals in 2005; icmje.org). The comparative analysis compared the most recent time frame ([?]2010) with the time frame before that (<2010) using risk ratios (RRs) with 95% CI. Second, we assessed the impact factor (IF) of journals using the Web of Science Journal Citation Report 2019 and pre-specified a 'high' IF at a cut-off of [?]4.0 and 'low' IF when <4.(15) This cut-off was chosen to include the top specialty journals in obstetrics and gynaecology in the 'high' group, that generally have an IF between 4 and 6 (e.g., Obstetrics & Gynecology, British Journal of Obstetrics and Gynaecology). We expected the most recent RCTs ([?]2010) published in high impact journals to perform better in some of the usefulness features compared to those in low impact journals. Third, we stratified usefulness features by whether or not there was a statistically significant result (P<0.05) in (at least one) primary outcome. Finally, we counted for each trial for how many of the 13 usefulness items it scored 'positive' to provide an overview on the number of trials meeting at least half (7/13) of the items.

Role of funding source

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

Results

We identified 57 eligible Cochrane systematic reviews and 1 IPD meta-analysis focussing on primary or secondary PTB prevention in pregnant women containing 373 potentially eligible RCTs (Figure 1). These Cochrane reviews were published between 2006 and 2019 (median year 2017). From the 373 eligible RCTs,

we were able to include 350 RCTs for data extraction (Figure 1) coming from 56 reviews (See Table S1 for overview of SRs and RCTs included). These 350 RCTs were published between 1967 and 2019, with a gradual increase in publications over time untill 2015, followed by lower numbers of studies in the years 2016-2019 due to an expected delay in the uptake of RCTs in systematic reviews (Figure S1). The 350 RCTs randomised a total of 400,903 participants in all continents, with a higher cumulative number of trials performed in North America (n=92 trials, n=82,241 randomised women) and Europe (n=93 trials, n=56,653 women) and an increase of trials coming from Asia (n=71 trials, n=114,573 women) with a steep rise after the millennium (Figure S1).

There were 108 RCTs published before 2000 (including 122,742 randomised women), 104 RCTs between 2000-2009 (including 118,928 randomised women) and 138 RCTs between 2010-2019 (including 159,233 randomised women) showing no significant difference in randomised women per published study between the three time frames. General characteristics of the total RCTs sample and stratified for publications between the three time frames are shown in Table 1. Some potential increase over time was seen in the number of multicenter trials (93/212 [43.9%] before 2010 and 74/138 [53.6%] [?]2010; RR 1.22, 95%CI 0.98-1.52, with a stable median of 5 to 6 participating centers). A substantial increase was observed in the reporting of primary outcomes (139/212 [65.6%] before 2010 and 120/138 [87%] [?]2010; RR 1.33, 95%CI 1.18-1.49), power calculations (141/212 [66.5%] before 2010 and 109/138 [79.0%] [?]2010; RR 1.19, 95%CI 1.04-1.35) and use of a Data Safety Monitoring Board (43/212 [20.3%] before 2010 and 42/138 [30.4%] [?]2010; RR 1.50, 95%CI 1.04-2.17) (Table 1). Stratification by low/high impact factor and by publication year before 2010 and [?]2010 among high impact factor publications are shown in supplementary material (Table S2).

The median sample size in the 350 trials was 224 (IQR 106 to 803; Table 1). Median inclusion time was 2.3 years (range 1 month to 10.8 years). From the total of 400,903 randomised women, data was analysed for 363,417 (90.6%) women.

Interventions

Out of 350 trials, 104 (29.7%) assessed treatments primarily on PTB prevention (e.g. progesterone in n=45 trials (43.3%), cervical cerclage in n=14 (13.5%), periodontal interventions in n=14 (13.5%), uterine home-monitoring in n=9 (8.7%) and smaller groups of trials on bed-rest, pessary, antibiotic and antihypertensive use (Table S3). A total of 246 trials (70.3%) stated PTB as one of their secondary aims (e.g. nutritional supplements/multi micronutrients/minerals in n=103 trials (41.9%), aspirin n=24 (9.8%), lifestyle (n=15) and behavioural interventions (n=15) (Table S3). The most common comparators were usual care n=151 (43.1%) or placebo n=146 (41.7%).

