Journal of Clinical Epidemiology 159 (2023) 225−234

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Methods and results of studies on reporting guideline adherence are poorly reported: a meta-research study

Tiffany Dal Santo[a,b], Danielle B. Rice[c,d], Lara S.N. Amiri[a], Amina Tasleem[a], Kexin Li[a], Jill T. Boruff[e], Marie-Claude Geoffroy[b,f,g], Andrea Benedetti[h,i,j,k], Brett D. Thombs[a,b,h,i,k,l,*]

[a]*Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada*
[b]*Department of Psychiatry, McGill University, Montreal, Quebec, Canada*
[c]*Department of Psychology, St. Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada*
[d]*Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada*
[e]*Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, Quebec, Canada*
[f]*Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada*
[g]*McGill Group for Suicide Studies, Douglas Research Centre, Montreal, Quebec, Canada*
[h]*Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada*
[i]*Department of Medicine, McGill University, Montreal, Quebec, Canada*
[j]*Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Quebec, Canada*
[k]*Biomedical Ethics Unit, McGill University, Montreal, Quebec, Canada*
[l]*Department of Psychology, McGill University, Montreal, Quebec, Canada*

Accepted 22 May 2023; Published online 2 June 2023

Abstract

Objectives: We investigated recent meta-research studies on adherence to four reporting guidelines to determine the proportion that provided (1) an explanation for how adherence to guideline items was rated and (2) results from all included individual studies. We examined conclusions of each meta-research study to evaluate possible repetitive and similar findings.

Study Design and Setting: A cross-sectional meta-research study. MEDLINE (Ovid) was searched on July 5, 2022 for studies that used any version of the Consolidated Standards of Reporting Trials, Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Standards for the Reporting of Diagnostic Accuracy Studies, or Strengthening the Reporting of Observational Studies in Epidemiology reporting guidelines or their extensions to evaluate reporting.

Results: Of 148 included meta-research studies published between August 2020 and June 2022, 14 (10%, 95% confidence interval [CI] 6%−15%) provided a fully replicable explanation of how they coded the adherence ratings and 49 (33%, 95% CI 26%−41%) completely reported individual study results. Of 90 studies that classified reporting as adequate or inadequate in the study abstract, six (7%, 95% CI 3%−14%) concluded that reporting was adequate, but none of those six studies provided information on how items were coded or provided item-level results for included studies.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. T.D.S. was supported by a Canadian Institutes of Health Research Masters Award, L.S.N.A. by a Mitacs Globalink internship, M.C.G. (Tier 2) and B.D.T. (Tier 1) by Canada Research Chairs, and A.B. by a Fonds de recherche du Québec−Santé senior researcher salary award, all outside of the present work.

Author Contributions: T.D.S., D.B.R., M.C.G., A.B., and B.D.T. were responsible for the study conception and design. J.T.B. and B.D.T. were responsible for the design of the database search. K.L. carried out the search. T.D.S., L.S.N.A., A.T., and B.D.T. contributed to data extraction, coding, and evaluation of included studies. T.D.S. conducted the analyses. T.D.S. drafted the manuscript, and D.B.R., L.S.N.A., A.T., K.L., J.T.B., M.C.G., A.B., and B.D.T. provided critical review and approved the final manuscript.

Data statement: All data extracted from included studies and used in our study are available in the manuscript and its tables or appendices. Additionally, a raw data file is available on the Open Science Framework (https://osf.io/gtm4z/).

Author statement: All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

* Corresponding author. Jewish General Hospital, 3755 Cote Ste Catherine Road, Montreal, Quebec H3T 1E2, Canada. Tel.: +1-514-340-8222x24244.

E-mail address: brett.thombs@mcgill.ca (B.D. Thombs).

https://doi.org/10.1016/j.jclinepi.2023.05.017
0895-4356/© 2023 Elsevier Inc. All rights reserved.

Conclusion: Almost all included meta-research studies found that reporting in health research is suboptimal. However, few of these reported enough information for verification or replication. © 2023 Elsevier Inc. All rights reserved.

Keywords: Research waste; Reproducibility; Replicability; Checklist; Checklists; Research-on-research

1. Introduction

Meta-research studies are conducted to identify areas where research design, conduct, or reporting could be improved and, thus, reduce research waste [1−6]. Meta-research itself, however, can be wasteful if it is poorly designed or reported or does not add substantively to knowledge.

Many meta-research studies evaluate reporting in health research studies based on checklists from reporting guidelines [7,8], such as the Consolidated Standards of Reporting Trials (CONSORT) [9], Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [10], Standards for the Reporting of Diagnostic Accuracy Studies (STARD) [11], or Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [12].

There are no reporting guidelines for meta-research, but many meta-research studies use methods closely aligned with systematic review methods [13−18]. The PRISMA statement for systematic reviews and meta-analyses stipulates that data collection and coding methods are defined and that results of all individual included studies are provided [10].

