

RESEARCH

Quality of reporting in systematic reviews of adverse events: systematic review

 OPEN ACCESS

Liliane Zorzela *PhD candidate*¹, Su Golder *MRC fellow in health services research*², Yali Liu *lecturer*³, Karen Pilkington *senior research fellow*⁴, Lisa Hartling *assistant professor*⁵, Ari Joffe *clinical professor*⁶, Yoon Loke *clinical senior lecturer*⁷, Sunita Vohra *professor*⁸

¹Department of Pediatrics, 4-548 Edmonton Clinic Health Academy, University of Alberta, Edmonton, Canada; ²Centre for Reviews and Dissemination, University of York, York, UK; ³Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Chengguan District, Lanzhou, Gansu, People's Republic of China; ⁴School of Life Sciences, University of Westminster, London, UK; ⁵Department of Pediatrics, 4-472 Edmonton Clinic Health Academy, University of Alberta, Canada; ⁶Stollery Children's Hospital, Department of Pediatrics, University of Alberta, Edmonton, Canada; ⁷University of East Anglia Medical School, Norwich, UK; ⁸Department of Pediatrics, University of Alberta, Canada, T5K 0L4

Abstract

Objectives To examine the quality of reporting of harms in systematic reviews, and to determine the need for a reporting guideline specific for reviews of harms.

Design Systematic review.

Data sources Cochrane Database of Systematic Reviews (CDSR) and Database of Abstracts of Reviews of Effects (DARE).

Review methods Databases were searched for systematic reviews having an adverse event as the main outcome, published from January 2008 to April 2011. Adverse events included an adverse reaction, harms, or complications associated with any healthcare intervention. Articles with a primary aim to investigate the complete safety profile of an intervention were also included. We developed a list of 37 items to measure the quality of reporting on harms in each review; data were collected as dichotomous outcomes ("yes" or "no" for each item).

Results Of 4644 reviews identified, 309 were systematic reviews or meta-analyses primarily assessing harms (13 from CDSR; 296 from DARE). Despite a short time interval, the comparison between the years of 2008 and 2010-11 showed no difference on the quality of reporting over time ($P=0.079$). Titles in fewer than half the reviews (proportion of reviews 0.46 (95% confidence interval 0.40 to 0.52)) did not mention any harm related terms. Almost one third of DARE reviews (0.26 (0.22 to 0.31)) did not clearly define the adverse events reviewed, nor did they specify the study designs selected for inclusion in their methods section. Almost half of reviews ($n=170$) did not consider patient risk factors or length of follow-up when reviewing harms of an intervention. Of 67 reviews of complications related to surgery or other procedures, only four (0.05 (0.01 to 0.14)) reported professional qualifications of the individuals involved. The overall, unweighted, proportion of reviews with

good reporting was 0.56 (0.55 to 0.57); corresponding proportions were 0.55 (0.53 to 0.57) in 2008, 0.55 (0.54 to 0.57) in 2009, and 0.57 (0.55 to 0.58) in 2010-11.

Conclusion Systematic reviews compound the poor reporting of harms data in primary studies by failing to report on harms or doing so inadequately. Improving reporting of adverse events in systematic reviews is an important step towards a balanced assessment of an intervention.

Introduction

A balanced assessment of interventions requires analysis of both benefits and harms. Systematic reviews or meta-analyses of randomised controlled trials are the preferred method to synthesise evidence in a comprehensive, transparent, and reproducible manner. Randomised controlled trials rarely assess harms as their primary outcome; therefore, they typically lack the power to detect differences in harms between groups (table 1). Usually designed to evaluate treatment efficacy or effectiveness, randomised controlled trials are often done over a short period of time, with a relatively small number of participants. These trials are known to be poor at identifying and reporting harms, which can lead to a misconception that a given intervention is safe, when its safety is actually unknown.¹⁻⁸ Systematic reviews with a primary objective to assess harms represent fewer than 10% of all systematic reviews published yearly.⁹⁻¹⁰ Systematic reviews of harms can provide valuable information to describe adverse events (frequency, nature, seriousness), but they are hampered by a lack of standardised

Correspondence to: S Vohra svohra@ualberta.ca

Extra material supplied by the author (see <http://www.bmj.com/content/348/bmj.f7668?tab=related#webextra>)

Web appendix: Search strategy

methods to report these events and the fact that harms are not usually the primary outcome of included studies.⁹⁻²⁸

Several studies have identified challenges when developing a systematic review of adverse events.⁹⁻²⁸ These include: the poor quality information on harms reported on original studies, difficulties in identifying relevant studies on adverse events when using standard systematic searches techniques, and the lack of a specific guideline to perform a systematic review of adverse events. The need for better reporting on harms in general¹⁻⁸—and in systematic reviews⁹⁻²⁸ in particular—has been voiced. In a previous review²⁸ of systematic reviews from the Cochrane Database of Systematic Reviews (CDSR) and Database of Abstracts of Reviews of Effects (DARE), our team identified a significantly increased number of reviews of adverse events over the past 17 years ($P < 0.001$); however, the proportion of these reviews out of the total number of reviews was unchanged at 5%.^{11 14 28} Some positive points were noted—for example, the increased number of databases searched per review and the reduction in number of systematic reviews limiting their search strategies by date or language—but appropriate reporting of search strategies was still a problem.²⁸

The PRISMA³⁰ (preferred reporting items for systematic review and meta-analysis) statement was developed to deal with suboptimal reporting in systematic reviews. Thus far, PRISMA has mainly focused on efficacy and not on harms. A reporting guideline specific for systematic reviews of harms is crucial to provide a better assessment of adverse events of interventions. The first step for successful guideline development is to document the quality of reporting in published research articles to justify the need for the guideline.³⁰ The goal of this review was to determine whether there is a need for a guideline specific for reviews of harms³⁰ through assessment of the quality of reporting in systematic reviews of harms published between January 2008 and April 2011, from two major databases.

