Journal of Clinical Epidemiology 183 (2025) 111818

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

Do peer reviewers comment on reporting items as instructed by the journal? A secondary analysis of two randomized trials

Hillary Wnfried Ramirez[a,b,c], Malena Chiaborelli[a,b,c], Christof M. Schönenberger[a], Katie Mellor[d,e], Alexandra N. Griessbach[a], Paula Dhiman[d,f], Pooja Gandhi[g], Szimonetta Lohner[h,i], Arnav Agarwal[j,k], Ayodele Odutayo[d,l], Michael M. Schlussel[d,f], Philippe Ravaud[m,n], David Moher[o,p], Matthias Briel[a,j], Isabelle Boutron[m,n], Sally Hopewell[d], Sara Schroter[q,r], Benjamin Speich[a,d,*]

[a]*CLEAR Methods Center, Division of Clinical Epidemiology, Department Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland*
[b]*Swiss Tropical and Public Health Institute (Swiss TPH), Basel, Switzerland*
[c]*University of Basel, Basel, Switzerland*
[d]*Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom*
[e]*Clinical Outcomes Assessment, Clarivate, London, United Kingdom*
[f]*The EQUATOR Network, Oxford, United Kingdom*
[g]*Department of Communication Sciences and Disorders, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Alberta, Canada*
[h]*Cochrane Hungary, Medical School, University of Pécs, Pécs, Hungary*
[i]*MTA−PTE Lendület "Momentum" Evidence in Medicine Research Group, Department of Public Health Medicine, Medical School, University of Pécs, Pécs, Hungary*
[j]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada*
[k]*Division of General Internal Medicine, Department of Medicine, McMaster University, Hamilton, Ontario, Canada*
[l]*Division of Nephrology, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada*
[m]*Centre d'Épidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique Hôpitaux de Paris (APHP), Paris F-75004, France*
[n]*Université de Paris, CRESS, Inserm, INRA, Paris F-75004, France*
[o]*Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada*
[p]*Faculty of Medicine, School of Epidemiology and Public Health, University of Ottawa, Canada*
[q]*The BMJ, London, United Kingdom*
[r]*Faculty of Public Health & Policy, London School of Hygiene & Tropical Medicine, London, United Kingdom*

Accepted 30 April 2025; Published online 8 May 2025

Abstract

Objectives: Two studies randomizing manuscripts submitted to biomedical journals have previously shown that reminding peer reviewers about key reporting items did not improve the reporting quality in published articles. Within this secondary analysis of peer reviewer reports we aimed to assess at what stage the intervention failed.

Study Design and Setting: We exploratively analyzed peer reviewer reports from two published randomized controlled trials (RCTs) conducted at biomedical journals. The first RCT (CONSORT-PR) assessed adherence to the Consolidated Standards of Reporting Trials (CONSORT) guideline in manuscripts presenting primary RCT results. The second RCT (SPIRIT-PR) included manuscripts presenting RCT protocols and assessed adherence to the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guideline. In both RCTs the control group consisted of peer reviewers receiving no reminder, whereas all reviewers in the intervention group received a reminder of the 10 most important reporting items. For this secondary analysis, we extracted from peer reviewer reports which of the ten key reporting items were mentioned by reviewers as requiring clarification. The main outcome of this secondary analysis was the difference in the mean

Funding: Benjamin Speich was supported by a Return Postdoc Mobility (P4P4PM_194496) grant from the Swiss National Science Foundation. Christof M Schönenberger was funded by the Janggen Pöhn Foundation and the Swiss National Science Foundation (MD-PhD grant Number: 323530_221860). Szimonetta Lohner was supported by the Hungarian Academy of Sciences (MTA) within the framework of the Lendület Programme. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

* Corresponding author. CLEAR Methods Center, Division of Clinical Epidemiology, Department of Clinical Research, University Hospital Basel, University of Basel, Totengässlein 3, Basel 4051, Switzerland.
E-mail address: benjamin.speich@usb.ch (B. Speich).

https://doi.org/10.1016/j.jclinepi.2025.111818

proportion of these ten reporting items for which at least one peer reviewer requested clarification. Furthermore, we assessed how this difference changed (i) if only published manuscripts were considered and (ii) when only requested changes that were implemented by authors were considered.

Results: We assessed peer reviewer reports from 533 manuscripts (n = 265 intervention group; n = 268 control group). Among the manuscripts in the intervention group, 21.1% (95% CI, 18.6%−23.6%) of the ten reporting items were requested for clarification, compared to 13.1% (95% CI, 18.6%−23.6%) in the control group, resulting in a mean difference of 8.0% (95% CI, 4.9%−11.1%). However, this difference diminished to 4.2% when assessing solely accepted and published manuscripts and was even further reduced to 2.6% when accounting for changes actually implemented by authors.

Conclusion: Reminding peer reviewers to check reporting items increased their focus on reporting guidelines, leading to more reporting-related requests in their reviews. However, the effect was strongly diluted during the peer review process due to rejected articles and requests not implemented by authors.

