

Reporting Guidelines for systematic reviews: PRISMA (and beyond)

Jesse A. Berlin, ScD
VP, Epidemiology

Johnson & Johnson Pharmaceutical Research and Development

IOM Committee on Standards for Systematic Reviews of Clinical Effectiveness Research

26 October 2009

PRISMA motivation

- Systematic reviews and meta-analyses are essential to summarize evidence relating to efficacy and safety of health care interventions accurately and reliably. The clarity and transparency of these reports, however, is not optimal. Poor reporting of systematic reviews diminishes their value to clinicians, policy makers, and other users.
 - (From the E&E paper)

More PRISMA motivation

- “Since the development of the QUOROM (*Q*uality *O*f *R*eporting *O*f *M*eta-*a*nalysis) Statement—a reporting guideline published in 1999—there have been several conceptual, methodological, and practical advances regarding the conduct and reporting of systematic reviews and meta-analyses.”
- TRANSLATION:
 - The thinking has evolved (whose thinking?)
 - There have been some new publications relevant to this area
 - Generally time for an update

There are OTHER guidelines

- (As a reminder, noting that our charge is NOT limited to randomized, parallel group studies)
- Individual Studies
 - Randomized clinical trials: CONSORT
 - Epidemiologic studies: STROBE
 - Other (randomized) study designs: CONSORT “extensions”
- Systematic reviews and meta-analyses
 - Randomized studies: QUOROM (pre-PRISMA)
 - Epidemiologic studies: MOOSE
- For a central clearinghouse see: website
<http://www.equator-network.org/>

PRISMA

Annals of Internal Medicine

ACADEMIA AND CLINIC

The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration

Alessandro Liberati, MD, DrPH; Douglas G. Altman, DSc; Jennifer Tetzlaff, BSc; Cynthia Mulrow, MD, MSc; Peter C. Gøtzsche, MD, DrMedSci, MSc; John P.A. Ioannidis, MD; Mike Clarke, BA, DPhil; P.J. Devereaux, MD, BSc, PhD; Jos Kleijnen, MD, PhD; and David Moher, PhD

PRISMA Overview

- Checklist
 - Rationale
 - Objectives
 - Was the MA registered? (Was there a protocol?)
 - Were eligibility criteria for studies prespecified?
 - Risk of bias in individual studies assessed?
 - Analysis: methods and results specified?
 - Discussion, including Limitations clearly stated and fair?

What Evidence Supports PRISMA?

- Let's look at an individual item as an example (drawn directly from the E&E document)

Risk of Bias

- **Item 12: Risk of Bias in Individual Studies**
- Describe methods used for assessing risk of bias in individual studies (including specification of whether this was done at the study or outcome level, or both), and how this information is to be used in any data synthesis.
- **Examples**
- “To ascertain the validity of eligible randomized trials, pairs of reviewers working independently and with adequate reliability determined the adequacy of randomization and concealment of allocation, blinding of patients, health care providers, data collectors, and outcome assessors; and extent of loss to follow-up (i.e. proportion of patients in whom the investigators were not able to ascertain outcomes)” (85).
- “To explore variability in study results (heterogeneity) we specified the following hypotheses before conducting the analysis. We hypothesised that effect size may differ according to the methodological quality of the studies” (86). **JB NOTE – PET PEEVE**

Risk of Bias - Explanation

- **Explanation**
- The likelihood that the treatment effect reported in a systematic review approximates the truth depends on the validity of the included studies, as certain methodological characteristics may be associated with effect sizes (87, 88). For example, trials without reported adequate allocation concealment exaggerate treatment effects on average compared to those with adequate concealment (88). Therefore, it is important for authors to describe any methods that they used to gauge the risk of bias in the included studies and how that information was used (89). Additionally, authors should provide a rationale if no assessment of risk of bias was undertaken. The most popular term to describe the issues relevant to this item is “quality,” but for the reasons that are elaborated in **Box 4** we prefer to name this item as “assessment of risk of bias.”

