

Reproducible research practices are underused in systematic reviews of biomedical interventions

Matthew J Page^{1,*}, Douglas G Altman², Larissa Shamseer^{3,4}, Joanne E McKenzie¹, Nadera Ahmadzai⁵,
Dianna Wolfe⁵, Fatemeh Yazdi⁵, Ferrán Catalá-López^{5,6}, Andrea C Tricco^{7,8}, David Moher^{3,4}

1. School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria 3004, Australia
2. UK EQUATOR Centre, Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford OX3 7LD, United Kingdom
3. Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, K1H 8L6, Canada
4. School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, K1H 8M5, Canada
5. Knowledge Synthesis Group, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, K1H 8M5, Canada
6. Department of Medicine, University of Valencia/INCLIVA Health Research Institute and CIBERSAM, Valencia, 46010, Spain
7. Knowledge Translation Program, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Ontario, M5B 1W8, Canada
8. Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, M5T 3M7, Canada

Correspondence to: Dr. Matthew Page, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria, 3004, Australia. Email address: matthew.page@monash.edu

Keywords: Reproducibility, Reporting, Systematic Reviews, Methodology, Quality

- 1
- 2 **Word count:** 4657 (main text), 2 figures, 3 tables, 1 appendix

119
120
121 1 **Abstract**
122

123 2 **OBJECTIVES:** To evaluate how often reproducible research practices, which allow others to recreate
124
125 3 the findings of studies, given the original data, are used in systematic reviews (SRs) of biomedical
126
127 4 research.
128

129 5 **STUDY DESIGN AND SETTING:** We evaluated a random sample of SRs indexed in MEDLINE® during
130
131 6 February 2014, which focused on a therapeutic intervention and reported at least one meta-analysis.
132
133 7 Data on reproducible research practices in each SR were extracted using a 26-item form by one
134
135 8 author, with a 20% random sample extracted in duplicate. We explored whether the use of
136
137 9 reproducible research practices was associated with a SR being a Cochrane review, as well as with
138
139 10 the reported use of the PRISMA Statement.
140
141

142 11 **RESULTS:** We evaluated 110 SRs of therapeutic interventions, 78 (71%) of which were non-Cochrane
143
144 12 SRs. Across the SRs there were 2,139 meta-analytic effects (including subgroup meta-analytic effects
145
146 13 and sensitivity analyses), 1,551 (73%) of which were reported in sufficient detail to recreate them.
147
148 14 Systematic reviewers reported the data needed to recreate all meta-analytic effects in the SR in 72
149
150 15 (65%) SRs only. This percentage was higher in Cochrane than in non-Cochrane SRs (30/32 [94%]
151
152 16 versus 42/78 [54%]; risk ratio 1.74, 95% confidence interval 1.39-2.18). Systematic reviewers who
153
154 17 reported imputing, algebraically manipulating, or obtaining some data from the study
155
156 18 author/sponsor infrequently stated which specific data were handled in this way. Only 33 (30%) SRs
157
158 19 mentioned access to datasets and statistical code used to perform analyses.
159
160

161 20 **CONCLUSIONS:** Reproducible research practices are underused in SRs of biomedical interventions.
162
163 21 Adoption of such practices facilitates identification of errors and allows the SR data to be re-
164
165 22 analysed.
166
167
168
169
170
171
172
173
174
175
176
177

1. Introduction

Biomedical researchers are increasingly encouraged to use reproducible research practices, which allow others to recreate the findings of studies, given the original data (1-3). Such practices include providing a detailed description of the data collected and used for analysis, along with descriptive metadata, clearly reporting the analysis methods and results, and sharing the dataset and statistical code used to perform analyses (1, 4). There are many benefits of performing such practices in the context of systematic reviews (SRs) of studies. For example, users can check for possible data entry errors when summary statistics for each included study are reported in sufficient detail. Transparent reporting of meta-analyses also makes it possible for others to reanalyse published meta-analyses using different inclusion criteria or statistical methods, or to perform additional analyses that address secondary research questions (5). For example, readers may reanalyse a published meta-analysis by restricting it to the subset of studies conducted in the setting where they work. Also, sharing of datasets and statistical analysis code allows other researchers to cumulatively add new data that are published, thus keeping meta-analytic effect estimates up-to-date (6, 7).

The limited data on use of reproducible research practices in SRs comes from studies that have recorded how well SRs adhere to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement. The PRISMA Statement includes an item recommending that for all outcomes considered, systematic reviewers report, for each study, “simple summary data for each intervention group and effect estimates and confidence intervals” (8). However, not all studies evaluating PRISMA adherence have provided data on adherence to this item, opting to present a total score summed across all PRISMA items instead (9). Further, many studies that have identified low adherence to this item have assessed SRs in a single clinical specialty (e.g. (10-14)), which limits generalisability of the findings. To our knowledge, no study has quantified how often systematic reviewers report the data needed to recreate all meta-analytic effect estimates in an SR (including subgroup meta-analytic effects and sensitivity analyses), nor investigated whether completeness of reporting varies by type of outcome (i.e. primary or other).

Efforts to increase transparent reporting of SR articles have existed for many years (e.g. the PRISMA Statement was disseminated in top medical journals in 2009); however, little attention has been given to the sharing of data collected as part of SRs (15). For example, since 2015 the BMJ encourages authors of all research articles to link their articles to the raw data from their studies, but requires data sharing on request as a minimum for clinical trials only (16). No study has investigated how often sharing of datasets and statistical analysis code is done by authors of SRs.

We investigated how often research practices that facilitate reproducibility of analyses were used in a cross-sectional sample of SRs of therapeutic interventions. We also explored whether the use of such reproducible research practices was associated with whether a SR was a Cochrane review, and with the systematic reviewers' reported use of the PRISMA Statement.

2. Methods

We conducted this project in accordance with a study protocol, which is available on the Open Science Framework (RRID:SCR_003238): <https://osf.io/523bq/>. This study was conducted concurrently with another project evaluating the application and interpretation of statistical methods in SRs. The results of the other project will be described elsewhere.

2.1. Selection of articles

We selected articles from a database of SRs we assembled previously (17), which consists of SRs of various study designs that were indexed in MEDLINE® during the month of February 2014. The eligibility criteria and search strategy to identify these SRs have been described elsewhere (17). Briefly, the database includes published, English-language articles that met the PRISMA for systematic review protocols (PRISMA-P) definition of a SR (18, 19). That is, articles that explicitly stated methods to identify studies (i.e. a search strategy), explicitly stated methods of study selection (e.g. eligibility criteria and selection process), and explicitly described methods of synthesis

(or other type of summary). The SRs were identified from a search of Ovid MEDLINE® using the search strategy employed by Moher et al. to retrieve SRs (20), which was restricted to February 2014.

In the previous study (17), all titles and abstracts were screened using the method of liberal acceleration (i.e. where two authors needed to independently exclude a record for it to be excluded, while only one author needed to include a record for it to be included). Subsequently, two authors independently screened each potentially relevant full text article retrieved. Any discrepancies in screening of full text articles were resolved via discussion, with adjudication by a third author when necessary. There were 684 SRs that met the inclusion criteria. A random sample of 300 SRs was drawn, and data on their epidemiological and reporting characteristics (e.g. clinical focus, screening and data collection methods used, funding) were collected. For the current study, we restricted inclusion to the 110 SRs of randomized trials or non-randomized studies of therapeutic (i.e. treatment or prevention) interventions, which reported at least one meta-analysis.

2.2. Data collection and verification

In the current study, we collected data using a standardized data collection form created in DistillerSR (Evidence Partners, Ottawa, Canada) (Appendix). We included 26 items that characterised reproducible research practices relating to the data and analyses. Selection and wording of items was influenced by recommendations in the PRISMA Statement (8, 21), the Methodological Expectations of Cochrane Intervention Reviews (22), and relevant papers on research reproducibility (2, 4, 5). Items included how many meta-analytic effect estimates were reported, which (if any) summary statistics for each individual study included in a meta-analysis were reported, whether it would be possible to recreate a particular subgroup analysis based on the data presented, and whether access to datasets and statistical analysis code was mentioned.