The translation of individual guideline items into evaluations of reporting and the results from individual included studies are core elements of studies on reporting guideline adherence. Many reporting guideline items are multifaceted [19]. Not defining how these items are translated into ratings in meta-research creates risk of unreliable or invalid coding and poses a barrier to replication. Similarly, not reporting individual study-level results does not allow verification or permit users to identify studies of interest.

We evaluated recent meta-research studies on reporting in health research studies that used the CONSORT [9], PRISMA [10], STARD [11], or STROBE [12] guidelines or one of their extensions and determined the proportion that provided (1) an explanation for how guideline items were translated into adherence ratings and (2) results from each included study. Additionally, we evaluated the studies' conclusions to assess whether they are likely generating new knowledge vs. addressing questions to which the answer is already known.

2. Materials and methods

We conducted a cross-sectional evaluation of recently published meta-research studies that evaluated adequacy of health research study reporting. We posted our study protocol on the Open Science Framework (https://osf.io/gtm4z/) before initiation. The present study is reported consistent with applicable PRISMA [10] items as these most closely align with our study design.

2.1. Eligibility

Studies published in any language were eligible if they used any version of the CONSORT [9], PRISMA [10], STARD [11], or STROBE [12] reporting guidelines or their extensions (e.g., CONSORT-ROUTINE [20], PRISMA-DTA [21], STROBE-MR [22]) to evaluate reporting in human health research publications. We selected these reporting guidelines for our study based on a prestudy review of citations to reporting guideline publications listed on the Enhancing the QUAlity and Transparency Of health Research website [19]; these guidelines were by far the most highly cited.

Included studies must have mentioned the name of an eligible guideline in their abstract. Studies that evaluated reporting using multiple reporting guidelines were eligible if at least one of the guidelines was eligible. Studies that investigated reporting as one of multiple research questions or assessed reporting as part of another research question were eligible. For consistency, we excluded studies that evaluated reporting based on checklists that included modified items from an otherwise eligible reporting guideline checklist (i.e., changed, removed, or added item content), added items to a checklist, or evaluated fewer than half of items in a checklist as this could create subsets of items or checklists with a different level of coding complexity than the original checklist. We excluded studies that evaluated <10 publications to avoid including studies that targeted single studies or small groups of studies to illustrate known reporting deficiencies.

2.2. Search and study selection method

We searched MEDLINE (ALL) via Ovid using the search strategy: (((quality or complete* or adequat* or transparen*) adj3 reporting) AND (CONSORT* or PRISMA* or STROBE* or STARD* or ''Consolidated Standards of Reporting Trials'' or ''Preferred Reporting Items for Systematic Reviews'' or ''Standards for Reporting Diagnostic accuracy studies'' or ''Strengthening the Reporting of Observational Studies in Epidemiology'')).tw,kf. The principal investigator (B.D.T.) worked with an experienced health sciences librarian (J.T.B.) to

What is new?

Key findings

- Ten percent of 148 meta-research studies included enough information on how they coded adherence ratings to understand how studies were rated or to replicate studies.
- Thirty three percent provided results for individual included studies.
- Almost all studies reached the conclusion that reporting is not adequate.

What this adds to what is known?

- Meta-research on reporting guideline adherence may be contributing to research waste due to poor reporting and repetitive results.

What is the implication and what should change now?

- Meta-researchers should shift focus away from further documenting poor reporting to developing, testing, and disseminating effective strategies to improve reporting.

develop the search. The search was run by a trained research assistant (K.L.) on July 5, 2022. See Appendix A for complete details on our search strategy. To include the most recently published meta-research studies, which would reflect relatively current practices, we reviewed citations identified in the search in reverse chronological order based on their PubMed Unique Identifier until we obtained our targeted sample size. Citations were uploaded to DistillerSR (Evidence Partners, Ottawa, Canada). Two reviewers (T.D.S. and L.S.N.A.) independently assessed study eligibility at the title and abstract level. If either reviewer deemed a study potentially eligible, two reviewers (T.D.S., L.S.N.A., or A.T.) independently assessed eligibility via full-text review. Discrepancies at the full-text level were resolved by consensus between reviewers, with a third reviewer (B.D.T.) consulted as necessary. Appendix B includes coding guides for determining eligibility.

2.3. Sample size calculation

Our experience, before initiating this study, in reviewing studies on adherence to reporting guidelines suggested that few studies provide coding definitions or report individual study results. We therefore hypothesized that the proportion of included articles that provided either would be small. Thus, we set our sample size to have a 95% confidence interval (CI) width of 15% around a percentage reporting of 33%. Based on CIs calculated using Agresti and Coull's method [23], we sought to obtain 148 studies.

