Journal of Clinical Epidemiology 162 (2023) 107−117

ORIGINAL ARTICLE

Few randomized trials in preterm birth prevention meet predefined usefulness criteria

Janneke van 't Hooft[a,b,*], Charlotte E. van Dijk[b], Cathrine Axfors[a,c], Zarko Alfirevic[d], Martijn A. Oudijk[b,e], Ben W.J. Mol[f,g], Patrick M. Bossuyt[h], John P.A. Ioannidis[a,i]

[a]*Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Palo Alto, CA, USA*
[b]*Department of Obstetrics and Gynecology, Amsterdam UMC, University of Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, Netherlands*
[c]*Department for Women's and Children's Health, Uppsala University, Uppsala, Sweden*
[d]*Center for Women's Health Research, Liverpool Women's Hospital, Liverpool, United Kingdom*
[e]*Amsterdam Reproduction and Development Institute, Amsterdam, The Netherlands*
[f]*Department of Obstetrics and Gynecology, Monash University, Melbourne, Victoria, Australia*
[g]*Aberdeen Centre for Women's Health Research, School of Medicine, University of Aberdeen, Aberdeen, UK*
[h]*Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam AUMC, University of Amsterdam, Amsterdam, The Netherlands*
[i]*Departments of Medicine, of Epidemiology and Population Health of Biomedical Data Science, and of Statistics, Stanford University, Palo Alto, CA, USA*

Accepted 24 August 2023; Published online 30 August 2023

Abstract

Objectives: We operationalized a research usefulness tool identified through literature searches and consensus and examined if randomized controlled trials (RCTs) addressing preterm birth prevention met predefined criteria for usefulness.

Study Design and Setting: The usefulness tool included eight criteria combining 13 items. 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 change over time was assessed using ≧ 2010 as a cutoff.

Results: Among 347 selected RCTs, published within 56 preterm birth Cochrane reviews, only 36 (10%, 95% CI = 7−14%) met more than half of the usefulness criteria. Compared to trials before 2010, recent trials used composite or surrogate (less informative) outcomes more often (13% vs. 25%, relative risk 1.91, 95% CI = 1.21−3.00). Only 16 trials reflected real practice (pragmatism) in design (5%, 95% CI = 3−7%), with no improvements over time. No trials reported involvement of mothers to reflect patients' research priorities and outcomes selection. Recent trials were more transparent.

Conclusion: Few preterm birth prevention RCTs met more than half of the usefulness criteria but most of usefulness criteria are improving after 2010. Use of informative outcomes, patient centeredness, pragmatism and transparency should be key targets for future research planning.

Keywords: Randomized controlled trials; Usefulness; Preterm birth; Clinical research; Pragmatism; Transparency

1. Introduction

Research varies in value for clinical practice, with substantial waste in the evidence pipeline [1,2]. Most current tools focus on completeness of reporting [3,4] which helps the readers in critical appraisal after the research has been conducted and analyzed. However, this does not capture whether research has the potential to be useful. It can be argued that usefulness needs to be considered at the design phase, so that study findings have the best potential to benefit patients.

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

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.

* Corresponding author. Department of Obstetrics and Gynecology, Amsterdam UMC, University of Amsterdam, Vrije Universiteit Amsterdam, Meibergdreef 9, Amsterdam, the Netherlands. Tel.: +31-61-3477-818; fax: +31-20-444-4645.

E-mail address: J.vanthooft@gmail.com (J. van 't Hooft).

https://doi.org/10.1016/j.jclinepi.2023.08.016

What is new?

Key findings

- One tenth of trials in preterm birth meets at least half of the usefulness items evaluated.

What this adds to what was known?

- Recent trials (published from 2010 onwards) have an increase in composite or surrogate (less informative) outcomes.

What is the implication and what should change now?

- Use of informative outcomes, patient centeredness, pragmatism and transparency should be key targets for future research planning.

A 2016 article [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. Our group previously proposed a comprehensive, simple tool that captures usefulness features in eight criteria with 13 items [6].

A search in Medline and Google Scholar (March 2021) using the terms "Usefulness" and "Randomised Controlled Trial" provided 15,473 results. After title and abstract screening we found only one publication discussing usefulness in medical research [5] but no other publication discussing the usefulness of the randomized controlled trials (RCTs) themselves, rather the usefulness of the intervention under study.

Here, we aimed to operationalize the developed research usefulness tool collating published criteria and examined if RCTs addressing preterm birth (PTB) could be considered potentially useful. By evaluating these eight criteria consisting of 13 items, we aimed to assess the current status of research regarding its usefulness, the evolution of its usefulness over time, and where potential improvement is needed.

PTB is a biomedical research topic that has a major global health burden [7,8] and a fairly mature track record of research 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 reported to provide only limited guidance on how to deal best with PTB [9]. Mapping their usefulness would be helpful for understanding the status quo to give guidance on specific elements that need attention in designing future PTB trials.

Box 1 Usefulness tool and details on the eight criteria combining 13 items [6].