Keywords: Randomized controlled trial; Reporting; Transparency; Publishing journals; Peer reviewers; Biomedical journals

Plain Language Summary

When new research is submitted to a journal, other experts in the field (peer reviewers) check the research to make sure it's reliable and clear. Among others, one important part of this process is ensuring that researchers follow reporting guidelines about what information should be included in their papers so that the readers can understand how the research was conducted. We wanted to find out if reminding peer reviewers to focus on the key parts of these guidelines (ie, 10 most important items) would help to improve the reporting quality of published research papers. For this purpose, we conducted two studies in which we randomized manuscripts to either an intervention group or a control group. In the intervention group, the peer reviewers from half of the included manuscript received such a reminder (ie, asking them to check whether the 10 most important reporting items are well described in the manuscript), whereas peer reviewers in the control group did not receive a reminder. Within our previously published main results of these studies we saw that the reporting quality of the published articles did not improve with this intervention. To find out why this approach did not work, we looked closer at the individual reports from peer reviewers and checked how often reviewers asked for these important details and whether authors made the necessary changes. We found that reminders did lead to more requests about reporting items from peer reviewers. However, as a high proportion of peer-reviewed articles is rejected during the peer review process and because not all requests for improvements are addressed by authors, this effect was not visible anymore (ie, ''diluted'') when assessing published research articles.

1. Introduction

Transparent reporting in published biomedical articles is of key importance to ensure that they are useful for clinicians, patients, researchers, and systematic reviewers [1,2]. To ensure transparent reporting, so called reporting guidelines were developed, specifying a list of information that must be reported in published articles [3,4]. The Enhancing the Quality and Transparency Of health Research network and many biomedical journals have promoted and enforced the use of these reporting guidelines [2,5−7]. Despite these efforts, most studies assessing the adherence to reporting guidelines still conclude that reporting quality is insufficient [8−10]. Hence, additional efforts with the potential to improve reporting quality should be considered.

To test if reminding peer reviewers of the 10 most important reporting items improves the reporting quality in published articles, two randomized controlled trials (RCTs) in collaboration with seven publishing journals (5 from the British Medical Journal [BMJ] Publishing Group and two from the Public Library of Science [PLOS]) were conducted from our research group [11]. For the first trial we included manuscripts presenting primary results of RCTs and assessed the reporting quality using the Consolidated Standards of Reporting Trials (CONSORT) [12] reporting guideline [11]. The second RCT included manuscripts presenting RCT protocols and assessed the reporting quality using the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) [13] reporting guideline [11]. In both trials, manuscripts were randomized either to the intervention group (where peer reviewers were sent an additional email from the journal reminding them of the most important reporting items and asking them to check whether they are adequately reported) or to the control group (usual journal practice; ie, no reminder about reporting items). In both studies, we found that the intervention of reminding peer reviewer of the most important reporting items was not

What is new?

Key findings

- Two previous randomized trials conducted at journal level showed that reminding peer reviewers of key reporting items did not increase reporting quality in published articles.
- Within this secondary analysis of the peer reviewers' reports in the two randomized trials, we found that peer reviewers who received a reminder (the intervention group) did more frequently suggest improvements to reporting.
- However, this effect was diluted by the high proportion of manuscripts that were rejected and by reviewer requests that were not implemented by authors.

What this adds to what is known?

- Despite not having a direct impact on the reporting quality in published articles, we could show that reminding peer reviewers of reporting items did lead to more requests about reporting items from peer reviewers.

What is the implication and what should change now?

- Journal staff should ensure that authors adequately address all peer reviewers' requested clarifications on reporting guideline items in their revised manuscripts.

effective at improving the reporting quality in published articles [14].

In the present study, we conducted a secondary in-depth analysis of peer reviewer reports from these two trials to gain more knowledge about the mechanisms of our tested intervention and assess at what stage the intervention failed. We aimed to assess (i) how often peer reviewers asked authors to clarify the reporting of the 10 most important reporting items; (ii) whether such requests were made more frequently in the intervention group compared with the control group; and (iii) whether the requested changes were implemented by authors in the published articles.

2. Methods

2.1. Study design

This study builds on data from two previously published RCTs conducted at journal level using submitted manuscripts as the unit of randomization [14]. In brief, in the first trial called CONSORT-PR we assessed if reminding peer reviewers of the 10 most important and underreported CONSORT items had a positive effect on reporting quality in published articles presenting primary results of RCTs. Within the second trial (SPIRIT-PR) we tested if reminding peer reviewers of the 10 most important and underreported SPIRIT items had a positive effect on reporting quality in published articles that contained RCT protocols [14]. Within both RCTs the intervention was sent out during the first peer-review round only (additional peer-review rounds were possible based on editors' decision, following journal practice). The control groups in both RCTs consisted of usual journal practice (ie, no reminder). For CONSORT-PR, the 10 most important and poorly reported items were selected based on previous literature [15]. In the case of SPIRIT-PR, we examined all available reporting assessments [16–18] and determined 10 items through a consensus process within the study team. For this preplanned exploratory secondary analysis [11] of available peer reviewer reports, we examined peer reviewer comments in detail, analyzing all comments for each manuscript, to identify at what stage the intervention may have failed. Notably, peer reviewer comments were not previously assessed for the main results of the two original RCTs [14].