Usefulness features

Figure 2 provides a bird's eye view of binary usefulness features in proportion to the total body of PTB trials (n=350) and according to publication year and impact factor. In supplementary material (Table S4) more details on the exact numbers and proportions are shown, split in subgroups.

Problem base. The incidence of PTB <37weeks in the control or placebo groups of the trials varied from 0-100% with a median of 13.7% (IQR 7-30%) (Table S4 and Figure S2). This median is comparable with the worldwide incidence of PTB of 11.1% (6), but the range covers the full spectrum, including very low risk, high risk and very high risk populations (i.e. triplet pregnancies).

Context placement. A total of 185/350 trials (53%, 95%CI 48-58%) justified the importance of their study in context of previous systematic reviews (n=101/138 [73%] 95%CI 65-80% published [?]2010) and 18/350 trials (5%, 95%CI 3-8%) performed a systematic review as part of their study or included and updated meta-analyses.

Information gain. Information gain was deemed to be present in 192/350 trials (55%, 95%CI 50-60%). Absence of power calculations decreased from 71/212 trials (34%) among those published before 2010 to 29 /138 trials (21%) among those published [?]2010 (RR 0.63, 95%CI 0.43-0.91) (Table 1). However, from the 250 trials reporting a power calculation, calculations were incomplete and thus un-informative in 107 (43%,

95%CI 40-50%). Examples of non-informative power calculations: '*ωϵ αιμϵδ φορ 80% ποωϵρ ωιτη αν α οφ 0.05 ανδ α β οφ 0.20′* , incoherent information '*32 infants are necessary to reach a power of 0.05% at the 80% confidence leve* l', or incomplete reporting on the expected (absolute or relative) proportions of the primary outcome with and without the intervention '*we aimed for a 20% difference'* . The aimed differences for prolongation of gestational age varied from 3 days till 14 days; and birthweight, from 80g till 1500g (Figure S2 and Table S5). Use of composite or surrogate outcomes increased from 28/212 trials (13%) before 2010 to 34/138 trials (25%) [?]2010 (RR 1.87, 95%CI 1.19-2.93)(Table S4).

Pragmatism: A total of 17/350 trials (5%, 95%CI 3-8%) employed a pragmatic design, with no difference over time in their relative frequency.

Patient centeredness and value for money: No trials reported involvement of mothers to reflect patients' top priorities in research questions or outcomes used. No value of information analysis was reported in any final manuscript.

Feasibility: There were 113/350 trials (32%, 95%CI 28-37%) that did not report their intended sample size, while 34/350 trials (10%, 95%CI 7-13%) were unable to recruit their intended sample size (not counting DSMB interference because of clear benefit or harm n=5).

Transparency: A total of 95/138 trials (69%, 95%CI 61-75%), published [?]2010 were registered and a total of 38/138 (28%, 95%CI 21-36) were preregistered (registration before randomization of first patient). Change over time using [?] 2010 as a cut-off shows an increase from 0.5% vs 28%, RR 58, 95% CI 8-420%. Protocols were available in 21/138 trials (15%, 95%CI 10-22%). Change over time shows an increase from 0.5% to 15%, RR 32, 95% CI 4-237%. Data sharing statement was reported in 11/138 (8%, 95%CI 5-14%) trials [?]2010 with an increase from 2% to 8%, RR 3, 95% CI 1-10%.

A total of 255/350 trials (73%, 95%CI 68-77%) reported funding sources and 151/350 (43%, 95%CI 38-48%) reported on conflicts of interest.