Some items in the form pertained to the whole SR (e.g. "How many meta-analyses were reported?"), while others were directed at one meta-analysis per SR, known as the index meta-analysis (e.g. "Were effect estimates and measures of precision of each individual study reported?"). The index meta-analysis was the first reported meta-analysis of the primary outcome. If no primary outcome was defined, we selected the first outcome listed in the Objectives section of the SR. If no outcome was listed in the Objectives, we selected the first reported meta-analysis, which may have been identified from the Abstract or Results section of the SR, depending on which was first reported in the article. We classified index meta-analysis outcomes as all-cause mortality, other objective outcome not requiring judgement (e.g. pregnancy, live births, laboratory outcomes), clinician-assessed outcome requiring judgement (e.g. events determined by clinical examination, cause-specific mortality), or patient-reported outcome (e.g. pain, mental health outcomes).

To count the number of meta-analyses in a SR, we summed the number of meta-analytic effect estimates presented in every forest plot, table, text, or web-based appendix. If a particular meta-analytic effect was presented in multiple locations, we only counted it once. For subgroup analyses with an overall effect reported (which synthesised data across all subgroups), we counted each subgroup effect as well as the overall effect, as long as each subgroup included at least two studies. We counted each meta-analytic effect regardless of whether it was completely reported. For example, if the systematic reviewers presented a meta-analytic effect estimate and 95% confidence interval (CI) and then stated that both were similar in a particular sensitivity analysis (without reporting the revised effect estimate and 95% CI), we counted each as a separate meta-analytic effect (i.e. two in this instance). However, we did not attempt to estimate the number of meta-analytic effects if the total number was unclear. For example, systematic reviewers may have stated in the Methods section, "We planned several sensitivity analyses based on study quality and intensity of intervention" (without providing definitions of each variable), yet only included a single statement in the Results section such as, "All results were robust in sensitivity analyses (data available on request)". In these cases, it was unclear how sensitivity analyses were actually

performed (e.g. a meta-analysis may have been subjected to multiple sensitivity analyses using various definitions of “intervention intensity”), or whether they were performed for all or only some outcomes. Therefore, we only counted analyses that were evidently performed.

We evaluated the completeness of reporting of results of all included studies to determine whether it would be possible to recreate each meta-analysis, subgroup analysis or sensitivity analysis based on the data presented. We did not attempt to recreate the analyses ourselves (e.g. by extracting the relevant data from each meta-analysis and re-calculating the meta-analytic effect). Analyses needed to meet the following criteria to be judged as “recreatable”:

1. effect estimates (e.g. mean difference or risk ratio) and measures of precision (e.g. 95% CIs or standard errors) were presented numerically for each individual study included in the analysis, or could be calculated from the study-level summary statistics reported (such as means and standard deviations (SDs) per intervention arm);
2. it was clear which studies were included in the analysis;
3. for subgroup analyses and sensitivity analyses, it was clear which level of the covariate was assigned to each study.

For example, if systematic reviewers presented a meta-analysis on a forest plot, which displayed the mean difference and 95% CI of each included study, then we recorded that this meta-analysis could be recreated from the data presented. However, if the systematic reviewers also stated that they performed a sensitivity analysis for this particular meta-analysis, but did not report which studies were removed, we recorded that the sensitivity analysis could not be recreated. Not all of the relevant data needed to be presented in a single forest plot to be judged as recreatable. For example, when systematic reviewers reported subgroup meta-analytic effect estimates in text only, we considered the subgroup analyses to be recreatable if the respective study effect estimates and 95% CIs were presented on a forest plot, and it was clear from a characteristics table what level of the covariate was assigned to each study.

One author (MJP) collected data from all SRs, from both the article and any web-based appendices. Once data from all articles were collected, data from a 20% random sample of SRs (n=22) were collected independently by one of three author (NA, DW, FY). Comparison of the data collected revealed 11 of 26 items where a discrepancy existed between two authors on at least one occasion (items marked in Appendix). All discrepancies were resolved via discussion. All items with a discrepancy in the 20% sample of SRs were checked for accuracy by one author (MJP) across the remaining 80% of SRs.

2.3. Data analysis

Once all discrepancies between data collectors had been resolved, the dataset for all included SRs was exported from DistillerSR into Microsoft Excel, where data were cleaned (i.e. invalid characters were removed and text data were converted to numeric where appropriate). Data for some items collected in the previous study by Page et al. (17) on general characteristics of the SRs (e.g. clinical focus, country of corresponding author) were merged with the current dataset. We summarised data as frequency and percentage for categorical items and median and interquartile range (IQR) for continuous items. We analyzed characteristics of all SRs and of SRs categorized as Cochrane or non-Cochrane. We explored whether use of reproducible research practices was associated with a SR being a Cochrane review, and, in a post-hoc analysis, with self-reported use of the PRISMA Statement. For the latter analysis, we excluded Cochrane SRs because they are supported by software that promotes complete reporting. We quantified associations using the risk ratio, with 95% CIs. All analyses were performed using Stata version 14 software (RRID:SCR_012763) (23).

3. Results

3.1. General characteristics of SRs

We evaluated 110 SRs of therapeutic interventions. Nearly all (97/110 [88%]) were published in late 2013, and 68/110 (62%) were led by systematic reviewers based in China, UK, Canada or USA (Table 1). The majority of SRs (78/110 [71%]) were not Cochrane reviews. The SRs focused on interventions

for one of 19 different conditions, with diseases of the digestive system, diseases of the circulatory system, infectious and parasitic diseases, and neoplasms being the most common. The interventions studied were pharmacological in 55/110 (50%) SRs, non-pharmacological in 43/110 (39%), or both in 12/110 (11%). A median of 13 studies (IQR 7-23) were included in the SRs. Authors of 32/78 (41%) non-Cochrane SRs reported using the PRISMA Statement to guide conduct or reporting. Most SRs (58/110 [53%]) were funded by a non-profit source, but the funding source was not reported in 38/110 (35%). The SRs included a median of 13 (IQR 5-27) meta-analytic effects (including those from subgroup meta-analyses and sensitivity analyses).

Table 1. General characteristics of systematic reviews of therapeutic interventions

Characteristic	All (n=110) ^a
Total number of journals	63
Journal impact factor (Thomson ISI 2012)	
0.0 – 5.0	59 (54%)
5.1 – 10	40 (36%)
10.1 – 15	0 (0%)
>15	2 (2%)
No impact factor	9 (8%)
Year of publication	
2014	11 (10%)
2013	97 (88%)
2012	2 (2%)
Country of corresponding author	
China	23 (21%)
UK	17 (15%)
Canada	15 (14%)
USA	13 (12%)
Other (21 countries with <10 SRs per country)	42 (38%)
Type of condition addressed by the SR (ICD-10 category)	
Diseases of the digestive system	14 (13%)
Diseases of the circulatory system	13 (12%)
Certain infectious and parasitic diseases	13 (12%)
Neoplasms (including cancers, carcinomas, tumors)	11 (10%)
Other (15 other ICD-10 classifications)	59 (54%)
Types of interventions addressed	
Pharmacological	55 (50%)
Non-pharmacological	43 (39%)
Both	12 (11%)
Number of included studies	13 (7-23)
Number of included participants	1,851 (630-5,540)
Use of PRISMA Statement mentioned in non-Cochrane SRs	32/78 (41%)
SR protocol/registration mentioned	
SR registered (e.g. in PROSPERO)	6 (5%)
Protocol publicly available	36 (33%)
Source of funding	
Non-profit	58 (53%)
For-profit	1 (1%)
Mixed	1 (1%)
Systematic reviewers specified there was no funding	12 (11%)
Not reported	38 (35%)
Statistical analyses reported	
Number of meta-analytic effect estimates	13 (5-27)
Any subgroup analyses reported	52 (47%)
Any sensitivity analyses reported	55 (50%)

^aData given as number (percent) or median (IQR). The denominator of fractions indicates the number of reports where the variable concerned was considered relevant to the systematic review (SR). Illustrative binomial exact 95% confidence intervals for percentages when sample size is 110: 1% (0.02% to 5%); 5% (1% to 10%); 10% (5% to 17%); 25% (17% to 34%); 50% (39% to 59%); 75% (65% to 82%). ICD-10 = International Classification of Diseases, Tenth Revision.

