

Changing patterns in reporting and sharing of review data in systematic reviews with meta-analysis of the effects of interventions: cross sectional meta-research study

Participant-visible method context. Results, discussion, captions, and the target figure are excluded.

This study was conducted as one of a group of studies in the REPRISE (REProducibility and Replicability In Syntheses of Evidence) project. The REPRISE project is investigating various aspects relating to the transparency, reproducibility, and replicability of systematic reviews with meta-analysis of the effects of health, social, behavioural, and educational interventions. Methods for all studies were prespecified in the same protocol.²⁰ Deviations from the protocol for the current study are outlined in the supplemental data. Identification and selection of articles We included a random sample of systematic reviews with meta-analysis of the effects of a health, social, behavioural, or educational intervention (ie, any intervention designed to improve health (defined as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”²¹), promote social welfare and justice, change behaviour, or improve educational outcomes; see the supplemental data for full eligibility criteria). To be considered a systematic review, authors needed to have, at a minimum, clearly stated their review objective(s) or question(s); reported the source(s) (eg, bibliographic databases) used to identify studies meeting the eligibility criteria; and reported conducting an assessment of the validity of the findings of the studies included (eg, via an assessment of risk of bias or methodological quality). We did not exclude systematic reviews providing limited detail about the methods used. We only included systematic reviews that presented results for at least one pairwise metaanalysis of aggregate data. Systematic reviews with network meta-analyses were eligible if they included at least one direct (ie, pairwise) comparison that fulfilled the criteria mentioned above. Systematic reviews with only meta-analyses of individual participant data were excluded because all eligible systematic reviews in this study will be subjected to a reproducibility check in another component of the REPRISE project,²⁰ and we lacked the resources to reproduce these meta-analyses of individual participant data. Furthermore, only reviews written in English were included. Using search strategies created by an information specialist (SM), we systematically searched PubMed, Science Citation Index, and Social Sciences Citation Index via Web of Science, Scopus via Elsevier, and Education Collection via ProQuest for systematic reviews indexed from 2 November to 2 December doi: 10.1136/bmj-2022-072428 | BMJ 2022;379:e072428 | the bmj

BMJ: first published as 10.1136/bmj-2022-072428 on 22 November 2022. Downloaded from <https://www.bmj.com/> on 21 April 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

guidelines have been developed, with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement¹¹ being among the more widely used.¹² Reporting guidelines provide a structure for reporting a systematic review, along with recommendations of items to report.¹³ Originally released in 2009, PRISMA was recently updated (to PRISMA 2020) to reflect advances in systematic review methodology.¹⁴ The few studies examining the impact of PRISMA suggest that some items (eg, inclusion of a flow diagram) improved after its introduction, but that others (eg, mention of a review protocol) remained infrequently reported.¹⁵ These analyses are limited to reviews published before 2015, and therefore the influence of reporting guidelines on more recent systematic reviews is unclear. In addition to transparent reporting, advocates for research transparency¹⁶ 17 also recommend that authors share systematic review data files and analytical code used to generate meta-analyses.¹⁸ While all data for a meta-analysis are typically summarised in tables or forest plots, sharing an editable file containing extracted data (eg, CSV, RevMan (.rm5)) reduces the time and risk of errors associated with manual extraction of such data. This then facilitates the review’s reuse in future updates and replications, or its inclusion in overviews of reviews, clinical practice guidelines, educational materials, and meta-research studies.¹⁸ 19 Sharing review data files is relatively easier than sharing individual participant data from primary studies, and signals that review authors are committed to practices that they encourage from authors of primary studies, who are often asked to share their data. Infrequent sharing of data in systematic reviews in health research has been observed, but these findings might not be generalisable to all health topics⁴ or across journals.⁷ Moreover, the types of data shared (eg, unprocessed data extracted from reports, data included in meta-analyses) has not been examined, nor has the impact of journals’ data sharing policies on rates of sharing in systematic reviews. Without a current, comprehensive evaluation of the completeness of reporting of systematic reviews, we lack data on which

items are infrequently reported and require most attention from authors, peer reviewers, editors, and educators. Furthermore, without data on the frequency and type of materials review authors currently share, we lack insight into how receptive review authors are to calls to share data underlying research projects. To review these research gaps, we aimed to:

- Evaluate the completeness of reporting in a cross section of systematic reviews with meta-analysis published in 2020
- Evaluate the frequency of sharing review data, analytical code, and other materials in the same cohort of reviews
- Compare reporting in these reviews with a sample of reviews published in 2014;
- Investigate the impact of reporting guidelines on the completeness of reporting in reviews published in 2020

