Incorporation of assessments of risk of bias of primary studies in systematic reviews of randomised trials: a cross-sectional study

Sally Hopewell,[1,2,3,4,5] Isabelle Boutron,[1,2,3,4] Douglas G Altman,[5] Philippe Ravaud[1,2,3,4]

ABSTRACT

Objective: We examined how assessments of risk of bias of primary studies are carried out and incorporated into the statistical analysis and overall findings of a systematic review.

Design: A cross-sectional review.

Sample: We assessed 200 systematic reviews of randomised trials published between January and March 2012; Cochrane (n=100), non-Cochrane (Database of Reviews of Effects) (n=100).

Main outcomes: Our primary outcome was a descriptive analysis of how assessments of risk of bias are carried out, the methods used, and the extent to which such assessments were incorporated into the statistical analysis and overall review findings.

Results: While Cochrane reviews routinely reported the method of risk of bias assessment and presented their results either in text or table format, 20% of non-Cochrane reviews failed to report the method used and 39% did not present the assessment results. Where it was possible to evaluate the individual results of the risk of bias assessment (n=154), 75% (n=116/154) of reviews had ≥1 trial at high risk of bias; the median proportion of trials per review at high risk of bias was 50% (IQR 31% to 89%). Despite this, only 56% (n=65/116) incorporated the risk of bias assessment into the interpretation of the results in the abstract and 41% (n=47/116) (49%; n=40/81 Cochrane and 20%; n=7/35 non-Cochrane) incorporated the risk of bias assessment into the interpretation of the conclusions. Of the 83% (n=166/200) systematic reviews which included a meta-analysis, only 11% (n=19/166) incorporated the risk of bias assessment into the statistical analysis.

Conclusions: Cochrane reviews were more likely than non-Cochrane reviews to report how risk of bias assessments of primary studies were carried out; however, both frequently failed to take such assessments into account in the statistical analysis and conclusions of the systematic review.

[1]INSERM, U738, Paris, France
[2]Hôpital Hôtel Dieu, Centre d'Epidémiologie Clinique, Paris, France
[3]Faculté de Médecine, Univ. Paris Descartes, Sorbonne Paris Cité, Paris, France
[4]French Cochrane Center, Paris, France
[5]Centre for Statistics in Medicine, University of Oxford, Oxford, UK

Correspondence to
Dr Sally Hopewell;
sally.hopewell@htd.aphp.fr

ARTICLE SUMMARY

Article focus
- Assessment of the validity of individual studies included in a systematic review, and the risk that they might overestimate or underestimate the true intervention effect, is a critical part of the systematic review process.
- Authors should clearly describe the methods used to assess the validity of individual studies (ie, 'risk of bias'). However, there is limited evidence to show the extent to which such assessments are incorporated into the results of a systematic review.
- The objective of our study was to examine how assessments of risk of bias of primary studies are carried out and incorporated into the statistical analysis and overall findings of systematic reviews.

Key messages
- Cochrane reviews were more likely than non-Cochrane reviews to report how assessments of risk of bias of the primary studies were carried out. However, most largely failed to show how such assessments were incorporated into the statistical analysis and in the interpretation of the overall conclusions, suggesting that there was no overall improvement in the last 10 years.
- Despite all the valuable efforts to transparently report and display the potential risk of bias of primary studies, it is clear that their impact on the overall findings of a systematic review is rarely assessed formally.

Strengths and limitations of this study
- Our sample of non-Cochrane reviews was drawn from the Database of Reviews of Effects, which meets strict methodological criteria. It is possible that our findings might be an underestimate of the problem compared to systematic reviews identified from other sources.

INTRODUCTION

Problems in the design and conduct of individual studies can raise questions about the validity of their findings. For example, reports of randomised trials with inadequate allocation concealment are likely to show exaggerated treatment effects.[1][2] Similarly, participants who are aware of their

assignment status are more likely to report symptoms, leading to biased results.[2] [3] In addition, selective reporting means that significant trial outcomes are more likely to be reported than those with non-significant outcomes.[4] An assessment of the validity of individual studies included in a systematic review, and the risk that they might overestimate or underestimate the true intervention effect, is therefore a critical part of the systematic review process.