Methods

Development of the checklist items

To assess the quality of reporting in systematic reviews of harms, we developed a set of items (table 2) to be reported in these reviews.

The items were originally based on a draft generated from analysis of a systematic review of harms conducted previously.³³ During the development of the data extraction form for this current review, several items were added. The wording and content were further refined over telephone meetings, by a group of experts in systematic reviews and guideline development.^{8 31 32 34-39}

Not every PRISMA item has a corresponding harms item, and a few have more than one suggestion per item. PRISMA items 15, 19-25, and 27 did not have any specific harms related items.

Search strategy

We searched the CDSR (via the Cochrane Library) and DARE (via the Centre for Reviews and Dissemination and the Cochrane Library) databases for systematic reviews having an adverse event as the primary outcome measured. DARE is compiled through rigorous weekly searches of bibliographic databases (including Medline, Embase, PsycINFO, PubMed, and the Cumulative Index to Nursing and Allied Health Literature). It also involves less frequent searches of the Allied and Complementary Medicine Database and the Education Resource Information Center, hand searching of key journals, grey literature, and regular searches of the internet. CDSR includes

all the systematic reviews published by the Cochrane Collaboration. We selected the combination of these two databases because they are likely to represent the most comprehensive collection of systematic reviews published in healthcare.²⁸

The search was limited to a 40 month period between 1 January 2008 and 25 April 2011. The web appendix shows the search strategy used. The dates were selected to include recent reviews in order to describe the current state of reporting in systematic reviews of harms.

Eligibility criteria

Reviews were selected if the primary outcome investigated was exclusively an unintended effect or effects of an intervention. It could be an adverse event, adverse effect, adverse reaction, harms, or complications (table 1) associated with any healthcare intervention (such as pharmaceutical interventions, diagnostic procedures, surgical interventions, or medical devices). Articles with a primary aim to investigate the complete safety profile of an intervention were included. Reviews were not excluded on the basis of their results or conclusions.

We excluded reviews assessing both beneficial and harmful effects (reviews of both efficacy and harms), reviews of desirable side effects of drugs, or reviews of prevention or reduction of unintended or adverse effects. No limitations on interventions, patient groups, or language were applied.

Screening and data extraction

Relevant studies were screened by title and abstract (when available) independently by two authors (LZ and SG). Any disagreements were resolved by consensus; if consensus could not be reached, disagreements were resolved with a third author (SV). Abstracts that were identified as potentially meeting the DARE criteria but were not assessed were called “provisional abstracts.” The full text was retrieved for these abstracts.

The data extraction was based on the items developed, piloted, and refined (table 2). Each field received a “yes” if the item was reported as defined, or “no” if not reported. The data were extracted by one author (LZ) and verified by a second author (YL). Disagreements were resolved by consensus.

Outcome

The main outcome assessed was the quality of reporting in reviews of harms for each of the 37 items. We also measured the proportion of each “yes” response for each year of the search: 2008, 2009, and 2010-11 (12 reviews published in the first four months of 2011 were combined with the reviews published in 2010). The quality of reporting was compared between the earliest and latest years reviewed (2008 v 2010-11) to assess any improvement in quality of reporting during the study period.

We deliberately decided not to include an intermediate category (such as “unclear”). If the item was not clearly reported, it was considered as a “no” response and the unclear category would simply be duplication.

The present review did not aim to evaluate the reason behind the review author’s decisions to examine harms. Our goal was to measure the quality of reporting on those reviews—answering the question “is the item clearly reported?” The goal was not to judge whether a methodologically appropriate decision was made (for example, statistical tests used, data extraction, data pooling), but to ensure clarity in reporting regarding the choices made. Information on the types of interventions, nature of included study designs, as well as search

strategies and databases searched in systematic reviews published between 1994 and 2011 was previously reported by our team.²⁸

This study does not intend to measure the effect of the PRISMA statement, for two reasons. Firstly, PRISMA focuses on efficacy; thus measurement of its effect would reasonably focus on systematic reviews of efficacy and not specifically of harms. Secondly, the 37 items measured in the present review were new items and not those found in PRISMA.

Data analysis

Data were collected as dichotomous outcomes (“yes” or “no”) for each item, and presented as proportions of reviews for each category (reported from 0 to 1) and as proportions divided by the database that the reviews were identified from (CDSR and DARE). We also provided 95% confidence intervals for each proportion based on methods described by Wilson and Newcombe^{40 41} using a correction for continuity.

An overall reporting quality rate was provided through an unweighted average of proportions of items with good reporting. $P \leq 0.05$ was considered statistically significant. We did statistical calculations using StataIC-13.

Results

The search yielded 4644 unique references. After screening and retrieving full text articles, an extra 14 papers were excluded as they did not fulfil the inclusion criteria. A total of 309 reviews were identified as systematic reviews of meta-analyses primarily assessing harms, of which 13 were identified at CDSR and 296 at DARE (fig 1). Disagreements at the inclusion and exclusion criteria were discussed by LZ and SG and consensus was reached after discussion. Three of the included papers were published in Chinese, two in Spanish, and one in Portuguese. Table 3 provides detailed information on the reporting of each item.

