Relation of completeness of reporting of health research to journals’ endorsement of reporting guidelines: systematic review

Adrienne Stevens clinical research manager1, Larissa Shamseer PhD candidate12, Erica Weinstein medical student3, Fatemeh Yazdi senior clinical research associate1, Lucy Turner biostatistician1, Justin Thielman MSc candidate1, Douglas G Altman professor4, Allison Hirst project manager5, John Hoey retired professor6, Anita Palepu professor78, Kenneth F Schulz distinguished scientist9, David Moher senior scientist12

1Centre for Practice-Changing Research, Ottawa Hospital Research Institute, Ottawa, ON, Canada, K1H 8L6; 2Department of Epidemiology and Community Medicine, University of Ottawa, K1H 8M5 Ottawa, Canada; 3Albert Einstein College of Medicine, Yeshiva University, Bronx, NY 10461, USA; 4Centre for Statistics in Medicine, University of Oxford, Oxford OX3 7LD, UK; 5Nuffield Department of Surgical Sciences, University of Oxford, Oxford OX3 9DU, UK; 6Population and Public Health Initiative, Queen’s University, Kingston, ON, Canada, K7L 3N6; 7Centre for Health Evaluation and Outcome Sciences, St Paul’s Hospital, Vancouver, BC, Canada, V6Z 1Y9; 8Department of Medicine, University of British Columbia, Vancouver, BC, Canada, V5Z 1M9; 9International Clinical Sciences Support Center, FHI 360, Durham, NC 27713, USA

Abstract

Objective To assess whether the completeness of reporting of health research is related to journals’ endorsement of reporting guidelines.

Design Systematic review.

Data sources Reporting guidelines from a published systematic review and the EQUATOR Network (October 2011). Studies assessing the completeness of reporting by using an included reporting guideline (termed “evaluations”) (1990 to October 2011; addendum searches in January 2012) from searches of either Medline, Embase, and the Cochrane Methodology Register or Scopus, depending on reporting guideline name.

Study selection English language reporting guidelines that provided explicit guidance for reporting, described the guidance development process, and indicated use of a consensus development process were included. The CONSORT statement was excluded, as evaluations of adherence to CONSORT had previously been reviewed. English or French language evaluations of included reporting guidelines were eligible if they assessed the completeness of reporting of studies as a primary intent and those included studies enabled the comparisons of interest (that is, after versus before journal endorsement and/or endorsing versus non-endorsing journals).

Data extraction Potentially eligible evaluations of included guidelines were screened initially by title and abstract and then as full text reports. If eligibility was unclear, authors of evaluations were contacted; journals’ websites were consulted for endorsement information where needed. The completeness of reporting of reporting guidelines was analyzed in relation to endorsement by item and, where consistent with the authors’ analysis, a mean summed score.

Results 101 reporting guidelines were included. Of 15 249 records retrieved from the search for evaluations, 26 evaluations that assessed completeness of reporting in relation to endorsement for nine reporting guidelines were identified. Of those, 13 evaluations assessing seven reporting guidelines (BMJ economic checklist, CONSORT for harms, PRISMA, QUOROM, STARD, STRICTA, and STROBE) could be analyzed. Reporting guideline items were assessed by few evaluations.

Conclusions The completeness of reporting of only nine of 101 health research reporting guidelines (excluding CONSORT) has been evaluated in relation to journals’ endorsement. Items from seven reporting guidelines were quantitatively analyzed, by few evaluations each. Insufficient evidence exists to determine the relation between journals’ endorsement of reporting guidelines and the completeness of reporting of published health research reports. Journal editors and researchers should consider collaborative prospectively designed, controlled studies to provide more robust evidence.

Systematic review registration Not registered; no known register currently accepts protocols for methodology systematic reviews.
**Introduction**

Reporting of health research is, in general, bad. Complete and transparent reporting facilitates the use of research for a variety of stakeholders such as clinicians, patients, and policy decision makers who use research findings; researchers who wish to replicate findings or incorporate those findings in future research; systematic reviewers; and editors who publish health research. Reporting guidelines are tools that have been developed to improve the reporting of health research. They are intended to help people preparing or reviewing a specific type of research and may include a minimum set of items to be reported (often in the form of a checklist) and possibly also a flow diagram. An important role for editors is to ensure that research articles published in their journals are clear, complete, transparent, and as free as possible from bias. In an effort to uphold high standards, journal editors may feel the need to endorse multiple reporting guidelines without knowledge of their rigor or ability to improve reporting. The CONSORT statement is a well known reporting guideline that has been extensively evaluated. A 2012 systematic review indicated that, for some items of the CONSORT checklist, trials published in journals that endorse CONSORT were more completely reported than were trials published before the time of endorsement or in non-endorsing journals. A similar systematic review of other reporting guidelines may provide editors and other end users with the information needed to help them decide which other guidelines to use or endorse.

Our objective was to assess whether the completeness of reporting of health research is related to journals’ endorsement of reporting guidelines other than CONSORT by comparing the completeness of reporting in journals before and after endorsement of a reporting guideline and in endorsing journals compared with non-endorsing journals. For context, the box provides readers with definitions of terms used throughout this review.

**Methods**

Our methods are available in a previously published protocol. This systematic review is reported according to the PRISMA statement (appendix 1). Any changes in methods from those reported in the protocol are found in appendix 2.

**Identifying reporting guidelines**

We first searched for and selected reporting guidelines. We included reporting guidelines from Moher et al’s 2011 systematic review, and we screened guidelines identified through the EQUATOR Network (October 2011; reflects content from PubMed searches to June 2011). We included English language reporting guidelines for health research if they provided explicit text to guide authors in reporting, described how the guidance was developed, and used a consensus process to develop the guideline.

After removing any duplicate results from the search yield, we uploaded records and full text reports to Distiller SR. Two people (AS and LS) independently screened reporting guidelines. Disagreements were resolved by consensus or a third person. Where needed, we contacted authors of evaluations (n=66) or journal editors (n=48) for additional information. One person (from among a smaller working group of the team) processed evaluations with responses to queries to authors and journal editors and collated multiple reports for evaluations.

We first assessed each published study from within an included evaluation according to the journal in which it was published (fig 1)). We collected information on endorsement from evaluations or journal websites. If the journal’s “Instruction to authors” section (or similar) specifically listed the guideline, we considered the journal to be an “endorser.”

**Data extraction and analysis**

For included reporting guidelines, one person extracted characteristics of the evaluation and outcomes and did validity assessments; a second person verified 20% of the characteristics of studies and 100% of the remaining information. We contacted authors for completeness of reporting data for evaluations, where needed. Variables collected are reflected in the tables, figures, and appendices. As no methods exist for synthesizing validity assessments for methods reviews, we present information in tables and text for readers’ interpretation.

Our primary outcome was completeness of reporting, defined as complete reporting of all elements within a guidance checklist item. As not all authors evaluated reporting guideline checklist items as stated in the original guideline publications, we excluded any items that were split into two or more separate items or reworded (leading to a change in meaning of the item). Comparisons of interest were endorsing versus non-endorsing journals and after versus before endorsement. The first comparison functions as a cross sectional analysis, and years in which articles from endorsing journals were published depicted
the years of comparison with articles from non-endorsing journals. We used the publication date of the reporting guideline as a proxy if the actual date of endorsement was not known. For the second comparison, we included before and after studies from the same journal only if a specific date of endorsement was known. We also examined the publication years of included studies to ensure that years were close enough within a given arm for reasonable comparison. As a result, not all studies included in the evaluations were included in our analysis.

We analyzed the completeness of reporting in relation to journals’ endorsement of guidelines by item (number of studies within an evaluation completely reporting a given reporting item) and by mean summed score (the mean of completely reported guideline items for each study included in an evaluation and compared the mean of those sums across studies between comparison groups). We used a mean summed score only when evaluations also analyzed in this manner. We used risk ratios, standardized mean differences, and mean differences with associated 99% confidence intervals for analyses, as calculated using Review Manager software.20 In most cases, we reworked authors’ data to form our comparison groups of interest for the analysis.

Where possible, we used a random effects model meta-analysis to do a quantitative synthesis across evaluations for a given checklist item or for the mean summed score. We entered evaluations into Review Manager as the “studies,” whereas studies included within a given evaluation formed the unit of analysis, just as the number of patients would normally be entered. We entered the pooled effect estimate and confidence interval values from Review Manager for each checklist into Comprehensive Meta-Analysis to create summary plots depicting a “snapshot” view for each reporting guideline.21 Secondary outcomes were methodological quality and unwanted effects of using a guideline, as reported in evaluations. We present data for these outcomes in narrative form.