The assessment of the risk of bias of studies included in a systematic review has evolved over time. Initially authors of systematic reviews did not evaluate the risk of bias, rather they evaluated the overall ';quality' of the studies included in the review,[5] even though the quality cannot be clearly defined. Until recently, the most common tools[6] were scales in which various components of quality were scored and combined to give a summary score; however, this can be misleading and should be discouraged as the results and conclusions may differ depending on the type of scale used.[7] [8] In recent years, the recommended approach requires authors to specify which individual methodological components they will assess and to provide a description and judgement for each item. This approach is recommended by the Cochrane Collaboration and is part of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) Statement.[9] [10] Whichever approach is used, authors of systematic reviews should clearly describe the methods they used to assess the risk of bias and how these assessments are incorporated into the review findings.[9] [10] Although these principles apply to all types of primary study, by far the most empirical research and development of methods has been in relation to randomised trials.

The aim of this study was to examine how assessments of risk of bias of primary studies in systematic reviews of randomised trials are currently carried out, the methods used, and the extent to which such assessments are incorporated into the statistical analysis and overall interpretation of the review findings. While we use the term 'risk of bias' to mean any method of assessing the validity of individual studies included in a systematic review, this is not necessarily how the authors of the systematic review have referred to their assessment.

METHODS
Systematic review selection and inclusion criteria
We assessed a convenience sample of 200 systematic reviews that evaluated randomised trials assessing the effects of healthcare interventions published between January and March 2012. We sampled systematic reviews from two specialised databases: those published in the Cochrane Database of Systematic Reviews (n=100) in the Cochrane Library (http://www.cochranelibrary.org) and those from the Database of Reviews of Effects (DARE) (n=100) through the Centre for Reviews and Dissemination, University of York (http://www.crd.york.ac.uk/crdweb). Systematic reviews included in DARE must meet strict methodological criteria, and thus we deemed them to be of a similar methodological standard to Cochrane systematic reviews. We excluded updates of previously published systematic reviews and those published in languages other than English. We also excluded systematic reviews of diagnostic test accuracy, prognosis, economics evaluations, qualitative studies and non-randomised studies. Where systematic reviews included randomised and non-randomised studies, we focused our assessment only on the elements that were related to randomised trials.

Data extraction
Data extraction was carried out by teams of assessors working in pairs, and any uncertainties or disagreements were resolved by involving a third assessor. Systematic reviews were allocated at random such that each assessor extracted a similar number of Cochrane and non-Cochrane systematic reviews. Prior to startingdata extraction, the assessors received training on how to complete the data extraction form (see online supplementary appendix 1). For each systematic review, we recorded the systematic review type (ie, Cochrane or non-Cochrane), medical specialty, type of intervention(s) and the number of included randomised trials. We assessed the method used to assess risk of bias of the included trials (ie, whether they used a summary scale, checklist or assessment of individual methodological components), the type of tool used (eg, the Cochrane Risk of Bias tool, the Jadad scale, the Pedro scale, etc), how the risk of bias assessment was carried out, by whom and which individual methodological components were assessed. We also evaluated how systematic reviews summarised the risk of bias across individual trials, how many systematic reviews included≥1 trial at high risk of bias, how many systematic reviews included≥1 trial at unclear risk of bias and how such assessments were interpreted in the abstract, discussion and conclusions section of the systematic review. Finally, for those systematic reviews which included a meta-analysis, we assessed whether and how the risk of bias assessment was incorporated into the statistical analysis (eg, using sensitivity analysis or metaregression).

Data analysis
We performed a descriptive analysis of how assessments of risk of bias were carried out, the methods used and the extent to which such assessments were incorporated into the statistical analysis and overall review findings. We also compared any differences in the approach used between the sample of Cochrane and non-Cochrane reviews.

RESULTS
Searches of the Cochrane Database of Systematic Reviews and the DARE between 1 January and 31 March 2012 identified 281 reports of systematic reviews. We assessed the full texts of all articles to confirm eligibility and that they were systematic reviews of randomised