The 309 systematic reviews and meta-analyses with harms as a primary outcome focused on the following interventions:

- Drugs (223 studies, proportion of reviews 0.72 (95% confidence interval 0.66 to 0.77))
- Surgery or other procedures (67 studies, 0.21 (0.17 to 0.26))
- Devices (13 studies, 0.04 (0.02 to 0.07))
- Blood transfusion (two studies, 0.006 (0.001 to 0.022))
- Enteral nutrition (two studies, 0.006 (0.001 to 0.022))
- Isolation rooms (one study, 0.003 (0.001 to 0.01))
- Surgical versus medical treatment (one study, 0.003 (0.001 to 0.01)).

Titles and abstracts

Titles in close to half of included systematic reviews and meta-analyses of harms did not mention any harm related terms (proportion of reviews 0.46 (95% confidence interval 0.40 to 0.52)) and had no report of a patient population or condition under review (0.42 (0.36 to 0.48)). Twenty five reviews (0.08 (0.05 to 0.11)) used the word “safety” to identify a review of harms. In the abstract section, reviews often used harm related terms (0.84 (0.79 to 0.88)). For other terms, one in every 6.5 reviews did not have any harms related word in the abstract, and half of reviews (0.5 (0.44 to 0.56)) did not report the study designs sought or included.

Introduction and rationale

Introductions were well written overall, explaining the rationale for the review and providing information on harms being reviewed. Fifty one reviews (proportion of reviews 0.16 (95% confidence interval 0.12 to 0.21)) were performed to investigate any adverse event associated with an intervention, rather than focusing on a specific event. As per our definition, these reviews provided “an explicit statement of questions being asked with reference to harms” as this broad goal was reported.

Methods

Protocol and registration

Consistent with Cochrane requirements for authors, all included systematic reviews conducted through the Cochrane Collaboration (Cochrane reviews) had a protocol. Cochrane reviews did not refer to a protocol in their full text reviews, but this item was considered “yes” for the reviews for which a protocol could be found. By contrast, only 22 reviews (proportion of reviews 0.07 (95% confidence interval 0.04 to 0.11)) identified through the DARE database reported the use of a protocol. Reporting clinical expertise was not deemed necessary to receive “yes” for this item (table 2).

Eligibility criteria

Almost one third of DARE reviews (proportion of reviews 0.26 (95% confidence interval 0.22 to 0.31)) did not have a clear definition of the adverse events reviewed, nor did they specify the study designs selected for inclusion in their methods section. All Cochrane reviews had a clear report of their search strategy, eligible study designs, and methods of data extraction.

Information sources

Authors did not usually search outside the peer reviewed literature for additional sources of adverse events; for example, only 54 reviews (proportion of reviews 0.17 (95% confidence interval 0.13 to 0.22)) searched databases of regulatory bodies or similar sources. Seven of 13 Cochrane reviews (0.53 (0.25 to 0.80)) searched for data from regulatory bodies or industry, compared with 47 of 296 (0.15 (0.11 to 0.20)) DARE reviews.

Study selection and data items

At the screening phase, most reviews only included studies if the harms being searched were reported (proportion of reviews 0.70 (95% confidence interval 0.65 to 0.75)). Of 67 reviews of complications related to surgery or procedures, only four (0.05 (0.01 to 0.14)) reported professional qualifications of the individuals involved.

Study characteristics

Reports of any possible patient related risk factors—such as age, sex, or comorbidities—were sought in less than half of reviews (proportion of reviews 0.41 (0.36 to 0.47)). Furthermore, only 10 reviews adjudicated whether the adverse event could be biologically, pharmacologically, or temporally caused by the intervention, as measured by item 12a.

Results

Fewer than half the reviews (132 of 309; proportion of reviews 0.42 (95% confidence interval 0.37 to 0.48)) only included controlled clinical trials (randomised or not). Only 59 reviews (0.19 (0.14 to 0.23)) included both clinical trials and observational studies. Only observational studies (prospective

or retrospective) were included in 109 reviews (0.35 (0.29 to 0.40)), case series or case reports were included in 24 (0.07 (0.05 to 0.11)). Three of 13 Cochrane reviews (0.23 (0.05 to 0.53)) included observational studies compared with 165 of 296 DARE reviews (0.55 (0.49 to 0.61)). After reviewing the full text, nine reviews did not report the designs included anywhere in the text. Length of follow-up or patient demographics were only reported in just over half of reviews (170 of 309 reviews; 0.55 (0.49 to 0.60)).

In a retrospective analysis, we compared the quality of reporting between the years of 2008 and 2010-11 to measure any possible improvement over time on the quality of reporting. There were no significant differences ($P=0.079$) on the proportion of good reporting between the earlier and later years.

The 13 CDSR reviews had overall better reporting than DARE reviews in the abstracts, methods, and results sections; the other categories had similar levels of reporting quality. Almost half of DARE reviews (proportion of reviews 0.46 (95% confidence interval 0.40 to 0.51)) had poor reporting in the results section. Because the number of reviews was considerably different between databases, we considered it inappropriate to proceed with any formal tests to compare CDSR and DARE reviews.

Figure 2 provides a graphic trend in the proportion of reviews with good reporting over time (2008, 2009, and 2010-11), by each item reviewed. Reviews had a poor report quality on methods and results, with an average of half of items being poorly reported on those sections.

The overall proportion of reviews with good reporting was 0.56 (95% confidence interval 0.55 to 0.57). The corresponding proportions were 0.55 (0.53 to 0.57) in 2008, 0.55 (0.54 to 0.57) in 2009, and 0.57 (0.55 to 0.58) in 2010-11.

