

What Reviewers Mean by “Writing Quality”: An Analysis of Published Peer Review Reports in *BMJ Open*

Carlotta Fiammenghi¹

¹ Dept. of Foreign Languages and Literatures, University of Verona, Verona, Italy

Correspondence: Carlotta Fiammenghi, Dept. of Foreign Languages and Literatures, University of Verona, Verona, Italy.

Received: May 5, 2026 Accepted: July 24, 2026 Online Published: August 27, 2026

doi:10.5539/ijel.v16n5p1 URL: <https://doi.org/10.5539/ijel.v16n5p1>

Abstract

Recent developments in scholarly publishing have led to increasing adoption of open peer review (OPR), a set of practices intended to enhance transparency and participation in the peer review process. Despite the centrality of peer review to academic writing and its acknowledged relevance within English for Research Publication Purposes (ERPP) and EAP, reviewers' comments remain a largely occluded genre. This study examines published peer review reports from the *British Medical Journal (BMJ) Open*, a widely recognized journal in the medical field and one of the earliest adopters of fully open peer review. Drawing on a corpus of 109 research articles and their review histories from a full 2024 issue, the analysis focuses on the frequency and nature of language-related comments in reviewers' reports. Results show that reviewers frequently comment on text organization, word choice, clarity and compliance with “academic style” norms. These findings suggest that OPR reports offer a valuable resource for ERPP and EAP pedagogy by making visible the linguistic, rhetorical and epistemic expectations that shape publication-oriented academic writing.

Keywords: open peer review, language-related comments, academic style, academic writing, medical academic English

1. Introduction

Peer review is widely recognized as a cornerstone of scientific writing, functioning as a quality-control mechanism intended to safeguard the reliability and credibility of published research. Although the idea that academic work should be scrutinized by knowledgeable peers can be traced back to antiquity, peer review only began to take a formalized shape in the eighteenth century: the Royal Society of London is frequently cited for instituting one of the earliest formal peer review procedures in 1752 for the *Philosophical Transactions* (Räsänen, 1999). Widespread institutionalization, however, did not occur until after the Second World War, when the exponential growth of scientific research required more structured evaluative systems. Since the 1970s, the double-blind review model has become standard practice across disciplines, with both author and reviewer identities concealed in an effort to reduce bias (Chapelle, 2014).

Despite its clear benefits, the double-blind model faces longstanding criticisms (Benos, et al., 2007). Scholars across fields have noted that concealment of identities limits reviewer recognition and may discourage participation, particularly in a research environment shaped by “publish or perish” pressures. More broadly, the opacity of the process has led to concerns about limited accountability and potential biases that remain underexamined precisely because the system is largely closed to public scrutiny (Squazzoni, et al., 2020).

In response, open peer review (OPR) has attracted increasing attention since the early 2000s. While the term does not yet have a single standardized definition, it generally encompasses the disclosure of reviewer and/or author identities (“open identities”), the publication of peer review reports (“open reports”), and sometimes broader participation extending beyond invited experts (“open participation”, including, for example, patients' review of clinical research) (Ross-Hellauer, 2017). These varying definitions mean that it is difficult to calculate precise measures of how many journals adopt OPR practices. However, it seems clear that adoption has been growing rapidly, primarily in the medical and scientific domains (Polka, Kiley, Konforti, Stern, & Vale, 2018; Wolfram, Wang, Hembree, & Park, 2020), although OPR remains far from the norm in academic publishing (Ross-Hellauer & Horbach, 2024). Concerning effectiveness, OPR is understood by “OPR journals” to increase transparency, accountability and reviewer visibility

(Groves, 2010; Groves & Loder, 2014), without evidence of declining review quality (Wolfram, Wang, Hembree, & Park, 2020; Rooyen & al., 2010), despite the fact that concerns persist about professional repercussions, especially for researchers who may feel vulnerable when signing critical reviews (Ross-Hellauer, Deppe, & Schmidt, 2017).

For scholars working in English for Academic Purposes and English for Research Publication Purposes (ERPP) (Flowerdew & Habibie, 2022), the genre of the peer review has always been regarded as an interesting site where academic writing norms, expectations and practices intertwine. Some scholars who have been able to access peer review reports and articles' earlier drafts (thanks to the collaboration of journal editors and editorial boards) have carried out studies exploring both the form and the content of these texts. This body of research has been growing since the emergence of OPR, which unveiled this previously occluded academic genre.

Earlier studies into peer review reports of scientific research papers found that a substantial proportion of reviewer comments are "interactional" in nature, focusing on the rhetorical appropriateness of manuscripts rather than purely on content (Gosden, 2003). Belcher (2007) similarly observed frequent but often vague references to language in reports for an applied linguistics journal, and connected this with wider journal gate-keeping practices in Anglophone academic contexts. Mungra & Webber (2010), who examined medico-scientific papers written in Italy, reported that almost half of all reviewer comments addressed language use, ranging from grammar and lexis to text structure and rhetorical features. Paltridge (2017) conducted a comprehensive examination of academic peer review by focusing on reports from the academic journal *English for Specific Purposes*, grant and promotion applications, and book reviews, and by focusing on the pragmatic nature of reviewer language. He found that this often employs indirect speech acts and politeness strategies, which may make it difficult for authors, especially beginners, to discern the intended requirements for revision. He also found that comments about language aspects, such as word choice, grammar, and style, play a prominent role in peer review reports, as they appear frequently across different review categories (i.e., minor revisions, major revisions, and rejection), although they are generally not the dominant reason for a manuscript's rejection. Rather, such language comments serve as detailed instructions, primarily necessary to ensure the manuscript meets the high standards of presentation and clarity required for publication, even when the underlying research is already considered fundamentally sound. He concludes that it is often through adhering to these comments that a paper successfully moves past the revision stage.