2.4. Data extraction

For each eligible meta-research study, data were extracted in DistillerSR by a single reviewer (T.D.S. or L.S.N.A.) and validated by a second reviewer (T.D.S., L.S.N.A., or A.T.) using the DistillerSR Quality Control function. Discrepancies were resolved by consensus between reviewers with a third reviewer (B.D.T.) consulted as necessary. See Appendix C for the data extraction form. Reviewers extracted (1) publication characteristics (first author last name; publication year; journal and 2021 journal impact factor); (2) country of corresponding author affiliations; (3) research question (research question related to reporting only; main research question was related to reporting only with other nonreporting questions; there were multiple research questions, including questions related to reporting and nonreporting questions, and main one is unclear; main research question was not related to reporting, but an eligible reporting analysis was conducted); (4) reporting guideline(s) evaluated; (5) number of publications included in the study; (6) main eligibility criteria of included publications (by reporting guideline, study design, field of research, patient population, intervention type, journal, other); (7) number of raters; (8) independence of raters; (9) rating method used (e.g., yes/no, fully/partially/not reported); and (10) conclusion about reporting adequacy. We reviewed abstracts to extract conclusions as these are the most read, and in many cases, the only part of an article that is read [24].

If a study's supplementary material was not accessible via the publishing journal's website, we contacted the corresponding author and journal editorial manager or editor-in-chief to request access. We sent up to two follow-up e-mails per study to corresponding authors and journal staff; if we did not receive a response, we coded the study based on available information.

To answer our main research questions, reviewers extracted (1) whether the authors provided an explanation for translating items into adherence ratings with enough information to be replicated and (2) if the authors provided results for each individual study included in their report. We searched for this information in the main study text and tables, supplementary material, and via any internet links provided. Explanations for how they coded adherence ratings must have specifically reported which parts of each item were required for the item to be coded as adequately reported. We coded conclusions about adequacy as adequate, inadequate—implicit, inadequate—explicit, mixed, vague, or no mention. Definitions for each are in Appendix C. For individual study results, we coded whether authors reported results for each item for all studies, reported partially (e.g., an overall score but not

item ratings for each study), or did not report individual study results. See Appendices D and E for the coding manual.

2.5. Analysis

We calculated the proportions of meta-research studies that provided (1) a coding guide for translating reporting guideline items into ratings with enough information for replication and (2) results for each included study. All proportions are presented with 95% CIs using the method of Agresti and Coull [23]. We also present results by subgroups defined by country of corresponding author affiliations, 2021 journal impact factor, reporting guideline evaluated, and research question (main research question related or not related to reporting). When presenting outcomes by subgroups, we included guideline extensions (e.g., CONSORT-ROUTINE) with the main guideline (e.g., CONSORT). The four subgroup analyses were established a priori. For the only quantitative grouping (by journal impact factor), the subgroups were established based on frequency data. We did not conduct statistical tests to compare subgroups because our study was not designed or powered for that purpose.

3. Results

3.1. Search results and included study characteristics

Our search yielded 1,698 unique titles and abstracts. We excluded 182 titles and abstracts and 88 full texts, reviewing in reverse chronological order, until we obtained 148 included studies (Fig. 1). Reasons for exclusion at the full-text level and references are shown in Appendix F. We were initially unable to find or access supplementary files for two of 148 studies. We contacted the authors and journal editors for these missing supplementary files and successfully obtained one set of files.

Included studies were initially listed in MEDLINE between August 14, 2020 and June 30, 2022. They included