Results

Literature search results

Reporting guidelines

Eighty one reporting guidelines from Moher et al’s 2011 systematic review19 and 23 of 98 reporting guidelines identified by the EQUATOR Network were initially eligible for inclusion (fig 2). After removal of the CONSORT guidelines, we included a total of 101 reporting guidelines.19 20-121

Evaluations of reporting guidelines

Our literature search included evaluations of the CONSORT guidelines, but we excluded those during the screening process. We located 17 225 records through bibliographic databases and an additional 49 records from other sources (bibliographies, web search for full text reports of conference abstracts, and articles suggested by authors of reporting guidelines and members of the research team). After removing companion (known multiple publications) and duplicate reports, we screened a total of 15 249 title and abstract records. Of those, 1153 were eligible for full text review. After two rounds of full text screening, contacting authors, and seeking journal endorsement information, we included a total of 26 evaluations (fig 3).122-147 A list of potential evaluations written in languages other than English or French is provided in appendix 4.

Nine reporting guidelines were assessed among the 26 included evaluations: STARD 2003 for studies of diagnostic accuracy (n=8),122-124 126 148-150 PRISMA 2004 for systematic reviews and meta-analyses (n=5),149 151-153 QUOROM 1999 for meta-analyses of random trials (n=3),122 140-142 154 BMJ economics checklist 1996 (n=2 evaluations),122 123 STROBE 2007 for observational studies in epidemiology (n=2),140 147 CONSORT extension for journal and conference abstracts 2008 (n=1),146 CONSORT extension for herbal interventions 2006 (n=1),127 and STRICTA 2002 for controlled trials of acupuncture (n=1).139

Characteristics of included studies

Reporting guidelines

Appendix 5 descriptively summarizes included reporting guidelines according to the focus of the guideline and the content area the guideline covers. Among included guidelines were those covering general health research reports; animal, pre-clinical, and other basic science reports; a variety of health research designs and types of health research; and a variety of content areas.

Evaluations of reporting guidelines

Tables 1 and 2 show characteristics of the included evaluations. The most frequent content focuses of evaluations were diagnostic studies (7/26; 27%), drug therapies (6/26; 23%), and unspecified (5/26; 19%); evaluations spanned a variety of biomedical areas. Funding was most frequently either not reported (13/26; 50%) or provided by a government agency (7/26; 27%), and the role of the funder in the conduct of the evaluation was not reported in most evaluations (22/26; 85%). Two thirds of the evaluations provided a statement regarding competing interests or declared authors’ source(s) of support (17/26; 65%). Corresponding authors of evaluations were located in nine countries; 37% (10/27) of corresponding authors were in the United Kingdom.

For each included evaluation, tables 3 and 4 show the number of studies relevant to our assessments, their year(s) of publication, and the number of journals publishing the relevant studies. Tables 5 and 6 present information on the extent of journals’ endorsement and whether the date of endorsement was provided by evaluation authors, journal websites, or editors.

Validity assessment

Tables 3 and 4 show validity assessments for the comparisons; supports for those judgments are in appendix 6. Table 3 provides information on evaluations for the endorsing versus
After versus before journal endorsement

Analyzed by checklist item, Strobe (34 items) and Prisma (27 items) were the only reporting guidelines with all items evaluated; the Quorum (1/17 items), StarD (1/25 items), and STRICTA (17/20 items) guidelines were evaluated for a subset of items. All were assessed by one evaluation each with the exception of Prisma. Relatively few relevant studies were included in the assessments (median 20, interquartile range 19–64, studies; figures 10[11, 12, 13, and 14]). Analyses for completeness of reporting in relation to endorsement were statistically non-significant for each checklist item.

PRisma (all checklist items), STRICTA (item subset), and Strobe (all checklist items) reporting guidelines were analyzed by a mean summed score and by one or two evaluations each. Relatively few relevant studies were included in the assessments (median 20, interquartile range 18–50, studies), and analyses for completeness of reporting in relation to endorsement for mean summed scores were statistically non-significant (table 8[18]).

Assessment of study methodological quality within evaluations

Nine of 26 evaluations assessed the methodological quality of included studies (table 9[19]): one economics evaluation, two evaluation assessing randomized trials of herbal medicines, five systematic review evaluations, and two evaluations assessing diagnostic studies. Relatively few studies per evaluation were included in the assessments. The three more recently published systematic review evaluations used AMSTAR, whereas the older two evaluations used the Oxman and Guyatt index. The two diagnostic evaluations used separate, non-overlapping criteria. Given the different methodological areas and tools represented by the evaluations, a meaningful synthesis statement was not possible.

Unwanted effects of reporting guideline use

None of the included evaluations reported on unwanted effects of reporting guideline use.

Discussion

We reviewed the evidence on whether endorsement of reporting guidelines by journals is associated with more complete reporting of research. Although we identified a large number of reporting guidelines, very few evaluations of those reporting guidelines were located and provided information to enable an examination with respect to endorsement.

Strengths and weaknesses of systematic review

This is the first systematic review to comprehensively review a broad range of reporting guidelines. We sourced these reporting guidelines from the Equator Network and another systematic review characterizing known, high quality guidelines. We gave careful consideration to the parameters required to enable our comparisons of interest and made a considerable effort to locate evaluations, including the re-analysis of others’ data.

As exemplified by the volume of literature we had to screen, searching is complex with methods reviews. No search filters or established bibliographic database controlled vocabulary terms exist, especially for reporting guidelines. For many methods reviews, the particular studies of interest are often embedded in other studies. The time consuming task of
screening leads to a very low yield. Although systematic reviews are customarily current with the literature on publication, all such evidence pertains to comparative effectiveness reviews and not to methods reviews, such as ours. An updated search would yield more than 6000 records for us to screen with likely only a few relevant studies. We were aware of additional evaluations that have been published since the date of our literature search, and we have added these into our review. These additional studies have not led to a change in our conclusions. Other recently published articles did not meet our criteria.

We do not believe that an updated search would identify sufficient additional studies to change our results.

We limited our inclusion to evaluations written in English or French. This may be a limitation of our work, but we are unclear as to how many evaluations might exist in other languages given that few reporting guidelines are translated into other languages.

We did not include the main CONSORT reporting guideline here, and this decision was made after the initial protocol was written. The volume of evaluations for CONSORT is so large that we felt that detailed analysis would have overwhelmed the evidence from other reporting guidelines; furthermore, a systematic review solely evaluating the effect of CONSORT is available as recently as 2012.

Comparison with other reviews

The findings from the 2012 CONSORT systematic review show that, for some CONSORT checklist items, trials published in journals that endorse CONSORT were more completely reported than were trials published before the time of endorsement or in non-endorsing journals. CONSORT is by far the most extensively evaluated reporting guideline, in contrast to the reporting guidelines covered in this review. At least one other review evaluating CONSORT for harms has been published.

We examined this review, and studies included in that review but not in ours would not have met our eligibility criteria.

Meaning of review: explanations and implications

Although reporting guidelines might have sufficient face validity to convince some editors to endorse them, we found little evidence to guide this policy. This is in stark contrast, for example, to the evidence required to introduce a new drug in the marketplace. Here, empirical evidence in the form of pivotal randomized trials would be required. Although reporting guidelines are not drugs, they have become increasingly popular, their trajectory continues to increase very quickly, and journal editors and others are making policy decisions about encouraging their use in hundreds if not thousands of journals.

Evidence relating to CONSORT, STARD, MOOSE, QUOROM, and STROBE indicates that no standard way exists in which journals endorse reporting guidelines. Furthermore, other than including recommendations in their “Instructions to authors,” little is known about what else is done by individual journals to ensure adherence to reporting guidelines. This is a question of fidelity; the effect of endorsement is therefore plagued by different, and not well documented, processes as to the “strength” of endorsement. For example, some journals require a completed reporting guideline checklist as part of the manuscript submission, whereas others only suggest the use of reporting guidelines to facilitate writing of manuscripts. In both instances, whether or how journals check that authors adhere to journals’ recommendations/requirements is not known. One strategy would be to encourage peer reviewers to check adherence to the relevant reporting guideline. A 2012 survey of journals’ instructions to peer reviewers shows that reference to or recommendation to use reporting guidelines during peer review was rare (19 of 116 journals assessed). When mentioned, instructions on how to use reporting guidelines during peer review were entirely absent; most journals pointed to CONSORT but few other reporting guidelines. Specifically, surveys of journals’ instructions to authors with respect to endorsement of CONSORT show that guidance is inconsistent and ambiguous and does not provide authors with a strong indication of what is expected of them in terms of using CONSORT during the manuscript submission process. Evidence from this review and a similar CONSORT systematic review suggest much room for improvement in how journals seek to achieve adherence to reporting guidelines. Developers of reporting guidelines and editors could work together and agree on the optimal way to endorse and implement reporting guidelines across journals (brining some standardization to the implementation process).