Biomedical journals have been central to this line of research: studies of major medical journals have examined the characteristics of good reviewers and good reviews, the manuscript features associated with acceptance, and the extent to which peer review improves reporting practices, including data presentation, acknowledgement of limitations and calibration of claims (Black, van Rooyen, Godlee, Smith, & Evans, 1998; Lee, Boyd, Holroyd-Leduc, Bacchetti, & Bero, 2006; Schriger, Sinha, Schroter, Liu, & Altman, 2006; Kesserlioglu, Kilicoglu, & ter Riet, 2019). Within this tradition, the *British Medical Journal* (henceforth *BMJ*) and *BMJ Open* seem to occupy an especially relevant position: both have made peer review processes more visible through open review practices, thereby enabling corpus-based and text-oriented analyses of reviewers' language. Importantly for the present analysis, Delgado, Garretson, & Delgado (2019) carried out a lexical and a net sentiment analysis on reviewer reports written for the *BMJ* between 2014 and 2017. They found that *BMJ* reviewers frequently addressed the methodology (rather than the actual results) of the paper, and that they often referred to the clarity and writing quality of the texts, using evaluations such as "clear" and "well-written" or, conversely, "unclear" and suggestions that the text "would be clearer" if revised.

The present study contributes to this line of inquiry by analyzing the openly published peer review reports of *BMJ Open*, chosen as a disciplinary site where clarity, methodological transparency and epistemic caution are not merely stylistic concerns, but central features of responsible research communication. *BMJ Open* is an online, open access journal, dedicated to publishing medical research from all disciplines and therapeutic areas. It publishes all research study types, from protocols through phase I trials to meta-analyses, including small, specialist studies, and negative studies, and has practiced fully open peer review since 2014. The instructions for reviewers explicitly state that the journal adheres to a fully open peer review process, and ask reviewers to assess papers' scientific credibility, methodology, statistics, references, research and publication ethics. As far as language is concerned, guidelines specify that reviewers will be asked to provide judgements on presentation of data and standard of written English¹. Through a quali-quantitative, thematic analysis of the reports accompanying all articles published in a full 2024 issue, the present study thus aims to shed light on reviewers' expectations regarding clarity, organization, academic style, and language effectiveness more generally. The corpus arguably provides a particularly fertile context for examining how reviewers articulate linguistic and rhetorical expectations.

2. Method

The dataset for this study consists of all research articles and associated peer review reports published in the most

recent complete issue of *BMJ Open* available at the time of data collection, which was conducted manually on 5th November 2024. For the quantitative component of the analysis, a dataset was constructed recording, for each article, the topic of the research paper, the dates of submission and acceptance, the total number of reviewers involved, and the institutional affiliations of both authors and reviewers. In cases of multiple authorship, only the first and corresponding author was considered, as this author typically assumes primary responsibility for the conceptualization, drafting, and revision of the manuscript. Author and reviewer affiliations were subsequently classified according to Kachru's (1985) tripartite model of English use, distinguishing between inner-circle contexts, where English is traditionally used as a first language; outer-circle contexts, where English has an institutionalized second-language role; and expanding-circle contexts, where English is primarily used as a foreign language. Although Kachru's model has been criticized for simplifying complex multilingual realities and for treating linguistic contexts as relatively stable and geographically bounded, it remains a useful heuristic for the broad purposes of the present study (Jenkins & Morán Panero, 2024). This classification was therefore used not as an indicator of authors' or reviewers' actual linguistic backgrounds or English proficiency, but as a proxy for the main linguistic and academic environment in which they were institutionally located.

The qualitative component of the study consisted of a thematic analysis of the peer review reports, carried out using Atlas.ti software (25 for Windows). Codes were developed inductively and refined iteratively, following Braun and Clarke's (2006) five-phase framework of familiarization with the data, initial coding, theme generation, theme review, and theme definition. Coding focused specifically on language-related reviewer comments and captured several dimensions, including the evaluative polarity of the comment (positive or negative), the IMRaD section of the manuscript to which it referred (when explicitly stated), whether the issue was perceived as major or minor (when indicated by the reviewer), and the specific linguistic aspect involved (such as grammar, lexical choice, textual flow, or academic style). The analysis was then repeated by the same researcher at a six-months' interval to ensure consistency and replicability.

For the purposes of this study, "language-related comments" were defined as reviewer remarks that explicitly targeted the writing quality of the manuscript rather than its scientific content. These included comments addressing clarity, readability, grammatical accuracy, vocabulary, cohesion, rhetorical appropriateness, or adherence to conventions of academic writing. Requests for additional data, information, or methodological clarification were excluded unless the reviewer explicitly framed the issue as a problem of linguistic expression, for example by stating that a sentence, paragraph, or section was unclear or difficult to follow. This distinction was necessarily interpretative, given the close interdependence of form and content in academic discourse (and beyond). However, the iterative nature of the thematic analysis, involving continuous refinement of codes and themes, together with the fact that the analysis was conducted twice at a six-month interval, helped ensure the robustness of the theme generation phase. Only those passages in which reviewers explicitly evaluated or problematized the language used to convey the research were ultimately coded as language-related.