2.2. Eligibility criteria

This study included peer reviewer reports for manuscripts randomized in the CONSORT-PR and SPIRIT-PR trials from participating BMJ Publishing Group journals (ie, CONSORT-PR: *BMJ Open, The BMJ, British Journal of Sports Medicine, British Journal of Ophthalmology,* and *Heart*; SPIRIT-PR: *BMJ Open*). For the CONSORT-PR trial, we did not have access to peer reviewer comments for manuscripts submitted to *PLOS ONE* and *PLOS Medicine*. Therefore, we excluded these manuscripts along with manuscripts for which we did not have access to peer reviewer reports. We only accessed the first round of peer reviewer reports and did not consider potential additional peer reviewer reports after the first round.

2.3. Data extraction

We set up an electronic data capture tool (using REDCap [19]) for data extraction which was carried out in duplicate by blinded extractors (H.W.R., M.C., and B.S.) who were unaware of the manuscripts' original group assignments. Any discrepancies in extraction were resolved through discussion or involving a third extractor.

We extracted the number of peer reviewers, and whether reviewers requested clarification on the ten most important and poorly reported items. For items with multiple sub-items, we noted whether reviewers referenced specific sub-items or made general requests for improvements (see detailed list of items and sub-items for CONSORT-PR in Table S1 and for SPIRIT-PR in Table S2). In

addition, we recorded whether reviewers mentioned reporting guidelines and in what context they mentioned the reporting guidelines.

2.4. Outcomes

The primary outcome of this secondary analysis of peer reviewer reports was the mean proportion of the ten most important reporting items for which at least one peer reviewer requested clarification, assessed at the manuscript level (ie, aggregating peer reviewer comments at the manuscript level as they were randomized, rather than analyzing each reviewer separately). Secondary outcomes included (i) the mean proportion of the ten most important and poorly reported items for which at least one peer reviewer requested clarification, considering each sub-item as a separate item (requests for clarification were excluded if peer reviewers mentioned the item in general without specifying which sub-items needed adaptation); (ii) the mean proportion of the ten selected reporting items for which peer reviewers requested clarification, assessed at the peer reviewer level (ie, examining how frequently individual reviewers raised concerns); (iii) the mean proportion of the ten selected reporting items for which peer reviewers requested clarification, considering each sub-item as a separate item at the peer reviewer level (excluding general requests for clarification that could not be allocated to a sub-item); (iv) the mean proportion of reporting items that peer reviewers requested to be clarified, which were later adequately reported in the published articles (only considering manuscripts that were published as included in our main trial analysis) [14]; and (v) the mean proportion of specific mentions of reporting guidelines, analyzed at the manuscript level, including the context in which reporting guidelines were mentioned.

2.5. Analyses

The proportion of clarifications requested on reporting items were reported as means, including 95% CI. We generated box histograms and a forest plot to visually compare for how many reporting items peer reviewers requested a clarification (stratified by treatment arms). Categorical variables were described using frequencies and percentages. For the primary outcomes we estimated the difference between arms using the student's t-test and reported them with respective 95% CIs.

In the main analyses, all manuscripts for which we had all peer reviewers' comments were analyzed at the manuscript level (ie, unit of randomization; checking whether at least one reviewer mentioned the reported items in their report). In additional analyses we (i) included only manuscripts that were published and included in the primary outcome analyses of the CONSORT-PR and SPIRIT-PR trials [14] to assess how many of the items highlighted by at least one peer reviewer were adequately reported in published articles (ie, "implemented by authors"), and (ii) to explore the data at the peer reviewer level (instead of the manuscript level). All analyses were conducted stratified for the CONSORT-PR and the SPIRIT-PR trials. The main analysis was further stratified by accepted and published vs rejected manuscripts. As the content of the articles which were not published is confidential, we cannot present a detailed baseline table. A baseline table of accepted and published articles was published together with the main results [14].

3. Results

3.1. Included peer reviewer reports

From the original 754 randomized manuscripts, we had access to the full peer review report of 533 manuscripts, excluding manuscripts randomized to PLOS journals (n = 212) and manuscripts for which the full report was not stored any more in the editorial management system (n = 8) (Fig 1). From the 533 analyzed peer reviewer reports, 292 manuscripts derived from the CONSORT-PR trial and 241 from the SPIRIT-PR trial (Fig 1). A total of 1459 peer reviewers were involved, with 740 allocated to the intervention group (from 265 manuscripts) and 719 allocated to the control group (from 268 manuscripts; Table S3). Of those 1459 reports of individual peer reviewers, 887 derived from the CONSORT-PR trial (median of three peer reviewers per manuscript; IQR 2−4) and 572 from the SPIRIT-PR trial (median of two peer reviewers per manuscript; IQR 2−3; Table S3).