Risk of bias – more explanation

- Many methods exist to assess the overall risk of bias in included studies, including scales, checklists, and individual components (90, 91). As discussed in **Box 4**, scales that numerically summarize multiple components into a single number are misleading and unhelpful (92, 93). Rather, authors should specify the methodological components that they assessed. Common markers of validity for randomized trials include the following: appropriate generation of random allocation sequence (94); concealment of the allocation sequence (93); blinding of participants, health care providers, data collectors, and outcome adjudicators (95–98); proportion of patients lost to follow-up (99, 100); stopping of trials early for benefit (101); and whether the analysis followed the intention-to-treat principle (100, 102). The ultimate decision regarding which methodological features to evaluate requires consideration of the strength of the empiric data, theoretical rationale, and the unique circumstances of the included studies.
- **(By now you get the idea that there are legitimate references)**

The Points

- Each included item has an explicit rationale and, where available, references to support the idea that the item being specified is somehow related to level of understanding by the reader, transparency and replicability of the process, likelihood of bias, etc.
- KEY POINT: These are REPORTING guidelines. There are other evaluation guidelines and conduct guidelines.

Acceptance of PRISMA?

- PRISMA was published simultaneously in:
 - *Annals of Internal Medicine*
 - *PLOS Medicine*
 - *BMJ*
 - *Journal of Clinical epidemiology*
 - *Open Medicine*

Evaluations of QUOROM

(that I could find)

- Biondi-Zoccai et al. Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: case study. *BMJ*, doi:10.1136/bmj.38693.516782.7C (published 16 January 2006).
- Hind D, Booth A. Do health technology assessments comply with QUOROM diagram guidance? An empirical study. *BMC Medical Research Methodology* 2007,7:49.

Biondi-Zoccai et al.

- 10 published systematic reviews (Aug 2003 – March 2005) of acetylcysteine to prevent contrast nephropathy
- Median compliance with QUOROM was 16 (range 11-17) out of 18 items
- Only 3 titles indicated that the paper was a systematic review of randomized trials
- Flow diagram only included in 3 papers

Biondi-Zoccai et al. (2)

- 7 studies reported an assessment of methodologic quality of primary studies
- Papers with more pages had higher QUOROM scores on average (Pearson $R = 0.73$, $p = 0.016$)
- Prior not for profit funding was associated with higher Guyatt-Oxman scores (6 for “yes” [range 3-7] vs. 2 for “no” [range 1-4], $p = 0.037$)
 - But NOT with QUOROM scores (16 [range 11-17] vs. 14 [range 12-16])
- QUOROM scores and Guyatt-Oxman scores were not associated ($R = -0.06$, $p = 0.86$)

Length Matters (Biondi-Zoccai)

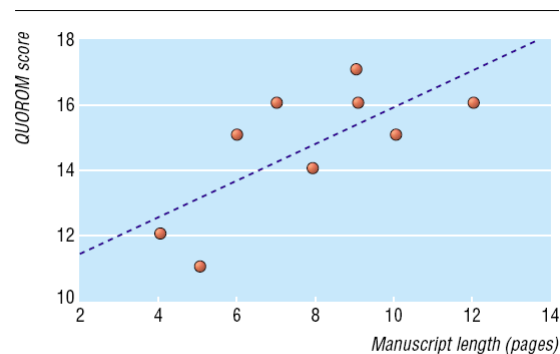


Fig 2 Scatter plot showing relation between manuscript length and quality of reporting of meta-analyses (QUOROM) score. Thorough and complete reports achieved higher scores

Agreement on eligibility of component studies

Biondi-Zoccai

Table 2 Pooled randomised controlled trials (RCTs) for each of included systematic reviews*