Usefulness criteria

For appraisal of randomized controlled trials

Goal: To provide a checklist that can help to inform researchers, patients, peer reviewers, journal editors, guideline developers and policy makers about the usefulness of an intervention study to inform clinical practice.

Practical use: For each criterion, there is one question to be answered, except for the second criterion, with two questions, and the last criterion, for which there are six separate questions to be answered. There will be no combined scoring as all individual criteria have their own perspective of usefulness information.

Note: Some criteria are operationalized here for specific application in preterm birth. Some aspects may need to be modified for operationalization in other areas of clinical trials.

1. Problem base: Is there a health problem that is big/important enough to fix?[a]

   What proportion is born <37 weeks in the study population?

2. Context placement: Has prior evidence been systematically assessed to inform (the need for) new studies?

   Is there already a systematic review available being cited in the paper?
   or
   Have the authors done a systematic review as part of the paper?

3. Information gain: Is the proposed study large and long enough to be sufficiently informative?

   Does it have 80% power for a clinically relevant primary outcome (clinically relevant: not being a composite or surrogate outcome) for a difference that can be deemed to be clinically important?[b]

4. Pragmatism: Does the research reflect real life? If it deviates does this matter?

   Do authors claim pragmatism (words in the manuscript could be "pragmatic, naturalistic, real life, generalizable, simple large trials, mega trial") without an obvious violation of pragmatism?[c]

5. Patient centeredness: Does the research reflect top patient priorities?

Does the trial state details of patient research prior and/or during trial design to influence chosen interventions and/or outcomes?

6. Value for money: Is the research worth the money?

   Is there a ''value of information'' analysis or ''budget impact calculation'' that was done before starting the trial?

7. Feasibility: Can this research be done?

   Has the trial reached ≥90% of the calculated sample size?

8. Transparency: Are methods, data, and analyses verifiable and unbiased in terms of transparency?

   8a. Has the trial been preregistered online?
   8b. Has the full trial protocol been published prior to trial commencement?
   8c. (if 8b is yes) Has the trial protocol been adhered to (or, if not, changes and reasons have been detailed/explained)?
   8d. Has the trial's funding (or none if not applicable) been clearly stated?
   8e. Have the authors' conflicts of interest (or none if not applicable) been clearly stated?
   8f. Are the study's raw data freely available (there is a statement in the manuscript about individual data being available—not necessarily online-)?

[a]This question would have to be different in each type of condition/disease, taking note that the common denominator is the burden of disease. In case of preterm birth the incidence of babies born < 37 weeks is a close surrogate in the study population involved. In case of an intervention study, one should consider the proportion in the control arm. Note: there is no definite ''good'' or ''bad'' in this usefulness item

[b]In cases of clearly post hoc power calculations this can be seen as no information gain. Surrogate outcomes are defined as outcomes that reflect a prior state to the main outcome of interest (e.g., cervical dilation instead of delivery, blood levels instead of relevant condition/disease/health state.

[c]Trials are obviously not pragmatic if they are (1) single-center; (2) blinded (patients and physicians, except for blinding assessors of outcome) or placebo controlled; (3) addressing a new intervention or new indication. In most trials it is very difficult or impossible to assess the full PRagmatic Explanatory Continuum Indicator Summary (PRECIS) scale of items without involving the authors of the original article. In case the PRECIS scale has been fully assessed for a trial by its authors, it will be recorded. Note: there is no definite ''good'' or ''bad'' in this usefulness item as not all clinical research questions require a pragmatic trial design approach, and typically it is reasonable to do some explanatory trials before venturing into proving usefulness through pragmatism.

2. Methods

2.1. Development of the usefulness tool and assessment of clinical research value using this tool

We searched 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 article highlighting usefulness of clinical research itself (rather than usefulness of the intervention examined in the trial). Eight specific criteria of usefulness were elaborated on/developed in this identified article [5]. These eight features were then discussed within a steering group consisting of two epidemiologists (P.B. and J.I.) and four clinicians in the field of obstetrics and gynecology trained in epidemiology (J.H., Z.A., M.O., B.W.M.). After discussion and consensus among all group members, a usefulness appraisal checklist consisting of eight criteria combining 13 items was developed (Appendix B). A commentary article provides more detailed rationale behind each item [6]. Details on how we have operationalized the 13 items can be seen in Box 1.

2.2. Study sample randomized trials in preterm birth

Our protocol and details on the search strategy can be found in PROSPERO, International prospective register of systematic reviews (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 (Appendix A), performed on November 2, 2017 [9]. We last updated this search on November 14, 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 labor and irrespective of risk status for PTB or comorbidities. 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 excluded.

Progesterone is the most studied drug intervention on PTB prevention. However, we found no update of the 2013 Cochrane review on this subject in singleton pregnant women [10]; therefore, we decided to include 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 [11].

After selection of Cochrane reviews, we included the RCTs regardless of language, provided that full-text articles were available. 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 noneligible trials.