trials. We excluded 44 non-Cochrane reviews and 23 Cochrane reviews (see figure 1 for reasons for exclusion). After exclusions, we selected at random 100 non-Cochrane and all remaining 95 Cochrane reviews (five additional Cochrane reviews were selected at random from the April 2012 issue of the Cochrane Database of Systematic Reviews to increase this sample to 100). The most common medical specialties of the included reviews were cardiology (n=20/200; 10%), neurology (n=19/200; 9.5%), obstetrics and gynaecology (n=19/200; 9.5%) and endocrinology (n=18/200; 9%) (table 1). Just over half (n=109/200; 54.5%) of all systematic reviews assessed drug interventions, one-fifth (n=38/200; 19%) assessed surgical or procedural interventions, with the remaining assessing counselling or lifestyle interventions (n=41/200; 20.5%) or types of equipment (n=12/200; 6%). The number of included randomised trials in Cochrane and non-Cochrane reviews was similar with a median of seven trials per systematic review (IQR 4–17).
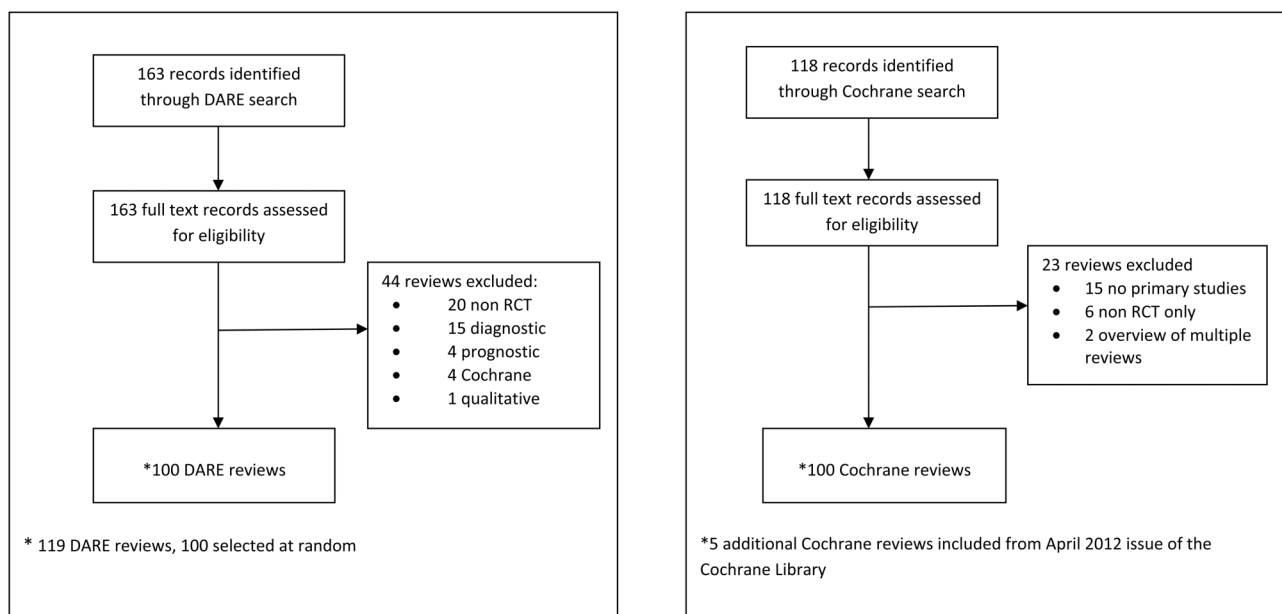
Method of risk of bias assessment

All 200 systematic reviews included some kind of assessment of risk of bias (table 1); however, the nature and extent of this assessment varied considerably. Cochrane reviews were much more likely to assess individual methodological components (Cochrane: 90%; non-Cochrane: 34%), whereas non Cochrane reviews were more likely to report using a quality assessment scale (Cochrane: 9%; non-Cochrane: 38%); 20% of non-Cochrane reviews did not report the method used to assess risk of bias. The majority (n=86/105; 82%) of Cochrane reviews reported using the Cochrane risk of bias tool; five reported using more than one tool. Tools used in non-Cochrane reviews were much more diverse: 20% (n=21/104) reported using the Cochrane risk of bias tool, 18% (n=19/104) the Jadad scale and 30% (n=31/104) used other methods of assessment, the most common being the Pedro scale (developed for assessing the quality of randomised trials in physiotherapy); four reported using more than one tool. A quarter (26%) of non-Cochrane reviews did not report the tool used for assessing risk of bias. Most systematic reviews reported in the methods section how the assessment of risk of bias was carried out, but only 5% (n=10) of systematic reviews reported using the assessment of risk of bias as part of their eligibility criteria.