3.2. General characteristics of index meta-analyses

Of the index meta-analysis outcomes examined, 68/110 (62%) were described explicitly as a “primary” outcome; no primary outcome was specified in the remaining SRs (Table 2). Most outcomes were classified as “other objective outcome not requiring judgement” or “clinician-assessed outcome requiring judgement” (each in 39/110 [35%] SRs). A median of six (IQR 3-11) studies were included in the index meta-analyses; however, the median number of studies was smaller in Cochrane (4, IQR 2-6) compared with non-Cochrane (7, IQR 4-12) SRs. The most common effect measures used in the index meta-analyses were the risk ratio (36/110 [33%]) and the mean difference (25/110 [23%]). A subgroup analysis accompanied 42/110 (38%) index meta-analyses, and a sensitivity analysis accompanied 51/110 (46%) index meta-analyses. All of these characteristics were similar in Cochrane and non-Cochrane SRs except for the reporting of a subgroup analysis, which was less frequent in Cochrane SRs.

Table 2. General characteristics of the index meta-analysis in each systematic review

Characteristic	All (n = 110) ^a	Cochrane (n = 32) ^a	Non-Cochrane (n = 78) ^a
Type of outcome ^b			
All-cause mortality	3 (3%)	0 (0%)	3 (4%)
Other objective outcome not requiring judgement	39 (35%)	11 (34%)	28 (36%)
Clinician-assessed outcome requiring judgement	39 (35%)	10 (31%)	29 (37%)
Patient-reported outcome	29 (26%)	11 (34%)	18 (23%)
Described as a primary outcome	68 (62%)	23 (72%)	45 (58%)
Number of included studies	6 (3-11)	4 (2-6)	7 (4-12)
Number of included participants	593 (309-2,444)	419 (218-1,604)	817 (341-2,952)
Effect measure			
Risk ratio	36 (33%)	14 (44%)	22 (28%)
Odds ratio	20 (18%)	4 (13%)	16 (21%)
Risk difference	2 (2%)	0 (0%)	2 (3%)
Mean difference	25 (23%)	6 (19%)	19 (24%)
Standardized mean difference	10 (9%)	4 (13%)	6 (8%)
Hazard ratio	6 (5%)	2 (6%)	4 (5%)
Rate ratio	2 (2%)	2 (6%)	0 (0%)
Other (e.g. proportion, rate)	9 (8%)	0 (0%)	9 (12%)
Random-effects model used	62 (56%)	13 (41%)	49 (63%)
Subgroup analysis reported	42 (38%)	9 (28%)	33 (42%)
Sensitivity analysis reported	51 (46%)	14 (44%)	37 (47%)

^aData given as number (percent) or median (IQR). Illustrative binomial exact 95% confidence intervals for percentages when sample size is 110: 1% (0.02% to 5%); 5% (1% to 10%); 10% (5% to 17%); 25% (17% to 34%); 50% (39% to 59%); 75% (65% to 82%).

^bExamples of "other objective outcome not requiring judgement" include pregnancy, live births, and laboratory outcomes such as biochemical measurements or serologic tests. Examples of "clinician-assessed outcome requiring judgement" include events determined by clinical examination, and cause-specific mortality.

3.3. Reproducible research practices in SRs

Across the SRs there were 2,139 meta-analytic effects (including those from subgroup meta-analyses and sensitivity analyses), 1,551 (73%) of which were reported in sufficient detail to recreate them. In 72/110 (65%) SRs, systematic reviewers reported the data needed to recreate all meta-analytic effect estimates in the SR, including subgroup meta-analytic effects and sensitivity analyses (Table 3). This proportion rose to 85/110 (77%) when considering only the core meta-analyses (i.e. excluding subgroup meta-analytic effects and sensitivity analyses), and to 102/110 (93%) when considering only the index meta-analysis.

1
2 In 79/110 (72%) of the index meta-analyses, summary statistics (e.g. means and SDs) for each
3 included study were reported, and in 101/110 (92%), effect estimates (e.g. mean difference) and
4 measures of precision (e.g. 95% CIs) of each included study were reported (Table 3). Very few (2/110
5 [2%]) systematic reviewers stated that some data in the index meta-analyses had been imputed (e.g.
6 missing SDs). However, it was clear which data were imputed and how this was conducted for only
7 one of these two meta-analyses. Slightly more systematic reviewers (16/110 [15%]) reported that
8 some data in the index meta-analysis had been algebraically manipulated prior to inclusion (e.g. 95%
9 CIs of the mean difference were converted to SDs). However, it was clear which data were
10 manipulated and how so in only 5/16 (31%) of these meta-analyses. Some unpublished data had
11 been obtained from the study author/sponsor for a small proportion of index meta-analyses (7/110
12 [6%]). However, systematic reviewers clearly specified which particular data had been obtained only
13 in three of these seven (43%) meta-analyses.

14
15 Of the 62 index meta-analyses that were analysed using a random-effects method, information on
16 the specific method was lacking in 11/62 (18%). Specifically, systematic reviewers had not reported
17 (or it could not be inferred from the statistical package used) which between-trial variance estimator
18 (e.g. DerSimonian and Laird (24), Sidik and Jonkman (25)), and which method to calculate the
19 confidence interval, had been used. Of the 42 index meta-analyses with at least one accompanying
20 subgroup analysis, the data needed to recreate all subgroup analyses was available in 29/42 (69%).
21 Of the 51 index meta-analyses with at least one accompanying sensitivity analysis, the data needed
22 to recreate all sensitivity analyses was available in 34/51 (67%).

23
24 In 32/110 (29%) SRs – all of which were Cochrane SRs – systematic reviewers provided access to a
25 downloadable Review Manager (RevMan, RRID:SCR_003581) file (26), which contains editable data
26 included all meta-analyses and in-built software to perform reanalysis. None of the non-Cochrane

827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885

- 1 SRs provided access to such a dataset, but one indicated that the dataset and statistical code used to
- 2 perform analyses was available on request from the corresponding author.