2.3. Data extraction

For each trial, we extracted the journal name, dates of publication and extracted general characteristics reported in the trial and details on the study group. Two reviewers (among J.H., L.D. and C.A.) independently assessed each full text using a standardized usefulness data extraction form. After duplicate extraction of 50 RCTs the overall interagreement was 98% (J.H. and L.D.) and 95% (J.H. and C.A.), with the item ''information gain'' scoring the lowest interagreement (93%). Any discrepancies were resolved by consensus and discussion with a fourth team member (J.I.). After a second round of 40 duplicate extractions the interagreement 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.

2.4. Statistical analysis

Descriptive analyses report percentages (with 95% confidence interval [CI]), medians (with interquartile range [IQR]) or means (with standard deviation). We performed two subgroup analyses. First, we compared the eight individual usefulness criteria in studies reported before 2010 and between 2010 and 2019. These time cutoffs 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 both time frame 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 prespecified a ''high'' IF at a cut-off of ≧4.0 and ''low'' IF when <4 [12]. This cutoff was chosen to include the top specialty journals in obstetrics and gynecology in the ''high'' group (e.g., Obstetrics and Gynecology, British Journal of Obstetrics and Gynecology). We expected the most recent RCTs (reported between 2010 and 2019) published in high impact journals to perform better in some of the usefulness features compared to those in lower impact journals. Third, we stratified articles 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 satisfied to provide an overview on the number of trials meeting at least half (7/13) of the items.

3. Results

We identified 57 eligible Cochrane systematic reviews and 1 IPD meta-analysis focusing on primary or secondary PTB prevention in pregnant women containing 373 potentially eligible RCTs (Fig. 1). These Cochrane reviews were published between 2006 and 2019 (median year 2017). From the 373 eligible RCTs, we were able to include 347 RCTs for data extraction (Fig. 1) coming from 56 reviews (See Appendix C for overview of included systematic reviews and RCTs). These 347 RCTs were published between 1967 and 2019. The 347 RCTs randomized 381,675 participants in all continents, with a higher number of trials performed in North America ($n = 92$ trials, $n = 82{,}241$ randomized women) and Europe ($n = 93$ trials, $n = 56{,}653$ women) and an increase of trials coming from Asia ($n = 68$ trials, $n = 95{,}345$ women) with a steep rise after the millennium (Appendix D, Figure 1).

There were 212 RCTs published before 2010 ($n = 241{,}675$ randomized women), 135 RCTs between 2010 and 2019 ($n = 140{,}005$ randomized women). General characteristics of the total RCTs sample and stratified by the three time periods are shown in Table 1. Stratification by impact factor and by publication year before 2010 and ≧2010 among high impact factor publications are shown in supplementary material (Appendix E, Table 1).

4. Interventions

Out of 347 trials, 104 (30%) assessed treatments primarily on PTB prevention and 243 trials (70%) stated PTB as one of their secondary aims (Appendix F, Table 2).

4.1. Usefulness features

Table 2 provides an overview of binary usefulness features among all PTB trials ($n = 347$) and according to publication year <2010 and ≧2010. In supplementary material (Appendix G, Table 3) more details are shown, split by different subgroups (impact factor, primary outcome with significant P-value or not)

4.1.1. Problem base

The incidence of PTB <37 weeks in the control or placebo groups of the trials varied from 0 to 100% with a median of 13.7% (IQR: 7−30%). This median is comparable with the worldwide incidence of PTB of 11.1% [7], but

the range covers the full spectrum, including very low risk, high risk, and very high risk populations (i.e., triplet pregnancies).

4.1.2. Context placement

A total of 183 of 347 trials (53%, 95% CI = 47–58%) justified the importance of their study in context of previous systematic reviews (n = 99/135 [73%] 95% CI = 65–80% published ≧2010). Among which 18 of 347 trials (5%, 95% CI = 3–8%) performed a systematic review as part of their study or included and updated meta-analyses.

4.1.3. Information gain

Information gain was deemed to be present in 189 of 347 trials (55%, 95% CI = 49–60%). Absence of power calculations decreased from 67 of 212 trials (32%) among those published before 2010 to 28 of 135 trials (21%) among those published ≧2010 (RR: 0.64, 95% CI = 0.44–0.93) (Table 1). However, from the 247 trials reporting a power calculation, calculations were incomplete and thus uninformative in 106 (43%, 95% CI = 40–50%). Examples of noninformative power calculations: ''we aimed for 80% power with an α of 0.05 and a β of 0.20'', incoherent information ''32 infants are necessary to reach a power of 0.05% at the 80% confidence level'', 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 (Appendix H, Table 4). Use of composite or surrogate outcomes increased from 28 of 212 trials (13%) before 2010 to 34 of 135 trials (25%) ≧2010 (RR: 1.91, 95% CI = 1.21–3.00).

4.1.4. Pragmatism

A total of 16 of 347 trials (5%, 95% CI = 3–7%) employed a pragmatic design, with no difference over time in their relative frequency.

4.1.5. 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.