3.1.1. Main results

Reviewers in the intervention group (who received an email reminder highlighting the ten most important reporting items) requested clarification on a greater number of reporting items compared with those in the control group (Fig 2; Fig S1). Combining the data from CONSORT-PR and SPIRIT-PR trials, peer reviewers in the intervention group asked for clarification of 21.1% (95% CI, 18.6%−23.6%) of the ten reporting items compared with 13.1% (95% CI, 18.6%−23.6%) in the control group (mean difference of 8.0%; 95% CI, 4.9%−11.1%; Table 1). In the CONSORT-PR trial, peer reviewers requested clarifications for 20.9% (95% CI, 17.5%−24.3%) of items in the intervention group and 14.6% (95% CI, 12.1%−17.1%) in the control group (mean difference of 6.3%, 95% CI, 2.0%−10.5%; Table 1). In the SPIRIT-PR trial, peer reviewers requested a clarification of 21.4% (95% CI, 17.7%−25.2%) of items in the intervention group and 11.3% (95% CI, 8.7%−13.9%) in the control group (mean difference of 10.1%, 95% CI, 5.6%−14.6%). Results were similar when assessing each sub-item as a separate item (Table 1). Furthermore, the effect was consistent when analyzing requested items stratified by the manuscript outcome status

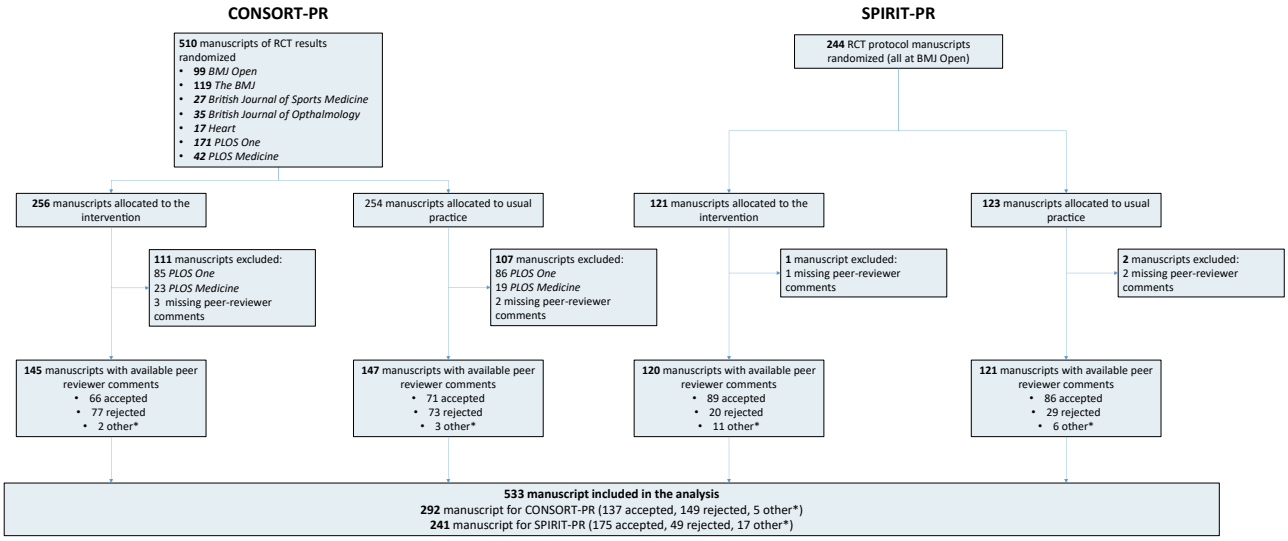

Figure 1. Flow chart of RCTs included in the secondary in-depth analysis of peer reviewer reports.

(ie, accepted and published vs rejected), but with slightly more items with requests for clarification in rejected manuscripts (Table S4).

3.1.2. *Dilution of effect during the peer review process*

For CONSORT-PR the mean difference between intervention groups decreases from 6.3% to 2.2% when analyzing only manuscripts that were later accepted and published (Table S5). When assessing the number of items that improved in the published version compared to the submitted version (based on reviewer requests), this difference decreased further to 1.3% (Table S5). For SPIRIT-PR, a similar pattern was observed (Table S6). The overall mean difference (combining data from CONSORT-PR and SPIRIT-PR) between the intervention group and control group decrease from 8.0% to 4.2% when only assessing accepted and published manuscripts. This difference decreased further to 2.6% when only considering changes that were then implemented by authors (meaning items that were requested in peer reviewer reports and adequately reported in published articles; Table 2). Approximately 55% (ie, 56.2% [173/308] intervention group and 53.3% [105/197]) of the reporting items criticized by peer reviewers were later adequately reported in the published article (Table 2).