3. Results

3.1 Description of the Corpus

The corpus is made of a total of 109 research articles and related peer review histories, spanning 40 distinct medical and health-related topics, which reflects the multidisciplinary scope of *BMJ Open*. The number of reviewers per article ranges from one to five, with a mean of 2.25 reviewers and a median of two, yielding a total of 245 single reviewers, indicating a relatively stable reviewing structure across the dataset. Review times range from 2 to 17 months, with a mean of 7.7 months ($SD \approx 2.9$) and a median of 7 months; most manuscripts therefore complete the peer review process within approximately one academic year, although a limited number experience substantially longer review trajectories.

Regarding authorship, the majority of first authors are affiliated with institutions located in expanding-circle contexts (71 articles, 65.14%), followed by authors from the inner circle (28 articles, 25.7%) and a smaller group from the outer circle (10 articles, 9.17.3%). This distribution highlights the global reach of the journal and confirms the prominence of authors working in contexts where English is typically used as a foreign language, with a particularly dominant role played by China (29 different single authors). Figure 1 maps the distribution of authors' affiliations across the world.

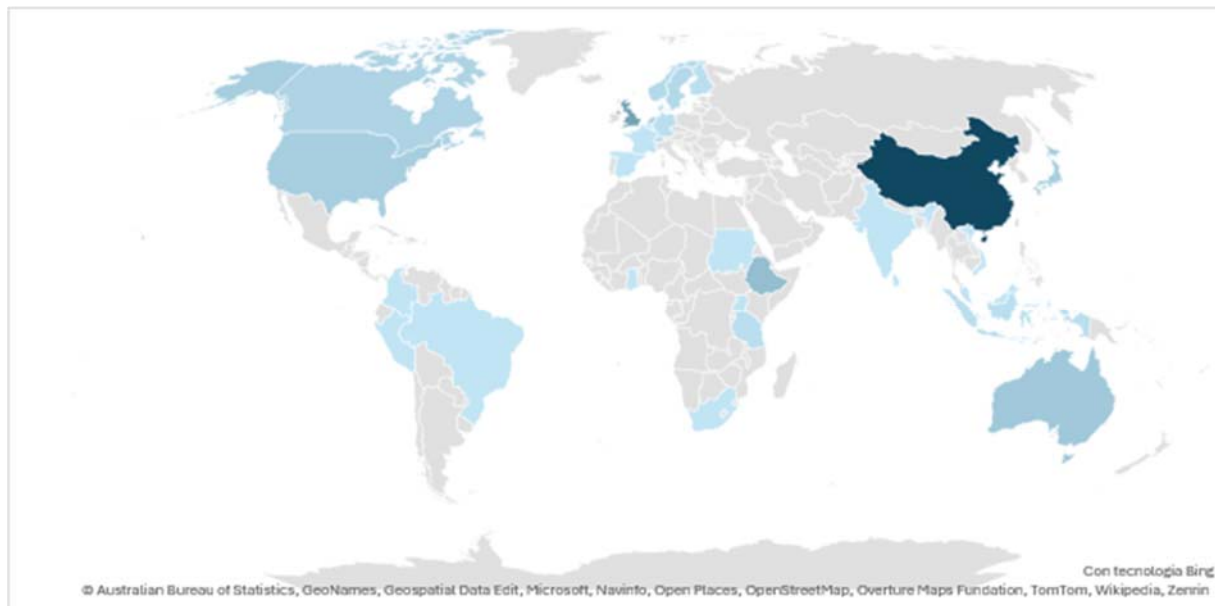


Figure 1. Map of authors' affiliations across the world

The peer review process at *BMJ Open* similarly shows a markedly international character, as the 245 single reviewers in the dataset are affiliated with institutions in 58 different countries. From a linguistic perspective, reviewers are distributed across all three Kachruvian circles: expanding-circle reviewers constitute the largest group (110, 44.9%), closely followed by inner-circle reviewers (98, 40%), while outer-circle reviewers form a smaller group (28, 11.43%). For nine reviewers, affiliation details were not provided.

Compared to the distribution of authors, the reviewer pool shows a more balanced representation between inner- and expanding-circle contexts, although an important role is played by inner-circle territories such as the United States of America (39 reviewers), the United Kingdom (26 reviewers), Australia and Canada (15 reviewers each), followed by China (19 reviewers). Figure 2 maps the distribution of reviewers' affiliations across the world.