- 3
- 4

Table 3. Reproducible research practices in systematic reviews of therapeutic interventions

Reproducible research practice	All (n = 110) ^a	Cochrane (n = 32) ^a	Non-Cochrane (n = 78) ^a
Reported the data needed to recreate all meta-analytic effect estimates in the SR ^b	72 (65%)	30 (94%)	42 (54%)
Reported the data needed to recreate all core meta-analytic effect estimates in the SR ^{b,c}	85 (77%)	32 (100%)	53 (68%)
Reported the data needed to recreate the index meta-analytic effect estimate ^b	102 (93%)	32 (100%)	70 (90%)
Reported summary statistics for each individual study in the index meta-analysis ^d	79 (72%)	29 (91%)	50 (64%)
Reported effect estimates and measures of precision for each individual study in the index meta-analysis	101 (92%)	32 (100%)	69 (88%)
Reported that some data in the index meta-analysis had been imputed	2 (2%)	1 (3%)	1 (1%)
Clear which data were imputed and how	1/2 (50%)	1/1 (100%)	0/1 (0%)
Reported that some data in the index meta-analysis had been algebraically manipulated ^e	16 (15%)	4 (13%)	12 (15%)
Clear which data were manipulated and how	5/16 (31%)	3/4 (75%)	2/12 (17%)
Reported that some data in the index meta-analysis had been obtained from the study author/sponsor	7 (6%)	4 (13%)	3 (4%)
Clear which data were obtained	3/7 (43%)	3/4 (75%)	0/3 (0%)
Reported the type of random-effects method used for the index meta-analysis (e.g. between-trial variance estimator stated or inferred)	51/62 (82%)	13/13 (100%)	38/49 (78%)
Reported the data needed to recreate each subgroup analysis for the index meta-analysis			
For all subgroup analyses	29/42 (69%)	9/9 (100%)	20/33 (61%)
For some subgroup analyses	5/42 (12%)	0 (0%)	5/33 (15%)
Not for any subgroup analysis	8/42 (19%)	0 (0%)	8/33 (24%)
Reported the data needed to recreate each sensitivity analysis for the index meta-analysis			
For all sensitivity analyses	34/51 (67%)	11/14 (79%)	23/37 (62%)
For some sensitivity analyses	9/51 (18%)	1/14 (7%)	8/37 (22%)
Not for any sensitivity analysis	8/51 (16%)	2/14 (14%)	6/37 (16%)
Mention of access to datasets and statistical analysis code used to perform analyses ^f	33 (30%)	32 (100%)	1 (1%)
Provided access to a file including all data included in meta-analyses and in-built software for reanalysis (e.g. RevMan file)	32 (29%)	32 (100%)	0 (0%)
Indicated that dataset and statistical analysis code were available upon request	1 (1%)	0 (0%)	1 (1%)

^aData given as number (percent). The denominator of fractions indicates the number of reports where the variable concerned was considered relevant to the systematic review (SR). Illustrative binomial exact 95% confidence intervals for percentages when sample size is 110: 1% (0.02% to 5%); 5% (1% to 10%); 10% (5% to 17%); 25% (17% to 34%); 50% (39% to 59%); 75% (65% to 82%).

^bBy "data needed to recreate all meta-analytic effect estimates", we mean that it is clear which studies were included in the meta-analysis, and effect estimates (e.g. mean difference or risk ratio)

and measures of precision (e.g. 95% confidence intervals) were reported numerically or could be calculated from summary statistics for each individual study.

^cExcludes subgroup meta-analytic effects and sensitivity analyses.

^dBy "summary statistics" we mean the number of events and sample size for binary outcomes, and the mean, standard deviation and sample size for continuous outcomes.

^eExamples include when 95% CIs of the mean difference were converted to standard deviations, or percentages were converted to frequencies.

^fExamples include when a data file (e.g. in RevMan or Microsoft Excel format) and code file (e.g. text file containing Stata commands) is uploaded as a supplementary file with the paper, or an online link to such files is provided, or information on how to request such files is provided. By analyses we mean meta-analyses, subgroup analyses and sensitivity analyses.

3.4. Influence of Cochrane status and self-reported use of PRISMA on reproducible research practices

All reproducible research practices were observed more often in Cochrane SRs compared with non-Cochrane SRs (Table 3, Figure 1). For example, Cochrane SRs were 74% more likely to report the data needed to recreate all meta-analytic effect estimates in the SR (30/32 [94%] Cochrane versus 42/78 [54%] non-Cochrane; risk ratio (RR) 1.74, 95% CI 1.39-2.18). The RR associations had 95% CIs that excluded the null in all cases except one – the reporting of data needed to recreate all sensitivity analyses for the index meta-analysis (RR 1.26, 95% CI 0.87-1.83).

Figure 1. Unadjusted risk ratio (RR) associations between type of systematic review (Cochrane versus non-Cochrane therapeutic systematic review) and reproducible research practices.

The association between self-reported mention of the PRISMA Statement and use of reproducible research practices in non-Cochrane SRs was less clear (Figure 2). Only two RR associations that favoured the use of PRISMA had 95% CIs that excluded the null. These included the reporting of summary statistics for each individual study included in the index meta-analysis (25/32 [78%] PRISMA versus 25/46 [54%] no PRISMA; RR 1.44, 95% CI 1.04-1.98) and the reporting of effect estimates and 95% CIs for each individual study included in the index meta-analysis (31/32 [97%] PRISMA versus 38/46 [83%] no PRISMA; RR 1.17, 95% CI 1.01, 1.36). Fewer SRs mentioning PRISMA

reported the data needed to recreate *all* meta-analytic effect estimates in the SR (15/32 [47%] PRISMA versus 27/46 [59%] no PRISMA; RR 0.80, 95% CI 0.51, 1.24). This was largely driven by the less frequent reporting of data needed to recreate all subgroup analyses in the SRs mentioning PRISMA (Figure 2).

Figure 2. Unadjusted risk ratio (RR) associations between self-reported use of PRISMA and reproducible research practices in non-Cochrane therapeutic systematic reviews.

4. Discussion

The use of reproducible research practices in SRs of therapeutic interventions was suboptimal in our sample. Systematic reviewers reported the data needed to recreate all meta-analytic effect estimates in the SR, including subgroup meta-analytic effects and sensitivity analyses, in only 65% of SRs. This percentage was higher in Cochrane than in non-Cochrane SRs (94% versus 54%). In contrast, the data needed to recreate the index (primary or first reported) meta-analysis was available in nearly all SRs (93%). Despite being recommended by PRISMA, summary statistics (e.g. means and standard deviations) of each individual study were not reported for 28% of index meta-analyses. Systematic reviewers who reported imputing, algebraically manipulating, or obtaining some data from the study author/sponsor infrequently stated which specific data were handled in this way. Only 30% of SRs mentioned access to datasets and statistical code used to perform analyses.

4.1. Strengths and limitations of the study

There are several strengths of our methods. We included SRs that had been identified and selected previously using rigorous methods (i.e. a systematic search and screening by two authors independently) (17). We did not restrict inclusion based on the clinical focus of the SR or type of SR (Cochrane or non-Cochrane). Therefore, we could collect data on a broad cross-section of SRs, which enhances the generalisability of our results. We collected data from both the original articles and

1 supplementary appendices. Unlike previous studies (e.g. (10-14)), we could quantify which particular
2 types of data were reported for meta-analytic effects (i.e. summary statistics or effect estimates or
3 both for each included study), and recorded whether this varied by type of outcome (i.e.
4 primary/first reported or other).

5
6 Some limitations of our study also exist. We included only SRs that had to meet some minimum
7 methodological quality criteria (as outlined in the PRISMA-P definition of an SR (18)), so our findings
8 may not generalise to SRs that are of poorer quality than the ones we examined (27). Our findings
9 may also not generalise to SRs published in a language other than English or indexed outside of
10 MEDLINE®. Our sample size is small, so the 95% confidence intervals of risk ratios associations are
11 imprecise. Two authors independently and in duplicate extracted data from a 20% random sample of
12 SRs; one author only evaluated the remainder. We attempted to minimize data collection errors by
13 verifying data for 11 of 26 items where there was at least one discrepancy between two authors in
14 the 20% random sample. In some SRs, it was unclear how particular subgroup/sensitivity analyses
15 were performed, or whether they were performed for all or only some outcomes. By not contacting
16 the systematic reviewers to resolve these uncertainties, we were only able to determine whether
17 analyses that were *evidently* performed had the data necessary for users to recreate them. We could
18 not determine whether systematic reviewers presented the data needed to recreate all analyses
19 that were *actually* performed.