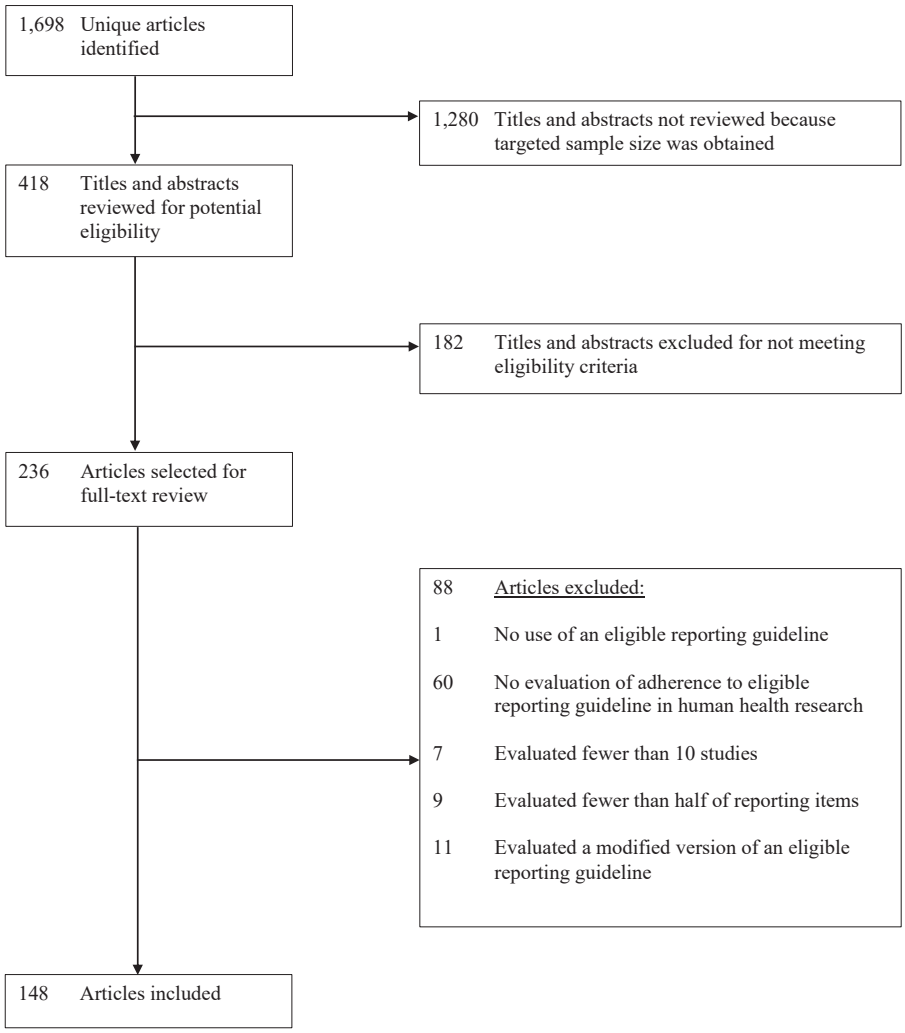

Fig. 1. PRISMA flow diagram.

between 10 and 2,844 studies (median = 52; interquartile range = 24–120). Affiliations of the corresponding authors of studies were from China (*N* = 51; 34%), the United States (*N* = 27; 18%), the United Kingdom (*N* = 9; 6%), Canada (*N* = 8; 5%), and 22 other countries (*N* = 53; 36%). Most assessed adherence to CONSORT (*N* = 61; 41%) or PRISMA (*N* = 59; 40%) or their extensions. The research question was only related to reporting in 46 (31%) studies, included a main question equal related to reporting plus multiple other questions not related to reporting in 13 (9%) studies, was one of multiple questions with no clear primary question in 65 (44%) studies, and had a main question not related to reporting in 24 (16%) studies. Most studies (103 [70%]) came from journals with journal impact factor >2.9. See Table 1 and Appendix G for individual study characteristics.

Of the 148 included studies, three (2%, 95% CI 1%–6%) used one rater, 10 (7%, 95% CI 4%–12%) used one rater with validation from a second rater, 113 (76%, 95% CI 69%–83%) used two or more independent raters, nine (6%, 95% CI 3%–11%) used two or more raters but did not state whether they were independent, three (2%, 95% CI 1%–6%) used other methods, and 10 (7%, 95% CI 4%–12%) did not report how many raters were used.

For classifying adherence to reporting checklist items, 66 (45%, 95% CI 37%–53%) classified items dichotomously, 61 (41%, 95% CI 34%–49%) used a multilevel approach (e.g., "fully reported", "partially reported", or "not reported"), two (1%, 95% CI 0%–5%) classified some items dichotomously and others with a multilevel approach, and 19 (13%, 95% CI 8%–19%) did not report how they classified items. See Appendix H.

3.2. Main outcomes

Of the 148 studies, 14 (10%, 95% CI 6%–15%) provided a fully replicable explanation of how they coded the adherence ratings, five (3%, 95% CI 2%–8%) provided a partially replicable explanation, and 129 (87%, 95% CI 81%–92%) did not provide enough information to know how coding decisions had been made (Table 2). Forty nine studies (33%, 95% CI 26%–41%) completely reported individual study results, 26 (18%, 95% CI 12%–25%) reported partial results for all studies, three (2%, 95% CI 1%–6%) reported results for some studies but not others, and 70 (47%, 95% CI 39%–55%) did not provide any individual study results (Table 3). Only four (3%, 95% CI 1%–7%) studies provided both fully replicable explanations of how they coded the adherence ratings and complete individual study results.

Reporting was mentioned in 122 abstract conclusions, and 90 of these classified reporting as either adequate or inadequate. Of these 90 studies, 6 (7%, 95% CI 3%–14%) concluded that reporting was adequate, 29 (32%, 95% CI 24%–42%) implicitly concluded that reporting was inadequate, and 55 (61%, 95% CI 51%–71%) did so

Table 1. Study characteristics (*N* = 148)

Study characteristics	*N* (%)
Year Published	
2020	21 (14)
2021	60 (41)
2022	56 (38)
Online only	11 (7)
Country of Corresponding Author Affiliations	
Canada	8 (5)
China	51 (34)
United Kingdom	9 (6)
United States	27 (18)
Other (all with ≤five studies)[a]	53 (36)
Journal Impact Factor[b]	
≤2.9	45 (30)
2.9 < JIF ≤ 4.9	55 (37)
>4.9	48 (32)
Included Study Eligibility Criteria[c]	
Study design	137 (93)
Patient population	68 (46)
Intervention type	65 (44)
Journal	17 (11)
Included in specified guidelines	13 (9)
Field of research	13 (9)
Other[d]	6 (4)
Research Question	
Only research question related to reporting	46 (31)
Main question related to reporting among multiple research questions	13 (9)
Multiple research questions with main question unclear	65 (44)
Main research question not related to reporting	24 (16)
Reporting Guideline[e]	
CONSORT	61 (41)
PRISMA	59 (40)
STARD	10 (7)
STROBE	18 (12)
Number of Included Publications Reviewed	
≤50	72 (49)
>50	76 (51)

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STARD, Standards for the Reporting of Diagnostic Accuracy Studies; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

[a] Australia (3); Brazil (3); Chile (1); Croatia (1); France (3); Germany (4); Greece (2); India (3); Iran (2); Ireland (2); Italy (3); Korea (4); Macao (2); Mexico (1); Portugal (1); Qatar (2); Saudi Arabia (2); South Africa (1); South Korea (4); Spain (3); Switzerland (1); the Netherlands (5).