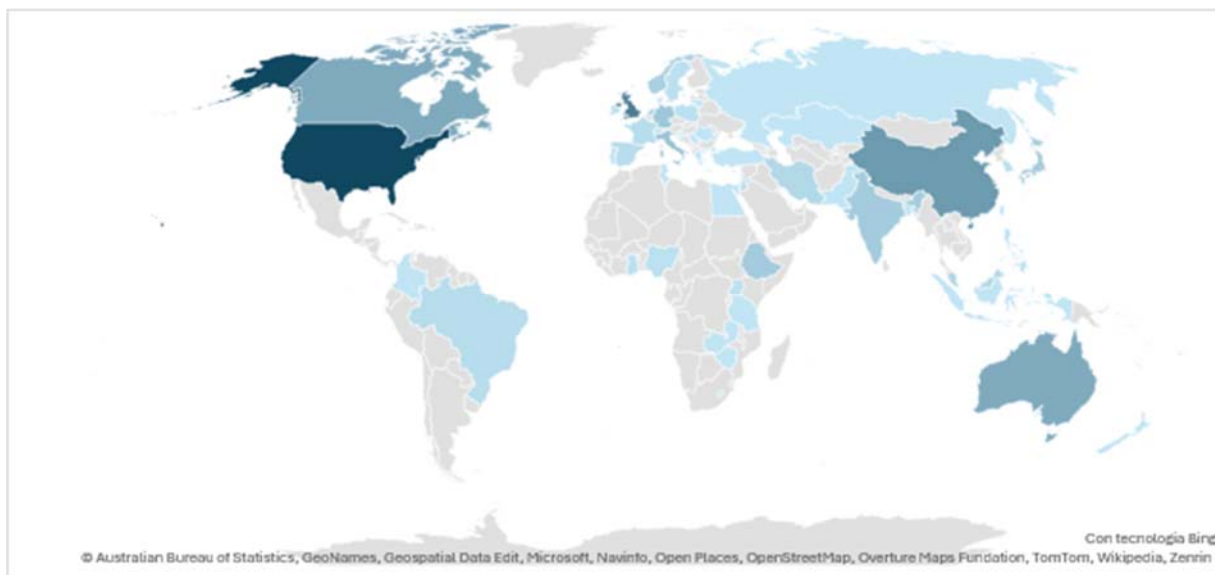


Figure 2. Map of reviewers' affiliations across the world

3.2 Language-Related Feedback

Language-related feedback varies considerably across articles. The total number of language-related comments in the corpus is 848, but language-related comments per article range from 0 to 49, with a mean of 7.7 comments ($SD \approx 7.3$) and a median of 6 comments. This reveals a markedly skewed distribution: while many articles receive little or no linguistic feedback, others attract substantial attention to language and style. Similarly, reviewer engagement, measured by the number of comments produced per reviewer, shows substantial dispersion, as individual reviewers produce between 0 and 44 comments, with a mean of 3.5 comments and a median of 2. Thus, while many reviewers contribute only a limited number of comments, a minority provide very detailed feedback. Importantly, negative language-related comments ($n = 726$) are far more frequent than positive ones ($n = 107$) both at the article and at the reviewer level, suggesting that reviewer engagement with language is predominantly corrective. Only 15 language-related comments are used as genuine requests for clarification, without a negative or positive evaluation, as in the following:

- I am not a native English speaker and had some difficulty to understand the term “remit” that is widely used. My dictionary did not help me much but I interpret the use of the word as if it intends to mean mandate or range of responsibility – correct?

3.2.1 Thematic Analysis of Language-Related Comments

The thematic analysis of the peer review reports yielded five positive language-related codes and ten negative language-related codes. These are listed in order of frequency in Table 1 and Table 2.

Table 1. Type and frequency of positive language-related comments

Comment	n. of quotations
Well-written	61
Clear	33
Well-organized	7
Concise	5
Appropriate hedging/boosting strategies	1
Total	107

Table 2. Type and frequency of negative language-related comments

Comment	n. of quotations
Information is messy/disorganized	177
Inappropriate or ineffective word choice/wording/phrasing	174
Unclear/Difficult to read	121
Long/repetitive/unnecessary paragraph/sentence	58
Problems with grammar/language of the manuscript	53
Typos/spelling/formatting mistakes	48
Inappropriate hedging/boosting strategies/causal language	46
Inappropriate use of acronyms/abbreviations	31
Vague	13
Not compliant with academic style/norms	5
Total	726

Positive comments generally mirror the dimensions highlighted by the negative comments, with reviewers evaluating the language of the manuscript overall (“well-written” vs. “problems with grammar/language”; notes about “typos, spelling, and formatting mistakes” also refer to surface-level language aspects). Reviewers also frequently evaluate the way information is structured and presented (“well-organized” vs. “messy/disorganized”) or focus on aspects of clarity (“clear” vs. “unclear”), conciseness (“concise” vs. “long/repetitive”), and style (“appropriate” vs. “inappropriate use of hedging and boosting strategies”). Negative comments about “vagueness” and “inappropriate use of acronyms and abbreviations” can also be listed as aspects of compliance to academic style and norms. Five comments criticized non-compliance with “academic style”, but without specifying which aspect was being highlighted, as in the following:

- The English language should be more academic and standard.

Only two comments (both by an expanding circle reviewer to an expanding circle author) explicitly asked for the text to be revised by a native speaker of English:

- A native speaker should check the text.
- The manuscript would benefit from being edited by a native English writer.

The ensuing subsections further elaborate on and provide examples of comments for both positive and negative themes.

3.2.1.1 Positive Comments

The most frequent positive comment related to the general language of the manuscript, evaluated as “well-written”, often in combination with assessments about the clarity and structuring of the text. Interestingly, however, most of these comments ($n = 53$, 87%) are placed at the beginning of the review, and are followed directly by one or more negative comments, as in the following:

- The study manuscript is generally well-written with a clear objective, methods, results, discussion and conclusion. However, the following areas need attention before the article can be considered for publication.
- Overall, I find the manuscript well-written and well-structured – I do, however, have some suggestions for improvement.
- Although the manuscript is well-written and clear, some areas could be strengthened.

Only 8 comments in the dataset are used specifically to praise one well-written section of the paper (for example: “the methods section was well-written”). These praise/criticism pairs (Hyland, 2000; Belcher, 2007) thus seem to configure the “well-written” positive comment as a politeness strategy, a face-saving act, rather than genuine language feedback. These results align closely with Paltridge’s (2017) examination of peer review reports for the journal *English for Specific Purposes*. This is significant, as it suggests that politeness strategies are a stable feature of peer review discourse across different review models: they appear not only in occluded peer review reports such as those of *English for Specific Purposes*, but also in the open peer review reports published by *BMJ Open*. Importantly, this pattern emerges across disciplinary boundaries, encompassing both reviewers whose expertise lies in language and linguistics and those working in medicine, health, and the biological sciences. Taken together, these findings suggest that politeness strategies, especially the use of face-saving praise–criticism pairings at the opening of review reports, represent a recurrent characteristic of the academic peer review genre.