3.1.3. *Requests for clarification from separate peer reviewers*

When analyzing all data from the CONSORT-PR and the SPIRIT-PR trial at the peer reviewer level, 9.0% (95% CI, 7.8%−10.0%) of the ten most important and underreported reporting items in the intervention group were requested to be clarified, compared to 5.6% (95% CI, 4.9%−6.3%) in the control group (Table 3). Results were

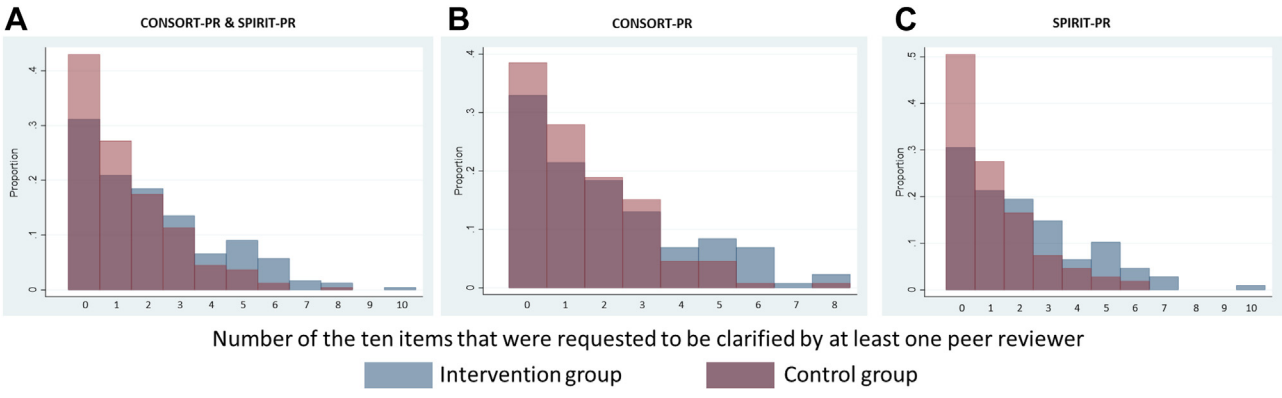

Figure 2. Proportion of manuscripts for which peer reviewers requested clarification for the ten most important reporting items, stratified by intervention arms. The number of requested clarifications in reporting when assessing the 10 most important and underreported reporting items in (A) the CONSORT-PR and SPIRIT-PR trials combined, (B) only the CONSORT-PR trial, and (C) only the SPIRIT-PR trial.

Table 1. Comparison of proportions of the ten most important reporting items for which at least one peer reviewer requested clarification between intervention and control groups (assessed at the manuscript level)

Outcomes	Intervention group[a]	Control group[a]	Overall	Mean difference
CONSORT-PR and SPIRIT-PR	*n* = 265	*n* = 268	*n* = 533	
Mean proportion of ten selected reporting items for which at least one peer reviewer requested clarification (95% CI)	21.1% (18.6–23.6)	13.1% (11.3–14.9)	17.1% (15.5–18.7)	8.0% (4.9–11.1)
Mean proportion of selected reporting items for which at least one peer reviewer requested clarification, considering each sub-item as a separate item (95% CI)	10.7% (9.2–12.1)	5.9% (4.9–7.0)	8.2% (7.4%–9.1%)	4.7% (3.1–6.5)
CONSORT-PR	*n* = 145	*n* = 147	*n* = 292	
Mean proportion of ten selected reporting items for which at least one peer reviewer requested clarification (95% CI)	20.9% (17.5–24.3)	14.6% (12.1–17.1)	17.7% (15.6–19.9)	6.3% (2.0–10.5)
Mean proportion of selected reporting items for which at least one peer reviewer requested clarification, considering each sub-item as a separate item (95% CI)	11.3% (9.3–13.3)	6.8% (5.4–8.1)	9.0% (7.8–10.2)	4.5% (2.1–6.9)
SPIRIT-PR	*n* = 120	*n* = 121	*n* = 241	
Mean proportion of ten selected reporting items for which at least one peer reviewer requested clarification (95% CI)	21.4% (17.7–25.2)	11.3 (8.7–13.9)	16.3% (14.0–18.7)	10.1% (5.6–14.6)
Mean proportion of selected reporting items for which at least one peer reviewer requested clarification, considering each sub-item as a separate item (95% CI)	9.9% (7.8–12.0)	4.8% (3.6–6.0)	7.3% (6.1–8.6)	5.1% (2.7–7.5)

CONSORT-PR, Consolidated Standards of Reporting Trials for peer review; n, sample size (number of observations or participants); *P* value, probability value indicating the statistical significance of the result; SPIRIT-PR, Standard Protocol Items: Recommendations for Interventional Trials for peer review.

[a] *Intervention Group*: Peer reviewers received a reminder of the 10 most important and underreported reporting items, *Control Group*: Peer reviewers followed the regular review process without receiving any reminders.

consistent when analyzing each sub-item as a separate reporting item and when considering the CONSORT-PR and SPIRIT-PR trials separately (Table 3).

3.2. General references to reporting guidelines

Across the CONSORT-PR and SPIRIT-PR trials, at least one peer reviewer mentioned reporting guidelines in 47.2% (125/265) of manuscripts in the intervention group compared to 24.3% (65/268) in the control group (Table S7). Specifically, 18.6% of the referrals to reporting guidelines suggested a specific improvement in the manuscript (24.9% in the intervention vs 12.3% in the control group), 12.0% mentioned that they checked the reporting items (18.9% intervention vs 5.2% control), and 9.4% stated that all items were adequately reported (14.3% intervention vs 4.5% control).