Discussion

Principal findings

We conducted this systematic review to assess the quality of reporting in systematic reviews of harms as a primary outcome using a set of proposed reporting items. This is the first step for the development of PRISMA Harms, a reporting guideline specifically designed for reviews of harms.³⁰ There is a substantial difference between the number of systematic reviews measuring an adverse event as the main outcome identified through DARE and those identified through CDSR. CDSR reviews comprised only a small fraction of the total reviews included. This distinction may compromise any direct comparison between the two databases. Reviews of harms as a secondary outcome or a coprimary outcome were not included in this review, which could be one reason for the large dissimilarity. Despite the small number of reviews of harms published in the CDSR, they were better reported overall than DARE reviews, probably owing to the clear guidelines provided by the Cochrane Collaboration²⁹ and the more flexible word limits allowed than those in other peer reviewed journals.

Several items were poorly reported in the included reviews, but a few are especially important when reviewing harms, because the lack of reporting on these could lead to misinterpretations of findings. In a systematic review, the screening phase is crucial and the exclusion of studies due to the absence of harms could overestimate the events and perhaps generate a biased review. Two thirds of reviews of harms only included studies if at least one adverse event was reported in the included studies. In a review of harms, “zero” is an important value, and studies with “no adverse events” are possibly as relevant to the review as those with reported adverse events. Nevertheless, zero events

in studies require careful interpretation, because the lack of reported harms may have different reasons: they may not have occurred (that is, a zero event), they may not have been investigated (that is, unknown if zero or no events occurred), or they may have been detected but not reported (that is, unknown if zero or no events occurred). The lack of reporting can be thought of as a measurement bias or reporting bias and should be considered as such.^{1-12 24-26}

All these scenarios have different implications for readers who need to judge whether an intervention may cause harm. Almost half of the included reviews of harms as a primary outcome did not consider patient risk factors or length of follow-up when reviewing adverse events of an intervention. Readers cannot properly judge whether there is an association between intervention and harms if these critical data are not reported. Most of the poorly reported items were identified at methods and results. The clarity on methods and results are essential to provide a clear picture of author’s intentions and limitations of findings and transparency is warranted.

Strengths and limitations

This review was unique by including more than 300 reviews, from two major databases, that looked at harms as a main outcome; each review was evaluated in depth using a novel set of 37 items to measure the quality of reporting. A limitation of this review was the lack of a reporting guideline specific for systematic reviews of harms; different formats of reporting were found, and assessing whether the reporting was adequate was challenging. The reviewers were generous in their assessment, accepting a range of reports as a “yes” to the presence of an item, which could have underestimated the degree of the problem.

Our review was limited exclusively to systematic reviews where harms were the primary focus. We believed that measuring the quality of reporting of adverse events in reviews specifically designed to evaluate adverse events would generate a pure sample focusing on the reporting of such events. The documenting of poor reporting on these reviews would imply poor reporting on adverse events in general. At this stage, we also decided to be inclusive and measure all potentially relevant items. In the future PRISMA Harms extension, we will limit these to the minimum set of essential items for reporting harms in a systematic review.

Comparison with other studies

Hopewell and colleagues¹⁴ reviewed a sample of 59 Cochrane reviews, of which only one was a harms review. The remaining 58 reviews focused primarily on benefit, with adverse events as secondary or tertiary considerations. Hopewell and colleagues reported that 32 (54%) reviews had fewer than three paragraphs of information on adverse events, and 11 (19%) had fewer than five sentences on adverse events.

Hammad and colleagues⁴² reviewed a sample of 27 meta-analyses primarily assessing drug safety and identified that more than 85% of the PRISMA items were reported in the majority of the reviews. However, most reviews did not report the 20 items specifically developed by the authors to address drug safety assessment. Several of the items considered important by Hammad and colleagues were similar to the ones developed by our team. The present review adds to the voice of many authors highlighting the poor reporting in reviews of adverse events.^{9-28 42}

Policy implications

Systematic reviews may compound the poor reporting of harms data in primary studies by failing to report on harms or doing so inadequately.⁹⁻²⁸ We recognise the need to optimise quality of reporting on harms in primary studies, and attempts to enhance it have already been made.⁶ At this phase, we measured the report quality in systematic reviews. As a future step, we intend to compare the quality of reporting in included studies with the reporting in the systematic review.

Despite their status as the preferred method for knowledge synthesis, systematic reviews can present an incomplete picture to readers by not representing a reliable assessment of a given treatment. Authors of the systematic review have a unique vantage point and can evaluate the entire evidence base under review, including deficits in reporting harms at the primary study level. This vantage point should be used to flag deficits in primary reporting. The goal over time should be to improve the quality and clarity of reporting in systematic reviews as well as the primary studies they evaluate. Although we are glad to see increasing numbers of systematic reviews of harms being published, we are also aware that the validity of the results can be heavily influenced by reviewers' decisions during conduct of the review—even more so than in reviews of benefit because we are dealing with sparse data and secondary outcomes.⁴³⁻⁴⁵ Hence, we emphasise here the crucial importance of transparent reporting of the methods used in systematic reviews of harms.

Conclusions

Systematic reviews of interventions should put equal emphasis on efficacy and harms. Improved reporting of adverse events in systematic reviews is one step towards providing a balanced assessment of an intervention. Patients, healthcare professionals, and policymakers should base their decisions not only on the efficacy of an intervention, but also on its risks. Guidance on a minimal set of items to be reported when reviewing harms is needed to improve transparency and informed decision making, thereby greatly enhancing the relevance of systematic reviews to clinical practice.

This review used a set of proposed reporting items to assess reporting, and the findings indicate that specific aspects of harms reporting could be improved. The items will be further refined with the aim of developing a final set of criteria that would constitute the PRISMA Harms. The PRISMA statement³⁰ is a living document, open to criticism and suggestions; these same principles will be shared by the PRISMA Harms. The development of a standardised format for reporting harms in systematic reviews will promote clarity and help ensure that readers have the basic information necessary to make an informed assessment of the intervention under review.