A fundamental outcome used by evaluators was the completeness of reporting according to items from the reporting guideline. Ideally, this means that all concepts were reported about a particular reporting guideline checklist item. For example, in the STARD statement, one checklist item covers the “technical specifications of material and methods involved using how and when measurements were taken, and/or cite references for the index tests and reference standard.” For this item, some evaluations separated and tracked reporting information for the index test separately from the reference standard. We had to exclude nine evaluations that did not have any original, unmodified checklist items (that is, guidance items that were split into subcomponents or written with modified interpretation). Furthermore, as noted in tables 1 and 2, more than half of the included evaluations applied modifications to one or more items of the original guideline, negating the inclusion of those items in our analyses.

Evaluating the completeness of reporting of reporting guidelines in relation to journals’ endorsement might seem straightforward. However, in reality, it is complex. One problem in approaching our analysis is that only three evaluations considered endorsement as the “intervention” of interest, of which two could be included in our quantitative analysis. As a result, we had to rework authors’ data to facilitate the comparisons of interest and track down journals’ endorsement information, requiring considerable time and effort. Evaluators of reporting guidelines, in general, have not considered endorsement as an “intervention” that has the potential to affect the completeness of reporting. Although evaluations in this review do not provide conclusive evidence, the CONSORT review provides some evidence that simple endorsement of reporting guidelines has the potential to affect the completeness of reporting.

One design used in the literature is the comparison of complete reporting before and after the publication of a reporting guideline. In thinking about this as an intervention and then considering endorsement, endorsement would likely serve as a “stronger” intervention given the need for manuscripts to adhere to a journal’s “Instructions to authors” and subsequent editorial process. However, as mentioned above, the strength of endorsement is crucial and varies across journals. Thus, although not ideal, a journal’s statement about endorsement of a guideline is the best available proxy indicator of a journal’s policy and perhaps authors’ behavior around use of reporting guidelines. In terms of experimental designs, randomizing journals to endorse a reporting guideline or continue with usual editorial policy would be difficult, if not impossible. One method of intervening and evaluating can be with peer reviewers, as
mentioned above. To our knowledge, at least one randomized trial by Cobo et al in 2011 has examined the use of reporting guidelines in the peer review process within a single journal that did not endorse any reporting guidelines; it found that manuscripts reviewed using reporting guidelines were of better quality than those that did not use reporting guidelines.156 Although these findings are applicable only to a single journal, more trials like this can provide journals with their own evidence on completeness of reporting and better inform editors as to whether efforts on endorsement and, further, implementation, are having their intended effects.

Beyond simple publication of a guideline, little effort is dedicated to knowledge translation (implementation) activities. As defined by the Canadian Institutes of Health Research, the crux of knowledge translation is that it is a move beyond the simple dissemination of knowledge into the actual use/implementation of knowledge.19 The EQUATOR Network has gone some way in providing a collated home and network of reporting guidelines and resources. However, knowledge producers/guideline developers are responsible for ensuring appropriate and widespread use of a particular guideline by knowledge users. Developers and interested researchers may wish to think about studying the behaviors of target users (for example, prospective journal authors) and developing, carrying out, and evaluating strategies that have the potential to affect behavior change around guideline use, similar to ongoing work in implementation of clinical research.160,161

Future research

Future evaluations of reporting guidelines should assess unmodified reporting items. Non-experimental designs on the basis of journal endorsement status can help to supplement the evidence base. However, researchers in this area, such as guideline developers, should consider carrying out prospectively designed, controlled studies, like the study by Cobo et al.,160 in the context of the journal’s editorial process to provide more robust evidence.

Conclusions

The completeness of reporting of only nine of 101 rigorously developed reporting guidelines has been evaluated in relation to journal endorsement status. Items from seven reporting guidelines were quantitatively analyzed by few evaluations each. Insufficient evidence exists to determine the relation between journals’ endorsement of reporting guidelines and the completeness of reporting in published health research reports. Future evaluations of reporting guidelines can take the form of comparisons based on journal endorsement status, but researchers should consider prospectively designed, controlled studies conducted in the context of the journal’s editorial process.

We thank Shona Kirtley for information regarding the EQUATOR network search strategy for reporting guidelines, Andra Morrison for peer reviewing the search strategies developed for this review, Becky Skidmore for designing and conducting literature searches, Mary Gauthier and Sophia Tsiouras for assisting with screening, Kavita Singh for assisting with screening and data extraction, Misty Pratt for assisting with screening and data extraction and verification, Raymond Daniel for article acquisition and management of bibliographic records within Reference Manager 12 and Distiller SR, Hadeel AI’iacob for assisting with verification of data extraction and analyses, Iveta Simera for contributions to the design of the project and feedback on the manuscript, and the authors of reporting guidelines and journal editors who kindly responded to our requests for information. We thank Stefania Boccia, Nikola Panic, Julian Peron, Benoit You, and other authors of evaluations who responded to our queries and provided raw data for our analyses.

Contributors: AS, LS, DM, DGA, LT, AH, JP, AP, and KFS contributed to the conception and design of this review via the published protocol. All authors contributed to the screening, data extraction, analysis, or data interpretation phases of the review. AS and DM drafted the review, and all remaining authors revised it critically for important intellectual content. All authors approved the final manuscript. DM is the guarantor.

Funding. This work was supported by a grant from the Canadian Institutes of Health Research (funding research number KSD-111750). The Canadian Institutes of Health Research had no role in the design, data collection, analysis, or interpretation of the data; in the writing of the report; or in the decision to submit this work for publication. DGA is supported by Cancer Research UK. DM by a University of Ottawa research chair, and KFS by FH360. All researchers are independent from their relevant funding agencies. DGA, DM, and KFS are executive members of the EQUATOR network; AH served as an EQUATOR staff member during this project. The EQUATOR Network is funded by the National Health Service National Library of Health, National Health Service National Institute for Health Research, National Health Service National Knowledge service, United Kingdom Medical Research Council, Scottish Chief Scientist Office, and Pan American Health Organization.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: all authors maintained their independence from the agency that funded this work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This systematic review did not require ethics approval in Canada.

Data sharing: Datasets are available on request from the corresponding author.

Transparency: The lead author (guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

1 Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 2005;365(9469):1159-62.
2 Chan S, Bhandari M. The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. J Bone Joint Surg Am 2007;89:1970-8.
3 Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med 2007;4:447-55.
4 Smith BM, Lee HJ, Lee CH, Choi M, Jones DE, Bausell RB, et al. Quality of reporting randomized controlled trials (RCTs) in the nursing literature: application of the consolidated standards of reporting trials (CONSORT). Nurs Outlook 2008;56(6):31-7.
5 Yesupriya A, Evangelou E, Kavoura FK, Patopoulos NA, Clyne M, Walsh MC, et al. Reporting of Randomised Controlled Trials (RCTs) in the cardiovascular literature: a review. J Med Libr Assoc 2007;95(4):282-7.
6 Zhang D, Yin P, Freeman M, Jordan R, Zheng N, Cheng KK. An assessment of the quality of randomised controlled trials conducted in China. Trials 2008;9:22.
7 Glascoulis G, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet 2014;383:267-76.
8 Moher D, Schulz KF, Simeon I, Altman DG. Guide for developers of health research reporting guidelines. PLoS Med 2010;7(2).
9 Moher D, Weeks L, Ocampo M, Seely D, Sampson M, Altman DG, et al. Describing reporting guidelines for health research: a systematic review. J Clin Epidemiol 2011;64:718-27.
10 World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2004;291(10):106-7.
11 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. JAMA 1996;276:637-9.
12 Moher D, Schulz KF, Altman DG, Lépape L. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med Res Methodol 2010;11:1-7.
13 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting parallel randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663-94.
14 Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c382.
What is already known on this topic

Complete and transparent reporting of research, which is often inadequate and incomplete, enables readers to assess the internal validity and applicability of findings. The completeness of reporting of the CONSORT guideline in relation to endorsement by journals has been evaluated and was shown to be associated with more complete reporting for several checklist items. No systematic review has comprehensively reviewed evaluations of other reporting guidelines.