3.2.1.2 Negative Comments

The most frequent category of negative comments concerned the organization of information and the rhetorical structuring of manuscripts. Reviewers frequently noted that sections or arguments did not appear in expected locations or that ideas were not presented in a logical progression, as in:

- The discussion of strengths and limitations might be moved to the end of the discussion as with most manuscripts. It is awkward so early in the manuscript and before the discussion of the findings.

By invoking “most manuscripts”, the reviewer explicitly appeals to genre conventions, indicating that organization is not simply a matter of clarity but of conformity to disciplinary norms. A similar concern is evident in another reviewer’s observation that an introductory section is not coherent and cohesive:

- The introduction could be more logically organized. Starting with the problem of physical inactivity, then moving to the relevance of daily steps, followed by the current state of the research, and finally introducing the concept of umbrella reviews might improve the flow.

The comment recommends a sequential structure and articulates this expected order through expressions such as “starting with”, “then moving to”, “followed by”, and “finally”, thus providing an explicit roadmap to guide authors towards conventional academic patterns. Again, the language used in these comments confirms Paltridge’s (2017) notations on politeness strategies: both use modal verbs such as “might” and “could” to avoid face-threatening acts such as direct requests or overt suggestions, but at the same time clearly indicate ways in which the manuscript should be improved.

Beyond organizational issues, reviewers also comment extensively on the lexis of the manuscript, pinpointing inappropriate word choices or ineffective phrasings. These notations may refer to specialized vocabulary, in which case they are generally followed by a brief explanation, as in the following:

- Please use the term prediction interval, instead of confidence interval. Confidence interval is used with parameter estimation in general, rather than forecasting for the future. Prediction interval refers to the future and contains both reducible and irreducible errors. However, confidence interval contains only reducible error.
- I recommend caution when using the term “probabilistic forecasting”. A probabilistic forecasting normally consists

of two parts: (i) a list of possible future outcome and (ii) their respective likelihood. What has been offered here appears to be particular instances of probabilistic forecasts, with a focus on prediction intervals, or their corresponding quantile. I suggest that the authors think about this and consider choosing wording that accurately describes their work.

Specialized vocabulary, however, does not appear to be the domain where authors need most help (as has already been noted in previous studies, even considering students' and novice writers' work, see for example Hartle, 2023). Rather, these comments often refer to general language choices, which may sound unnatural, or carry unintended connotations, as in:

- Digital access of capability issues is unusual wording - suggest rephrase - maybe digital poverty and digital literacy.
- Avoid the use of "vulnerable" [...]. The word vulnerable comes off as disempowering and it is not clear what is makes reference to [sic] aside from the fact the patients are in hospital, which is always a position of vulnerability in some way.

The third most common negative language-related comment in the dataset concerns issues of readability and clarity. Some reviewers directly tie problems of clarity to grammatical issues: for instance, one report stated that "the writing is oftentimes unclear due to grammatical issues/concerns", and another report recommended a thorough review "from top to bottom for English grammar to support clarity". These formulations are telling because the problem, in the reviewers' framing, is not grammatical correctness in and of itself, but its role in facilitating understanding within the academic community. Similarly, unclarity may also be linked to issues with vocabulary, as in: "the terminology is not coherent through the manuscript, what leads to unclarity [sic]". In most cases, however, reviewers note that a section or a sentence is unclear without further elaborating, as in "The results section is unclear" or "Could you please rewrite this sentence for clarity".

Finally, a number of reviewers focus on features of academic style, mostly hedging and boosting strategies. Some comments evaluate whole sections or paragraphs (as in "the discussion section overstates the study findings" or "do not overreach conclusions in the abstract"). Others instead focus on individual language items such as modal verbs (e.g., by suggesting inserting the modal verb "may" in a sentence) or specific lexical items, as in:

- Caution about overreading the data from this cross sectional study. Suggest you change wording in the sentence on Page 17 line 383 from "to the potential effects of HL on health outcomes in YP" to "potential association of HL to health outcomes in YP."

As this excerpt exemplifies, these suggestions often intertwine with comments about inappropriate uses of causal language, particularly in cases where authors appear to imply causal relationships that are not warranted by the study design. Interestingly, reviewers are here concerned with the implicit meanings and nuances of given lexical choices. For example, one reviewer examined the sentence "adults who increased their cannabis use were 1.97 times more likely to experience suicidal ideation", observing that "*such phrasing implies that* increasing cannabis use causes suicidal ideation" (my emphasis). The reviewer recommended reframing the claim with "a more cautious use of language" and suggested referring to "associations" or "correlations" instead. Further illustrative cases are the following (my emphases):

- I would caution on using casual language as this study is not designed for that. The abstract states that physical activity improves adherence to medications, but this is not necessarily true - they are associated.
- Line 56–57: The conclusion that the results suggest that daily steps are a changeable indicator of health regarding mortality and CVD events is not appropriate. *The term "Changeable" assumes causality* and that interventions or trials were reviewed in this paper. Since all of the studies here are observational, I strongly suggest rewording this to reflect that daily steps are associated with a lower risk of mortality/CVD.

These examples arguably highlight the close interdependence of form and content in academic discourse, by illustrating the centrality of hedging and epistemic caution, which are core features of academic genres, to reviewers' expectations.

3.2.1.3 Co-Occurrence of Themes

The co-occurrence of themes in the corpus was analyzed using the dedicated function in Atlas.ti, in order to explore how different types of language-related comments intersect with author and reviewer characteristics, textual locations within the research article, and the perceived scope of the issues raised. Table 3 reports the raw co-occurrence counts across these dimensions.