Subgroup analyses

Publications in high impact journals (impact factor [?]4) after 2010 vs low impact journals after 2010 showed potential higher rates of context placement (52/67 [78%] vs 49/71 [69%]; RR 1.1, 95%CI 0.9-1.4), information gain (47/67 [70%] vs 32/71 [45%]; RR 1.56, 95%CI 1.15-2.10) and transparency features: trial registration (62/67 [93%] vs 33/71 [47%]; RR 1.99, 95%CI 1.54-2.58), published protocols (17/67 [25%] vs 4/71 [6%]; RR 4.50, 95%CI 1.60-12.7), and data availability (9/67 [13%] vs 2/71 [3%]; RR 4.77, 95%CI 1.07-21.27). However, only 36 of these 67 high impact journal publications (54%, 95%CI 42-65%) reported a complete power calculation and 17/67 (25%, 95%CI 16-37%) used a surrogate and/or composite outcomes as their primary outcome (Table S4). Some examples of surrogate outcomes: "cervical dilation" in women receiving home uterine monitoring for early PTB detection; "number of hospital antenatal visits" in women receiving routine doppler ultrasound; biochemical test results like HbA1c values in diabetic pregnant women or CRP values in pregnant women treated for a (vaginal) infection (Table S6).

Out of the 350 trials, 259 (74%) reported a primary outcome of which 97/259 (38%) reported a statistically significant result (P<0.05) in the primary outcome(s). Trials with a non-significant finding in the primary outcome(s) most frequently satisfied several usefulness criteria. This is most prominently for information gain (RR 1.27, 95%CI 1.03-1.56) and transparency features like trial registration (RR 1.31, 95%CI 0.97-1.78), preregistration (RR 1.34, 95%CI 0.73-2.4) and protocol availability (RR 2.0, 95%CI 0.76-5.34) (Table S4).

Usefulness overall

Among 350 selected RCTs, only 38 of trials (11%, 95% CI 8-15%) met half of the usefulness criteria (Figure 3).

Discussion

Main findings

A usefulness tool was developed with eight criteria combining 13 items identified through literature searches and consensus. Among 350 RCTs in PTB, many usefulness features were not met, with one tenth of trials meeting half of the items evaluated. Exploring the change in usefulness over time, most usefulness transparency features started to appear after the year 2000 and became more prominent after 2010. We found no substantial change in information gain, except for higher impact journals, which increased their information gain by reporting more complete power calculations, but in return more surrogate and composite outcomes as primary outcome were used. There was a remarkable absence of patient centeredness and value for money, and a very low percentage of pragmatic trial designs.

Strength and Limitations

This is the first study that provides a practical tool to assess usefulness of clinical research. At the same time, we have been able to demonstrate the use of the tool in 350 clinical trials in PTB. This assessment not only demonstrates the practical use of this tool, but also provided a very relevant overview on usefulness in the field of PTB research.

There are some limitations that need to be addressed. First, RCTs included in Cochrane reviews do not represent all RCTs on PTB prevention. However, pregnancy is the earliest field systematically addressed by Cochrane and its coverage of relevant trials is probably very high.

Second, usefulness data collection is dependent on the complete and faithful reporting of those features in published articles. One can, for example, argue that 'value for money' considerations might be described by the research group in their funding application and not in their published articles. Therefore, an underestimation of the prevalence of this item is possible. Conversely, some items may be over-estimated, e.g. power calculations may have been added post-hoc and some multi-center, unmasked trials of existing interventions may still violate pragmatism, contrary to authors' claims, and therefore our estimate of the proportion of pragmatic trials is an upper bound.

Third, the usefulness features are not meant as a 'checkbox' to ensure high quality and low bias. A study scoring 'high' in all usefulness items can still provide highly biased or even false data. Also, some usefulness items are not always 'good' or 'bad'. One such example is pragmatism. Not all clinical research questions require a pragmatic trial design (12) and typically, it is reasonable to do some explanatory trials before venturing into proving usefulness through pragmatism.

Fourth, for information gain we used an approach focused on power calculations and use of relevant outcomes. However, one can also measure how extensively the results of a study change prior perceptions of the evidence ("entropy change").(16) A well-powered study may not change our prior evidence much, if it fully agrees with what we already knew before running the study and if the evidence was already conclusive before the new study was run.

Fifth, we have operationalized the eight criteria of usefulness with the aim of applying them in a specific field, in this case PTB trials, for demonstration purposes. For most of the eight items, the same operationalized definitions can be applied to any other clinical research field. The one exception is definition of problem base. Depending on the clinical problem, different problem-specific and field-specific definitions would need to be conceived.