RESEARCH

Data collection Two authors (PN and either MJP, RK, or ZA) collected data independently and in duplicate from all of the 300 systematic reviews using a standardised form created in REDCap version 10.6.12, hosted at Monash University.²³ Any discrepancy in the data collected was

resolved via discussion or adjudication by the senior reviewer (MJP). Before data collection, a pilot test of the data collection form was performed on a random sample of 10 systematic reviews and the form was adjusted as necessary. The full data collection form (supplemental data) includes a subset of items used in previous evaluations of completeness of reporting^{9–10} along with additional items to capture some issues not previously examined. The wording of items in the data collection form was matched to previous evaluations^{9–10} to facilitate comparison. The form consisted of three sections (table 1). The first section captured general characteristics of the review, which were all extracted manually except for the country of the corresponding author, which was extracted using R code adapted from the easyPubMed package version 2.13.24–25. The interventions were classified as health, social, behavioural, or educational interventions (see definitions in the supplemental data). The second section consisted of items describing the review's reporting characteristics, the index metaanalysis (defined as the first meta-analysis mentioned in the abstract/results sections), and its data sharing characteristics. All of the reporting items evaluated are recommended in the 2009 PRISMA statement (in either the main checklist or the explanation and elaboration document²⁶), except for the items on whether search strategies for all bibliographic databases and nondatabase sources were reported. To facilitate our analysis of the impact of reporting guidelines, we also recorded whether the authors self-reported using a reporting guideline, defined as any document specifying essential items to report in a systematic review (eg, PRISMA, MECIR (Methodological Expectations of Cochrane Intervention Reviews), MECCIR (Methodological Expectations of Campbell Collaboration Intervention Reviews) standards). The final section captured the data sharing policy of the journal where the article was published. A data sharing policy refers to the journal's requirements and

Table 1 | Items for data collection and data sources (see the supplemental data (S4 appendix) for further details)

Source	Section	Systematic review*	General characteristics of the systematic review	Reporting characteristics of the systematic review
			Reporting characteristics of the index (first reported) meta-analysis	Sharing characteristics of the systematic review
			Journal website	

Reporting characteristics of the index (first reported) meta-analysis

Sharing characteristics of the systematic review

Journal website

Data items

Title; journal name; corresponding author's country; source of funding for the review; conflicts of interest of review authors; number of studies included in the review; types of participants and interventions investigated. Whether a reporting guideline was cited; whether a protocol or registration record for the review was cited; whether eligibility criteria for participants, interventions, comparators, outcomes, and study designs were reported; what details of the search methods were reported (including which databases, the interface used to search them, the years of coverage, date of the search, and whether a full Boolean search logic—using operators such as “AND,” “OR,” “NOT”—was reported); what method of study selection, data collection, and risk-of-bias assessment authors reported using; what software (and packages) authors reported using; whether the number of records yielded by the searches were reported overall and for each database; whether any full text articles excluded from the review were cited. Outcome domain investigated; number of included studies; effect measure used; whether methods required to prepare data for metaanalysis were reported; whether the meta-analysis model used was reported; whether the meta-analysis method used was reported; whether the between-study (heterogeneity) variance estimator used was reported; whether summary statistics for each included study were reported; whether effect estimates and measures of precision for each included study were reported. Whether a data or code availability

statement appeared in the review; which types of files were made publicly accessible either as a supplementary file or uploaded to a repository (eg, files containing data used in all analyses, analytical code used to generate results, files containing citations of all records that were screened and excluded); whether files shared had a persistent identifier (eg, doi) or license (eg, cc-by) applied to them. Journal name; whether the journal only publishes evidence syntheses (ie, systematic reviews and their protocols); whether the journal has a data or code sharing policy, or both; whether sharing data and/or issuing a data availability statement is mandatory for systematic reviews published by the journal.

*Includes the main report and any supplementary file(s), and the review protocol (if the authors specified that the relevant information was included).

the bmj | BMJ 2022;379:e072428 | doi: 10.1136/bmj-2022-072428

BMJ: first published as 10.1136/bmj-2022-072428 on 22 November 2022. Downloaded from <https://www.bmj.com/> on 21 April 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

2020. All searches were conducted on 3 December 2020. An example of the search strategy for PubMed was (meta-analysis[PT] OR meta-analysis[TI] OR systematic[sb]) AND 2020/11/02:2020/12/02[EDAT]). Search strategies for all databases are available in the supplemental data. We used Endnote version 9.3.3 for automatic deduplication of records, then randomly sorted unique records in Microsoft Excel using the RAND() function, and imported the first 2000 records yielded from the search into Covidence22 for screening. Two authors (MJP and either P-YN or RK) independently screened the titles and abstracts of the 2000 records against the eligibility criteria. We retrieved the full text of all records deemed potentially eligible, and two authors (P-YN and either MJP or RK) independently evaluated them in random order against the eligibility criteria until we reached our target sample size of 300 systematic reviews. Any disagreement at each stage of screening was resolved via discussion or adjudication by the senior reviewer (MJP). Because this study was primarily descriptive, we aimed to examine reporting across a range of practices. We selected our sample size of 300 systematic reviews as a balance of feasibility and precision. This sample size allowed us to restrict the width of a 95% two sided Wald type confidence interval around the estimated percentage of reviews reporting a particular practice to a maximum of 6%, assuming a prevalence of 50%. For a prevalence of less (or greater) than 50%, the absolute width will be smaller. This maximum confidence interval width was small enough such that our interpretation of the confidence interval limits would be generally consistent.