What this study adds

Apart from CONSORT, 101 rigorously developed reporting guidelines exist for reporting health research, only nine of which could be evaluated and regarding their only few endorsement levels. Few data are available to help editors regarding endorsement of specific reporting guidelines. Future evaluations of reporting guidelines based on journal endorsement status can help to supplement the evidence base. However, researchers should consider prospectively designated, controlled studies conducted in the context of the journal's editorial process to provide more robust evidence.
## Tables

### Table 1  | Characteristics of included evaluations for BMJ economics, CONSORT extension for abstracts, CONSORT extension for harms, CONSORT extension for herbal interventions, and PRISMA reporting guidelines

| Author, year* | Country of corresponding author | Sources of funding; role of funder; authors’ source(s) of support | Content focus | Specific medical or scientific specialty† | Extent of guideline assessed‡ |
|---------------|---------------------------------|---------------------------------------------------------------|--------------|----------------------------------------|-----------------------------|
| BMJ economics guideline, 1996 |                                |                                                              |              |                                        |                             |
| Herman, 2005*§ | United States                   | Government agency: grant from National Center for Complementary and Alternative Medicine; not reported; not reported (authors declare no competing interests) | Complementary medicine | Unspecified | All items                             |
| Jefferson, 1998** | United Kingdom                  | Not reported; not reported; not reported                      | Unspecified  | Unspecified                           | Subset of items¶            |
| CONSORT extension for abstracts, 2008** |                                |                                                              |              |                                        |                             |
| Ghimire, 2014* | South Korea                     | Not reported; not reported; not reported (authors declare no competing interests) | Unspecified  | Oncology                               | Subset of items¶            |
| CONSORT extension for harms, 2004** |                                |                                                              |              |                                        |                             |
| Haidich, 2011*§ | Greece                          | Not reported; not reported; not reported                      | Drug therapies | Several medical specialties††          | All items                  |
| Turner, 2011*§ | Canada                          | Government agency: National Center for Complementary and Alternative Medicine, National Institutes of Health; not reported; authors declare no competing interests | Complementary medicine | Unspecified | Subset of items‡‡                     |
| Peron, 2014*| France                          | Not reported; not reported; charitable foundation: Nuovo-Soldat Foundation (authors declare no competing interests) | Drugs therapies | Oncology                               | Subset of items¶            |
| Cornelius, 2013* | United Kingdom                  | Government agency: National Institute for Health Research Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London; not reported; not reported (authors declare no competing interests) | Drug therapies | Neurosciences                          | Subset of items¶            |
| Lee, 2008* | Canada                          | Government agency: Canadian Institutes of Health Research Chronic Disease New Emerging Team grant (joint sponsorship from Canadian Diabetes Association, Kidney Foundation of Canada, Heart and Stroke Foundation of Canada, and two other Canadian Institutes of Health Research Institutes); not reported; not reported | Drug therapies | Clinical neurology                     | Subset of items¶            |
| CONSORT extension for herbal interventions, 2006** |                                |                                                              |              |                                        |                             |
| Ernst, 2011** | United Kingdom                  | Not reported; not reported; not reported                      | Complementary medicine | Medicine, general and internal         | Subset of items             |
| PRISMA, 2009 |                                |                                                              |              |                                        |                             |
| Tunis, 2013*| Canada                          | No funding; not applicable; not reported (authors state no competing interests; authors have declared financial activities not related to article) | Unspecified  | Radiology, nuclear medicine, and medical imaging | All items                  |
| Panic, 2013*| Italy                           | Not reported; funder had no role in work; Academic: ERAWEB, Charitable: Fondazione Veronesi (authors declare no competing interests) | Unspecified  | Gastroenterology and hepatology        | All items                  |
| Fleming, 2013* | United Kingdom                  | Not reported; not reported; not reported.                    | Unspecified  | Dentistry, oral surgery, and medicine  | All items                  |

*All included evaluations were published as full reports.
†2011 journal impact factor categories used for classification.
‡If authors of evaluations deemed particular guidance item to be “not applicable” to literature they were assessing, those items were excluded from analysis; for evaluations with zero or one studies in one comparison arm, those evaluations were removed from synthesis because that one arm would determine direction of effect.
§Included in quantitative analysis.
¶As determined by authors of this review when comparing with published guidance.
Table 1 (continued)

| Author, year* | Country of corresponding author | Sources of funding; role of funder; authors’ source(s) of support | Content focus | Specific medical or scientific specialty† | Extent of guideline assessed‡ |
|---------------|---------------------------------|---------------------------------------------------------------|--------------|----------------------------------------|-------------------------------|

**Official extension of CONSORT reporting guideline; “official” defined as at least one author from original CONSORT reporting guideline on authorship of extension.

††Cardiac and cardiovascular systems, hematology, immunology, infectious diseases, obstetrics and gynecology, oncology, psychiatry, respiratory system, and rheumatology.

‡‡Evaluation’s authors indicated subset was assessed but authors of this review determined smaller subset was analyzed when comparing with published guidance.
| Author, year | Country of corresponding author | Sources of funding; role of funder; authors’ source(s) of support | Content focus | Specific medical or scientific specialty† | Extent of guideline assessed‡ |
|-------------|----------------------------------|---------------------------------------------------------------|--------------|------------------------------------------|-------------------------------|
| Hind, 2007§§ | United Kingdom                   | Not reported; not reported; not reported (authors declare they previously worked for UK NHS Health Technology Assessment Programme (source of included reports)) | Therapeutic interventions (generic) | Unspecified | Subset of items |
| Biondi-Zoccai, 2006†† | Italy                             | No funding; not applicable; not reported (authors declare no competing interests) | Drug therapies | Urology and nephrology | All items |
| Poolman, 2007 | Canada, Netherlands              | Not reported; not reported; academic: Canadian Institutes of Health Research Canada Research Chair; Industry: Merck Sharp and Dohme Netherlands, Biomet Netherlands, Zimmer Netherlands; other: Stichting Wetenschappelijk Onderzoek Orthopaedische Chirurgie Fellowship, Anna Fonds Foundation, Nederlandse Vereniging voor Orthopedische Traumatologie Fellowship | Surgery | Orthopedics | All items |
| STARD, 2003 | Freeman, 2009§§§ | United Kingdom | Government agency: European Commission funds allocated to Safe Activities For Everyone Network of Excellence under 6th Framework; not reported; not reported | Biochemical and laboratory research methods | Obstetrics and gynecology | All items |
| Mahoney, 2007§§§ | United States                   | Industry: LifeScan Inc; not reported; study funder | Diagnostic (glucose monitoring) | Endocrinology and metabolism | All items |
| Selman, 2011§§§ | United Kingdom               | Not reported; not reported; other: charitable foundation (Wellbeing of Women) and Medical Research Council/Royal College of Obstetricians and Gynaecologists Clinical Research Training Fellowship (authors declare no competing interests) | Diagnostic studies | Obstetrics and gynecology | Subset of items¶ |
| Smidt, 2006§§§ | Netherlands                    | Government agency: ZonMW; funder did not play role in study or manuscript**; authors declare no competing interests. | Diagnostic studies | Medicine, general and internal | Subset of items¶ |
| Coppus, 2006§ | Netherlands                     | Government agency: VIDI-program of ZonMW and charitable foundation: Scientific foundation of the Maxima Medical Center; not reported; not reported | Diagnostic studies | Reproductive biology | Subset of items§ |
| Johnson, 2007†† | United Kingdom                 | Not reported; not reported; not reported (authors declare no competing interests) | Diagnostic studies | Ophthalmology | Subset of items |
| Krzych, 2009∥∥ | Poland                           | Self financed; not applicable; not reported | Diagnostic studies | Cardiac and cardiovascular systems | Subset of items†† |
| Paranjathy, 2007∥∥ | United Kingdom                | No funding; not reported; authors state no information to disclose | Diagnostic studies | Ophthalmology | All items |
| STRICTA, 2002∥∥ | Hammerschlag, 2011∥∥ | United States | Not reported; not reported; personnel support from Oregon College of Oriental Medicine research department and Helfgott Research Institute of National College of Natural Medicine | Complementary Medicine | Unspecified | Subset of items¶ |
| STROBE, 2007 | Parsons, 2011∥∥∥               | United Kingdom | Not reported; not reported; not reported | Surgery | Orthopedics | All items |
| Delaney, 2010∥∥∥ | United States                 | Industry: Biomedical Excellence for Safer Transfusion collaborative (industry sponsored); not reported; authors declare no competing interests | Platelet transfusion | Hematology | Subset of items¶ |