20
21 Another possible limitation is that the majority of SRs were published in late 2013. It is possible that
22 more recent SRs use reproducible research practices more often, given the increasing number of
23 publications discussing scientific reproducibility (e.g. (2, 3, 28-31)). Further, more journals include
24 statements on data sharing nowadays. We analysed the data sharing policies of the 63 journals that
25 published the 110 SRs evaluated, and found that, as of September 2017, 27 journals encourage data
26 sharing for all research articles, while three require it (policies of each journal are available on the
27 Open Science Framework (RRID:SCR_003238): <https://osf.io/qda8f/>). Our study provides a useful

1 baseline against which the impact of these data sharing statements have on future SRs can be
2 assessed.

3

4 **4.2. Comparison with other studies**

5 Previous studies have also reported suboptimal reproducible research practices in SRs and other
6 study designs. For example, in a random sample of 441 biomedical journal articles published
7 between 2000 and 2014, mostly on uncontrolled human or animal studies, none made all raw data
8 directly available (1). Other studies have evaluated the adherence of SRs to PRISMA (e.g. (9-14)), and
9 reported low adherence to the item recommending that summary statistics, effect estimates and
10 measures of precision for each included study be reported. To our knowledge, ours is the first study
11 to explore the use of reproducible research practices in SRs in detail, by evaluating multiple practices
12 (several of which have not been examined previously), and multiple analyses in the SRs.

13

14 **4.3. Explanations of study results and implications**

15 It is reassuring that the data needed to recreate the index (primary or first reported) meta-analysis
16 was available in nearly all SRs, although disappointing that this was not the case for all analyses.

17 There are several possible reasons for this finding. Some systematic reviewers may consider it
18 sufficient to report data fully for the primary outcome only, given that this outcome is likely to be
19 the most important to clinical practice. Others may cite the space constraints of journals, since word
20 limits and restrictions on the numbers of tables and figures force systematic reviewers to be
21 selective about which data they present in an article. Nevertheless, several options exist to present
22 the relevant data for all analyses. For example, most journals now allow a web-based appendix for
23 extensive descriptions of methods and results, and we observed several examples of such
24 appendices providing complete data for hundreds of meta-analytic effect estimates (e.g. (32)). If
25 appendices are not allowed by a journal, systematic reviewers can upload the relevant data to public
26 repositories such as the Open Science Framework (RRID:SCR_003238, <https://osf.io/>) or the Dryad

1 Digital Repository (RRID:SCR_005910, <http://datadryad.org/>). With these options, reproducible

2 research practices should become routine in SRs.

3

4 Our findings suggest that some systematic reviewers may not recognise the benefits of reporting for

5 each meta-analysis the summary statistics for each intervention group in each study (e.g. means and

6 SDs), and the source of the data (e.g. extracted from the article, obtained from the study author, or

7 imputed). For example, odds ratios can be more easily interpreted if the event rates per group are

8 also presented (21). Also, when SDs are available for continuous outcomes, readers can examine

9 their consistency across studies and thus be reassured that standard error and SD have not been

10 confused by systematic reviewers (33). Further, knowing the source of the data facilitates the

11 conduct of meta-research (e.g. to quantify the rate of data extraction errors in a sample of SRs (34)).

12 We recognise that presenting summary statistics is not always possible with some meta-analysis

13 software packages. For example, users of RevMan (RRID:SCR_003581) (26) are unable to present

14 summary statistics when they synthesise study effect estimates that have been adjusted for a set of

15 confounders (using the generic inverse variance method). Further, for cluster randomized trials, if an

16 estimate of intervention effect and its standard were available from a model that adjusted for

17 clustering, then this should be used in preference to using the numerators and denominators and

18 making some adjustment to these. Nevertheless, in such instances readers should be informed of

19 the adjusted nature of the data (e.g. in an explanation provided in a footnote beneath the forest

20 plot, or in a linked table).

21

22 Our comparison of SRs mentioning use of PRISMA and those not suggests that the PRISMA

23 Statement could be revised to more strongly advocate the use of reproducible research practices.

24 PRISMA-users were statistically significantly more likely to report the data needed to recreate the

25 index meta-analysis, but not *all* meta-analyses in the SR. It is possible that users of PRISMA overlook

26 the guidance in the explanation and elaboration document that accompanies the checklist, which

27 explains why complete reporting of all meta-analyses is important, and includes recommendations

that are more detailed. New tools (e.g. such as writing software) that detail all the key elements of each PRISMA item may lead to more transparent reporting of SRs (35). Also, guidance on which datasets and statistical analysis code to share, and how to do so, could be added to the checklist.

It is perhaps not surprising that SR datasets were infrequently available in the SRs we examined, given that the sharing of data collected as part of SRs has received little attention from commentators (15). Even the SRs that were rated favourable in this regard (Cochrane SRs) share only a subset of all the data that would have been collected. That is, only the analysable study data needed to generate forest plots are available in the RevMan file that accompanies every published Cochrane SR. Systematic reviewers could provide access to an assortment of other types of data collected for SRs. For example, summary data for outcomes for which meta-analysis was not possible could be provided in a re-usable format (e.g. Microsoft Excel or CSV format), ready to be synthesised with new data once published. Also, data and code files used to convert statistics into a format required for meta-analysis (e.g. 95% CIs to SDs) could be shared to increase transparency (36). Further, data extraction files that clearly indicate where data were obtained from may reduce the need for study authors to handle repeated requests for data from different SR teams (15).

Infrastructure to share SR data are now available via repositories to store data collection forms, such as the Systematic Review Data Repository (SRDR) (37, 38), and results, via platforms such as the Open Science Framework (RRID:SCR_003238). However, doing so is not without its barriers. For example, authors may cite the additional time required to prepare data and metadata in a way that can be easily interpreted and used by other researchers. There may be a need for journals to require data sharing for SRs that they publish, and academic institutions to reward academics for open data practices, to incentivise these practices (39). Further, training courses for systematic reviewers should highlight the benefits of embracing data sharing initiatives such as these, to improve the quality and efficiency of future SRs.

5. Conclusions

Reproducible research practices in SRs of therapeutic interventions are suboptimal. Strategies are needed to facilitate the provision of detailed descriptions of data gathered and data used for analysis, transparent reporting of the analysis method and results, and sharing of datasets and statistical analysis code so that others can recreate the findings or perform secondary analyses.

Acknowledgments

We thank Sean Harrison (University of Bristol) for assistance with data analysis.

Competing Interests

We have read the journal's policy and the authors of this manuscript have the following competing interests: ACT is an Associate Editor for *Journal of Clinical Epidemiology* but had no involvement in the peer review process or decision for publication. MJP and JEM are affiliates of Cochrane Australia. MJP is a Co-Convenor of the Cochrane Bias Methods Group. JEM is a Co-Convenor of the Cochrane Statistical Methods Group. ACT is an author of two of the systematic reviews included in this study, but was not involved in eligibility assessment or data collection.

Funding

There was no direct funding for this study. MJP is supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535). DGA is a National Institute for Health Research Senior Investigator. JEM is supported by a NHMRC Australian Public Health Fellowship (1072366). FCL is supported by the Generalitat Valenciana (PROMETEOII/2015/021). ACT is funded by a Tier 2 Canada Research Chair in Knowledge Synthesis. DM is supported in part by a University Research Chair, University of Ottawa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

All authors declare to meet the ICMJE conditions for authorship. MJP, DGA and DM conceived the study design. LS and JEM provided input into the study design. MJP, DGA, LS, JEM and DM selected items for inclusion in the data collection form. MJP, NA, DW, and FY collected data. MJP undertook the statistical analyses. MJP wrote the first draft of the article. All authors contributed to revisions of the article. All authors approved the final version of the submitted article.