4.1.6. Feasibility

There were 113 of 347 trials (33%, 95% CI = 28–38%) that did not report their intended sample size, while 34 of 347 trials (10%, 95% CI = 7–13%) were unable to recruit their intended sample size (not counting *Data and Safety Monitoring Board* interference because of clear benefit or harm, n = 5).

4.1.7. Transparency

A total of 93 of 135 trials (69%, 95% CI = 61–75%) published 2010 were registered, and a total of 36 of 135 trials (27%, 95% CI = 20–36) were preregistered (registration before randomization of first patient). Change over time using ≧2010 as a cutoff shows an increase from 0.4% vs. 39%, RR: 57, 95% CI = 8–408%. Protocols were available in 21 of 135 trials (15%, 95% CI = 10–22%). Change over time shows an increase from 0.5% to 15%, RR: 33, 95% CI = 4–242%. Data sharing statement was reported in 10 of 135 (7%, 95% CI = 4–14%) trials ≧2010 with an increase from 2% to 7%, RR: 3, 95% CI = 1–9%.

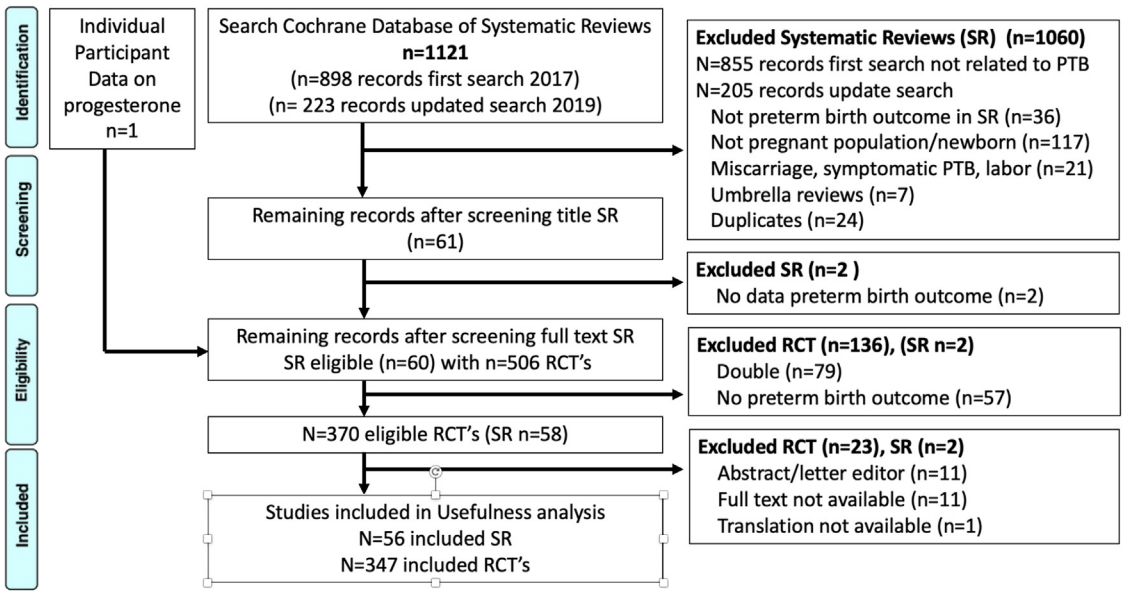

Fig. 1. Flow chart of included randomized controlled trials in preterm birth from (Cochrane) Systematic Reviews. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1. Characteristics of included randomized controlled trials (total) and stratified by publications yr