Methodological components assessed

Overall, the median number of individual methodological components assessed per systematic review was six (IQR 5 to 7), ranging from 1 to 27 items (table 2). Nearly all Cochrane reviews assessed the method of random sequence generation (100%), concealment of the allocation sequence once randomised (100%), blinding (99%) and incomplete outcome data (ie, missing outcome data due to attrition) (95%) compared to 62%, 60%, 69% and 61% of non-Cochrane reviews, respectively. Very few systematic reviews (Cochrane: 7%; non-Cochrane: 2%) assessed blinding separately for more than one outcome measure or incomplete outcome data for more than one outcome (eg, where the outcome was measured at different time points) (Cochrane: 8%; non-Cochrane: 1%). Evidence of selective outcome reporting was assessed in 86% of Cochrane reviews compared to only 20% of non-Cochrane reviews. A number of systematic reviews (Cochrane: 86%; non-Cochrane: 49%) also assessed other methodological items, the most common being whether trialists had carried out an intention-to-treat analysis (n=29),

Figure 1 Inclusion of systematic reviews (published between 1 January to 31 March 2012).

Table 1 General characteristics and method of risk of bias assessment in individual systematic reviews

	Overall (n=200)	Cochrane (n=100)	Non-Cochrane (n=100)	Difference between proportions (95% CI)
Common medical specialty				
	Cardiology 20 (10%)	Neurology 16 (16%)	Cardiology 15 (15%)	
	Neurology 19 (9.5%)	Obs/Gynae 15 (15%)	Oncology 15 (15%)	
	Obs/Gynae 19 (9.5%)	Infectious diseases 7 (7%)	Endocrinology 9 (9%)	
	Oncology 18 (9%)	Musculoskeletal 7 (7%)	Psychiat/psychol 6 (6%)	
	Endocrinology 13 (6.5%)	Cardiology 5 (5%)	Surgery 6 (6%)	
Type of intervention				
Drug	109 (54.5%)	55 (55%)	54 (54%)	1.0 (−12.8 to 14.8)
Surgery/procedure	38 (19%)	20 (20%)	18 (18%)	2.0 (−8.9 to 12.9)
Counselling/lifestyle	41 (20.5%)	17 (17%)	24 (24%)	−7.0 (−18.1 to 4.1)
Equipment	12 (6%)	8 (8%)	4 (4%)	4.0 (−2.6 to 10.6)
Number of included studies				
Median (IQR)	10 (5 to 23)	6.5 (4 to 14)	16.5 (7 to 33.5)	
Minimum/maximum	1 to 385	1 to 102	3 to 385	
Number of randomised trials				
Median (IQR)	7 (4 to 17)	6 (4 to 14)	8 (5 to 22)	
Minimum/maximum	1 to 104	1 to 86	1 to 104	
Number of meta-analyses				
Median (IQR)	4 (2 to 12)	11 (3 to 22)	3 (1 to 5)	
Minimum/maximum	0 to 367	0 to 277	0 to 367	
Method of assessing risk of bias*				
Single components	128 (62%)	92 (90%)	36 (34%)	55.9 (45.1 to 66.7)
Scale	49 (24%)	9 (9%)	40 (38%)	−29.2 (−40.0 to −18.4)
Checklist	9 (4%)	1 (1%)	8 (6%)	−6.6 (−12.0 to −1.2)
Not specified	21 (10%)	0 (0%)	21 (20%)	−20.0 (−27.6 to −12.3)
Specific tool used†				
Cochrane RoB tool	107 (51%)	86 (82%)	21 (20%)	55.9 (45.1 to 66.7)
Modified Cochrane RoB tool	15 (7%)	9 (9%)	6 (6%)	2.8 (−4.2 to 9.8)
Jadad scale	26 (12%)	7 (7%)	19 (18%)	−11.6 (−20.4 to −2.8)
Other	34 (16%)	3 (2%)	31 (30%)	−26.9 (−36.3 to −17.6)
Not specified	27 (13%)	0	27 (26%)	−25.9 (−34.3 to −17.5)
Who did the assessment?				
One person	5 (2.5%)	1 (1%)	4 (4%)	−3.0 (−7.3 to 1.3)
Two people	167 (83.5%)	98 (98%)	69 (69%)	29.0 (19.5 to 38.4)
Not reported	28 (14%)	1 (1%)	27 (27%)	−26.0 (−34.9 to −17.1)
Assessment used as eligibility criteria				
Yes	10 (5%)	2 (2%)	8 (8%)	−6.0 (−11.9 to 0.0)
No	182 (91%)	98 (98%)	84 (84%)	14.0 (6.3 to 21.7)
Not reported	8 (4%)	0	8 (8%)	−8.0 (−13.3 to −2.7)
Number of items assessed‡				
Median (IQR)	6 (5 to 7)	6 (6 to 7)	5 (3 to 7)	
Minimum/maximum	1 to 27	3 to 15	1 to 27	