[b] Journals for which we could not find a journal impact factor were coded as 0.

[c] Included reviews could be counted in more than one category.

[d] Studies reviewed included a specific questionnaire, were on acceptability of a specific intervention, were abstracts submitted to specific conferences, or were studies that used a specific database.

[e] Including extensions to specified reporting guidelines.

Table 2. Number and percent of studies that provided a fully or partially replicable explanation of how they coded the adherence ratings or did not provide such coding explanations for the overall sample ($N = 148$) and subgroups

	N % (95% CI)		
Subgroups	**Fully replicable**	**Partially replicable**	**Not replicable**
All	14 10% (6%, 15%)	5 3% (2%, 8%)	129 87% (81%, 92%)
Country of Corresponding Author Affiliations			
Canada	3 38% (14%, 69%)	2 25% (7%, 59%)	3 38% (14%, 69%)
China	3 6% (2%, 16%)	0 0% (0%, 7%)	48 94% (84%, 98%)
United Kingdom	1 11% (2%, 44%)	1 11% (2%, 44%)	7 78% (45%, 94%)
United States	2 7% (2%, 23%)	0 0% (0%, 13%)	25 93% (77%, 98%)
Other	5 9% (4%, 20%)	2 4% (1%, 13%)	46 87% (75%, 94%)
Journal Impact Factor			
≤2.9	1 2% (0%, 12%)	0 0% (0%, 8%)	44 98% (88%, 100%)
2.9 < JIF ≤ 4.9	6 11% (5%, 22%)	3 6% (2%, 15%)	46 84% (72%, 91%)
>4.9	7 15% (7%, 27%)	2 4% (1%, 14%)	39 81% (68%, 90%)
Reporting Guideline			
CONSORT and extensions	9 15% (8%, 26%)	3 5% (2%, 14%)	49 80% (69%, 88%)
PRISMA and extensions	1 2% (0%, 9%)	1 2% (0%, 9%)	57 97% (89%, 99%)
STARD and extensions	2 20% (6%, 51%)	1 10% (2%, 40%)	7 70% (40%, 89%)
STROBE and extensions	2 11% (3%, 33%)	0 0% (0%, 18%)	16 89% (67%, 97%)
Research Question			
The only research question was related to reporting or there were multiple research questions and the main one was related to reporting or not defined	14 11% (7%, 18%)	4 3% (1%, 8%)	106 86% (78%, 91%)
The main research question was not related to reporting	0 0% (0%, 14%)	1 4% (1%, 20%)	23 96% (80%, 99%)

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STARD, Standards for the Reporting of Diagnostic Accuracy Studies; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

explicitly. Of the 6 studies that concluded that reporting was adequate, none provided any explanation of how items were coded or item-level results for individual studies. The 4 studies with a fully replicable explanation of how they coded the adherence ratings and complete individual study results all concluded that reporting was inadequate (Table 4). Outcomes for individual meta-research studies are shown in Appendix I.

As shown in Tables 2–4, most subgroup results did not differ substantively from overall conclusions, excluding subgroups with very small numbers of meta-research studies (e.g., < 10 studies). One exception was among 124 studies where the main research question was related to reporting. Thirty five studies (28%, 95% CI 21%–37%) completely reported individual study results, compared to 14 of 24 studies (58%, 95% CI 39%–76%) where the main research question was not related to reporting.

4. Discussion

We examined 148 health research studies that evaluated reporting guideline adherence. Of these, only 10% provided enough information to understand how

Table 3. Level of reporting of included study results for overall sample ($N = 148$) and subgroups