4. Discussion

Our secondary in-depth analysis of peer reviewer reports from two trials revealed that reminding peer reviewers to check specific reporting items influenced their behavior. Peer reviewer reports for manuscripts receiving the intervention suggested improvements for the 10 targeted reporting items more frequently compared to those in the control group. However, this effect was strongly diluted by the facts that (i) a large proportion (43%; 231/533) of manuscripts was not published and (ii) that only approximately half of the requested changes from peer reviewers were adequately implemented by authors (Fig S2; Table 2). Hence, the effect disappeared when assessing the reporting quality in published articles, the focus of our trials [14].

Our study aligns with previous findings from Hopewell et al, showing that peer reviewers often fail to detect deficiencies in reporting (as seen in our control group) and that authors frequently fail to adequately address requested

Table 2. Comparison of proportions of reporting-items that peer reviewers requested clarification which were later adequately reported in the published articles for CONSORT-PR and SPIRIT-PR combined

	Requested to be clarified: Overall		Requested to be clarified: Accepted and published manuscripts		Requested to be clarified: Adequately reported in published articles	
Item	**Intervention group (n = 265)**	**Control group (n = 268)**	**Intervention group (n = 150)**	**Control group (n = 152)**	**Intervention group (n = 150)**	**Control group (n = 152)**
Overall CONSORT-PR and SPIRIT-PR combined, (%)	560/2650[a] (21.1%)	352/2680[a] (13.1%)	308/2650[a] (11.6%)	197/2680[a] (7.4%)	173/2650[a] (6.5%)	105/2680[a] (3.9%)
Difference in proportions between groups	**8.0%**		**4.2%**		**2.6%**	
Percentage of reporting items correctly reported in article when requested by peer-reviewer	-	-	-	-	173/308 (56.2%)	105/197 (53.3%)

CONSORT-PR, Consolidated Standards of Reporting Trials for peer review; SPIRIT-PR, Standard Protocol Items: Recommendations for Interventional Trials for peer review.

CONSORT-PR and SPIRIT-PR, referring to the randomized trials conducting at journal level to potential improve the reporting quality [12,13].

[a] Referring to the 10 most important and underreported reporting-items that were assessed in submitted manuscripts.

Table 3. Mean proportion of the ten most important reporting items for which peer reviewers asked for clarification (assessed on the peer reviewer level)

Outcomes	**Intervention group[a]**	**Control group[a]**	**Overall**	**Mean difference**
CONSORT-PR & SPIRIT-PR	n = 740	n = 719	n = 1459	
Mean proportion of ten selected reporting items for which peer reviewers requested clarification (95% CI)	9.0% (7.8–10.0)	5.6% (4.9–6.3)	7.3% (6.7–7.9)	3.4% (2.2–4.7)
Mean proportion of selected reporting items for which peer reviewers requested clarification, considering each sub-item as a separate item (95% CI)	4.3% (3.7–4.8)	2.3% (2.0–2.7)	3.3% (3.0%–3.6%)	2.0% (1.3–2.6)
CONSORT-PR	n = 448	n = 439	n = 887	
Mean proportion of ten selected reporting items for which peer reviewers requested clarification (95% CI)	8.3% (7.1–9.5)	5.6% (4.7–6.5)	7.0% (15.6–19.9)	2.7% (1.2–4.2)
Mean proportion of selected reporting items for which peer reviewers requested clarification, considering each sub-item as a separate item (95% CI)	4.2% (3.5–4.8)	2.4% (2.0–2.9)	3.3% (2.9–3.7)	1.7% (0.9–2.5)
SPIRIT-PR	n = 292	n = 280	n = 572	
Mean proportion of ten selected reporting items for which peer reviewers requested clarification (95% CI)	10.1% (8.3–11.8)	5.5 (4.4–6.6)	7.8% (6.8–8.9)	4.6% (2.4–6.7)
Mean proportion of selected reporting items for which peer reviewers requested clarification, considering each sub-item as a separate item (95% CI)	4.4% (3.4–5.3)	2.2% (1.7–2.7)	3.3% (2.7–3.9)	2.2% (1.1–3.3)

CONSORT-PR, Consolidated Standards of Reporting Trials for peer review; n, sample size (number of observations or participants); P value, probability value indicating the statistical significance of the result; SPIRIT-PR, Standard Protocol Items: Recommendations for Interventional Trials for peer review.

[a] *Intervention Group*: Peer reviewers received a reminder of the 10 most important and underreported reporting items, *Control Group*: Peer reviewers followed the regular review process without receiving any reminders.

changes [20]. A Cochrane review published in 2023 identified 10 RCTs showing that training peer reviewers lead to little or no improvement in the quality of the peer review [21]. Based on this evidence and the fact that reviewers are becoming fatigued [22] by an ever increasing volume of new journals and submitted research and by the fact that their services are not sufficiently valued and hardly acknowledged [23−25], a broader discussion is necessary to determine the structure and future of peer review. Open questions include whether it is acceptable to assign additional responsibilities to peer reviewers, whether they should be provided with appropriate incentives, or whether certain tasks, such as assessing reporting quality, should be handled by journal staff (eg, paid expert reviewers) or might be supported by artificial intelligence tools.