*Method of assessing risk of bias: seven reviews used more than one method; Cochrane n=2; non-Cochrane n=5.
†Type of tool used: nine reviews used more than one tool; Cochrane n=5; non-Cochrane n=4. Other tool n=34: Pedro n=8; Downs and Black n=4; own scale n=4; Schulz n=2; Maastricht criteria n=2; Critical Skills Appraisal Programme (CASP) n=1; Consolidated Standards for Reporting Trials (CONSORT) n=1; Centre for Reviews and Dessimination n=1; Detsky scale n=1; Grading of Recommendations Assessments, Development and Evaluation (GRADE) n=1; Heyland score n=1; Juni n=1; Modified Jadad scale n=1; PRISMA n=1.
‡Number of items assessed: nine reviews unclear number of items assessed; Cochrane n=0; non-Cochrane n=9.

evidence of baseline imbalance (n=27), funding source (n=26), small sample size (n=17), early stopping (n=12) and lack of reporting of a power calculation (n=11). Poor reporting was common across many non-Cochrane reviews, which meant that sometimes it was unclear whether the systematic review had assessed individual items, as shown in table 2.

Presentation and incorporation of risk of bias assessment into the analysis

We examined how the results of the risk of bias assessment were presented in individual systematic reviews (table 3). More than half (62%) of the Cochrane reviews used a combination of presentation formats including a text description, table, graph and/or figure.

Table 2 Methodological components assessed in individual systematic reviews

	Overall (n=200)	Cochrane (n=100)	Non-Cochrane (n=100)	Difference between proportions (95% CI)
Random sequence generation				
Yes	162 (81%)	100 (100%)	62 (62%)	38.0 (28.4 to 47.5)
No	23 (11.5%)	0	23 (23%)	−23.0 (−31.2 to −17.7)
Unclear	15 (7.5%)	0	15 (15%)	−15.0 (−21.9 to −8.0)
Allocation concealment				
Yes	160 (80%)	100 (100%)	60 (60%)	40.0 (30.4 to 49.6)
No	26 (13%)	0	26 (26%)	−26.0 (−34.6 to −17.4)
Unclear	14 (7%)	0	14 (14%)	−14.0 (−20.8 to −7.2)
Overall assessment of blinding*				
Yes	168 (84%)	99 (99%)	69 (69%)	30 (21.0 to 39.2)
No	22 (11%)	1 (1%)	21 (21%)	−20 −28.2 to −11.8)
Unclear	10 (5%)	0	10 (10%)	−10 (−15.9 to −4.1)
Blinding of participants, personnel, outcome assessors (combined)				
Yes	100 (50%)	57 (57%)	43 (43%)	14.0 (2.8 to 27.7)
No	90 (45%)	43 (43%)	47 (47%)	−4.0 (−17.8 to 9.8)
Unclear	10 (5%)	0	10 (10%)	−10.0 (−15.9 to −4.1)
Blinding of participants and personnel (separate)				
Yes	56 (28%)	38 (38%)	18 (18%)	20.0 (7.9 to 32.1)
No	130 (65%)	60 (60%)	70 (70%)	−10.0 (−23.1 to 3.1)
Unclear	14 (7%)	2 (2%)	12 (12%)	−10.0 (−16.9 to −3.0)
Blinding of outcome assessors (separate)				
Yes	66 (33%)	42 (42%)	24 (24%)	18.0 (5.2 to 30.8)
No	122 (61%)	58 (58%)	64 (64%)	−6.0 (−19.5 to 7.5)
Unclear	12 (6%)	0	12 (12%)	−12.0 (−18.4 to −5.6)
Assessed blinding >1 outcome				
Yes	9 (4.5%)	7 (7%)	2 (2%)	5.0 (0.7 to 10.7)
No	178 (89%)	91 (91%)	87 (87%)	4.0 (−4.6 to 12.6)
Unclear	13 (6.5%)	2 (2%)	11 (11%)	−9.0 (−5.4 to 13.3)
Incomplete outcome data				
Yes	156 (78%)	95 (95%)	61 (61%)	34.0 (23.5 to 44.4)
No	31 (15.5%)	4 (4%)	27 (27%)	−23.0 (−32.5 to −13.5)
Unclear	13 (6.5%)	1 (1%)	12 (12%)	−11.0 (−17.6 to −4.3)
Assessed incomplete outcome data for>1 outcome				
Yes	9 (4.5%)	8 (8%)	1 (1%)	7.0 (1.3 to 12.7)
No	178 (89%)	91 (91%)	87 (87%)	4.0 (−4.6 to 12.6)
Unclear	13 (6.5%)	1 (1%)	12 (12%)	−11.0 (−18.9 to −5.1)
Selective outcome reporting				
Yes	106 (53%)	86 (86%)	20 (20%)	66.0 (55.6 to 76.4)
No	81 (40.5%)	14 (14%)	67 (67%)	−53.0 (−64.4 to −41.5)
Unclear	13 (6.5%)	0	13 (13%)	−13.0 (−19.6 to −6.4)
Other sources of bias				
Yes	135 (67.5%)	86 (86%)	49 (49%)	37.0 (25.1 to 48.9)
No	56 (28%)	14 (14%)	42 (42%)	−28.0 (−39.8 to −16.2)
Unclear	9 (4.5%)	0	9 (9%)	−9.0 (−14.6 to −3.4)