	N % (95% CI)			
Subgroups	**Completely reported**	**Partially reported—all studies**	**Partially reported—some studies**	**Not reported**
All	49 33% (26%, 41%)	26 18% (12%, 25%)	3 2% (1%, 6%)	70 47% (39%, 55%)
Country of Corresponding Author Affiliations				
Canada	2 25% (7%, 59%)	0 0% (0%, 32%)	0 0% (0%, 32%)	6 75% (41%, 93%)
China	26 51% (38%, 64%)	2 4% (1%, 13%)	0 0% (0%, 7%)	23 45% (32%, 59%)
United Kingdom	4 44% (19%, 73%)	2 22% (6%, 55%)	0 0% (0%, 30%)	3 33% (12%, 65%)
United States	3 11% (4%, 28%)	13 48% (31%, 66%)	1 4% (1%, 18%)	10 37% (22%, 56%)
Other	14 26% (16%, 40%)	9 17% (9%, 29%)	2 4% (1% 13%)	28 53% (40%, 66%)
Journal Impact Factor				
≤ 2.9	16 36% (23%, 50%)	9 20% (11%, 34%)	1 2% (0%, 12%)	19 42% (29%, 57%)
$2.9 < \text{JIF} \leq 4.9$	15 27% (17%, 40%)	12 22% (13%, 34%)	0 0% (0%, 7%)	28 51% (38%, 64%)
>4.9	18 38% (25%, 52%)	5 10% (5%, 22%)	2 4% (1%, 14%)	23 48% (35%, 62%)
Reporting Guideline				
CONSORT and extensions	14 23% (14%, 35%)	6 10% (5%, 20%)	0 0% (0%, 6%)	41 67% (55%, 78%)
PRISMA and extensions	25 42% (31%, 55%)	15 25% (16%, 38%)	1 2% (0%, 9%)	18 31% (20%, 43%)
STARD and extensions	3 30% (11%, 60%)	1 10% (2%, 40%)	1 10% (2%, 40%)	5 50% (24%, 76%)
STROBE and extensions	7 39% (20%, 61%)	4 22% (9%, 45%)	1 6% (1%, 26%)	6 33% (16%, 56%)
Research Question				
The only research question was related to reporting or there were multiple research questions and the main one was related to reporting or not defined	35 28% (21%, 37%)	22 18% (12%, 25%)	2 2% (0%, 6%)	65 52% (44%, 61%)
The main research question was not related to reporting	14 58% (39%, 76%)	4 17% (7%, 36%)	1 4% (1%, 20%)	5 21% (9%, 41%)

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STARD, Standards for the Reporting of Diagnostic Accuracy Studies; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

individual checklist items were rated, and 33% reported results for all studies evaluated. We did not identify any substantive differences by subgroups. Of 90 studies that classified reporting as adequate or inadequate in their abstracts, 7% concluded that reporting was adequate; however, none of these studies provided an explanation of how they coded items or provided item-level results for individual studies. Only 3% of included meta-research studies provided both a fully replicable explanation of how they coded the adherence ratings and complete individual study results, and all of those studies concluded that reporting was inadequate.

No previous studies have examined the degree that meta-research studies on reporting guideline adherence adequately report key aspects of their own studies. Given that meta-research is done to scrutinize research

Table 4. Conclusions in abstracts of included studies on research reporting for overall sample ($N = 148$) and subgroups

	N % (95% CI)					
Subgroups	**Adequate**	**Inadequate—implicit**	**Inadequate—explicit**	**Mixed**	**Vague**	**No mention**
All	6 4% (2%, 9%)	29 20% (14%, 27%)	55 37% (30%, 45%)	10 7% (4%, 12%)	22 15% (10%, 22%)	26 18% (12%, 25%)
Country of Corresponding Author Affiliations						
Canada	0 0% (0%, 32%)	0 0% (0%, 32%)	2 25% (7%, 59%)	1 13% (2%, 47%)	3 38% (14%, 69%)	2 25% (7%, 59%)
China	1 2% (0%, 10%)	15 29% (19%, 43%)	21 41% (29%, 55%)	0 0% (0%, 7%)	6 12% (6%, 23%)	8 16% (8%, 28%)
United Kingdom	0 0% (0%, 30%)	2 22% (6%, 55%)	3 33% (12%, 65%)	1 11% (2%, 44%)	1 11% (2%, 44%)	2 22% (6%, 55%)
United States	1 4% (1%, 18%)	2 7% (2%, 23%)	12 44% (28%, 63%)	4 15% (6%, 33%)	4 15% (6%, 33%)	4 15% (6%, 33%)
Other	4 8% (3%, 18%)	10 19% (11%, 31%)	17 32% (21%, 46%)	4 8% (3%, 18%)	8 15% (8%, 27%)	10 19% (11%, 31%)
Journal Impact Factor						
≤ 2.9	2 4% (1%, 15%)	19 42% (29%, 57%)	9 20% (11%, 34%)	4 9% (4%, 21%)	7 16% (8%, 29%)	4 9% (4%, 21%)
$2.9 < \text{JIF} \leq 4.9$	2 4% (1%, 12%)	17 31% (20%, 44%)	11 20% (12%, 32%)	6 11% (5%, 22%)	6 11% (5%, 22%)	13 24% (14%, 36%)
>4.9	2 4% (1%, 14%)	19 40% (27%, 54%)	9 19% (10%, 32%)	0 0% (0%, 7%)	9 19% (10%, 32%)	9 19% (10%, 32%)
Reporting Guidelines						
CONSORT and extensions	3 5% (2%, 14%)	12 20% (12%, 31%)	29 48% (36%, 60%)	3 5% (2%, 14%)	9 15% (8%, 26%)	5 8% (4%, 18%)
PRISMA and extensions	1 2% (0%, 9%)	15 25% (16%, 38%)	16 27% (17%, 40%)	7 12% (6%, 23%)	11 19% (11%, 30%)	9 15% (8%, 27%)
STARD and extensions	0 0% (0%, 28%)	2 20% (6%, 51%)	4 40% (17%, 69%)	0 0% (0%, 28%)	2 20% (6%, 51%)	2 20% (6%, 51%)
STROBE and extensions	2 11% (3%, 33%)	0 0% (0%, 18%)	6 33% (16%, 56%)	0 0% (0%, 18%)	0 0% (0%, 18%)	10 56% (34%, 75%)
Research Question						
The only research question was related to reporting or there were multiple research questions and the main one was not related to reporting or not defined	6 5% (2%, 10%)	54 44% (35%, 52%)	25 20% (14%, 28%)	10 8% (4%, 14%)	20 16% (11%, 24%)	9 7% (4%, 13%)
The main research question was not related to reporting	0 0% (0%, 14%)	1 4% (1%, 20%)	4 17% (7%, 36%)	0 0% (0%, 14%)	2 8% (2%, 26%)	17 71% (51%, 85%)