Characteristic	Total (n = 347)	Yr <2010 (n = 212)	Yr 2010–2019 (n = 135)
Journal type			
General medical	88 (25.4%)	58 (27.4%)	30 (22.2%)
Specialty O and G	171 (49.3%)	114 (53.8%)	57 (42.2%)
Specialty other	88 (25.4%)	40 (18.9%)	48 (35.6%
Journal impact factor median (IQR)	5.2 (2.6–6.6)	5.2 (2.8–6.6)	5.0 (2.2–6.6)
High impact factor (≧4)	204 (58.8%)	139 (65.6%)	65 (48.1%)
Low impact factor (<4)	111 (32%)	58 (27.4%)	53 (39.3%)
No impact factor	32 (9.2%)	15 (7.1%)	17 (12.6%)
Pilot studies	10 (2.9%)	6 (2.8%)	4 (3%)
Continents patient inclusion			
Africa	27 (7.8%)	14 (6.6%)	13 (9.6%)
Asia	68 (19.6%)	32 (15.1%)	36 (26.7%)
Australia	25 (7.2%)	19 (9%)	6 (4.4%)
Europe	93 (26.8%)	52 (24.5%)	41 (30.4%)
North America	92 (26.5%)	67 (31.6%)	25 (18.5%)
South America	20 (5.8%)	14 (6.6%)	6 (4.4%)
International	22 (6.3%)	14 (6.6%)	8 (5.9%)
Number of authors median (IQR)	6 (4–9)	5 (4–7)	6 (5–11)
Group name only	6 (1.7%)	6 (2.8%)	0 (0%)
≦2	15 (4.3%)	12 (5.7%)	3 (2.2%)
3–10	265 (76.4%)	169 (79.7%)	96 (71.1%)
≧11	61 (17.6%)	25 (11.8%)	36 (26.7%)
Single-center[b]	179 (51.6%)	117 (55.2%)	62 (45.9%)
Multicenter	166 (47.8%)	93 (43.9%)	73 (54.1%)
Median (IQR)	5 (3–15)	6 (3–14)	5 (2–16)
Interventions			
Drug	222 (64%)	124 (58.5%)	98 (72.6%)
Surgery/procedure	15 (4.3%)	15 (7.1)	0 (0%)
Device	4 (1.2%)	1 (0.5%)	3 (2.2%)
Participative intervention	40 (11.5%)	21 (9.9%)	19 (14.1%)
Organization of care	54 (15.6%)	41 (19.3%)	13 (9.6%)
Screening	12 (3.5%)	10 (4.7%)	2 (1.5%)
Comparator(s)			
Placebo	144 (41.5%)	89 (42%)	55 (40.7%)
Usual care	151 (43.5%)	100 (47.2%)	51 (37.8%)
Drug	46 (13.3%)	20 (9.4%)	26 (19.3%)
Surgery/procedure	1 (0.3%)	1 (0.5%)	0 (0%)
Participative intervention	5 (1.4%)	2 (0.9%)	3 (2.2%)
Sample randomized median (IQR)	242 (106–802)	250 (108–966)	215 (105–747)
<50	18 (5.2%)	15 (7.1%)	3 (2.2%)
50–99	51 (14.7%)	29 (13.7%)	22 (16.3%)
100–499	154 (44.4%)	87 (41%)	67 (49.6%)
500–999	52 (15%)	30 (14.2%)	22 (16.3%)
1,000–4,999	60 (17.3%)	42 (19.8%)	18 (13.3%)
>5,000	12 (3.5%)	9 (4.2%)	3 (2.2%)
Inclusion time in days median (IQR)	823 (547–1,154)	823 (580–1,096)	822 (519–1,278)
No power calculation available	100 (28.8%)	71 (33.5%)	29 (21.5%)
Any power calculation available	247 (71.2%)	141 (66.5%)	106 (78.5%)
Complete	141 (57.1%)	87 (61.7%)	54 (50.9%)

(*Continued*)

Table 1. Continued

Characteristic	Total (n = 347)	Yr <2010 (n = 212)	Yr 2010–2019 (n = 135)
Relative/Absolute risk diff	38 (15.4%)	19 (13.5%)	18 (17%)
Differences means	58 (23.5%)	28 (19.9%)	30 (28.3%)
Power only	11 (4.5%)	7 (5.0%)	4 (3.7%)
Primary outcome reported	256 (73.8%)	139 (65.6%)	117 (86.7%)
Significant primary outcome[a]	151(43.5%)	97 (45.8%)	54 (40%)
DSMB reported	84 (24.2%)	43 (20.3%)	41 (30.4%)
Ethics approval reported	292 (84.1%)	108 (84.9%)	112 (83%)
Funding source			
Public/Not for profit	210 (60.5%)	116 (54.7%)	94 (69.6%)
Private/Industry	19 (5.5%)	16 (7.5%)	3 (2.2%)
Both	17 (4.9%)	14 (6.6%)	3 (2.2%)
No funding	5 (1.4%)	1 (0.5%)	4 (3%)
Unclear	3 (0.9%)	3 (1.4%)	0 (0%)
Nonreported	93 (26.8%)	62 (29.2%)	31 (23%)

Abbreviations: IQR, interquartile range; DSMB, data and safety monitoring board.
[a] No trial reported harm in their main conclusion.
[b] n = 2 (0.6%) of studies it was not described if it was a multicenter or single-center study.

A total of 253 of 347 trials (73%, 95% CI = 68–77%) reported funding sources and 149 of 347 (43%, 95% CI = 38–48%) reported on conflicts of interest.

4.2. Subgroup analyses

Publications in high impact journals (impact factor ≧4) on and after 2010 vs. low impact journals on and after 2010 showed higher rates of information gain (45/65 [69%] vs. 31/70 [44%]; RR: 1.56, 95% CI = 1.15–2.13) and transparency features: trial registration (60/65 [92%] vs. 33/70 [47%]; RR: 1.96, 95% CI = 1.51–2.53), and published protocols (17/65 [26%] vs. 4/70 [6%]; RR: 4.58, 95% CI = 1.62–12.9). However, only 34 of these 65 high impact journal publications (52%, 95% CI = 40–64%) reported a complete power calculation and 17/65 (26%, 95% CI = 17–37%) used a surrogate and/or composite outcomes as their primary outcome (Appendix G, Table 3). All examples of surrogate outcomes are listed in Appendix I.

A total of 256/347 trials (74%) reported a primary outcome of which 97/256 (38%) reported a statistically significant result ($P < 0.05$) in the primary outcome(s). Trials with a nonsignificant finding in the primary outcome(s) most frequently satisfied several usefulness criteria, compared to trials with a significant finding in the primary outcome. This is most prominently for protocol availability (RR: 2.0, 95% CI = 0.78–5.35) (Appendix G, Table 3).