*Overall assessment of blinding—this included personnel, outcome assessors (combined), participants and personnel (separate) and/or blinding of outcome assessors (separate).

In comparison, non-Cochrane reviews (39%) were more likely to present just a text description or table, although more than a third (39%) did not provide any presentation of the results of the risk of bias assessment. Where it was possible to evaluate the individual results of the risk of bias assessment (n=154), we examined the number of systematic reviews with one or more trials at a high or unclear risk of bias. Overall, 75% (n=116/154) of systematic reviews had one or more trials at high risk of bias; of these 116 systematic reviews, the median

proportion of trials per review at high risk of bias was 50% (IQR 31–89%). For just under half (46%) of the non-Cochrane reviews, it was not possible to evaluate the individual results of the risk of bias assessment based on the information reported in the systematic review. Of the 116 systematic reviews which had more than one trial, high risk of bias of just over half (56%; 65/116) incorporated the risk of bias assessment into the interpretation of the results in the abstract of the systematic review. This interpretation could have been a specific comment

Table 3 Presentation and incorporation of risk of bias assessment into the analysis in individual systematic reviews

	Overall (n=200)	Cochrane (n=100)	Non-Cochrane (n=100)	Difference between proportions (95% CI)
Presentation of risk of bias assessment				
Description or table only	39 (19.5%)	0	39 (39%)	−39.0 (−48.6 to −29.4)
Description and table	56 (28%)	38 (38%)	18 (18%)	20.0 (7.8 to 32.0)
Description, table and figure/graph	66 (33%)	62 (62%)	4 (4%)	58.0 (47.8 to 68.3)
Not reported	39 (19.5%)	0	39 (39%)	−39.0 (−48.6 to −29.4)
Proportion of trials at risk of bias per systematic review *				
≥1 Trial at high risk of bias	116/154 (75%)	81/100 (81%)	35/54 (65%)	
Median proportion per review (IQR)	50% (31% to 89%)	57% (33% to 89%)	50% (25% to 100%)	
≥1 Trial at unclear risk of bias	119/154 (77%)	99/100 (99%)	20/54 (37%)	
Median proportion per review (IQR)	85% (57% to 100%)	91.5% (69.5% to 100%)	63.5% (41% to 100%)	
Not reported	46	0	46	
≥1 Trial at high risk of bias and incorporated into interpretation of results				
Abstract	65/116 (56%)	51/81 (63%)	14/35 (40%)	
Plain language summary	–	34/81 (42%)	–	
Discussion	101/116 (87%)	78/81 (96%)	23/35 (66%)	
Conclusion	47/116 (41%)	40/81 (49%)	7/35 (20%)	
Assessment incorporated into GRADE				
Yes	51 (25.5%)	45 (45%)	6 (6%)	39.0 (28.2 to 49.8)
Not applicable (GRADE not used)	149 (74.5%)	55 (55%)	94 (94%)	
How assessment was incorporated into the results				
Descriptive only	174 (87%)	89 (89%)	85 (85%)	4.0 (−5.3 to 13.3)
Meta-analysis only	1 (0.5%)	0	1 (1%)	−1.0 (−2.9 to 0.9)
Both	18 (9%)	11 (11%)	7 (7%)	4.0 (−3.9 to 11.9)
Not performed	7 (3.5%)	0	7 (7%)	−7.0 (−12.0 to −1.9)