Contributors: All authors participated in the study conception and design, and analysis and interpretation of data; drafted the article or revisited it critically for important intellectual content; and approved the final version to be published. SV is the study guarantor.

Funding: No specific funding was given for this study.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations 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: No ethics approval was necessary because this was a review of published literature. No patient data or confidential information were used in this manuscript.

Data sharing: The dataset is available from the corresponding author at svohra@ualberta.ca.

This 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.

- Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch Intern Med* 2009;169:1737-9.
- Sigh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012;13:138.
- Papanikolaou PN, Ioannidis JP. Availability of large-scale evidence on specific harms from systematic reviews of randomized trials. *Am J Med* 2004;117:582-9.
- Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756-61.
- Loke YK, Derry S. Reporting of adverse drug reactions in randomized controlled trials—a systematic survey. *BMC Clin Pharmacol* 2001;1:3.
- Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-43.
- Yazici Y. Some concerns about adverse event reporting in randomized clinical trials. *Bull NYU Hosp Jt Dis* 2008;66:143-5.
- Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized clinical trials: an extension of the CONSORT statement. *Ann Int Med* 2004;141:781-8.
- McIntosh HM, Woolcott NF, Bagnall AM. Assessing harmful effects in systematic reviews. *BMC Med Res Methodol* 2004;4:19.
- Ernst E, Pittler MH. Assessment of therapeutic safety in systematic reviews: literature review. *BMJ* 2001;323:546.
- Ernst E, Pittler MH. Systematic reviews neglect safety issues. *Arch Intern Med* 2001;161:125-6.
- Aronson JK, Derry S, Loke YK. Adverse drug reactions: keeping up to date. *Fundam Clin Pharmacol* 2002;16:49-56.
- Moseley AM, Elkins MR, Herbert RD, Maher CG, Sherrington C. Cochrane reviews used more rigorous methods than non-cochrane reviews: survey of systematic reviews in physiotherapy. *J Clin Epidemiol* 2009;62:1021-30.
- Hopewell S, Wolfenden L, Clarke M. Reporting of adverse events in systematic reviews can be improved: survey results. *J Clin Epidemiol* 2007;61:597-602.
- Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.
- Golder S, Loke YK, McIntosh HM. Room for improvement? A survey of the methods used in systematic reviews of adverse effects. *BMC Med Res Methodol* 2006;6:3.
- Vandermeer B, Bialy L, Johnston B, Hooton N, Hartling L, Klassen TP, et al. Meta-analyses of safety data: a comparison of exact versus asymptotic methods. *Stat Methods Med Res* 2009;18:421-32.
- Loke YK, Golder S, Vandenbroucke J. Comprehensive evaluations of the adverse effects of drugs: importance of appropriate study selection and data sources. *Ther Adv Drug Saf* 2011;2:59-68.
- Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007;7:32.
- Golder S, Loke YK, McIntosh HM. Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. *J Clin Epidemiol* 2008;61:440-8.
- Golder S, McIntosh HM, Loke YK. Identifying systematic reviews of the adverse effects of health care interventions. *BMC Med Res Methodol* 2006;6:22.
- Derry S, Loke YK, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Med Res Methodol* 2001;1:7.
- Golder S, Loke Y. Is there evidence for biased reporting of published adverse effects data in pharmaceutical industry-funded studies? *Br J Clin Pharmacol* 2008;66:767-3.
- Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med* 2005;142:1090-9.
- Chou R, Aronson N, Atkins D, Ismaila AS, Santaguida P, Smith DH, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health care program. *J Clin Epi* 2010;63:502-12.
- Cornelius VR, Perrio MJ, Shakir SA, Smith LA. Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. *Pharmacoeconomics Drug Saf* 2009;18:1223-31.
- Warren FC, Abrams KR, Golder S, Sutton AJ. Systematic review of methods used in meta-analyses where a primary outcome is an adverse or unintended event. *BMC Med Res Methodol* 2012;12:64.
- Golder S, Loke YK, Zorzela L. Some improvements are apparent in identifying adverse effects in systematic reviews from 1994 to 2011. *J Clin Epidemiol* 2013;66:253-60.
- Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. Cochrane Collaboration, 2011. www.cochrane-handbook.org.
- Moher D, Liberati A, Tetzlaff J, Altman D. The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- Moher D, Schulz K, Simera I, Altman D. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217.
- Hill AB. The environment and disease: association or causation? *Proc Roy Soc Med* 1965;58:295-300.
- Pilkington K, Boshnakova A. Complementary medicine and safety: a systematic investigation of design and reporting of systematic reviews. *Complement Ther Med* 2012;20:73-82.
- Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;345:e5661.
- Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of non inferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152-60.
- Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.

What this paper adds

There is room for improved reporting in systematic of harms, including clear definition of the events measured, length of follow-up, and patient risk factors

Comparisons of reviews of harms have shown no improvement in the quality of reporting over time

Lack of detail or transparency in the reporting of systematic reviews of harms could hinder proper assessment of validity of findings

What is already known on this topic

The number of systematic reviews of adverse effects has increased significantly over the past 17 years

Harms are poorly reported in randomised controlled trials, but it is unclear whether there are weaknesses in the reporting of systematic reviews of harms

Although the PRISMA statement aims to provide guidance on transparent reporting in systematic reviews, its recommendations are mainly focused on studies of beneficial outcomes