4.3. Usefulness overall

Among 347 RCTs only 36 of trials (10%, 95% CI = 7–14%) met half of the usefulness criteria (Fig. 2) of which 2 of 212 (0.9%) trials were published before 2010 and 34 of 135 (25%) trials were published between 2010 and 2019 (Fig. 2).

5. Discussion

5.1. Main findings

We operationalized a research usefulness tool and examined if RCTs addressing preterm birth prevention met predefined criteria for usefulness. Among 347 RCTs in PTB, many usefulness features were not met, with only 10% of trials meeting half or more 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 more complete reporting of 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 assessments, and a very low percentage of pragmatic trial designs.

5.2. Strength and limitations

This is the first study that provides the empirical application of a practical tool to assess aspects related to usefulness of clinical research in an entire field. We could demonstrate the practical use of this tool, and also provide a 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 depends on the complete and faithful reporting of those features in published articles. One can, for example, argue that "value for

Table 2. Usefulness criteria overall and according to publication yr <2010 and between 2010 and 2019

Usefulness	Total (N = 347)	Yr <2010 (N = 212)	Yr 2010–2019 (N = 135)
1. Problem base[a]			
2. Context placement reference to SR	183 (52.7%)	84 (39.6%)	99 (73.3%)
SR reported same population	138 (75.4%)	62 (73.8%)	76 (76.8%)
SR reported different population	45 (24.6%)	22 (26.2%)	23 (23.2%)
Context placement SR and/or MA as part of paper	18 (5.2%)	10 (4.7%)	8 (5.8%)
3. Information gain present	189 (54.5%)	113 (53.3%)	76 (56.3%)
No information gain and reasons:	158 (45.1%)	99 (46.7%)	59 (43.7%)
No power calculation	84 (53.2%)	67 (67.7%)	28 (47.5%)
Retrospective power calculation	11 (7.0%)	11 (11.1%)	0 (0%)
Surrogate outcome	32 (20.3%)	14 (14.1%)	18 (30.5%)
Composite outcome	15 (9.5%)	5 (5.1%)	10 (16.9%)
Combination of the above[b]	16 (10.1%)	2 (2%)	3 (5.1%)
4. Pragmatism	16 (4.6%)	10 (4.7%)	6 (4.4%)
Obvious violation of pragmatism[c]	4 (25%)	3 (30%)	1 (16.7%)
5. Patient centeredness	0 (0%)	0 (0%)	0 (0%)
6. Value for money	0 (0%)	0 (0%)	0 (0%)
7. Feasibility			
Yes	195 (56.2%)	106 (50%)	89 (65.9%)
No info (no sample size reported)	113 (32.6%)	81 (38.2%)	32 (23.7%)
Not feasible and reasons	39 (11.2%)	25 (11.8%)	14 (10.4%)
Pilot study	4 (10.3%)	2 (8%)	2 (14.3%)
DSMB interference[d]	13 (33.3%)	8 (32%)	5 (35.7%)
Low recruitment speed	12 (30.8%)	8 (32%)	4 (28.6%)
Funding	5 (12.8%)	3 (12%)	2 (14.3%)
No reason given	5 (12.8%)	4 (16%)	1 (7.1%)
8. Transparency			
Trial registration[e]	120 (34.6%)	27 (12.7%)	93 (68.9%)
8a. Preregistration trial	37 (30.8%)	1 (3.7%)	36 (39.1%)
8b. Protocol published	22 (6.3%)	1 (0.5%)	21 (15.2%)
8c. Any comment on adherence to protocol	11 (50%)	1 (100%)	10 (50%)
8d. Funding registration	253 (72.9%)	149 (70.3%)	104 (77%)
8e. Conflict of interest stated	149 (42.9%)	43 (20.3%)	106 (78.5%)
Conflict of interest present[f]	35 (23.5%)	11 (25.6%)	24 (22.6%)
8f. Statement on the availability of data[g]	15 (4.3%)	5 (2.4%)	10 (7.4%)
Statement of availability of code	2 (0.6%)	0 (0%)	2 (1.5%)

Abbreviation: DSMB, data and safety monitoring board; SR, systematic review; MA, meta-analysis..

[a] Problem base: Incidence PTB <37 wk median (IQR): 13.7% (7.2–30%). Range: 0–100%, Incidence PTB <34 wk median (IQR): 0.3% (3.4–19.4%). Range 0–100%.

[b] Combination of surrogate or composite outcome with no or retrospective power calculation. Power calculations that were incomplete were scored as ''information gain present''.

[c] Violation of pragmatism: placebo controlled trial n = 1, single-center trial n = 2, blindation of assessors n = 1 trial.

[d] DSMB interference reasons: harm n = 3 (all tree trials showed in final analysis no signs for harm anymore), clear effectiveness n = 2, futility n = 8.