*All included evaluations were published as full reports.
†2011 journal impact factor categories used for classification.
‡If authors of evaluations deemed particular guidance item to be “not applicable” to literature they were assessing, those items were excluded from analysis; for evaluations with zero or one studies in one comparison arm, those evaluations were removed from synthesis because that one arm would determine direction of effect.
§Included in quantitative analysis.
Table 2 (continued)

| Author, year* | Country of corresponding author | Sources of funding; role of funder; authors’ source(s) of support | Content focus | Specific medical or scientific specialty† | Extent of guideline assessed‡ |
|---------------|----------------------------------|---------------------------------------------------------------|---------------|------------------------------------------|-------------------------------|

‡As determined by authors of this review when comparing with published guidance.
**Specifically, funding agency did not play role in design or conduct of study, collection, management, analysis, or interpretation of data; or preparation, review, or approval of manuscript.
††Authors of evaluations indicated subset was assessed, but authors of this review determined smaller subset was analyzed when comparing with published guidance.
‡‡Unofficial extension of CONSORT reporting guideline.
| Author, year | Relevant studies for assessment (endorsing v non-endorsing) | Year of publication of assessed studies | Journals that published assessed studies | Two or more assessors for completeness of reporting* | No of items assessed as reported in methods section* | Comprehensive search strategy* | Balance of studies per comparison groups† |
|--------------|-------------------------------------------------------------|----------------------------------------|------------------------------------------|-----------------------------------------------|-----------------------------------------------|--------------------------------|------------------------------------------|
| BMJ economic guidelines, 1996 | | | | | | | |
| Herman, 2005‡‡ | 2 v 11 | 2003-04 | 1 v 10 | Unclear | High | Low | High |
| Jefferson, 1998** | 1 v 5 | 1997-98† | 1 v 1 | Unclear | Unclear | High | High |
| CONSORT extension for abstracts, 2008 | | | | | | | |
| Ghimire, 2014*** | 74 v 234 | 2010-12 | 2 v 4 | High | Unclear | Low | Low |
| CONSORT extension for harms, 2004 | | | | | | | |
| Haidch, 2011‡‡ | 25 v 77 | 2006 | 2 v 3 | High | High | High | Low |
| Turner, 2011‡‡ | 5 v 189 | 2009 | 5 v 104 | Low | High | Low | Low |
| Peron, 2013‡† | 43 v 282 | 2007-11 | 2 v 8 | Unclear | Low | Low | Low |
| Cornelius, 2013‡† | 1 v 6 | 2009 | 1 v 5 | High | High | High | High |
| Lee, 2008*** | 1 v 1 | 2005 | 1 v 1 | High | High | High | High |
| CONSORT extension for herbal interventions, 2006 | | | | | | | |
| Ernst, 2011‡† | 1 v 4 | 2009 | 1 v 3 | Unclear | High | Low | High |
| PRISMA, 2009 | | | | | | | |
| Tunis, 2013‡† | 13 v 48 | 2010-11 | 1 v 8 | High | High | Low | Low |
| Panic, 2013‡† | 30 v 30 | Jan-Oct 2012 | 6 v 10 | High | High | Low | Unclear |
| Fleming, 2013‡† | 20 v 2 | 2009-11 v 2010-11 | 2 v 1 | High | High | Low | Low |
| QUOROM, 1999 | | | | | | | |
| Biondi-Zoccai, 2006‡‡ | 1 v 6 | 2004 | 1 v 6 | High | High | Low | High |
| Poolman, 2007‡† | 1 v 6 | 2006 v 2005 | 1 v 5 | High | Unclear | Low | High |
| STARD, 2003 | | | | | | | |
| Freeman, 2009‡‡ | 3 v 9 | 2004-05 | 2 v 7 | Unclear | High | High | High |
| Mahoney, 2007‡‡ | 6 v 20 | 2003-05 | 4 v 13 | High | High | Low | High |
| Selman, 2011‡‡ | 14 v 36 | 2003-06 | 6 v 22 | High | Low | Low | Low |
| Smidt, 2006‡‡ | 95 v 46 | 2004 | 7 v 5 | High | High | Low | Low |
| Cuppas, 2006‡† | 8 v 19 | 2004 | 1 v 1 | Low | High | Unclear | High |
| Johnson, 2007‡† | 1 v 10 | 2005 | 1 v 4 | High | High | Low | High |
| Krzych, 2009‡† | 4 v 21 | 2004-06 | 2 v 16 | Unclear | High | Low | High |
| Paranjothy, 2007‡† | 1 v 8 | 2005-06 | 1 v 4 | High | High | Low | High |
| STRICTA, 2002 | | | | | | | |
| Hammerschlag, 2011‡‡ | 17 v 130 | 2002-05 | 3 v 64 | Low | High | Low | Unclear |
| STROBE, 2007 | | | | | | | |
| Parsons, 2011‡‡ | 9 v 38 | 2008-10 | 2 v 6 | Low | Unclear | Low | Low |
| Delaney, 2010‡† | 1 v 4 | 2008 | 1 v 3 | High | Low | High | High |

*High=high validity; low=low validity; unclear=unclear validity.
†Assessed once authors’ data reorganized into comparison groups.
‡Included in quantitative synthesis.
§Estimated based on information provided in article.
| Author, year                                                                 | Relevant studies for assessment (after v before endorsement) | Year of publication of assessed studies | Journals that published assessed studies | Two or more assessors for completeness of reporting* | No of items assessed as reported in methods section* | Comprehensive search strategy* | Balance of studies per journal in comparison groups† | Sampling took place in period following publication of guideline† |
|----------------------------------------------------------------------------|----------------------------------------------------------------|----------------------------------------|------------------------------------------|-----------------------------------------------------|-----------------------------------------------------|------------------------------------------|-------------------------------------------------|--------------------------------------------------|
| BMJ economic guidelines, 1996                                               |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Jefferson, 1998                                                              | 1 v 8                                                            | 1997-98 v 1994-95§                       | 1                                        | Unclear                                            | Unclear                                            | High                                     | High                                            | Low                                              |
| CONSORT extension for abstracts, 2006                                        |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Ghimire, 2014                                                                | 74 v 16                                                         | 2010-12 v 2005-07                        | 2                                        | High                                               | Unclear                                            | Low                                      | Low                                             | Low                                              |
| CONSORT extension for harms, 2004                                           |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Lee, 2008                                                                   | 1 v 2                                                            | 2005 v 1999-2000                        | 1                                        | High                                               | High                                               | High                                     | High                                            | Low                                              |
| PRISMA, 2009                                                                 |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Panic, 2013‡                                                                | 27 v 26                                                          | 2012 v 2008-11                          | 6                                        | High                                               | High                                               | Low                                      | Low                                             | Unclear                                          |
| Fleming, 2013‡                                                              | 14 v 12                                                          | 2009-11 v 2006-09                       | 1                                        | High                                               | High                                               | Low                                      | High                                            | Low                                              |
| QUOROM, 1999                                                                |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Hind, 2007‡‡                                                                | 13 v 15                                                          | 2005 v 2003                             | 1                                        | Low                                                | High                                               | Low                                      | High                                            | High                                              |
| STARD, 2003                                                                 |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Smidt, 2006‡‡                                                               | 95 v 78                                                          | 2004 v 2000                             | 7                                        | High                                               | High                                               | Low                                      | Unclear                                          | Low                                              |
| Selman, 2011†                                                               | 3 v 1                                                            | 2005-06 v 2003                          | 1                                        | High                                               | Low                                                | Low                                      | Low                                             | High                                              |
| STRICTA, 2002                                                               |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Hammerschlag, 2011‡‡                                                        | 11 v 4                                                           | 2003-05 v 1999-2001                    | 2                                        | Low                                                | High                                               | Low                                      | Unclear                                          | Low                                              |
| STROBE, 2007                                                                |                                                                  |                                        |                                          |                                                     |                                                     |                                          |                                                  |                                                  |
| Parsons, 2011‡‡                                                             | 9 v 11                                                           | 2008-10 v 2005-08                      | 2                                        | Low                                                | Unclear                                            | Low                                      | Low                                             | Low                                              |