- 37 Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med* 2006;144:364-67.
- 38 Boutron I, Moher D, Altman DG, Schulz K, Ravaud P. Methods and processes of the CONSORT group: example of an extension for trials assessing nonpharmacologic treatments. *Ann Intern Med* 2008;148:W60-6.
- 39 MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W, White A, et al. STRICTA revision group. Revised standards for reporting interventions in clinical trials of acupuncture (STRICTA): extending the CONSORT statement. *PLoS Med* 2010;7:e1000261.
- 40 Newcombe, Robert G. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17:857-72.
- 41 Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927;22:209-21.
- 42 Hammad T, Neyarapally G, Pinheiro S, Iyasu S, Rochester G, Dal Pan G. Reporting of meta-analyses of randomized controlled trials with a focus on drug safety: An empirical assessment. *Clin Trials* 2013;10:389-97.
- 43 Diamond GA, Bax L, Kaul S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. *Ann Intern Med* 2007;147:578-81.
- 44 Hernandez AV, Walker E, Ioannides-Demos LL, Kattan MW. Challenges in meta-analysis of randomized clinical trials for rare harmful cardiovascular events: the case of rosiglitazone. *Am Heart J* 2008;156:23-30.
- 45 Shuster JJ, Jones LS, Salmon DA. Fixed vs random effects meta-analysis in rare event studies: the rosiglitazone link with myocardial infarction and cardiac death. *Stat Med* 2007;26:4375-85.

Accepted: 17 December 2013

Cite this as: *BMJ* 2014;348:f7668

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

Tables

Table 1 | Glossary of terms

Term	Definition
Adverse drug reaction ²⁹	An adverse effect specific to a drug
Adverse effect ^{25,29}	An unfavourable outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it
Adverse event ^{25,29}	An unfavourable outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it
Complication ²⁹	An adverse event or effect following surgical and other invasive interventions
Harm ⁸	The totality of possible adverse consequences of an intervention or therapy; they are the direct opposite of benefits
Safety ⁸	Substantive evidence of an absence of harm. The term is often misused when there is simply absence of evidence of harm
Side effect ²⁹	Any unintended effect, adverse or beneficial, of a drug that occurs at doses normally used for treatment
Toxicity ⁸	Describes drug related harms. The term may be most appropriate for laboratory determined measurements, although it is also used in relation to clinical events. The disadvantage of the term "toxicity" is that it implies causality. If authors cannot prove causality, the terms "abnormal laboratory measurements" or 'laboratory abnormalities' are more appropriate to use

Table 2| Definitions of reporting items

Item and number	Definition of "yes"
Title	
1a) Specifically mentions "harms" or other related term	Should contain a word or phrase related to harm, such as "adverse effect," "adverse event," "complications," and "risk"
1b) Clarifies whether both benefits and harms are examined or only harms	Mentions whether benefits are reviewed or not
1c) Mentions the specific intervention being reviewed	Mentions the intervention being reviewed
1d) Refers to specific patient group or conditions (or both) in which harms have been assessed	Clearly states the specific group of patients or conditions being reviewed
Abstract	
2a) Specifically mentions "harms" or other related terms	Should contain a word or phrase related to harm, such as "adverse effect," "adverse event," "complications," and "risk"
2b) Clarifies whether both benefits and harms are examined or only harms	Mentions whether benefits are reviewed or not
2c) Refers to a specific harm assessed	Clearly states the adverse event being reviewed. General descriptions are accepted, for example, "cardiovascular events" or "maternal complications." Also acceptable if the report is searching for any adverse event and it does not refer to a specific harm
2d) Specifies what type of data was sought	Clearly names the kind of studies searched (for example, randomised controlled trials, cohort studies, case reports only, all types of data)
2e) Specifies what type of data was included	Clearly names the kind of studies included in the analysis or results (that is, randomised controlled trials, cohort studies, case reports, all types of data)
2f) Specifies how each type of data has been appraised	Clearly states the method of quality appraisal for included studies (that is, Jadad score, Cochrane risk of bias tool, Newcastle-Ottawa scale)
Introduction	
3a) Explains rationale for addressing specific harm(s), condition(s) and patient group(s)	Reports reasons for proceeding with the systematic review
3b) Clearly defines what events or effects are considered harms in the context of the intervention(s) examined	Clearly states the adverse event being reviewed. General definitions are acceptable (for example, "cardiovascular events" or "maternal complications"). Also acceptable if the report is searching for any adverse events, as in a scoping review, and does not refer to a specific harm
3c) Describes the rationale for type of harms systematic review done: hypothesis generating versus hypothesis testing	Clearly states whether a specific adverse event is being reviewed (hypothesis testing review; for example, cardiovascular deaths) or the authors are searching for any adverse events related to the intervention (hypothesis generating)
3d) Explains rationale for selection of study types or data sources and relevance to focus of the review	Clearly states which study designs are included (for example, randomised controlled trials, cohorts, case reports). Rationale is not necessary for "yes"
Objectives	
4a) Provides an explicit statement of questions being asked with reference to harms	Clearly defines, preferably at the end of introduction section, the intervention and the adverse events being reviewed. For this review, it was considered acceptable if intervention and outcome were clearly stated
Methods	
Protocol and registration	
5a) Describes whether protocol was developed in collaboration with someone with clinical expertise for field or intervention under study	Mentions whether a protocol was developed previously. For this review, it was not necessary to state the clinical expertise of who developed it
Eligibility criteria	
6a) Clearly defines what events or effects are considered harms in the context of the intervention(s) examined	Provides a clear definition of the adverse event being reviewed. If a general description was provided previously (that is, "cardiovascular events"), now the specific events need to be defined
6b) Specifies type of studies on harms to be included	Defines which kind of study designs will be included (for example, randomised controlled trials, cohort studies, case-control studies)
Information sources	
7a) States whether additional sources for adverse events were searched (for example, regulatory bodies, industry); if so, describes the source, terms used, and dates searched	Reports whether any other sources of adverse events (other than regular peer reviewed journals) were searched. For this review, it was not necessary to have the terms used and dates searched
Search	
8a) Presents the full search strategy if additional searches were used to identify adverse events	Reports search strategy used. Adverse events terms and databases searched
Study selection	