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; STARD, Standards for the Reporting of Diagnostic Accuracy Studies; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

methodology [25], some might assume that these studies are rigorously conducted and reported. However, there are no consensus standards for conducting and reporting these studies. Our study shows that most meta-research studies find that reporting in health research is suboptimal, but few of these studies themselves reported enough information for verification or replication.

Good research asks important questions and uses methods that allow us to be confident in its conclusions [26]. Researchers considering initiating a study on adherence to reporting guidelines and editors who must decide whether to publish such studies should be able to clearly articulate how the studies might add to what is known about the state of research reporting. Evaluating reporting to understand the influence of new or updated reporting guidelines or to assess the effects of interventions designed to improve reporting would likely be justified. Simply documenting poor reporting guideline adherence in yet one

more subspecialty area, however, would likely be less useful.

Authors of any studies that evaluate reporting should clearly describe how reporting was evaluated and should provide study-level information so others can evaluate and validate findings. Reporting guidelines for meta-research studies do not exist, but a protocol for such guidelines has been published [14]. The authors of these proposed guidelines should ensure that meta-research studies on reporting, in addition to other important items, address the reporting gaps we have identified here.

To date, the only reporting guideline with a standardized tool to facilitate evaluation of reporting completeness is the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement [27]. Such assessment forms are necessary to ensure that adherence is evaluated in a consistent manner and can be replicated [28]. Similar forms would ideally be developed for other reporting guidelines. Meanwhile, researchers who do evaluate reporting can refer to examples of studies that we reviewed that provided fully replicable explanations of how they rated adherence and reporting of individual study results.[G23,G100,G111,G112] In each of those studies, for each reporting guideline item, the researchers delineated precise definitions of the information required for different ratings.

Rather than prioritizing additional studies on the poor quality of health research reporting, interventions are needed to help researchers, peer reviewers, and journal editors improve reporting. A 2019 scoping review identified 31 interventions created to improve reporting guideline adherence, but only 11 had been evaluated in any way [29]. Strategies varied on what step of the writing or publishing process they targeted, but most aimed to improve adherence at the journal level, such as editorial endorsement of specific reporting guidelines, or requiring authors to submit a completed reporting checklist. The scoping review found four randomized trials of interventions to enhance adherence; the only one that showed a statistically significant effect on reporting was the Consort-based WEB tool, which supports adherence at the manuscript writing stage [30]. The tool divides CONSORT items into bullet points and emphasizes key reporting elements that need to be reported for the main CONSORT checklist and selected extensions [30]. In the trial of the Consort-based WEB tool, which included 41 participants, the global score for completeness of reporting (0−10 scale) was 2.1 points higher (95% CI 1.5−2.7) in 123 CONSORT domains drafted with the tool compared to 123 domains drafted without using the tool [30]. Another intervention, published after the search period of the scoping review, in which a journal required authors to incorporate section headings that reflected CONSORT items into their manuscripts, also improved reporting [31]. Overall, however, there are few interventions that have been tested in randomized trials and found to be effective, and there is only limited evidence on interventions that have been tested [29]. Resources should be allocated to developing, testing, and disseminating effective interventions that address different aspects of the complex factors that contribute to how well research is reported [29].

4.1. Strengths and limitations

Strengths of our study include that we developed and posted a protocol before initiating the study, we have provided all coding manuals and individual study results in supplementary materials, and we included a large sample size of the most recently published studies based on an a priori power analysis.

There are some limitations that also need to be considered. First, we only searched MEDLINE and used a pragmatic search strategy; this could have led us to miss potentially eligible studies, although it is unlikely that health research studies in other databases or that were less clearly identified as studies on reporting would have been more completely reported. Second, we included meta-research studies that assessed adherence to four reporting guidelines listed in the Enhancing the QUAlity and Transparency Of health Research website based on how often they have been cited, but we did not assess others. We do not believe that including other reporting guidelines would have influenced results substantively considering that we assessed reporting in the meta-research studies themselves and not reporting levels of studies that used those reporting guidelines.