Finally, we did not yet examine how the 13 items are correlated to each other. Providing a total usefulness score might therefore not be appropriate as all individual criteria provide their own perspective of usefulness information and they are not interchangeable.

Interpretation

Previous empirical evaluations have focused on one or a few aspects of some of the items that were considered in our eight usefulness criteria. For example, there are several empirical studies examining the conduct of systematic reviews preceding a trial (17,18), the use of power calculations (19), pragmatism (12) and use of transparency practices such as protocol and data sharing, registration, disclosures of funding and conflicts

of interest.(20,21) However, our evaluation provides a composite assessment across multiple domains in a scale that is unprecedented and offers a wider view. For PTB research there is no prior empirical evaluation of most of these usefulness features, but there is definitely awareness of the problems arising from lack of these features.(22) One study for example evaluated the effect of pre-registration and its impact on reducing selective outcome reporting in trials and meta-analysis evaluating progesterone for PTB prevention. 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 pre-registered primary outcome only (n=22), compared to the totality of trials and reviews.(23) This example highlights the importance of addressing preregistration, protocol and data availability. Sadly, these transparency features still represent a minority of the 138 trials published between 2010 to 2019 assessed in the current project.

Current status and prospects of usefulness of research in preterm birth

The multidimensional assessment of usefulness can map the strengths and weaknesses of a large field. The remarkable absence of reference to patient centeredness in our assessment could be the first 'low hanging fruit' to improve usefulness in future trials. Luckily, an extensive patient/parents and clinicians research priority list of 15 research questions related to PTB has already been developed through the James Lind Alliance.(24) Participants have also been involved in a 'core outcome set' for PTB prevention studies.(25) The essential step is implementing these two tools in future PTB trials.

Different studies may require very different levels of financial investment and may differ substantially in how much we can learn from them.(5) An assessment of value of information through formal modelling (13) can help reduce the conduct of studies that do not convincingly change practice. Hopefully, this will help decrease the use of surrogate and composite outcomes and the conduct of underpowered studies.(26) Proper power calculations require clinicians' and women's input to define the minimal clinically important difference. For PTB research, this is a remarkable knowledge gap. We found only one study investigating the minimal clinically important difference for preventive PTB strategies (cerclage, progesterone).(27)

The very low number of pragmatic trials may also be a missed opportunity. Traditional trials recruit highly selected participants, seen in specialist environments and meet rigorous inclusion/exclusion criteria such as being free of comorbidities, which might influence generalizability of the results. In contrast, pragmatic trials support the generation of evidence relevant to real-world decision making.(28)

Conclusion

We have demonstrated that a checklist of the eight criteria of usefulness can be used successfully to explore the usefulness of clinical research. Focusing on preterm birth research, our usefulness assessment points out that those clinical trials so far lacked many usefulness features, with one tenth of trials meeting half of the items evaluated. Use of informative outcomes, patient centeredness, pragmatism and transparency should be key targets for future research planning. These usefulness criteria can be adopted across diverse domains of clinical investigations and may offer feedback to different stakeholders (researchers, patients, peer reviewers, journal editors, guideline developers and policy makers) to improve future study design.

Acknowledgments

We thank John Borghi for his help on the search strategy and Khalid Khan for providing guidance in writing the manuscript. Perrine Janiaud, Yan Min, Eugenia Khassina, Shervin Emani and Leonardo Tozzi for providing kind translations of the clinical studies in other languages. We thank Wentao Li and Mario Malicki for providing feedback on the project.

Disclosure of interest

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 postdoctoral grant from the Netherlands Organization for Health Research and Development (Rubicon grand 452182306). CA is supported by postdoctoral grants from the Knut and Alice Wallenberg Foundation (KAW 2019.0561), Uppsala University,

and the Sweden-America Foundation. BM is supported by a NHMRC Investigator grant (GNT1176437). BM reports consultancy for Guerbet, has been a member of the ObsEva advisory board and holds Stock options for ObsEva. The work of JI has been supported by an unrestricted gift from Sue and Bob O'Donnell.