Table 2 (continued)

Item and number	Definition of "yes"
9a) States what study designs were eligible and provide rationale for their selection	Reports the study designs included in the review (for example, randomised controlled trials, cohort studies, case-control studies). No rationale is required
9b) Defines if studies were screened on the basis of the presence or absence of harms related terms in title or abstract	Reports whether the study screening is based on the report of adverse events or not (for example, review of mortality associated with antihyperglycaemic drugs; only studies reporting on mortality were included)
Data collection process	
10a) Describes method of data extraction for each type of study or report	States whether a data extraction form is used, and how it is done (in duplicate, checked by a second author)
Data items	
11a) Lists and defines variables for which data were sought for individual therapies	Reports variable(s) sought for the intervention reviewed (for example, whether the outcome is kidney failure, what variable was used to define kidney failure—creatinine level, creatinine clearance, need of dialysis)
11b) Lists and defines variables for which data were sought for patient underlying risk factors	States whether any potential patient risk factors or confounders are sought (for example, age, sex, comorbidities, previous events)
11c) Lists and defines variables for which data were sought for practitioner training or qualifications	States any variable(s) sought for healthcare personnel (practitioner) (for example, relevant degree, years of experience, other qualifications)
11d) Lists and defines harms for individual therapies	Provides the definition for the harm(s) sought (for example, side effects of propranolol and atenolol defined as severe bradycardia (heart rate \leq 45 beats per min) and severe hypotension (systolic blood pressure $<$ 80 mm Hg)
Risk of bias in individual studies	
12a) For uncontrolled studies, describes whether causality between intervention and adverse event was adjudicated and if so, how	Specifies whether review authors adjudicate if the intervention can cause the harm (for example, use of Bradford Hill criteria for causality ³²)
12b) Describes risk of bias in studies with incomplete or selective report of adverse events	Describes how studies not reporting adverse events are handled regarding risk of bias. For example, adverse events reported in a randomised controlled trial investigating glucose control in type 2 diabetes included hypoglycaemia but not mortality. Did the authors assess whether the event did not occur (zero deaths), was not measured (and may have occurred), or was measured but not reported
Summary measures	
13a) If rare outcomes are being investigated, specifies which summary measures will be used (for example, event rate, events or person time)	Defines the summary measures used for rare events
Synthesis of results	
14a) Describes statistical methods of handling with the zero events in included studies	Clearly defines how studies with no adverse events reported are reported and analysed (for example, when zero was an outcome for a 2x2 table, reviewers added a 0.5 value to it)
Additional analysis	
16a) Describes additional analysis with studies with high risk of bias	Describes whether studies with high risk of bias are analysed separately (for example, subgroup analysis with high and low risk of bias studies analysed separately)
Results	
Study selection	
17a) Provides process, table, or flow for each type of study design	Provides a table (or text) containing included studies and a clear reason for exclusion of studies
Study characteristics	
18a) Reports study characteristics, such as patient demographics or length of follow-up that may have influenced the risk estimates for the adverse outcome of interest	Reports any potential confounder or patient risk factors that can affect the outcome (adverse event)
18b) Describes methods of collecting adverse events in included studies (for example, patient report, active search)	Reports how adverse events are investigated in included studies (for example, voluntary report, active search)
18c) For each primary study, lists and defines each adverse event reported and how it is identified	Clear definition of adverse event under investigation and how it is identified (method of measurement, assessment or identification)
Conclusion	
26a) Provides balanced discussion of benefits and harms with emphasis on study limitations, generalisability, and other sources of information on harms	Clearly discusses the benefits of the interventions and the harms identified in the review

Table 3| Summary of findings

Item and number	Proportion (%) and No of reviews with "yes" responses		
	Total reviews (n=309)	CDSR reviews (n=13)	DARE reviews (n=296)
Title			
1a) Specifically mentions "harms" or other related term	53 (n=165)	54 (n=7)	53 (n=158)
1b) Clarifies whether both benefits and harms are examined or only harms	76 (n=234)	92 (n=12)	75 (n=222)
1c) Mentions the specific intervention being reviewed	98 (n=303)	100 (n=13)	98 (n=290)
1d) Refers to specific patient group or conditions (or both) in which harms have been assessed	58 (n=178)	92 (n=12)	56 (n=166)
Abstract			
2a) Specifically mentions "harms" or other related terms	84 (n=261)	85 (n=11)	85 (n=250)
2b) Clarifies whether both benefits and harms are examined or only harms	92 (n=285)	92 (n=12)	92 (n=273)
2c) Refers to a specific harm assessed	97 (n=299)	100 (n=13)	97 (n=289)
2d) Specifies what type of data was sought	50 (n=156)	92 (n=12)	49 (n=144)
2e) Specifies what type of data was included	49 (n=151)	92 (n=12)	47 (n=139)
2f) Specifies how each type of data has been appraised	6 (n=19)	23 (n=3)	5 (n=16)
Introduction			
3a) Explains rationale for addressing specific harm(s), condition(s) and patient group(s)	99 (n=307)	100 (n=13)	99 (n=294)
3b) Clearly defines what events or effects are considered harms in the context of the intervention(s) examined	94 (n=289)	92 (n=12)	94 (n=277)
3c) Describes the rationale for type of harms systematic review done: hypothesis generating versus hypothesis testing	99 (n=306)	100 (n=13)	99 (n=293)
3d) Explains rationale for selection of study types or data sources and relevance to focus of the review	39 (n=120)	54 (n=7)	38 (n=113)
Objectives			
4a) Provides an explicit statement of questions being asked with reference to harms	86 (n=265)	100 (n=13)	85 (n=252)
Methods			
Protocol and registration			
5a) Describes if protocol was developed	11 (n=35)	100 (n=13)	7.4 (n=22)
Eligibility criteria			
6a) Clearly defines what events or effects are considered harms in the context of the intervention(s) examined	75 (n=231)	92 (n=12)	74 (n=219)
6b) Specifies type of studies on harms to be included	74 (n=230)	92 (n=12)	74 (n=218)
Information sources			
7a) States whether additional sources for adverse events were searched (for example, regulatory bodies, industry)	17 (n=54)	54 (n=7)	16 (n=47)
Search			
8a) Presents the full search strategy if additional searches were used to identify adverse events	69 (n=214)	100 (n=13)	68 (n=201)
Study selection			
9a) States what study designs were eligible	74 (n=229)	100 (n=13)	73 (n=216)
9b) Defines if studies were screened on the basis of the presence or absence of adverse events	71 (n=218)	92 (n=12)	70 (n=206)
Data collection process			
10a) Describes method of data extraction for each type of study or report	64 (n=199)	100 (n=13)	63 (n=186)
Data items			
11a) Lists and defines variables for which data were sought for individual therapies	50 (n=154)	23 (n=3)	51 (n=151)
11b) Lists and defines variables for which data were sought for patient underlying risk factors	42 (n=129)	23 (n=3)	43 (n=126)
11c) Lists and defines variables for which data were sought for practitioner training or qualifications	1 (n=4)	0 (n=0)	1 (n=4)
11d) Lists and defines harms for individual therapies	68 (n=210)	69 (n=9)	68 (n=201)
Risk of bias in individual studies			