*High=high validity; low=low validity; unclear=unclear validity.
†Assessed once authors' data reorganized into comparison groups.
‡Included in quantitative synthesis.
§Estimated based on information provided in article.
| Author, year | Endorsing journals that published assessed studies | Extent of endorsement | Date of endorsement provided |
|-------------|-------------------------------------------------|-----------------------|-----------------------------|
| BMJ economic guidelines, 1996                      |                                                  |                       |                             |
| Herman, 2005†‡ | BMJ                                             | Submit checklist       | By journal, email           |
| Jefferson, 1998‡ | BMJ                                             | Submit checklist       | By journal, email           |
| CONSORT extension for abstracts, 2008              |                                                  |                       |                             |
| Ghimire, 2014 ‡ | Lancet                                          | Suggests use           | By journal, email           |
| CONSORT extension for harms, 2004                  |                                                  |                       |                             |
| Haidich, 2011 ‡ | Annals of Internal Medicine                     | Submit checklist       | By journal, email           |
|            | The Lancet                                      | Submit checklist       | By journal, email           |
| Turner, 2011 ‡ | The American Journal of Gastroenterology        | Submit checklist       | By journal, email           |
|            | American Journal of Kidney Diseases             | Suggests use           | By journal, email           |
|            | Applied Health Economics and Health Policy      | Suggests use           | By journal, email           |
|            | JAMA                                             | Submit checklist       | Not provided                |
|            | Phytomedicine                                   | Suggests use           | Not provided                |
| Peron, 2014 ‡ | Lancet                                           | Submit checklist       | By journal, email           |
|            | Lancet Oncology                                 | Submit checklist       | By journal, email           |
| Cornelius, 2013 ‡ | Lancet Oncology                               | Submit checklist       | By journal, email           |
| Lee, 2008 ‡ | BMJ                                             | Submit checklist       | By journal, email           |
| CONSORT extension for herbal interventions, 2006  |                                                  |                       |                             |
| Ernst, 2011 ‡ | Annals of Internal Medicine                     | Suggests use           | Not provided                |
| PRISMA, 2009                                       |                                                  |                       |                             |
| Panic, 2013 ‡ | Radiology                                       | Suggests use           | Unknown based on information given |
|            | Alimentary Pharmacology and Therapeutics        | Extent of endorsement at time of author’s analysis unknown (all journals) | Provided by author (all journals) |
|            | American Journal of Gastroenterology            |                       |                             |
|            | BMC Gastroenterology                            |                       |                             |
|            | Colorectal Disease                              |                       |                             |
|            | Diseases of the Colon and Rectum               |                       |                             |
|            | Gut                                             |                       |                             |
|            | Gut Pathogens                                   |                       |                             |
|            | Hepatitis Monthly                               |                       |                             |
|            | HPB                                             |                       |                             |
| Fleming, 2013 ‡ | American Journal of Orthodontics and Dentofacial Orthopedics | Submit checklist | By journal, email           |
|            | Angle Orthodontist                              | Suggests use           | Not provided                |
|            | European Journal of Orthodontics                | Submit checklist       | By journal, email           |
|            | Journal of Orthodontics                         | Suggests use           | By journal, email           |

*Evaluations included in quantitative analysis.
†Endorsing versus non-endorsing journals comparison only.
| Author, year | Endorsing journals that published assessed studies | Extent of endorsement | Date of endorsement provided |
|--------------|-----------------------------------------------|----------------------|----------------------------|
| **QUOROM, 1999** | | | |
| Hind, 2007**‡† | UK NHS Health Technology Assessment Programme | Submit checklist | By evaluation |
| Biondi-Zoccai, 2006**‡‡ | Clinical Cardiology | Unknown based on information given | Unknown based on information given |
| Poolman, 2007**‡‡ | BMJ | Suggests use | Not provided |
| **STARD, 2003** | | | |
| Freeman, 2009**‡‡ | American Journal of Obstetrics and Gynecology | Submit checklist | Unknown based on information given |
| | Molecular Diagnosis§ | Suggests use | Not provided |
| Mahoney, 2007**‡‡ | Archives of Disease in Childhood (including Fetal and Neonatal Edition) | Suggests use | Unknown based on information given |
| | Clinical Biochemistry | Suggests use | Not provided |
| | Emergency Medicine Journal | Suggests use | Unknown based on information given |
| | Journal of the Medical Association of Thailand | Suggests use | Not provided |
| Selman, 2011**‡‡ | American Journal of Obstetrics and Gynecology† | Submit checklist | Unknown based on information given |
| | Cancer† | Suggests use | Not provided |
| | Clinical Radiology† | Suggests use | Not provided |
| | Journal of the Medical Association of Thailand† | Suggests use | Not provided |
| | Obstetrics and Gynecology | Suggests use | By journal, email |
| | Radiology† | Suggests use | By journal website |
| Smitd, 2006**‡‡ | Annals of Internal Medicine | Suggests use | Journal website or by evaluation (all journals) |
| | BMJ | Suggests use | |
| | Clinical Chemistry | Submit checklist | |
| | JAMA | Suggests use | |
| | The Lancet | Submit checklist | |
| | Neurology | Submit checklist | |
| | Radiology | Suggests use | |
| Coppus, 2006**‡‡ | Human Reproduction | Journal no longer endorses guideline | |
| Johnson, 2007**‡‡ | Ophthalmic and Physiologic Optics | Submit checklist | By journal, email |
| Krzych, 2009**‡‡ | Clinical Chemistry** | Submit checklist | Reported in another evaluation |
| | Heart | Suggests use | Not provided |
| Paranjathy, 2007**‡‡ | British Journal of Ophthalmology | Suggests use | Not provided |
| **STRICTA, 2002** | | | |
| Hammerschlag, 2011**‡‡ | Acupuncture in Medicine | Suggests use | By journal, email |
| | Journal of Alternative and Complementary Medicine | Suggests use | By journal, email |
| | Medical Acupuncture† | Suggests use | By journal, email |
| **STROBE, 2007** | | | |
| Parsons, 2011**‡‡ | Clinical Orthopaedics and Related Research | Suggests use | By journal, email |
| | The Journal of Bone and Joint Surgery (American) | Suggests use | By journal, email |
| Delaney, 2010**‡‡ | Annals of Surgery | Suggests use | Not provided |

*Evaluations included in quantitative analysis.
†After versus before journal endorsement comparison only.
‡Endorsing versus non-endorsing journals comparison only.
§Now published as Molecular Diagnosis and Therapy.
¶In quantitative analysis for endorsing versus non-endorsing journals only.
**Reported in another included evaluation.
Table 7 | Analysis by mean summed score of items for reporting guideline checklists, endorsing versus non-endorsing journals

| Reporting guideline† | No of evaluations‡ | No of studies (total) | Effect estimate (99% CI) |
|----------------------|-------------------|-----------------------|-------------------------|
| CONSORT extension for harms, 2004 | 1§ | 25 v 77 (102) | Mean difference 0.04 (~1.50 to 1.58) |
| PRISMA, 2009 | 3¶ | 63 v 80 (143) | Standardized mean difference 0.53 (0.02 to 1.03) |
| STARD, 2003 | 3** | 23 v 65 (88) | Standardized mean difference 0.52 (~0.11 to 1.16) |
| STRICTA, 2002 | 1†† | 17 v 130 (147) | Mean difference 1.42 (~0.04 to 2.88) |
| STROBE, 2007 | 1§ | 9 v 38 (47) | Mean difference 1.55 (~3.19 to 6.29) |

*Individual forest plots depicting these summary data are shown in appendix 7.
†QUOROM (two evaluations) was not estimable because of one study in one comparison arm per assessed evaluation.
‡Only evaluations that calculated summed score for report were included.
§All checklist items summed.
¶Subset of items was summed for one evaluation.
**Subset of items was summed for two of three evaluations.
††Subset of items was summed.
Table 8 | Analysis by mean summed score for reporting guideline checklists, after versus before journal endorsement*

| Reporting guideline | No of evaluations† | No of studies (total) | Effect estimate (99% CI) |
|---------------------|-------------------|----------------------|-------------------------|
| PRISMA, 2009        | 2‡                | 41 v 38 (79)         | Standardized mean difference 0.49 (–0.10 to 1.08) |
| STRICTA, 2002       | 1§                | 11 v 4 (15)          | Mean difference 1.82 (–2.49 to 6.13) |
| STROBE, 2007        | 1‡                | 9 v 11 (20)          | Mean difference 1.16 (–3.97 to 6.29) |