Data availability

The study protocol, data collection form, and the raw data and statistical analysis code for this study are available on the Open Science Framework: <https://osf.io/523bq/>

References

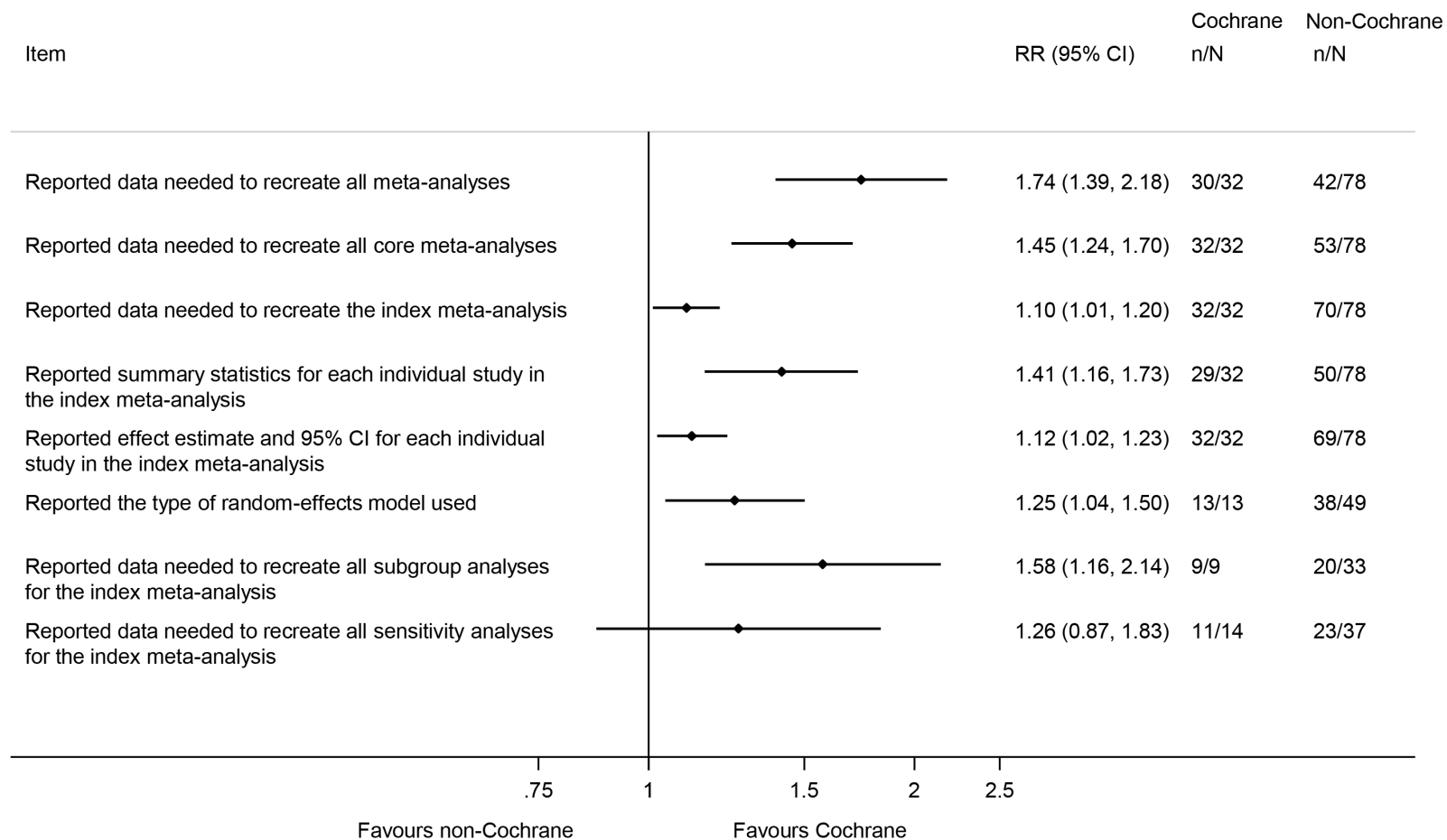
1. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biol.* 2016;14(1):e1002333.
2. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. SCIENTIFIC STANDARDS. Promoting an open research culture. *Science.* 2015;348(6242):1422-5.
3. Errington TM, Iorns E, Gunn W, Tan FE, Lomax J, Nosek BA. An open investigation of the reproducibility of cancer biology research. *Elife.* 2014;3:e04333.
4. Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? *Sci Transl Med.* 2016;8(341):341ps12.
5. Lakens D, Hilgard J, Staaks J. On the reproducibility of meta-analyses: six practical recommendations. *BMC Psychol.* 2016;4(1):24.
6. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med.* 2014;11(2):e1001603.
7. Crequit P, Trinquart L, Ravaud P. Live cumulative network meta-analysis: protocol for second-line treatments in advanced non-small-cell lung cancer with wild-type or unknown status for epidermal growth factor receptor. *BMJ Open.* 2016;6(8):e011841.
8. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* 2009;6:e1000097.
9. Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, Thielman J, et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. *BMJ.* 2014;348:g3804.
10. Tam WW, Lo KK, Khalechelvam P. Endorsement of PRISMA statement and quality of systematic reviews and meta-analyses published in nursing journals: a cross-sectional study. *BMJ Open.* 2017;7(2):e013905.