However, because the author and reviewer groups are unevenly distributed in the corpus, raw co-occurrence counts are only partially informative when comparing inner-, outer- and expanding-circle categories. Expanding-circle authors constitute the largest author group, while the reviewer pool is more balanced but still unequal across Kachruvian

categories. For this reason, raw author- and reviewer-related co-occurrences were also normalized in relation to the number of articles in each author group and the number of reviewers in each reviewer group. Normalized values were calculated as co-occurrence rates per 100 articles for author categories and per 100 reviewers for reviewer categories. These values are reported in Table 4.

The normalized figures show that grammar-related comments are markedly more frequent for expanding-circle authors, with 67.6 co-occurrences per 100 articles, compared with 30.0 for outer-circle authors and 7.1 for inner-circle authors. This suggests that comments explicitly addressing grammar or general language review are disproportionately associated with manuscripts whose first or corresponding authors are institutionally located in expanding-circle contexts. However, other categories show a more distributed pattern. Organization-related comments are frequent across all author groups, while word-choice comments are proportionally highest for inner-circle authors. Similarly, comments on clarity are relatively frequent across inner-, outer- and expanding-circle authors. These findings suggest that reviewer attention to language cannot be interpreted solely as a response to presumed non-native English proficiency. Rather, several recurrent categories, especially organization, wording and clarity, appear to reflect broader expectations concerning genre competence, rhetorical effectiveness and disciplinary appropriateness. With respect to reviewer characteristics, normalized rates show that inner-circle reviewers produce the highest rates of several negative language-related comments, particularly word choice, organization, clarity, grammar, and hedging. Nevertheless, expanding-circle reviewers also contribute substantially to language-related feedback. This confirms that language-focused reviewing is not exclusively associated with inner-circle reviewers, even though inner-circle reviewers appear particularly active in formulating such comments in the present dataset. Taken together, these author- and reviewer-related patterns support an interpretation of language-focused peer review as associated not only with linguistic background, but also with researchers' academic and discursive socialization.

The raw co-occurrence table also reveals patterns concerning the textual location of language-related feedback. Negative comments are most frequently associated with the Discussion, Results, Methods and Introduction sections, particularly in relation to organization, clarity and wording. These are sections in which argumentation, interpretation, evidential positioning, and disciplinary framing are especially salient. Finally, most language-related comments are framed as minor rather than major issues, regardless of category. This is especially evident for grammar and wording, which are overwhelmingly classified as minor, while the distinction is less marked in high-frequency categories such as organization and clarity. Requests for clarification are exclusively coded as minor, further underscoring their dialogic rather than evaluative function.

Table 3. Raw co-occurrence counts between language-related themes and coded dimensions

Theme	Author inner	Author outer	Author expanding	Reviewer inner	Reviewer outer	Reviewer expanding	Section Title	Section Abstract	Section Introduction	Section Methods	Section Results	Section Discussion	Section Conclusion	Major	Minor
(N) Academic style	1	0	4	3	0	1	0	1	0	0	0	1	0	0	0
(N) Grammar	2	3	48	31	2	17	0	4	2	2	6	2	1	1	11
(N) Hedging	10	2	33	35	2	8	1	4	1	2	3	8	1	6	3
(N) Repetitions	16	4	40	33	9	14	2	5	14	4	9	11	0	4	3
(N) Organization	50	18	111	93	18	59	1	13	23	26	28	29	5	10	11
(N) Typos, spelling, formatting	11	6	33	23	2	20	0	6	2	4	9	7	0	1	8
(N) Unclear	34	11	76	83	10	24	4	15	12	14	11	8	2	11	12
(N) Acronyms, abbreviations	9	1	22	21	2	9	2	4	2	3	0	1	0	1	4
(N) Vague	5	2	5	8	2	2	0	3	1	0	0	3	0	0	0
(N) Word choice	58	14	102	111	17	40	11	18	23	16	15	10	4	7	37
(P) Clear	13	2	20	16	4	14	2	3	4	4	1	1	0	0	0
(P) Concise	3	0	2	3	1	1	1	2	0	0	0	0	0	0	0
(P) Hedging	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
(P) Organization	3	0	4	4	0	3	0	0	0	0	1	1	0	0	0
(P) Well-written	21	4	36	31	7	22	0	1	3	1	0	0	0	0	0

Table 4. Normalized co-occurrence rates by author and reviewer group, per 100 articles/reviewers

Theme	Author inner	Author outer	Author expanding	Reviewer inner	Reviewer outer	Reviewer expanding
(N) Academic style	3.6	0.0	5.6	3.1	0.0	0.9
(N) Grammar	7.1	30.0	67.6	31.6	7.1	15.5
(N) Hedging	35.7	20.0	46.5	35.7	7.1	7.3
(N) Repetitions	57.1	40.0	56.3	33.7	32.1	12.7
(N) Organization	178.6	180.0	156.3	94.9	64.3	53.6
(N) Typos, spelling, formatting	39.3	60.0	46.5	23.5	7.1	18.2
(N) Unclear	121.4	110.0	107.0	84.7	35.7	21.8
(N) Acronyms, abbreviations	32.1	10.0	31.0	21.4	7.1	8.2
(N) Vague	17.9	20.0	7.0	8.2	7.1	1.8
(N) Word choice	207.1	140.0	143.7	113.3	60.7	36.4
(P) Clear	46.4	20.0	28.2	16.3	14.3	12.7
(P) Concise	10.7	0.0	2.8	3.1	3.6	0.9
(P) Hedging	0.0	0.0	1.4	1.0	0.0	0.0
(P) Organization	10.7	0.0	5.6	4.1	0.0	2.7
(P) Well-written	75.0	40.0	50.7	31.6	25.0	20.0

Note. Values represent co-occurrence rates per 100 articles for author groups and per 100 reviewers for reviewer groups. They are not percentages, since a single article or reviewer may be associated with multiple language-related comments.