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

Contribution to Authorship

JH: was involved in the conception, planning, carrying out, analysis and writing of the work.

CD: was involved in the conception, planning, carrying out, analysis and writing of the work.

CA: was involved in carrying out, analysis and writing of the work.

NH: was involved in planning, analysis and writing of the work.

ZA: was involved in the conception, planning and writing of the work.

MO: was involved in the conception, planning and writing of the work.

PB: was involved in the conception, planning and writing of the work.

JI: was involved in the conception, planning, carrying out and writing of the work.

Details of Ethical Approval

No approval was needed for this project as it did not involve human or animal subjects.

Funding

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

Data sharing

Data as well as code will be shared at an open access platform within 3 months after publication (e.g. osf.io)

References

1. 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.

2. Hackbarth AD. Eliminating Waste in US Health Care. JAMA. 2012 Apr 11;307(14):1513.

3. Pandis N, Chung B, Scherer RW, Elbourne D, Altman DG. CONSORT 2010 statement: extension checklist for reporting within person randomised trials. BMJ. 2017 Jun 30;j2835.

4. 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.

5. Ioannidis JPA. Why Most Clinical Research Is Not Useful. PLOS Med. 2016 Jun 21;13(6):e1002049.

6. Blencove H, Cousens S, Oesergaard M, Chou D, Moller AB, Narewal R, et al. National, regional and worldwide estimates of preterm birth. Lancet. 2012;379(9832):2162-72.

7. Aarnoudse-Moens CSH, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-Analysis of Neurobehavioral Outcomes in Very Preterm and/or Very Low Birth Weight Children. PEDIATRICS. 2009 Aug 1;124(2):717–28.

8. Medley N, Vogel JP, Care A, Alfirevic Z. Interventions during pregnancy to prevent preterm birth: an overview of Cochrane systematic reviews. Cochrane Pregnancy and Childbirth Group, editor. Cochrane Database Syst Rev [Internet]. 2018 Nov 14 [cited 2020 Aug 27]; Available from: http://doi.wiley.com/10.1002/14651858.CD012505.pub2

9. Dodd JM, Jones L, Flenady V, Cincotta R, Crowther CA. Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth. Cochrane Pregnancy and Childbirth Group, editor. Cochrane Database Syst Rev [Internet]. 2013 Jul 31 [cited 2020 Sep 2]; Available from: http://doi.wiley.com/10.1002/14651858.CD004947.pub3

10. the EPPPIC group. Evaluating Progestogens for Preventing Preterm birth International Collaborative (EPPPIC): meta-analysis of individual participant data from randomised controlled trials. Lancet. 2021;397(10280).

11. 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 May;62(5):464–75.

12. Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic? BMC Med. 2018 Dec;16(1):49.

13. 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-E22.

14. Mauskopf JA, Earnshaw S, Mullins CD. Budget impact analysis: review of the state of the art. Expert Rev Pharmacoecon Outcomes Res. 2005 Feb;5(1):65–79.

15. Djamel D. Latest Impact Factors Journal List, Thomson Reuters based on 2019 Journal Citation Reports. 2019 [cited 2020 Aug 27]; Available from: http://rgdoi.net/10.13140/RG.2.2.15078.06720

16. Evangelou E, Siontis KC, Pfeiffer T, Ioannidis JPA. Perceived information gain from randomized trials correlates with publication in high–impact factor journals. J Clin Epidemiol. 2012 Dec;65(12):1274–81.

17. 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. 2007;100:4.

18. Heneghan C, Aronson JK. Why reports of clinical trials should include updated meta-analyses. BMJ Evid-Based Med. 2020 Feb;25(1):1–2.

19. Copsey B, Thompson JY, Vadher K, Ali U, Dutton SJ, Fitzpatrick R, et al. Sample size calculations are poorly conducted and reported in many randomized trials of hip and knee osteoarthritis: results of a systematic review. J Clin Epidemiol. 2018 Dec;104:52–61.

20. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible Research Practices and Transparency across the Biomedical Literature. Vaux DL, editor. PLOS Biol. 2016 Jan 4;14(1):e1002333.

21. Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. Dirnagl U, editor. PLOS Biol. 2018 Nov 20;16(11):e2006930.

22. Townsend R, Duffy JMN, Khalil A. Increasing value and reducing research waste in obstetrics: towards woman-centered research. Ultrasound Obstet Gynecol. 2020 Feb;55(2):151–6.

23. 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 Int J Obstet Gynaecol. 2017 Jun;124(7):1008–15.

24. Uhm S, Crowe S, Dowling I, Oliver S. The process and outcomes of setting research priorities about preterm birth – a collaborative partnership. Infant. 2014;10(6):178–81.

25. van 't Hooft J, Alfirevic Z, Asztalos E, Biggio J, Dugoff L, Hoffman M, et al. CROWN initiative and preterm birth prevention: researchers and editors commit to implement core outcome sets. BJOG Int J Obstet Gynaecol. 2018 Jan;125(1):8–11.

26. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011 Dec 28;343(dec28 1):d7995–d7995.

27. Ross S, Milne J, Dwinnell S, Tang S, Wood S. Is it possible to estimate the minimal clinically important treatment effect needed to change practice in preterm birth prevention? Results of an obstetrician survey used to support the design of a trial. BMC Med Res Methodol. 2012 Dec;12(1):31.

28. Tunis SR, Stryer DB, Clancy CM. Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy. JAMA [Internet]. 2003 Sep 24 [cited 2020 Aug 27];290(12). Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.290.12.1624

Supporting information

Box S1. Search strategy

Box S2. Usefulness criteria checklist

Box S3. Details on the interpretation of the eight usefulness criteria

Table S1. List of systematic reviews and randomised controlled trials included

Figure S1. a. Distribution of studies published by year

b. Cumulative number of studies per continent over time.

Table S2. Characteristics of included randomised controlled trials (total) and stratified by Impact Factor (IF) below and above 4, and publication before and after 2010 in high impact factor journals.

Table S3. Details on intervention and controls by study focus (preterm birth prevention primary or secondary aim)

Table S4 . Usefulness criteria overall and stratified on publications before 2000, between 2000-2009 and between 2010-2019; impact factor below and above 4; and publications with impact factor above 4 published after 2010.

Figure S2. Distribution of number of trials (y-axis) with the reported incidence (%) of preterm birth (birth <37 weeks of gestation).

Figure S3. Histograms with relative risk reduction (A) and absolute risk reduction (B) for the 143 studies with a complete power calculation with a negative relative or absolute risk reduction for an increase in the incidence (e.g., increase in the proportion preterm birth >37 weeks), and a positive relative or absolute risk reduction for a decrease in the incidence (decrease in the proportion of preterm birth <37 weeks).

Table S5. Details on minimal important differences used for power calculations expressed as absolute difference in ordinal or continuous outcome variables

Table S6. List of surrogate outcomes.

Hosted file

`Figure 1. PRISMA Flow chart of included randomised controlled trials.docx` available at https://authorea.com/users/421146/articles/527288-value-of-clinical-research-usefulness-tool-development-and-systematic-review-of-350-randomised-controlled-trials-in-preterm-birth

Hosted file

`Table 1 Baseline characteristics.docx` available at https://authorea.com/users/421146/articles/527288-value-of-clinical-research-usefulness-tool-development-and-systematic-review-of-350-randomised-controlled-trials-in-preterm-birth

Hosted file

`Figure 2. Usefulness criteria overall and according to publication year and impact factor of journal.d` available at https://authorea.com/users/421146/articles/527288-value-of-clinical-research-

usefulness-tool-development-and-systematic-review-of-350-randomised-controlled-trials-in-preterm-birth

Hosted file

Figure 3. Number of trials meeting 0 to 13 usefulness criteria.docx available at https://authorea.com/users/421146/articles/527288-value-of-clinical-research-usefulness-tool-development-and-systematic-review-of-350-randomised-controlled-trials-in-preterm-birth