*Individual forest plots depicting these summary data are shown in appendix 7.
†Only evaluations that calculated summed score for report were included.
‡All checklist items were summed.
§Subset of items was summed.
| Author, year | Methodological quality assessment |
|-------------|----------------------------------|
| BMJ economic guidelines, 1996 | Evaluated economic evaluations on four criteria: randomization; prospective economic data collection; comparison group was usual care; and study was not blinded or mandatory regarding participation. Both studies in endorsing arm met all four criteria compared with 5/11 studies in non-endorsing arm. |
| Herman, 2005 | Assessed studies by using Cochrane risk of bias tool. Only study from endorsing journal was assessed as at moderate risk of bias. Studies from non-endorsing journals were assessed at high (n=2) or moderate (n=2) risk of bias. |
| Ernst, 2005 | Assessed studies by using the Oxman and Guyatt index (range of 1 (minimal flaws) to 7 (extensive flaws)). Only study from endorsing journal scored 2 on index; studies (n=6) from non-endorsing journals scored range of 1-6 points. |
| ERNST, 2011 | Assessed reviews by using AMSTAR. Data by item are not presented. Endorsing versus non-endorsing journals: using data provided by author, mean summed score was 7.3 (range 3-9, n=27 articles) after journal endorsement and 6.0 (range 0-9, n=26 articles) before endorsement. |
| PRISMA, 2009 | Assessed studies by using the Oxman and Guyatt index (range of 1 (minimal flaws) to 7 (extensive flaws)). Only study from endorsing journal scored 7 points. Studies from non-endorsing journals (n=6) scored range of 1-6 points; four studies scoring 1 or 2 points are considered to have “major flaws” according to index. |
| Tunis, 2013 | Assessed reviews by using AMSTAR. Data by item are not presented. Endorsing versus non-endorsing journals: using data provided by author, mean summed score was 7.6 of 11 points. |
| Panic, 2013 | Assessed reviews by using AMSTAR. Data by item are not presented. Endorsing versus non-endorsing journals: using data provided by author, mean summed score was 6.4 (range 1-9). After versus before journal endorsement: using data provided by author, mean summed score was 7.6 of 11 points. |
| Fleming, 2013 | Authors assessed reviews by using AMSTAR tool but analyzed across all included studies. |
| QUOROM, 1999 | Assessed studies by using the Oxman and Guyatt index (range of 1 (minimal flaws) to 7 (extensive flaws)). Only study from endorsing journal scored 2 on index; studies (n=6) from non-endorsing journals scored range of 1-6 points. |
| Biondi-Zoccai, 2006 | Assessed studies by using the Oxman and Guyatt index (range of 1 (minimal flaws) to 7 (extensive flaws)). Only study from endorsing journal scored 7 points. Studies from non-endorsing journals (n=6) scored range of 1-6 points; four studies scoring 1 or 2 points are considered to have “major flaws” according to index. |
| Poolman, 2007 | Used the Oxman and Guyatt index (maximum score 7 points). Only study from endorsing journal scored 7 points. Studies from non-endorsing journals (n=6) scored range of 1-6 points; four studies scoring 1 or 2 points are considered to have “major flaws” according to index. |
| STARD, 2003 | Assessed eight aspects that authors state address internal and external validity of included studies: selective participant sampling; lack of reporting number of replicates, if done, that were used for overall study outcome; lack of reporting failure rate; lack of including reported failure rate in analysis; difference in reported and adjusted accuracy; lack of controlling for presence of fetal DNA; and lack of known genotypes in study as control. Raw data provided in tabular form without summary in text. Studies (n=3) from endorsing journals ranged from 2 to 4 of 8 flaws. Studies (n=8) from non-endorsing journals ranged from 2 to 6 flaws, and information from one study was not interpretable. |
| Freeman, 2009 | Assessed studies by using QUADAS tool but analyzed across all included studies. |
| Kczycz, 2009 | Authors assessed studies by using QUADAS tool but analyzed across all included studies. |
Figures

Fig 1 Schematic depicting relation among evaluation of reporting guideline, studies contained within it, and determination of comparison groups according to journal endorsement status.

Fig 2 PRISMA flow diagram for selecting reporting guidelines for health research. RG=reporting guideline.
Fig 3 PRISMA flow diagram for selecting evaluations of relevant reporting guidelines. RG=reporting guideline; SR=systematic review
**Fig 4** Completeness of reporting summary plot for BMJ economics checklist, endorsing versus non-endorsing journals. Summary plots in this and other related figures were generated in Comprehensive Meta-analysis. In brief, summary effect estimates for each checklist item are shown, and those estimates were previously calculated in Review Manager. For example, checklist item "economic importance of question" was assessed in only one evaluation, which had 13 studies (2 studies from endorsing journal and 11 studies from non-endorsing journals; appendix 7) that provided information on whether study had reported on that checklist item. Appendix 7 shows analyses for each checklist item conducted in Review Manager.

**Fig 5** Completeness of reporting summary plot for CONSORT extension for harms checklist, endorsing versus non-endorsing journals.
Fig 6 Completeness of reporting summary plot for PRISMA checklist, endorsing versus non-endorsing journals. Although all evaluations assessed all items, one evaluation was excluded from analysis of two checklist items because of zero or one studies for analysis.
### Fig 7
Completeness of reporting summary plot for STARD checklist, endorsing versus non-endorsing journals. Effect estimate for checklist item “Test methods: definition of cut-offs of index test and reference standard” was not estimable during quantitative analysis because of zero events in each arm (one evaluation in analysis).

| STARD checklist items | No of evaluations | No of studies | Risk ratio (99% CI) |
|-----------------------|-------------------|---------------|---------------------|
| Title, abstract, keywords | 2 | 62 | 1.13 (0.55 to 2.33) |
| Introduction | 3 | 88 | 1.06 (0.84 to 1.33) |
| Participants - study population | 3 | 88 | 1.36 (0.55 to 3.38) |
| Participants - recruitment | 3 | 88 | 1.06 (0.67 to 1.68) |
| Participants - sampling | 3 | 88 | 0.78 (0.45 to 1.36) |
| Participants - data collection | 3 | 88 | 0.94 (0.74 to 1.19) |
| Test methods - reference standard | 3 | 88 | 1.02 (0.79 to 1.32) |
| Test methods - technical specifications | 2 | 38 | 1.00 (0.61 to 1.64) |
| Test methods - persons reading test and reference | 2 | 38 | 1.67 (0.60 to 4.64) |
| Test methods - readers blinded to other result? | 2 | 38 | 0.83 (0.02 to 38.04) |
| Statistical methods - measures and uncertainty | 2 | 60 | 2.42 (1.05 to 5.59) |
| Statistical methods - test reproducibility | 2 | 38 | 0.56 (0.12 to 2.63) |
| Results - recruitment | 3 | 87 | 0.81 (0.51 to 1.39) |
| Results - participant characteristics | 3 | 88 | 1.73 (1.13 to 2.64) |
| Results - participant flow | 4 | 227 | 1.33 (0.86 to 2.05) |
| Results - time interval from test to reference | 3 | 85 | 0.64 (0.26 to 1.57) |
| Results - disease severity | 3 | 88 | 0.73 (0.08 to 6.66) |
| Results - cross-tabulation of test by reference results | 3 | 88 | 1.26 (0.71 to 2.22) |
| Results - any adverse events | 3 | 68 | 1.42 (0.34 to 5.98) |
| Results - diagnostic accuracy estimates and uncertainty | 2 | 60 | 1.85 (0.72 to 4.76) |
| Results - how indeterminate results, missing data, outliers handled | 3 | 79 | 1.41 (0.82 to 2.42) |
| Results - variability between subgroups | 2 | 29 | 1.26 (0.75 to 2.11) |
| Results - test reproducibility | 2 | 38 | 0.56 (0.12 to 2.63) |
| Discussion | 3 | 88 | 1.00 (0.89 to 1.13) |