RCT	Availability†		Systematic reviews									
	Source	Report type	Birk et al [§]	Isehaber et al [§]	Alonso et al [§]	Kabiragar et al [§]	Pannu et al [§]	Gara and Frenes [§]	Bagshaw and Ghali [§]	Misra et al [§]	Nallamothu et al [§]	Duong et al [§]
Tepel et al (2000)	Medline	Article	+	+	+	+	+	+	+	+	+	+
Adelman et al (2002)‡	ACC only	Abstract	EX	—	EX	+	—	—	—	—	—	EX
Allagaband et al (2002)	Medline	Article	+	+	+	+	+	+	+	+	+	+
Aznus et al (2002)	TCT (Medline since 2005)	Abstract (article since 2005)	—	—	+	—	+	—	—	—	—	+
Brigioni et al (2002)	Medline	Article	+	+	+	+	+	+	+	+	+	+
Diaz-Sandoval et al (2002)	Medline	Article	+	+	+	+	+	+	+	+	+	+
Durham et al (2002)	Medline	Article	+	+	+	+	+	+	+	—	+	+
Goldenberg et al (2002)	ACC (Medline since 2004)	Abstract (article since 2004)	—	—	—	+	—	—	—	+	—	+
Kahlon et al (2002)	AHA only	Abstract	—	—	—	+	+	—	—	—	—	+
Octoa et al (2002)	ACC (Medline since 2004)	Abstract (article since 2004)	—	—	+	—	+	—	—	—	—	+
Shyu et al (2002)	Medline	Article	+	+	+	+	+	+	+	+	+	+
Valero et al (2002)§	Medline	Article	EX	—	—	+	+	—	+	—	—	+
Baker et al (2003)	Medline	Article	—	—	+	—	+	+	+	—	—	+
Boccalandro et al (2003)‡	Medline	Article	—	—	EX	+	+	—	—	—	—	—
Efrati et al (2003)	Medline	Article	—	—	EX	—	—	—	+	—	—	EX
El Mahmoud et al (2003)§	Medline	Article	—	—	—	—	—	—	—	—	—	+
Gomes et al (2003)	AHA (Medline since 2005)	Abstract (article since 2005)	—	—	—	—	—	—	—	—	—	+
Kay et al (2003)	Medline	Article	+	+	+	+	+	+	+	+	+	+
Kefer et al (2003)	Medline	Article	—	—	—	—	—	—	+	—	—	—
Le Feuvre et al (2003)	ACC only	Abstract	—	—	—	—	+	—	—	—	—	EX
Loutraniakis et al (2003)	ACC only	Abstract	—	—	+	+	+	—	—	—	—	+
MacNeil et al (2003)	Medline	Article	—	—	—	—	—	+	+	—	—	+
Noqareda et al (2003)‡	ACC only	Abstract	—	—	EX	+	—	—	—	—	—	EX
Oudemeyer et al (2003)	ACC (Medline since late 2003)	Abstract (article since late 2003)	—	—	+	+	+	+	+	—	—	+
Tadros et al (2003)‡	Medline	Article	—	—	EX	—	—	—	—	—	—	+
Agarwal et al (2004)	—	Unpublished	—	—	—	—	—	—	—	—	—	—
Fung et al (2004)	Medline	Article	—	—	—	—	—	—	+	—	—	+
Miner et al (2004)	Medline	Article	—	—	—	—	—	—	—	—	—	+
Webb et al (2004)	Medline	Article	—	—	—	—	—	—	—	—	—	+

ACC—American College of Cardiology conference proceedings; AHA—American Heart Association conference proceedings; EX—explicitly excluded from systematic review; TCT—Transcatheter Cardiovascular Therapeutics conference proceedings.
 †References to RCT are available from pertinent meta-analyses.
 ‡Study data available either as complete reports quoted in Medline, complete reports quoted in Embase, or conference proceedings.
 §Non-RCT.
 ¶In Italian.
 ††In French.

Hind and Booth

- Focus on study flow diagrams
- Raise the issue of citations vs. studies
- Examined published monographs from the UK NHS Health Technology Assessment (HTA) programme (2001 – 2005) (using Medline)

Study selection (Hind and Booth)

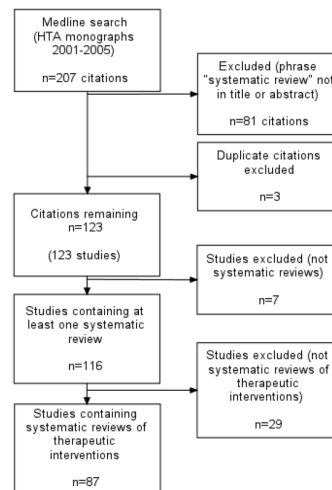


Figure 2
Study selection for empirical study of QUOROM compliance.

Hind and Booth (2)

- 49% of systematic reviews of clinical effectiveness illustrated the study selection process with a diagram
- Nearly doubled from 2002 (36%) -2004 (68%), but declined in 2005
- For those containing a formal meta-analysis, 32% had a diagram (14% in 2001 to 56% in 2004, declined in 2005)

(Jill said to be provocative, so ...)