4. Discussion

Taken together, the findings presented in this paper reinforce existing ERPP scholarship highlighting the importance of genre awareness and rhetorical competence in navigating academic publishing. The prominence of comments dealing with structure, information sequencing, hedging and causal language suggests that reviewers place great emphasis on the discursive norms of the research article genre. In this sense, language-related feedback in peer review should not be understood as concerning only grammatical correctness or surface-level textual accuracy; rather, the findings indicate that reviewers for *BMJ Open* frequently use language-related comments to regulate the ways in which authors organize information, construct arguments, calibrate claims, and align their writing with the epistemic conventions of medical research articles.

This is especially evident in comments concerning hedging, boosting and causal language, which show that reviewers' language-related feedback participates in the regulation of epistemic responsibility. The findings suggest that in medical research writing, lexical choices implying causality and/or probability may alter the perceived strength of evidence provided in the manuscript. Thus, reviewers' requests to replace causal formulations with more cautious alternatives function as actual interventions into the epistemological positioning of the manuscript. From an EAP and ERPP perspective, this reinforces the idea that successful research publication depends not only on linguistic accuracy, but also on the authors' ability to align their claims with the study design, the evidential scope, and disciplinary expectations (Hyland, 2000). The medical context of the corpus is particularly significant in this respect. In medical academic discourse, questions of clarity, precision, and epistemic caution are not simply stylistic preferences: they are closely connected to the responsible communication of evidence (Ravinetto & Amir Singh, 2023). Overstated causal claims, ambiguous terminology, or insufficiently cautious conclusions may affect how research findings are interpreted by clinicians, journalists, policy makers, and patients. Reviewers' comments on wording and phrasing can therefore be interpreted as part of broader disciplinary concerns with risk and ambiguity, ultimately revolving around issues of trust, reliability, and the ethical circulation of medical research findings. In this sense, language-related peer review in *BMJ Open* appears to operate at the intersection between linguistic form, disciplinary norms, and health communication.

The frequent comments on organization and information sequencing also seem to support this interpretation: reviewers often indicate where information should appear, how sections should be ordered, and ultimately which rhetorical progression would better conform to the expectations of the research article genre. Thus, these comments implicitly and explicitly construct the medical research article as a highly conventionalized genre in which meaning depends both on what information is included and on where and how it is presented (Li & Ge, 2009). For this reason, such comments can be understood as comments about genre competence, showing that publishable academic writing requires authors to anticipate the readers' expectations, to guide their interpretation, and to organize knowledge according to recognizable disciplinary patterns.

What is more, these examples show how reviewers' feedback implicitly teaches academic writing by modelling

expected rhetorical structures, clarifying epistemic norms, and signaling what counts as “appropriate” or “standard” language within the discipline. In this sense, published OPR reports offer a valuable pedagogical resource for ERPP, ESP, and EAP research and instruction, providing insight into actual evaluative practices and expectations. Arguably, their pedagogical value lies in their authenticity: unlike invented materials or decontextualized lists of academic writing conventions, OPR reports reveal how such conventions are negotiated and enforced in real publication contexts. For ERPP and EAP pedagogy, these reports could thus be used to teach novice academic writers how reviewers formulate criticism, how to construct “clarity” in their writing, how to soften or strengthen their claims, how to express causal statements in relation to study design and scope. They could also help students and early-career researchers understand the rhetorical functions of the different sections of the academic article, and particularly the Introduction, Results, and Discussion sections (where reviewers in this corpus frequently commented on aspects of organization, wording, and epistemic stance) (Hyland, 2003; 2004). For ESP instruction, and especially for medical academic English, OPR reports may be particularly useful in showing how disciplinary and linguistic expertise intersect, as they demonstrate that medical academic writing involves not only accurate reporting of the data, but also careful management of certainty, implication, relevance, and limitation (see also Fiammenghi & Hartle, 2025 for an analysis of how these aspects are negotiated by experienced, non-native writers of academic medical English, too).

Furthermore, the OPR dimension has potential implications also for how reports may be written, as reviewers writing in an open review environment are likely conscious that they are potentially addressing not only the authors and editors, but also a wider audience of readers who may consult the review history after publication. Previous studies have shown that OPR does not negatively affect the quality of the peer review process (Wolfram, Wang, Hembree, & Park, 2020); the present study suggests that it may affect how reviewers use language to construct their own disciplinary identity. Indeed, this public or semi-public dimension may help explain the recurrent use of praise–criticism pairings identified in the corpus. Positive comments such as “well-written” frequently occur at the beginning of reports and are followed by more specific criticism or suggestions for revision. As discussed above, such comments may function less as precise linguistic evaluation than as face-saving acts, in line with previous work on politeness strategies in peer review discourse, even when it is occluded. In the context of OPR, however, these strategies may acquire an additional function: they allow reviewers to display fairness, professionalism and collegiality in a text that will become publicly visible.