[e] Trial registration: n = 66 (54.1%) Clinicaltrials.gov, n = 26 (21.3%) ISRCTN, n = 8 (6.6%) ACTNR, n = 8 (6.6%) IRCT, n = 7 (5.7) multiple registries, n = 7 (5.7%) other registries like Clinical Trials Registry of India (CTRI) Dutch Trial Register (NTR), EU Clinical trial Register, German trial registry (DRKS), Pan African Clinical Trial Registry (PACTR), Randomized Clinical Trial Latin-American Register. No traceable trial registry (n = 1).

[f] Conflict of interest present: ''yes'' in case authors state a conflict of interest in the ''conflict of interest'' section, or by report in ''authors contribution or financial discloser'' or evident by affiliation of the author (e.g., working directly in private sector related to the research topic).

[g] Details on statement of availability of data: No sharing n = 2, Through authors n = 7, Through authors and universities after 3 years n = 5, Through authors with ethical approval n = 2.

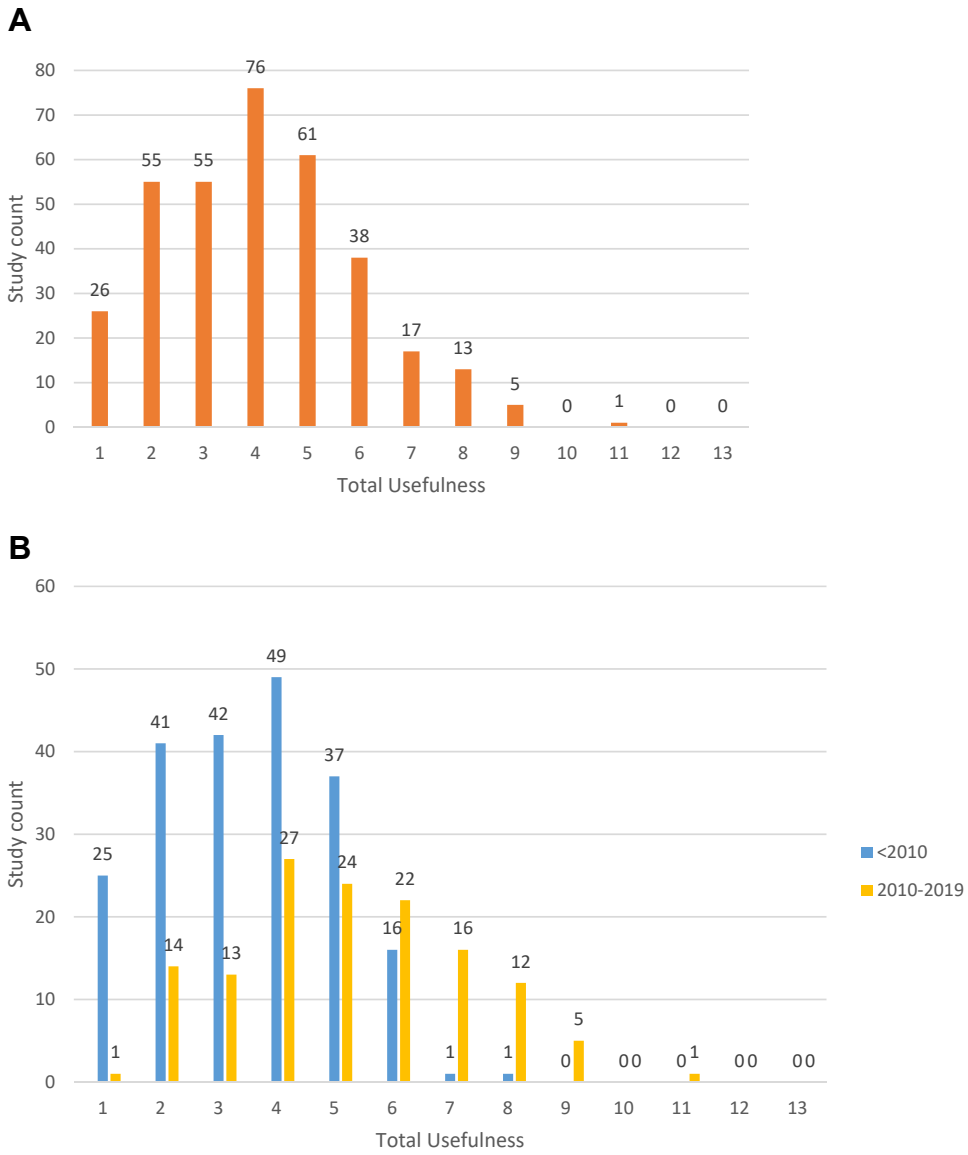

Fig. 2. (A), Number of trials meeting 0–13 usefulness criteria. (B), Number of trials meeting 0–13 usefulness criteria stratified by year <2010 and 2010–2019. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

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 overestimated, for example, power calculations may have been added post hoc and some multicenter, 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 criteria 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 [13] and typically, it is reasonable to do some explanatory trials before venturing into proving usefulness through pragmatism [6].