*Proportion of trials at risk of bias per systematic review: based on approach or scoring system used by authors of systematic review and where it was possible to evaluate (eg, Cochrane ≥ one key domain not adequate (high risk of bias) or not reported (unclear risk of bias); Jadad ≥ three high quality (low risk of bias) ≤ two low quality (high risk of bias); Pedro≥six high quality (low risk of bias)≤five low quality (high risk of bias).
GRADE, Grading of Recommendations Assessments, Development and Evaluation.

in the results or conclusions section of the abstract (eg, X studies were at high risk of bias, were not blinded or had inadequate methods of allocation concealment) or a more general comment about the overall quality of the evidence. Most Cochrane reviews (96%; n=78/81) incorporated the risk of bias assessment into the interpretation of the results in the discussion section of the systematic review, compared to 66% (n=23/35) of non-Cochrane reviews. Just under half (49%; n=40/81) of the Cochrane reviews incorporated the risk of bias assessment into the interpretation of the conclusions section of the systematic review compared to only 20% (n=7/35) of non-Cochrane reviews.

We also looked at whether and how the risk of bias assessment was incorporated into the analysis of individual systematic reviews. In total, 166 (83%) systematic reviews included a meta-analysis of which only 19 (n=19/166; 11%) (Cochrane n=11; non-Cochrane n=8) incorporated the risk of bias assessment into the statistical analysis; 15 of the 19 meta-analysis had one or more trials at high risk of bias. The most common type of analysis performed was a sensitivity analysis (n=14) whereby studies at high or unclear risk of bias were excluded from the meta-analysis to determine if the size of the overall effect estimate changed as a result of excluding high-risk studies. Other analysis included subgroup analysis, whereby studies at high or unclear risk of bias were analysed separately from those at low risk of bias, and meta-regression. Overall, 45% of Cochrane reviews used the Grading of Recommendations Assessments, Development and Evaluation (GRADE) approach[11] as a means of interpreting the overall quality of the body of evidence, compared to only 6% of non-Cochrane reviews (the risk of bias assessment is a key component of the GRADE approach[12]).

DISCUSSION
Summary of main findings
Our study provides a current and comprehensive view of how assessments of risk of bias of primary studies are carried out in a recent sample of systematic reviews, the methods used and the extent to which these assessments are incorporated into the statistical analysis and overall review findings. Our findings show that Cochrane reviews are more likely to assess individual methodological components,[13] whereas non-Cochrane reviews were more likely to report using a quality assessment scale such as the Jadad scale[14] or other such scale, contrary to recommendations warning of the hazards of using such an approach.[7 8] Irrespective of the approach chosen, most systematic reviews included the items sequence generation, allocation concealment, blinding and incomplete outcome data as part of their assessment of risk of bias, although poor reporting meant that sometimes it was unclear whether some non-Cochrane reviews had assessed specific items as they did not report the individual results of the risk of bias assessment.

On the basis of the assessment carried out by the authors of the systematic review, three quarters of the reviews had one or more trials at high risk of bias, with the median proportion of trials per review at high risk of bias being 50% (ranging from 31% to 89%). Despite this, only around half of these systematic reviews incorporated the risk of bias assessment into the interpretation of the results in the abstract or conclusions of the systematic review. There were very few systematic reviews which conducted a meta-analysis incorporating the results of the risk of bias assessment into the statistical analysis, for example by performing sensitivity analysis to determine if the overall effect estimate changed as a result of excluding studies at high risk of bias.