5. Conclusion

We found that of the 148 studies we assessed, 10% provided a fully replicable explanation of how they coded the adherence ratings, 33% completely reported individual study results, and 7% of those that categorized reporting as being adequate or inadequate concluded that adherence to reporting guidelines was adequate, although none of the studies that rated reporting as adequate were themselves well reported. Meta-research is done to reduce research waste by improving how research is performed, communicated, and used [25], but our study shows that meta-research on reporting may be a significant contributor to waste. Most recent studies on reporting guideline adherence do not appear to have added meaningfully to what we know about the problem of research reporting. Poor reporting of key elements in most of these studies does not allow for conclusions beyond that overall reporting continues to be suboptimal or provide an understanding of how to address the most salient reporting gaps. New studies on adherence should only be conducted if there is a specific and justified rationale to address a well-defined, nonredundant research question. Rather than more research on poor reporting in another subspecialty area, research is needed that develops effective

interventions to improve reporting, tests them in randomized trials, and disseminates them via support and training tools.

Declaration of competing interest

All authors have completed the ICJME uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organization for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years. All authors declare no relationships or activities that could appear to have influenced the submitted work. No funder had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. D.B.R. and B.D.T. declared that they were named or group authors of three included studies (G100,G111,G112) conducted to benchmark reporting before publishing a new reporting guideline.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2023.05.017.

References

[1] Ioannidis JPA, Fanelli D, Dunne DD, et al. Meta-research: evaluation and improvement of research methods and practices. PLoS Biol 2015;13:e1002264.
[2] Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet 2014;383:156−65.
[3] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374:86−9.
[4] Glasziou P, Chalmers I. Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. BMJ 2018;363:k4645.
[5] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet 2014;383:267−76.
[6] Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, et al. Biomedical research: increasing value, reducing waste. Lancet 2014;383:101−4.
[7] Jin Y, Sanger N, Shams I, Luo C, Shahid H, Guowei L, et al. Does the medical literature remain inadequately described despite having reporting guidelines for 21 years? - a systematic review of reviews: an update. J Multidiscip Healthc 2018;11:495−510.
[8] Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, Stevens A, et al. Systematic review adherence to methodological or reporting quality. Syst Rev 2017;6:131.
[9] Schulz KF, Altman DG, Moher D, the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18.
[10] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
[11] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527.
[12] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335:806−8.
[13] Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med 2017;22:139−42.
[14] Lawson DO, Puljak L, Pieper D, Schandelmaier S, Collins GS, Brignardello-Petersen R, et al. Reporting of methodological studies in health research: a protocol for the development of the Methodological STudy reportIng Checklist (MISTIC). BMJ Open 2020;10:e040478.
[15] Zhang Z. Meta-epidemiological study: a step-by-step approach by using R. J Evid Based Med 2016;9:91−7.
[16] Gundogan B, Dowlut N, Rajmohan S, Borrelli MR, Millip M, Iosifidis C, et al. Assessing the compliance of systematic review articles published in leading dermatology journals with the PRISMA statement guidelines: a systematic review. JAAD Int 2020;1:157−74.
[17] Ritchie A, Seubert L, Clifford R, Perry D, Bond C. Do randomised controlled trials relevant to pharmacy meet best practice standards for quality conduct and reporting? A systematic review. Int J Pharm Pract 2020;28:220−32.
[18] Canagarajah NA, Porter GJ, Mitra K, Chu TSM. Reporting quality of surgical randomised controlled trials in head and neck cancer: a systematic review. Eur Arch Otorhinolaryngol 2021;278:4125−33.
[19] The EQUATOR Network. Enhancing the QUAlity and transparency of health research. Available at https://www.equator-network.org/. Accessed February 4, 2023.
[20] Kwakkenbos L, Imran M, McCall SJ, McCord KA, Fröbert O, Hemkens LG, et al. CONSORT extension for the reporting of randomised controlled trials conducted using cohorts and routinely collected data (CONSORT-ROUTINE): checklist with explanation and elaboration. BMJ 2021;373:n857.
[21] McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 2018;319:388−96.
[22] Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: the STROBE-MR statement. JAMA 2021;326:1614−21.
[23] Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. Am Stat 1998;52:119−26.
[24] Dogan RI, Muray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. Database 2009;2009:bap018.
[25] Ioannidis JPA. Meta-research: why research on research matters. PLoS Biol 2018;16:e2005468.
[26] Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984;3:409−22.
[27] Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, Collins GS, et al. Uniformity in measuring adherence to reporting guidelines: the example of TRIPOD for assessing completeness of reporting of prediction model studies. BMJ Open 2019;9:e025611.
[28] Logullo P, MacCarthy A, Kirtley S, Collins GS. Reporting guideline checklists are not quality evaluation forms: they are guidance for writing. Health Sci Rep 2020;09:e165.
[29] Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. BMJ Open 2019;9:e026589.
[30] Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (consort-based WEB tool) randomized controlled trial. BMC Med 2015;13:221.
[31] Koletsi D, Fleming PS, Behrents RG, Lynch CD, Pandis N. The use of tailored subheadings was successful in enhancing compliance with CONSORT in a dental journal. J Dent 2017;67:66−71.