It is therefore important to acknowledge that a study is not useless if only few usefulness criteria are met. A study scoring low in some or many usefulness criteria may very well add information to the evidence pipeline. However, the added information of the study might be substantially higher when the usefulness items are taken into consideration.

Fourth, the scoring of usefulness might not be always a representation of the reality of the trial. Although the three reviewers were very consistent and with a high interagreement, the researchers were dependent on the reporting in the manuscripts.

Fifth, 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'') [14]. 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.

Finally, we did not yet examine how the 13 items are correlated. Correlation patterns may vary in different research fields and may change over time and in different settings. 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.

6. Interpretation

Previous empirical evaluations have focused on one or a few aspects of some of the items that were considered in our usefulness criteria. For example, there are several empirical studies examining the conduct of systematic reviews preceding a trial [15,16], the use of power calculations [17], pragmatism [13] and use of transparency practices such as protocol and data sharing, registration, disclosures of funding, and conflicts of interest [18,19]. However, our evaluation provides a composite assessment across multiple domains in a scale that offers a wider view. For PTB research there is no prior empirical evaluation of most of these usefulness criteria, but there is definitely awareness of some of the problems arising from lack of these features [20]. One study for example evaluated the effect of preregistration and its impact on reducing selective outcome reporting in trials and meta-analysis evaluating progesterone for PTB prevention [21]. 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, compared to the totality of trials and reviews [21].

7. Conclusion

We have demonstrated that a set of usefulness criteria can be used successfully to map the usefulness of clinical research. Clinical trials in PTB typically lacked most usefulness features, with only one-10th of trials meeting at least half of the items evaluated. Nevertheless, most of usefulness criteria are improving after 2010, mostly among the highest impact journals. 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.

Author contributions

J.H. was involved in the conception, planning, carrying out, analysis and writing of the work. C.D. was involved in the conception, planning, carrying out, analysis and writing of the work. C.A. was involved in carrying out, analysis and writing of the work. Z.A. was involved in the conception, planning and writing of the work. M.O. was involved in the conception, planning and writing of the work. P.B. was involved in the conception, planning and writing of the work. J.I. was involved in the conception, planning, carrying out and writing of the work.

Declaration of competing 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 452,182,306). C.A. is supported by postdoctoral grants from the Knut and Alice Wallenberg Foundation (K.A.W. 2019.0561), Uppsala University, and the Sweden-America Foundation. B.M. is supported by an NHMRC Investigator grant (GNT1176437). B.M. reports consultancy for Guerbet, has been a member of the ObsEva advisory board and holds Stock options for ObsEva. The work of J.I. has been supported by an unrestricted gift from Sue and Bob O'Donnell. J.I. is a team member of the editorial board of JCE.

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

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.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2023.08.016.

References

[1] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, et al. Biomedical research: increasing value, reducing waste. Lancet 2014;383:101−4.

[2] Hackbarth AD. Eliminating waste in US health care. JAMA 2012; 307:1513.
[3] Pandis N, Chung B, Scherer RW, Elbourne D, Altman DG. CONSORT 2010 statement: extension checklist for reporting within person randomised trials. BMJ 2017;357: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;13(6):e1002049.
[6] van 't Hooft J, van Dijk CE, Axfors C, Alfirevic Z, Oudijk MA, Khan KS, et al. Assessing the usefulness of randomised trials in obstetrics and gynaecology. BJOG 2023;130:695−701.
[7] Blencove H, Cousens S, Oesergaard M, Chou D, Moller AB, Narewal R, et al. National, regional and worldwide estimates of preterm birth. Lancet 2012;379:2162−72.
[8] 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;124(2): 717−28.
[9] Medley N, Vogel JP, Care A, Alfirevic Z. Interventions during pregnancy to prevent preterm birth: an overview of cochrane systematic reviews. Cochrane Database Syst Rev 2018;11:CD012505.
[10] 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 Database Syst Rev 2013:CD004947.
[11] The EPPPIC Group. Evaluating Progestogens for Preventing Preterm Birth International Collaborative (EPPPIC): meta-analysis of individual participant data from randomised controlled trials. Lancet 2021; 397:1183−94.
[12] Djamel D. Latest impact factors journal list, thomson reuters based on 2019 journal citation reports 2019. Available at http://rgdoi.net/10.13140/RG.2.2.15078.06720. Accessed August 27, 2020.
[13] Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? BMC Med 2018;16(1):49.
[14] 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;65:1274−81.
[15] 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.
[16] Heneghan C, Aronson JK. Why reports of clinical trials should include updated meta-analyses. BMJ Evid Based Med 2020;25(1):1−2.
[17] 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;104:52−61.
[18] Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible research practices and transparency across the biomedical literature. PLoS Biol 2016;14(1):e1002333.
[19] Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015−2017. PLoS Biol 2018;16(11):e2006930.
[20] Townsend R, Duffy JMN, Khalil A. Increasing value and reducing research waste in obstetrics: towards woman-centered research. Ultrasound Obstet Gynecol 2020;55(2):151−6.
[21] 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; 124:1008−15.