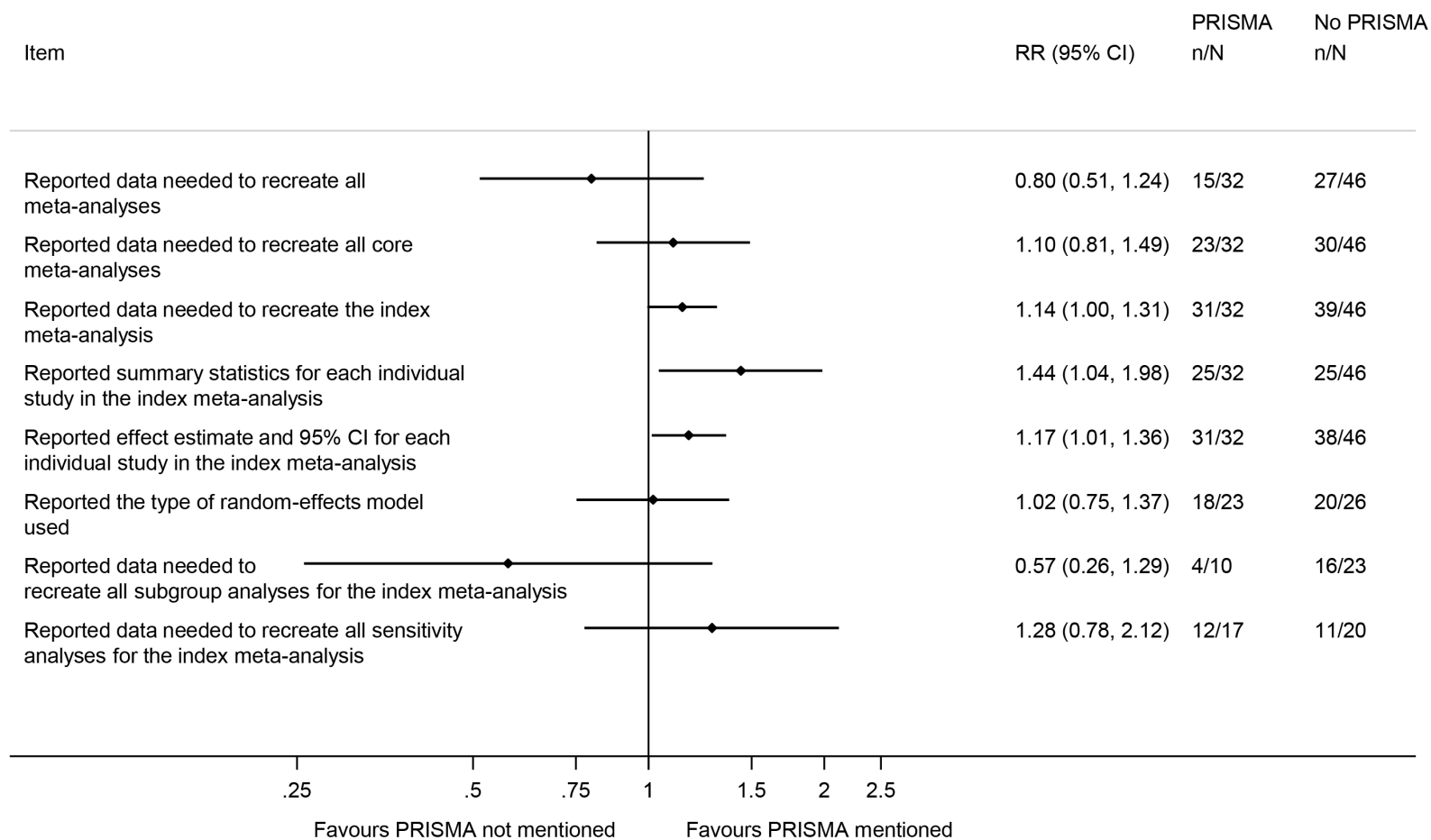
1476
1477
1478 1 11. Peters JP, Hooft L, Grolman W, Stegeman I. Reporting Quality of Systematic Reviews and
1479
1480 2 Meta-Analyses of Otorhinolaryngologic Articles Based on the PRISMA Statement. PLoS One.
1481
1482 3 2015;10(8):e0136540.
1483
1484 4 12. Liu Y, Zhang R, Huang J, Zhao X, Liu D, Sun W, et al. Reporting quality of systematic
1485
1486 5 reviews/meta-analyses of acupuncture. PLoS One. 2014;9(11):e113172.
1487
1488 6 13. Tunis AS, McInnes MD, Hanna R, Esmail K. Association of study quality with completeness of
1489
1490 7 reporting: have completeness of reporting and quality of systematic reviews and meta-
1491
1492 8 analyses in major radiology journals changed since publication of the PRISMA statement?
1493
1494 9 Radiology. 2013;269(2):413-26.
1495
1496 10 14. Gagnier JJ, Kellam PJ. Reporting and methodological quality of systematic reviews in the
1497
1498 11 orthopaedic literature. J Bone Joint Surg Am. 2013;95(11):e771-7.
1499
1500 12 15. Wolfenden L, Grimshaw J, Williams CM, Yoong SL. Time to consider sharing data extracted
1501
1502 13 from trials included in systematic reviews. Syst Rev. 2016;5(1):185.
1503
1504 14 16. Loder E, Groves T. The BMJ requires data sharing on request for all trials. BMJ.
1505
1506 15 2015;350:h2373.
1507
1508 16 17. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and
1509
1510 17 Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional
1511
1512 18 Study. PLoS Med. 2016;13(5):e1002028.
1513
1514 19 18. Moher D, Shamseer L, Clarke M, Ghera D, Liberati A, Petticrew M, et al. Preferred reporting
1515
1516 20 items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst
1517
1518 21 Rev. 2015;4(1):1.
1519
1520 22 19. Shamseer L, Moher D, Clarke M, Ghera D, Liberati A, Petticrew M, et al. Preferred reporting
1521
1522 23 items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and
1523
1524 24 explanation. BMJ. 2015;349:g7647.
1525
1526 25 20. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting
1527
1528 26 characteristics of systematic reviews. PLoS Med. 2007;4:e78.
1529
1530
1531
1532
1533
1534

1535
1536
1537 1 21. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA
1538
1539 2 statement for reporting systematic reviews and meta-analyses of studies that evaluate
1540
1541 3 healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
1542
1543 4 22. Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. *Methodological Expectations of*
1544
1545 5 *Cochrane Intervention Reviews*. London: Cochrane; 2016.
1546
1547 6 23. StataCorp. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP; 2015.
1548
1549 7 24. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88.
1550
1551 8 25. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results
1552
1553 9 of studies. *Stat Med*. 2007;26(9):1964-81.
1554
1555 10 26. The Nordic Cochrane Centre (The Cochrane Collaboration). *Review Manager (RevMan)*. 5.1.
1556
1557 11 Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2011.
1558
1559 12 27. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic
1560
1561 13 Reviews and Meta-analyses. *Milbank Q*. 2016;94(3):485-514.
1562
1563 14 28. Kidwell MC, Lazarevic LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg LS, et al.
1564
1565 15 Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing
1566
1567 16 Transparency. *PLoS Biol*. 2016;14(5):e1002456.
1568
1569 17 29. McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, et al. How open science helps
1570
1571 18 researchers succeed. *Elife*. 2016;5:e16800.
1572
1573 19 30. Nosek BA, Errington TM. Making sense of replications. *Elife*. 2017;6.
1574
1575 20 31. Open Science Collaboration. *PSYCHOLOGY*. Estimating the reproducibility of psychological
1576
1577 21 science. *Science*. 2015;349(6251):aac4716.
1578
1579 22 32. Hill J, Hoyt J, van Eijk AM, D'Mello-Guyett L, Ter Kuile FO, Steketee R, et al. Factors affecting
1580
1581 23 the delivery, access, and use of interventions to prevent malaria in pregnancy in sub-Saharan
1582
1583 24 Africa: a systematic review and meta-analysis. *PLoS Med*. 2013;10(7):e1001488.
1584
1585 25 33. Altman DG, Cates C. The need for individual trial results in reports of systematic reviews.
1586
1587 26 *BMJ*. 2001;323:776.
1588
1589
1590
1591
1592
1593

1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652

1	34.	Gotzsche PC, Hrobjartsson A, Maric K, Tendam B. Data extraction errors in meta-analyses that
2		use standardized mean differences. <i>JAMA</i> . 2007;298(4):430-7.
3	35.	Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online
4		writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool)
5		randomized controlled trial. <i>BMC Med</i> 2015;13:221.
6	36.	Shamseer L, Roberts J. Disclosure of data and statistical commands should accompany
7		completely reported studies. <i>J Clin Epidemiol</i> . 2016;70:272-4.
8	37.	Ip S, Hadar N, Keefe S, Parkin C, Iovin R, Balk EM, et al. A Web-based archive of systematic
9		review data. <i>Syst Rev</i> . 2012;1:15.
10	38.	Li T, Vedula SS, Hadar N, Parkin C, Lau J, Dickersin K. Innovations in data collection,
11		management, and archiving for systematic reviews. <i>Ann Intern Med</i> . 2015;162(4):287-94.
12	39.	Moher D, Goodman SN, Ioannidis JP. Academic criteria for appointment, promotion and
13		rewards in medical research: where's the evidence? <i>Eur J Clin Invest</i> . 2016;46(5):383-5.
14		





Competing Interests

We have read the journal's policy and the authors of this manuscript have the following competing interests: ACT is an Associate Editor for *Journal of Clinical Epidemiology* but had no involvement in the peer review process or decision for publication. MJP and JEM are affiliates of Cochrane Australia. MJP is a Co-Convenor of the Cochrane Bias Methods Group. JEM is a Co-Convenor of the Cochrane Statistical Methods Group. ACT is an author of two of the systematic reviews included in this study, but was not involved in eligibility assessment or data collection.

Supplementary File

Data collection form

Data in each systematic review were collected using the following form by one author. A 20% random sample of SRs was extracted in duplicate, which revealed several items that were discrepant between two reviewers on at least one occasion. Data for these items (indicated by an asterisk * below) were verified for all systematic reviews by one author.