SOME QUESTIONS FOR US TO CONSIDER

Question 1

- Do we need additional guidelines or simply expansion on and clarification about those that exist?
 - We already have many sets of principles
 - QUOROM has evidently not been universally adopted (but the evaluations haven't been extensive)
 - How do (would) we enforce PRISMA, anyway?

Is PRISMA Enough?

- In some respects, but ... additional “context” may be needed for comparative effectiveness
- MOOSE is part of the picture
- What about combining across randomized and non-randomized studies?
- What about mixed treatment comparisons?
 - Simplest example: Use A vs. placebo and B vs. placebo to get A vs. B
- Need for prospective MA (before knowing results of component studies)

Question 2

- How do we help the public (including patients and physicians) understand what level of evidence supports conclusions
 - Reported in the media
 - Presented as treatment guidelines
 - Used to make reimbursement decisions
- How much do we know about how much current research is understood (or trusted)?
 - There ARE people who are expert in health communication, but level of evidence may be a more subtle problem

Follow-on to question 2

- Who cares about PRISMA?
 - Even if PRISMA were strictly enforced (same for CONSORT, MOOSE, etc.), reporting guidelines, **at best**, allow evaluation of the validity of the meta-analysis. What about the actual performance of the M-A?
 - Beyond a finite (small?) number of methodologists, who is even capable of making such evaluations?
 - There is an element (a large element) of subjectivity in making judgments about adequacy of methods. How reliable (reproducible) would some kind of PRISMA assessment be?

More questions (paraphrased from Jill)

- Does PRISMA have the demonstrated potential to elevate the quality of clinical guidance?
 - same question for a “conduct” guideline
- To truly inform patient decision making?
 - Or physician, for that matter
 - If only an expert can even begin to apply the criteria, then do we need a way to “translate” the expert evaluation?
- If every systematic review of a clinical effectiveness question met PRISMA standards, what would we gain?
- Should PRISMA be mandatory in government funded research?

Oh yeah – we’re including non-randomized studies in our remit

- Q: How many epidemiologists does it take to change a light bulb?
- A: 5, one to change the bulb, and 4 to critique the methods
- POINT: Lots of issues around confounding (by indication), selection bias, etc.
 - How will we ever surround the topic of internal validity assessment in a widely-acceptable way? Others have tried, but have crashed and burned

So what?

So this committee needs to articulate

- Why a set (or sets) of guidelines would improve
 - Validity of reported meta-analyses
 - The ability of physicians and patients to trust that information
- How the information contained in these “certified, prime grade” meta-analyses would be made both physically accessible and interpretable to patients and their healthcare providers

BACKUP SLIDES

Structured Summary (the abstract)

- **Item 2: Structured Summary**
- Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; funding for the systematic review; and systematic review registration number.

Example of a good abstract

- **Example**
- *“Context:* The role and dose of oral vitamin D supplementation in nonvertebral fracture prevention have not been well established.
- *Objective:* To estimate the effectiveness of vitamin D supplementation in preventing hip and nonvertebral fractures in older persons.

More of the example abstract

- *Data Sources:* A systematic review of English and non-English articles using MEDLINE and the Cochrane Controlled Trials Register (15), and EMBASE (15). Additional studies were identified by contacting clinical experts and searching bibliographies and abstracts presented at the American Society for Bone and Mineral Research (14). Search terms included randomized controlled trial (RCT), controlled clinical trial, random allocation, double-blind method, cholecalciferol, ergocalciferol, 25-hydroxyvitamin D, fractures, humans, elderly, falls, and bone density.
- *Study Selection:* Only double-blind RCTs of oral vitamin D supplementation (cholecalciferol, ergocalciferol) with or without calcium supplementation vs calcium supplementation or placebo in older persons (60 years) that examined hip or nonvertebral fractures were included.