### Fig 8
Completeness of reporting summary plot for STRICHTA checklist, endorsing versus non-endorsing journals.

| STRICHTA checklist items | No of evaluations | No of studies | Risk ratio (99% CI) |
|--------------------------|-------------------|---------------|---------------------|
| Style of acupuncture | 1 | 146 | 0.70 (0.35 to 1.39) |
| Rationale for treatment | 1 | 147 | 1.04 (0.72 to 1.51) |
| Sources to justify rationale | 1 | 147 | 1.41 (1.00 to 1.99) |
| Points used unilateral/bilateral | 1 | 146 | 0.98 (0.72 to 1.34) |
| Numbers of needles inserted | 1 | 146 | 1.05 (0.72 to 1.53) |
| Depths of insertion | 1 | 145 | 1.17 (0.62 to 2.21) |
| Responses elicited | 1 | 143 | 1.15 (0.72 to 1.84) |
| Needle stimulation | 1 | 146 | 1.60 (1.25 to 2.04) |
| Needle retention time | 1 | 147 | 1.11 (0.86 to 1.44) |
| Needle type | 1 | 147 | 1.48 (0.93 to 2.36) |
| Number of treatment sessions | 1 | 147 | 1.04 (0.92 to 1.17) |
| Frequency of treatment | 1 | 120 | 1.06 (0.86 to 1.31) |
| Other interventions | 1 | 29 | 0.87 (0.29 to 2.59) |
| Duration of relevant training | 1 | 147 | 1.39 (0.67 to 2.89) |
| Length of clinical experience | 1 | 147 | 1.70 (0.65 to 4.43) |
| Expertise in specific condition | 1 | 146 | 1.52 (0.34 to 6.76) |
| Explanations given regarding treatment and control interventions | 1 | 147 | 1.13 (0.34 to 3.78) |
| Sources that justify choice of control | 1 | 145 | 1.11 (0.59 to 2.09) |
**STROBE checklist items**

| STROBE checklist items                        | No of evaluations | No of studies | Risk ratio (99% CI) | Risk ratio (99% CI) |
|-----------------------------------------------|-------------------|---------------|---------------------|---------------------|
| Title or abstract                              | 1                 | 47            | 0.40 (0.08 to 2.09) |                    |
| Abstract                                       | 1                 | 47            | 1.02 (0.73 to 1.43) |                    |
| Introduction - background and rationale        | 1                 | 47            | 1.02 (0.82 to 1.26) |                    |
| Introduction - objectives                      | 1                 | 47            | 1.09 (0.64 to 1.86) |                    |
| Methods - study design                         | 1                 | 47            | 0.94 (0.68 to 1.29) |                    |
| Methods - setting, location, dates            | 1                 | 47            | 1.06 (0.53 to 2.11) |                    |
| Methods - participant eligibility              | 1                 | 47            | 0.99 (0.71 to 1.39) |                    |
| Methods - participant matching                 | 1                 | 24            | 1.50 (0.55 to 4.10) |                    |
| Methods - outcome, exposure, other variables   | 1                 | 47            | 0.88 (0.38 to 2.03) |                    |
| Methods - data sources and measurement         | 1                 | 47            | 0.91 (0.67 to 1.24) |                    |
| Methods - addressing sources of bias           | 1                 | 47            | 0.84 (0.15 to 4.87) |                    |
| Methods - study size                           | 1                 | 47            | 1.58 (0.73 to 3.43) |                    |
| Methods - handling of quantitative variables   | 1                 | 47            | 0.92 (0.56 to 1.51) |                    |
| Methods - statistical methods                  | 1                 | 47            | 0.92 (0.56 to 1.51) |                    |
| Methods - subgroups and interactions           | 1                 | 47            | 1.54 (0.48 to 4.93) |                    |
| Methods - loss to follow-up, matching, sampling| 1                 | 42            | 0.81 (0.14 to 4.76) |                    |
| Methods - sensitivity analyses                 | 1                 | 47            | 1.21 (0.19 to 7.53) |                    |
| Results - participant flow                     | 1                 | 47            | 0.99 (0.71 to 1.39) |                    |
| Results - reasons for nonparticipation         | 1                 | 47            | 0.84 (0.06 to 12.01)|                    |
| Results - flow diagram                         | 1                 | 47            | 2.81 (0.33 to 24.14)|                    |
| Results - participant characteristics          | 1                 | 47            | 1.61 (0.99 to 2.61) |                    |
| Results - missing data                         | 1                 | 47            | 0.70 (0.13 to 3.93) |                    |
| Results - follow-up time (cohort studies)      | 1                 | 10            | 1.08 (0.56 to 2.08) |                    |
| Results - outcome data                         | 1                 | 47            | 0.95 (0.58 to 1.56) |                    |
| Results - estimates and precision              | 1                 | 47            | 0.82 (0.43 to 1.56) |                    |
| Results - boundaries for continuous variable categories | 1            | 44            | 0.49 (0.04 to 6.27) |                    |
| Results - translating relative into absolute risks | 1           | 35            | 1.04 (0.02 to 61.88)|                    |
| Results - other analyses                       | 1                 | 46            | 1.23 (0.30 to 5.02) |                    |
| Discussion - key results                       | 1                 | 47            | 0.99 (0.81 to 1.21) |                    |
| Discussion - limitations                       | 1                 | 47            | 1.51 (1.05 to 2.18) |                    |
| Discussion - interpretation                    | 1                 | 47            | 1.26 (0.93 to 1.70) |                    |
| Discussion - generalizability                  | 1                 | 47            | 1.35 (0.88 to 2.07) |                    |
| Other - funding                                | 1                 | 47            | 1.47 (0.93 to 2.32) |                    |

**Fig 9** Completeness of reporting summary plot for STROBE checklist, endorsing versus non-endorsing journals. Effect estimate for checklist item “Methods: missing data” was not estimable during quantitative analysis because of zero events in each arm.
Fig 10 Completeness of reporting summary plot for PRISMA checklist, after versus before journal endorsement. Although all evaluations assessed all items, one evaluation was excluded from analysis of one checklist item because of zero and one studies for comparison arms.

Fig 11 Completeness of reporting summary plot for QUOROM checklist, after versus before journal endorsement.

Fig 12 Completeness of reporting summary plot for STARD checklist, after versus before journal endorsement.
### Fig 13 Completeness of reporting summary plot for STRICTA checklist, after versus before journal endorsement

| STRICTA checklist items                              | No of evaluations | No of studies | Risk ratio (99% CI) | Risk ratio (99% CI) |
|------------------------------------------------------|-------------------|---------------|---------------------|---------------------|
| Style of acupuncture                                 | 1                 | 15            | 0.91 (0.19 to 4.31) |                     |
| Rationale for treatment                              | 1                 | 15            | 1.64 (0.43 to 6.25) |                     |
| Sources to justify rationale                         | 1                 | 15            | 1.64 (0.43 to 6.25) |                     |
| Points used (uni/bilateral)                          | 1                 | 15            | 1.00 (0.66 to 1.51) |                     |
| Numbers of needles inserted                          | 1                 | 15            | 1.82 (0.49 to 6.75) |                     |
| Depths of insertion                                  | 1                 | 15            | 1.82 (0.17 to 19.64)|                     |
| Responses elicited                                   | 1                 | 15            | 1.64 (0.43 to 6.25) |                     |
| Needle stimulation                                   | 1                 | 15            | 3.19 (0.55 to 18.60)|                     |
| Needle retention time                                | 1                 | 15            | 0.97 (0.60 to 1.56) |                     |
| Needle type                                           | 1                 | 15            | 1.64 (0.43 to 6.25) |                     |
| Number of treatment sessions                         | 1                 | 15            | 1.00 (0.66 to 1.51) |                     |
| Frequency of treatment                               | 1                 | 12            | 0.97 (0.54 to 1.75) |                     |
| Duration of relevant training                        | 1                 | 15            | 1.27 (0.31 to 5.22) |                     |
| Length of clinical experience                        | 1                 | 15            | 1.45 (0.12 to 17.24)|                     |
| Expertise in specific condition                      | 1                 | 15            | 2.92 (0.08 to 109.09)|                    |
| Explanations given regarding treatment and control   | 1                 | 14            | 0.55 (0.04 to 7.72) |                     |
| Sources that justify choice of control               | 1                 | 15            | 0.73 (0.26 to 2.04) |                     |
**Fig 14** Completeness of reporting summary plot for STROBE checklist, after versus before journal endorsement. Effect estimate for checklist item “Methods: missing data” was not estimable during quantitative analysis because of zero events in each arm.