Question	Response options
RefID	[Free text]
Your initials	[Free text]
1. Were any meta-analyses performed?	Yes No
2. *How many meta-analyses were reported (either in the main paper or online supplement)? Sum the number of meta-analyses reported either in forest plots, tables or text. If the same meta-analytic estimate is presented in a forest plot and in text, only count this estimate once. For subgroup analyses with an overall effect reported, count each subgroup meta-analytic effect as well as the overall effect. Count each meta-analysis regardless of whether data for it were fully reported. For example, if the authors present a meta-analysis on a forest plot and then state that they performed a sensitivity analysis for this particular meta-analysis, count each as a separate meta-analysis (i.e. n=2 meta-analyses in this instance).	[Free text]
3. *For how many meta-analyses were summary statistics or effect estimates and measures of precision of each individual study reported numerically (in either the main paper and online supplement), such that meta-analyses could be recreated? Sum the number of meta-analyses reported either in forest plots, tables or text. If the same meta-analytic estimate is presented in a forest plot and in text, only count this estimate once. By “effect estimates”, we mean estimates such as mean difference for continuous outcomes, or risk ratio, odds ratio or risk difference for binary outcomes. By “measures of precision”, we mean a standard error or confidence interval.	[Free text]

Question	Response options
For subgroup analyses with an overall effect reported, count each subgroup meta-analytic effect as well as the overall effect. Count each meta-analysis only if it could be recreated from the reported study-level data. For example, if the authors present a meta-analysis on a forest plot (showing study-level effect estimates and confidence intervals) and then state that they performed a sensitivity analysis for this particular meta-analysis, but it is unclear which studies were removed in the sensitivity analyses, only count the meta-analysis which can be recreated (i.e. the one on the forest plot). Note that a forest plot does not need to be presented for each sensitivity analysis; it is sufficient if study-level data are available for the primary meta-analysis and it is clear which studies were removed in the sensitivity analysis.	
4. Were any treatment-by-covariate interaction analyses (e.g. subgroup analyses or meta-regression) reported?	Yes No
5. Were any sensitivity analyses reported?	Yes No
Answer the remaining questions for only one meta-analysis per review (the index meta-analysis) Select the first reported meta-analysis of the primary outcome. If no primary outcome was defined, select the first outcome listed in the Objectives. If no outcome is listed in the Objectives, select the first reported meta-analysis. Note that the first result may be identified from the Abstract or Results section of the review, depending on where it is first reported in the publication.	
6. What is the outcome (e.g. all-cause mortality, anxiety)?	[Free text]
7. Was the outcome labelled as 'primary' by the review authors?	Yes No
8. How many studies were included in the meta-analysis? State "Not reported" if not reported	[Free text]
9. What was the total number of participants included in the meta-analysis?	[Free text]

Question	Response options
State "Not reported" if not reported	
10. What effect measure was used in the meta-analysis?	<p>Risk ratio</p> <p>Odds ratio</p> <p>Risk difference</p> <p>Hazard ratio</p> <p>Rate ratio</p> <p>Mean difference</p> <p>Standardized mean difference</p> <p>Ratio of means</p> <p>Other (please specify)</p>
<p>11. *Were summary statistics of each individual study reported (in text, table or figure) such that effect estimates of primary studies could be recreated?</p> <p>By "summary statistics", we refer to the mean, standard deviation and sample size for continuous outcomes, or the number of events and sample size for binary outcomes.</p>	<p>Yes</p> <p>No</p>
<p>12. Were effect estimates and measures of precision of each individual study reported (in text, table or figure)?</p> <p>By "effect estimates", we mean estimates such as mean difference for continuous outcomes, or risk ratio, odds ratio or risk difference for binary outcomes.</p> <p>By "measures of precision", we mean a standard error or confidence interval.</p>	<p>Yes, study effect estimates and measures of precision reported numerically (in text, tables or forest plots)</p> <p>Yes, but study effect estimates and measures of precision were only shown graphically (i.e. no numerical values)</p> <p>No</p>
13. *Did review authors report whether or not they imputed any missing statistics (e.g. standard	Review authors reported that they DID impute

Question	Response options
<p>deviations) for studies included in this particular meta-analysis?</p> <p>For example, do the authors report in the main text, Characteristics table, or in a footnote to the forest plot that they imputed missing statistics for particular studies, and then cite those studies?</p>	<p>missing statistics</p> <p>Review authors reported that they DID NOT impute missing statistics</p> <p>No such statement was made</p>
14. *Is it clear which values were imputed and how?	<p>Yes</p> <p>No</p>
<p>15. *Did review authors report whether or not they algebraically manipulated data (e.g. calculated a standard error from a P-value) for studies included in this particular meta-analysis?</p> <p>For example, do the authors report in the main text, Characteristics table, or in a footnote to the forest plot that they algebraically manipulated data for particular studies, and then cite those studies?</p>	<p>Review authors reported that they DID algebraically manipulate data</p> <p>Review authors reported that they DID NOT algebraically manipulate data</p> <p>No such statement was made</p>
16. *Is it clear which values were algebraically manipulated and how?	<p>Yes</p> <p>No</p>
<p>17. *Did review authors report whether or not they obtained unpublished outcome data from the trialists or sponsor, which were then included in this particular meta-analysis?</p> <p>For example, do the authors report in the main text, Characteristics table, or in a footnote to the forest plot that they retrieved data for particular studies, and then cite those studies?</p>	<p>Review authors reported that they DID obtain unpublished data</p> <p>Review authors reported that they DID NOT obtain unpublished data</p> <p>No such statement was made</p>
18. *Is it clear which values were retrieved from the trialists or sponsor?	<p>Yes</p>

Question	Response options
	No
19. Was the meta-analysis performed using the random-effects model?	Yes
	No
	Unclear/not stated
20. Which between-trial variance estimator was used?	DerSimonian and Laird
	Sidik and Jonkman
	Paule and Mandel
	Restricted Maximum Likelihood (REML)
	Other (please specify)
	Not reported
21. Which method was used to calculate the confidence interval for the summary effect?	Not reported, but Wald-type method assumed (select when DerSimonian and Laird approach stated but there is no specific mention of an alternative method to calculate the confidence interval)
	Wald-type method
	t-distribution
	Hartung-Knapp
	Other (please specify)
	Not reported
22. Was a treatment-by-covariate interaction analysis (e.g. subgroup analysis or meta-regression) reported for this particular meta-analysis?	Yes
	No

Question	Response options
	Unsure
<p>23. *Were sufficient data to enable replication of each interaction analysis reported? Only consider the interaction analyses relating to the index meta-analysis. By “sufficient data”, we mean that for each study, the stratum to which it belonged and the summary statistics or effect estimates and a measure of precision (standard error or confidence interval) were reported. Note that such information may be reported across multiple sections within a paper (e.g. characteristics table, forest plot).</p>	<p>Yes, for all interaction analyses</p> <p>Yes, but only for some interaction analyses</p> <p>No, not for any interaction analysis</p>
24. Was a sensitivity analysis reported for this particular meta-analysis?	<p>Yes</p> <p>No</p> <p>Unsure</p>
<p>25. *Were sufficient data to enable replication of each sensitivity analysis reported? Only consider the sensitivity analyses relating to the index meta-analysis. By “sufficient data”, we mean that it is clear which studies were included in the analysis, and the summary statistics or effect estimates and a measure of precision (standard error or confidence interval) for each included study were reported. Note that such information may be reported across multiple sections within a paper (e.g. characteristics table, forest plot).</p>	<p>Yes, for all sensitivity analyses</p> <p>Yes, but only for some sensitivity analyses</p> <p>No, not for any sensitivity analysis</p>
<p>26. Do authors report any mention of access to datasets that stand behind the analyses presented in the paper? For example, the complete dataset is uploaded as a supplementary file (e.g. as an Excel, RevMan or .dta file), or information on how to request a complete dataset is provided, or an online link to the complete dataset is provided. By analyses we mean meta-analyses, treatment-by-covariate interaction analyses, sensitivity analyses, and analyses of small-study effects.</p>	<p>Yes</p>

Question	Response options
	No
General comments	[Free text]