RESEARCH

Secondary use of data collected on systematic reviews from 2014 We obtained the dataset previously collated by Page et al,¹⁰ which included data on completeness of reporting and sharing of review data in a random sample of 110 systematic reviews of health interventions indexed in Medline in February 2014. The reviews included in the 2014 dataset were drawn from a random sample of 300 systematic reviews of health research that answered questions of intervention efficacy, diagnostic test accuracy, epidemiology, or prognosis, 110 of which evaluated the effect of health interventions and met the same eligibility criteria that the 2020 reviews had to meet (apart from year of publication). We extracted individual review data from the 2014 dataset for all reporting and sharing items that were worded the same or similarly as the items collected in the 2020 sample. Where necessary, we recoded data in the 2014 sample to ensure harmonisation with the 2020 sample. We did not collect any additional data on the systematic reviews (or the journals they were published in) in the 2014 sample. Given the systematic reviews in 2014 were identified via Medline only, whereas the systematic reviews in 2020 were identified via five databases (PubMed, Science Citation Index, Social Sciences Citation Index, Scopus, and Education Collection), we determined how many of the included reviews from 2020 happened also to be indexed in Medline, to ensure the comparison between years was appropriate. Data analysis We summarised general and reporting characteristics of the included systematic reviews using descriptive statistics (eg, frequency and percentage for categorical items, median and interquartile range for continuous items). We calculated risk ratios to quantify differences in the percentage of reviews meeting indicators of “completeness of reporting” and “sharing of review materials” between the following groups: • Reviews published in 2020 in an evidence synthesis journal (defined as a journal which has a strong or exclusive focus on systematic reviews and their protocols, as identified from the journal website’s aims and scope sections) versus reviews published elsewhere • Reviews of health interventions published in 2020 versus reviews of health interventions published in 2014 4

- Reviews published in 2020 reporting use of a

reporting guideline (eg, PRISMA) versus reviews published the same year not reporting such use • Reviews published in 2020 in journals with a data sharing policy versus journals without one • Reviews published in 2020 in journals with a policy that mandates either data sharing or declaration of data availability, irrespective of whether the policy applies universally to all studies or specifically to systematic reviews, versus journals without such a policy. Risk ratios and Wald type normal 95% confidence intervals were calculated using the epitool package version 0.5-10.1 (R version 4.0.3).²⁷ Where the numerators were small (<5) in either group, or the outcome was very rare (<5%) in either group, we instead used penalised likelihood logistic regression (implemented via the logistf package version 1.24 in R).²⁸ Penalised likelihood logistic regression has been shown to improve estimation of the odds ratio and its confidence interval for rare events or unbalanced samples.^{29 30} The odds ratios from these models can be interpreted as risk ratios when the events are rare in both groups.³¹ The risk ratios and their 95% confidence intervals were displayed using forest plots (implemented via the forestplot package version 1.10.1 in R).³² Rather than relying on statistical significance when interpreting risk ratio associations (ie, claiming that an association exists when the 95% confidence interval did not include the null), we defined an equivalence range for all comparisons as 0.9-1.1. Any risk ratio less than 0.9 or greater than 1.1 (ie, a 10% difference in rate of reporting in either direction) was deemed to be an important difference. Since no previous study has identified a meaningful threshold for important changes in reporting in systematic reviews, this equivalence range was determined based on consensus between investigators. Assuming an item was reported by 50% of reviews in 2014, a risk ratio of 1.1 reflects that the item was reported by 55% of reviews in 2020 (a difference of five percentage points). If the reporting rate in 2014 is higher than 50% (eg, 80%), the threshold to be considered an important difference will be higher (ie, eight percentage points). We conducted two post hoc sensitivity analyses. The first was conducted by excluding Cochrane reviews because they are subjected to strict editorial processes to ensure adherence to methodological conduct and reporting standards, and the second by excluding reviews on covid-19 owing to concerns about short publication turnarounds, which could have an impact on reporting quality.³³

Patient and public involvement We did not involve patients or members of the public directly when we designed our study, interpreted the results, or wrote the manuscript, because our focus was to identify problems in how researchers report their work in scientific journals with a predominantly scientific readership. However, the idea for our study arose from our concerns as people who interact with doi: 10.1136/bmj-2022-072428 | BMJ 2022;379:e072428 | the bmj

BMJ: first published as 10.1136/bmj-2022-072428 on 22 November 2022. Downloaded from <https://www.bmj.com/> on 21 April 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

expectations regarding public sharing of data and code used in the review. Web archives (<https://web.archive.org/>) were used to retrieve the version of the policy published before 1 November 2020. We collected data from the main report of the systematic review, any supplementary file provided on the journal server or any cited repository, the review protocol (if the authors specified that the relevant information was contained therein), and journal websites (table 1). In the event of discrepancies between the protocol and the main report, we gave preference to data from the main report.

RESEARCH