The reason why authors failed to take into account the risk of bias assessment in the statistical analysis and interpretation of the review findings is not clear, but it could be due to a lack of specific guidance on how this should be performed. For example, a study by Lundh and Gotzsche[15] examining the Instruction to Authors of 50 Cochrane Review Groups found that only half had specific recommendations for using the risk of bias assessment of studies analytically in a systematic review. The Cochrane Handbook recommends that the assessment of the risk of bias within each trial should inform the statistical analysis.[13] The two preferred analytical strategies are to either restrict the primary meta-analysis to studies at low risk of bias or to present the meta-analysis stratified according to risk of bias. It is recommended that the choice between these strategies should be based on the context of the particular systematic review and the balance between the potential for bias and the loss of precision when studies at high or unclear risk of bias are excluded.[16] However, it is unclear to what extent such restrictions should include all methodological components at high risk of bias, given the evidence that some components might be more susceptible to bias than others.[2] Even when risk of bias assessments are not incorporated into the statistical analysis, it is still possible to present a meta-analysis for all studies while providing a summary of the risk of bias across studies. However, there is then a danger that any risk of bias will be downplayed in the discussion and conclusions of the systematic review.[16]

Comparison with other studies
The findings from our study are consistent with those of an earlier study by Moja et al[17] who compared the assessment methodological quality in 809 Cochrane reviews and 156 systematic reviews in paper-based journals published between 1995 and 2002. Their study also showed that only 10% of systematic reviews incorporated the assessment of risk of bias of the primary studies into the statistical analysis (eg, by performing a sensitivity analysis), suggesting no overall improvement in the last 10 years. It is clear that despite all the valuable efforts to transparently report and display the potential risk of bias of primary studies (which in itself can be very time consuming), the impact on the overall findings of the systematic review is rarely assessed

formally. This is despite the growing number of systematic reviews being published,[18] improvements in systematic review methodology,[6 16] and methods of reporting systematic review findings.[9 10]

Limitations

A limitation of our study is that our sample of non-Cochrane reviews was drawn from DARE, which includes only systematic reviews which meet strict methodological criteria (http://www.crd.york.ac.uk/crdweb). It is likely, therefore, that the findings from our sample of non-Cochrane reviews might give an underestimate of the problem compared to systematic reviews identified from other sources. For example, in a study of 213 systematic reviews of randomised trials published in 2004 and identified by searching MEDLINE, all Cochrane reviews reported information about quality assessment (risk of bias) compared to only half of the non-Cochrane reviews.[19] This is similar to a study by Jadad *et al*[20] of 75 systematic reviews published in 1995 which found that all Cochrane reviews reported information about quality assessment compared to only a third of the non-Cochrane reviews.

Conclusions

Our study shows that overall the Cochrane reviews performed better than non-Cochrane reviews in the reporting of how assessments of risk of bias of the primary studies were carried out; however, both largely failed to show how such assessments were incorporated into the statistical analysis and in the interpretation of the overall conclusions of the systematic review. It is not sufficient to present the analysis and interpretation of a systematic review based on all included studies and ignore the flaws identified during the assessment of risk of bias.[16] The higher the proportion of studies assessed at high risk of bias, the more cautious the authors should be in the analysis and interpretation of the results.[2] From our study, it is clear that these recommendations are not always followed;, the reasons for this are unclear and would warrant further investigation.

REFERENCE

1. Pildal J, Hrobjartsson A, Jorgensen KJ, *et al*. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36:847–57.
2. Savovic J, Jones HE, Altman DG, *et al*. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012.
3. Schulz KF, Grimes DA. *The Lancet hadbook of essential concepts in clinical research*. London: Elsevier 2006.
4. Dwan K, Altman DG, Arnaiz JA, *et al*. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e30–81.
5. Boutron I, Ravaud P. Classification systems to improve assessment of risk of bias. *J Clin Epidemiol* 2012;65:236–8.
6. Methods guide for effectiveness and comparative effectiveness reviews. AHRQ Publication No 10(12)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. 2012. http://www.effectivehealthcare.hrq.gov.
7. Juni P, Witschi A, Bloch R, *et al*. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
8. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.
9. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
10. Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;151:W65–94.
11. Guyatt GH, Oxman AD, Vist GE, *et al*. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
12. Guyatt GH, Oxman AD, Vist G, *et al*. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
13. Higgins JP, Altman DG, Gotzsche PC, *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
14. Jadad AR, Moore RA, Carroll D, *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.
15. Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 2008;8:22.
16. Higgins JP, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S. eds *Cochrane handbook for systematic reviews of interventions* 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. http://www.cochrane-handbook.org and http://www.cochrane-handbook.org
17. Moja LP, Telaro E, D'Amico R, *et al*. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 2005;330:1053.
18. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?. *PLoS Med* 2010;7:e1000326.
19. Moher D, Tetzlaff J, Tricco AC, *et al*. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.
20. Jadad AR, Cook DJ, Jones A, *et al*. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;280:278–80.