Our in-depth analysis of peer reviewer reports has the strength that we had access to peer reviewer reports from manuscripts from two RCTs conducted at the journal level [14]. Assessing these reports allowed us to study in detail at what stage the intervention of these two RCTs failed. However, our analysis has some limitations worth mentioning. First, we do not know what happened to manuscripts that were not published (ie, rejected manuscripts). In theory, the impact of the intervention (ie, an increase in requested changes on reporting items by peer reviewers) could have contributed to improve reporting in manuscripts that were potentially submitted to other journals. Second, we cannot be certain that the effect at the peer reviewer level, diminished entirely in published articles. However, in our sample size calculation for the trials, we defined a relevant effect as the adequate reporting of at least one additional reporting item in the intervention group compared to the control group. Hence, we would conclude that if a small effect of the intervention was present at the level of published articles, it would probably not be sufficiently relevant to justify the additional peer reviewer burden. Third, we did not have access to peer reviewer reports from the PLOS journals due to confidentiality restraints. However, we would not expect a different result in those journals. Fourth, a few peer reviewers in the intervention group (52/740; 7%) did not receive the reminder email because they handed in their peer review report before the reminder email could be sent out. Hence, it is possible that our calculated effect sizes slightly underestimate the true effect.

In conclusion, our in-depth analysis of two randomized trials showed that reminding peer reviewers to check reporting items increased their focus on reporting guidelines leading to more reporting-related requests in their review reports. However, the effect was strongly diluted during the peer review process (particularly due to rejected articles and requests not being implemented by authors). This suggests that simple and low-cost interventions may not be enough, and other strategies like using expert reviewers, might be needed to better adhere to reporting guidelines and make research more transparent. Furthermore, journals should make sure that requested clarifications are adequately addressed in revised manuscripts.

Ethics statement

Both trials (i.e., CONSORT-PR and SPIRIT-PR) received ethical approval from the Medical Sciences Interdivisional Research Ethics Committee of the University of Oxford (R62779/RE001), and were prospectively registered on Open Science Framework (https://osf.io/c4hn8 and https://osf.io/z2hm9). In addition, they are registered on ClinicalTrials.gov (NCT05820971 and NCT05820984).

CRediT authorship contribution statement

Hillary Wnfried Ramirez: Writing − original draft, Project administration, Formal analysis, Data curation. **Malena Chiaborelli:** Writing − review & editing, Project administration, Data curation. **Christof M. Schönenberger:** Writing − review & editing, Data curation. **Katie Mellor:** Writing − review & editing, Data curation. **Alexandra N. Griessbach:** Writing − review & editing, Data curation. **Paula Dhiman:** Writing − review & editing, Data curation. **Pooja Gandhi:** Writing − review & editing, Data curation. **Szimonetta Lohner:** Writing − review & editing, Data curation. **Arnav Agarwal:** Writing − review & editing, Data curation. **Ayodele Odutayo:** Writing − review & editing, Conceptualization. **Michael M. Schlussel:** Writing − review & editing, Methodology, Formal analysis, Conceptualization. **Philippe Ravaud:** Writing − review & editing, Methodology, Conceptualization. **David Moher:** Writing − review & editing, Methodology, Conceptualization. **Matthias Briel:** Writing − review & editing, Methodology, Conceptualization. **Isabelle Boutron:** Writing − review & editing, Methodology, Conceptualization. **Sally Hopewell:** Writing − review & editing, Supervision, Project administration, Methodology, Conceptualization. **Sara Schroter:** Writing − review & editing, Supervision, Resources, Project administration, Methodology, Data curation, Conceptualization. **Benjamin Speich:** Writing − original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

B.S. and M.B. received unrestricted grants from Moderna (2021/22) for studies unrelated to the presented work. S.S. is employed by BMJ Publishing Group. K.M. is employed by Clarivate. D.M., S.H., and I.B. are members of the CONSORT executive and authors of the CONSORT 2010 statement. D.M. is an author of the SPIRIT 2013 Statement. D.M., M.M.S., P.D., and P.R. are members of the Enhancing the Quality and Transparency of Research network. There are no competing interests for any other author.

Acknowledgments

The authors are grateful to the seven publishing journals from the BMJ Publishing Group and the Public Library of Science which allowed us to randomize manuscripts to the intervention or control group. For this secondary in-depth analysis of peer reviewer reports, the authors would like to specifically thank the five journals which granted us access to peer reviewer reports (ie, *BMJ Open, The BMJ, British Journal of Sports Medicine, British Journal of Ophthalmology, Heart*).

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2025.111818.

Data availability

A completely de-identified dataset is provided in the supplementary appendix

References

[1] Moher D, Altman DG, Schulz KF, Simera I, Wager E. Guidelines for Reporting Health Research: A User's Manual. 1st ed. Wiley. 2014. Available at: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118715598. Accessed March 31, 2025. https://doi.org/10.1002/9781118715598.