And still more abstract

- *Data Extraction:* Independent extraction of articles by 2 authors using predefined data fields, including study quality indicators.
- *Data Synthesis:* All pooled analyses were based on random-effects models. Five RCTs for hip fracture (n = 9294) and 7 RCTs for nonvertebral fracture risk (n = 9820) met our inclusion criteria. All trials used cholecalciferol. Heterogeneity among studies for both hip and nonvertebral fracture prevention was observed, which disappeared after pooling RCTs with low-dose (400 IU/d) and higher-dose vitamin D (700-800 IU/d), separately. A vitamin D dose of 700 to 800 IU/d reduced the relative risk (RR) of hip fracture by 26% (3 RCTs with 5572 persons; pooled RR, 0.74; 95% confidence interval [CI], 0.61-0.88) and any nonvertebral fracture by 23% (5 RCTs with 6098 persons; pooled RR, 0.77; 95% CI, 0.68-0.87) vs calcium or placebo. No significant benefit was observed for RCTs with 400 IU/d vitamin D (2 RCTs with 3722 persons; pooled RR for hip fracture, 1.15; 95% CI, 0.88-1.50; and pooled RR for any nonvertebral fracture, 1.03; 95% CI, 0.86-1.24).
- *Conclusions:* Oral vitamin D supplementation between 700 to 800 IU/d appears to reduce the risk of hip and any nonvertebral fractures in ambulatory or institutionalized elderly persons. An oral vitamin D dose of 400 IU/d is not sufficient for fracture prevention” (23).

The Explanation

- **Explanation**
- Abstracts provide key information that enables readers to understand the scope, processes, and findings of a review and to decide whether to read the full report. The abstract may be all that is readily available to a reader, for example, in a bibliographic database. The abstract should present a balanced and realistic assessment of the review’s findings that mirrors, albeit briefly, the main text of the report.
-
- We agree with others that the quality of reporting in abstracts presented at conferences and in journal publications needs improvement (24, 25). While we do not uniformly favor a specific format over another, we generally recommend structured abstracts. Structured abstracts provide readers with a series of headings pertaining to the purpose, conduct, findings, and conclusions of the systematic review being reported (26, 27). They give readers more complete information and facilitate finding information more easily than unstructured abstracts (28 –32).

What are they referencing?

- 26. **Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited.** Ann Intern Med. 1990;113:69-76. [PMID: 2190518]
- 27. **Mulrow CD, Thacker SB, Pugh JA. A proposal for more informative abstracts of review articles.** Ann Intern Med. 1988;108:613-5. [PMID: 3348568]
- 28. **Froom P, Froom J. Deficiencies in structured medical abstracts.** J Clin Epidemiol. 1993;46:591-4. [PMID: 8326342]
- 29. **Hartley J. Clarifying the abstracts of systematic literature reviews.** Bull Med Libr Assoc. 2000;88:332-7. [PMID: 11055300]
- 30. **Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: are structured abstract more efficient?** Journal of Information Science. 1996; 22:349-56.
- 31. **Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals.** N Engl J Med. 1987;317:426-32. [PMID: 3614286]
- 32. **Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association.** CMAJ. 1994;150:1611-5. [PMID: 8174031]

Study Selection

- **Item 9: Study Selection**
 - State the process for selecting studies (i.e., for screening, for determining eligibility, for inclusion in the systematic review, and, if applicable, for inclusion in the meta-analysis).
- **Example**
 - “Eligibility assessment . . . [was] performed independently in an unblinded standardized manner by 2 reviewers. . . . Disagreements between reviewers were resolved by consensus” (74).

Study Selection - Explanation

- Efforts to enhance objectivity and avoid mistakes in study selection are important. Thus authors should report whether each stage was carried out by one or several people, who these people were, and, whenever multiple independent investigators performed the selection, what the process was for resolving disagreements. The use of at least two investigators may reduce the possibility of rejecting relevant reports (75). The benefit may be greatest for topics where selection or rejection of an article requires difficult judgments (76). For these topics, authors should ideally tell readers the level of inter-rater agreement, how commonly arbitration about selection was required, and what efforts were made to resolve disagreements (e.g., by contact with the authors of the original studies).

Study Selection - References

- 75. **Edwards P, Clarke M, DiGuseppi C, Pratap S, Roberts I, Wentz R.** Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Stat Med.* 2002;21:1635-40. [PMID: 12111924]
- 76. **Cooper HM, Ribble RG.** Influences on the outcome of literature searches for integrative research reviews. *Science Communication.* 1989;10:179-201.