Table 3 (continued)

Item and number	Proportion (%) and No of reviews with "yes" responses		
	Total reviews (n=309)	CDSR reviews (n=13)	DARE reviews (n=296)
Title			
12a) For uncontrolled studies, describes whether causality between intervention and adverse event was adjudicated	3 (n=10)	23 (n=3)	2 (n=7)
12b) Describes risk of bias in studies with incomplete or selective report of adverse events	2 (n=6)	0 (n=0)	2 (n=6)
Summary measures			
13a) If rare outcomes are being investigated, specifies which summary measures will be used (for example, event rate, events or person time)	73 (n=227)	100 (n=13)	72 (n=214)
Synthesis of results			
14a) Describes statistical methods of handling with the zero events in included studies	13 (n=41)	46 (n=6)	12 (n=35)
Additional analysis			
16a) Describes additional analysis with studies with high risk of bias	20 (n=61)	77 (n=10)	17 (n=51)
Results			
Study selection			
17a) Provides process, table, or flow for each type of study design	79 (n=244)	92 (n=12)	78 (n=232)
Study characteristics			
18a) Reports study characteristics, such as patient demographics or length of follow-up that may have influenced the risk estimates for the adverse outcome of interest	55 (n=171)	85 (n=11)	54 (n=160)
18b) Describes methods of collecting adverse events in included studies (for example, patient report, active search)	62 (n=190)	85 (n=11)	61 (n=179)
18c) For each primary study, lists and defines each adverse event reported and how it is identified	22 (n=68)	23 (n=3)	22 (n=65)
Discussion			
26a) Provides balanced discussion of benefits and harms with emphasis on study limitations, generalisability, and other sources of information on harms	83 (n=257)	92 (n=12)	83 (n=245)

Figures

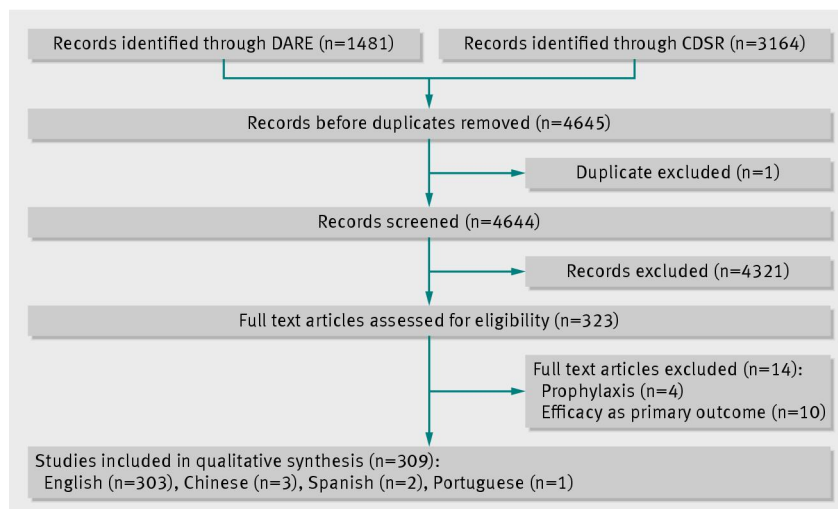


Fig 1 PRISMA flow diagram

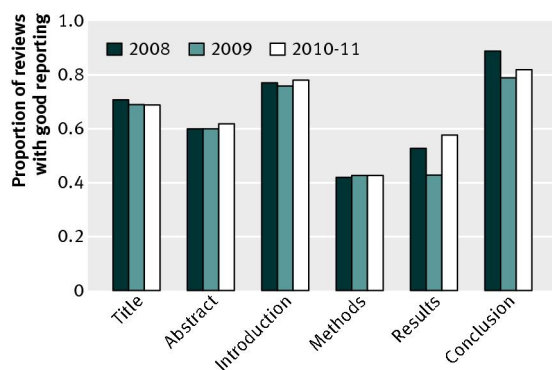


Fig 2 Proportion of good reporting by subheadings