[2] Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. BMC Med 2010;8:24. https://doi.org/10.1186/1741-7015-8-24.

[3] Group TSoRT. A proposal for structured reporting of randomized controlled trials. JAMA 1994;272(24):1926−31.

[4] Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. J R Soc Med 2016;109:67−77. https://doi.org/10.1177/0141076815625599.

[5] Network E. Enhancing the QUAlity and Transparency of health Research. What is a reporting guideline? Available at: https://www.equator-network.org/about-us/what-is-a-reporting-guideline/. Accessed March 31, 2025.

[6] Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014. Trials 2016;17:301. https://doi.org/10.1186/s13063-016-1408-z.

[7] Smith TA, Kulatilake P, Brown LJ, Wigley J, Hameed W, Shantikumar S. Do surgery journals insist on reporting by CONSORT and PRISMA? A follow-up survey of 'instructions to authors. Int J Surg 2014;12:429−32. https://doi.org/10.1016/j.amsu.2014.12.003.

[8] Jin Y, Sanger N, Shams I, Luo C, Shahid H, Li G, 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. https://doi.org/10.2147/JMDH.S155103.

[9] Samaan Z, Mbuagbaw L, Kosa D, Borg Debono V, Dillenburg R, Zhang S, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. J Multidiscip Healthc 2013;6:169−88. https://doi.org/10.2147/JMDH.S43952.

[10] Yin Y, Shi F, Zhang Y, Zhang X, Ye J, Zhang J. Evaluation of reporting quality of randomized controlled trials in patients with COVID-19 using the CONSORT statement. PLoS One 2021;16. https://doi.org/10.1371/journal.pone.0257093.

[11] Speich B, Schroter S, Briel M, Moher D, Puebla I, Clark A, et al. Impact of a short version of the CONSORT checklist for peer reviewers to improve the reporting of randomised controlled trials published in biomedical journals: study protocol for a randomised controlled trial. BMJ Open 2020;10:e035114. https://doi.org/10.1136/bmjopen-2019-035114.

[12] Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:100−7. https://doi.org/10.1136/bmj.c332.

[13] Chan A, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 Statement: defining standard protocol items for clinical trials. Ann Intern Med 2013;158(3):200−7. https://doi.org/10.7326/0003-4819-158-3-201302050-00583.

[14] Speich B, Mann E, Schonenberger CM, Mellor K, Griessbach AN, Dhiman P, et al. Reminding peer reviewers of reporting guideline items to improve completeness in published articles: primary results of 2 randomized trials. JAMA Netw Open 2023;6:e2317651. https://doi.org/10.1001/jamanetworkopen.2023.17651.

[15] Hopewell S, Boutron I, Altman DG, Barbour G, Moher D, Montori V, et al. Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. BMC Med 2016;14:199. https://doi.org/10.1186/s12916-016-0736-x.

[16] Kyte D, Duffy H, Fletcher B, Gheorghe A, Mercieca-Bebber R, King M, et al. Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. PLoS One 2014;9:e110229. https://doi.org/10.1371/journal.pone.0110229.

[17] Speich B, Odutayo A, Peckham N, Ooms A, Stokes JR, Saccilotto R, et al. A longitudinal assessment of trial protocols approved by research ethics committees: the Adherance to SPIrit REcommendations in the UK (ASPIRE-UK) study. Trials 2022;23:601. https://doi.org/10.1186/s13063-022-06516-1.

[18] Gryaznov D, von Niederhausern B, Speich B, Kasenda B, Ojeda-Ruiz E, Blumle A, et al. Reporting quality of clinical trial protocols: a repeated cross-sectional study about the Adherence to SPIrit Recommendations in Switzerland, Canada and Germany (ASPIRE-SCAGE). BMJ Open 2022;12:e053417. https://doi.org/10.1136/bmj.g4145.

[19] Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208.

[20] Hopewell S, Collins GS, Boutron I, Yu LM, Cook J, Shanyinde M, et al. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. BMJ 2014;349. https://doi.org/10.1136/bmj.g4145.

[21] Hesselberg JO, Dalsbo TK, Stromme H, Svege I, Fretheim A. Reviewer training for improving grant and journal peer review. Cochrane Database Syst Rev 2023;11:MR000056. https://doi.org/10.1002/14651858.MR000056.pub2.

[22] Publons. Global state of peer review 2018. Available at: https://clarivate.com/lp/global-state-of-peer-review-report/. Accessed March 31, 2025.

[23] Jia P, Xie W, Zhang G, Wang X. Do reviewers get their deserved acknowledgments from the authors of manuscripts? Scientometrics 2023;128:5687−703. https://doi.org/10.1007/s11192-023-04790-7.

[24] Kitchen TS. Peer review — recognition wanted. Available at: https://scholarlykitchen.sspnet.org/2015/01/08/peer-review-recognition-wanted/. Accessed March 31, 2025.

[25] Blog TOP. Authors aren't the only ones: acknowledging peer reviewers. Available at: https://theplosblog.plos.org/2022/01/acknowledging-reviewers. Accessed March 31, 2025.