Finally, studies into OPR practices can also inform debates on the role of English as “the default academic language” (Bocanegra-Valle, 2014) and the ideological and power struggles it implies. Indeed, as noted by Di Ferrante, Bernstein, & Gironzetti (2019), a “centripetal pull” toward English persists in academic publishing, evidenced by higher submission rates, citation counts, and the choices of authors who publish in English even when it is not their native tongue. The authors note that these trends are fueled by institutional pressures and career requirements that prioritize high-impact English publications over local language dissemination, and that even editorial commitment to multilingualism and to supporting non-English scholarship does little to change this direction. The findings of the present study contribute to this debate in a nuanced way. On the one hand, explicit native-speakerist comments are rare in the corpus: only two comments directly recommend revision by a native speaker of English. This suggests that language-related gatekeeping in the reports analyzed here does not usually operate through overt appeals to nativeness. On the other hand, the relative absence of explicit native-speakerism does not necessarily mean that linguistic inequality is absent. Rather, gatekeeping may operate through tacit expectations concerning what counts as “clear”, “concise”, appropriately hedged, logically organized, and “academically standard”. These expectations may be particularly demanding for authors working in contexts where English is not the dominant language, especially if such expectations remain implicit, or are formulated in the reports only briefly, through comments such as “unclear”, “not academic”, or “needs rephrasing/re-wording”. The data also shows that inner-circle authors also receive negative language-related comments, and that expanding-circle reviewers also actively produce such feedback. This supports an interpretation of peer review language practices in terms of academic and discursive socialization rather than linguistic background alone: ultimately, what reviewers evaluate is not English proficiency per se, but familiarity with the discursive conventions of publishable (medical) research writing. Nevertheless, because access to such conventions is often unevenly distributed across institutions, languages, and academic systems, these evaluative practices remain relevant to discussions of linguistic justice in global academic publishing.

In this sense, analyses of publicly available, transparent peer review reports can be used as pedagogical materials for novice academic English writers, and also constitute evidence of actual language practices in the global academia, and therefore have the potential to foster linguistic justice in global academic publishing.

4.1 Limitations and Conclusions

This study examined the language-related comments found in openly accessible peer review reports accompanying

109 research articles published in a full *BMJ Open* issue. The analysis shows that reviewers frequently comment on organizational clarity, rhetorical appropriateness and academic style, in addition to concerns about grammar and wording. The nature of the comments evidences that reviewers' expectations are grounded not solely in language proficiency but in familiarity with the genre-specific conventions of research writing.

These findings contribute to EAP and ERPP research by offering a systematic description of how reviewers articulate concerns about linguistic and rhetorical features in an open peer review environment. They also reveal the pedagogical potential of OPR materials: published review reports can help both instructors and learners better understand the discursive norms that guide academic evaluation.

Several limitations should be noted. First, because the corpus is limited to published articles, it does not account for papers rejected during review, and thus cannot determine whether language issues contribute to rejection. Open peer review cannot offer insight into the characteristics of rejected papers which remain confined within editorial offices. Second, authors' linguistic backgrounds were approximated through institutional affiliation rather than individual proficiency, but this has several important limitations: it only represents the environment in which these authors and reviewers were working at the time of writing/reviewing the article, but it may not represent the linguistic background in they were born or educated. Perhaps more importantly, it was not possible to assess whether they had benefited from formal academic writing instruction (although they were all experts in the medical, and therefore not in the language, field). Third, in cases where multiple authors were credited, only the first and corresponding author was analyzed, but this may have overlooked the active involvement of other authors in the writing process. As far as the findings are concerned, the most important limitations concern co-occurrence analysis: while it offers a useful overview of patterned associations among comment types, author and reviewer characteristics, and textual locations, it does not imply causality. Although normalized co-occurrence rates were calculated to reduce the effect of unequal group sizes among inner-, outer-, and expanding circle reviewers and authors, these figures should still be interpreted cautiously. The qualitative thematic analysis helped capture the sequential organization and rhetorical embedding of comments within peer review reports, such as the ways in which praise may be used as a face-saving strategy to mitigate criticism, but recurrent intersections would require triangulation with inferential statistics before stronger claims about group-based reviewing patterns could be made.

The analysis presented in the article could also be expanded in several ways. Future examinations will consider the authors' responses to reviewers, to assess whether their language-related comments were accepted, and to gauge to what extent their texts were modified to comply with the reviewers' requests. Furthermore, it will be necessary to expand the research to also include the possible use of Artificial Intelligence to carry out peer review and to draft peer review reports (if and where publishers allow).

Despite its limitations, the present study supports EAP and ERPP research emphasizing the linguistic demands placed on authors from a medico-scientific background, and suggests implications for academic writing pedagogy and interdisciplinary collaboration aimed to foster linguistic equity in global academic writing.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process:

During the preparation of this work the author used ChatGPT Version 5.1 (EDU license) for language editing, for clarity, and for stylistic refinement. This is the typical prompt used: "read this paragraph/sentence and (1) check it for grammatical accuracy (2) if necessary, highlight any unclear words/phrases/sentences and suggest possible revisions". After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

- Belcher, D. D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16, 1–22. <https://doi.org/10.1016/j.jslw.2006.12.001>
- Benos, D., Bashari, E., Chaves, J. M., Gaggar, A., Kapoor, N., LaFrance, M., ... Zotov, A. (2007). The ups and downs of peer review. *Advances in Physiological Education*, 31(2), 145–152. <https://doi.org/10.1152/advan.00104.2006>
- Black, N., van Rooyen, S., Godlee, F., Smith, R., & Evans, S. (1998). What makes a good reviewer and a good review for a general medical journal? *JAMA*, 280(3), 231–233. <https://doi.org/10.1001/jama.280.3.231>
- Bocanegra-Valle, A. (2014). 'English is my default academic language': Voices from LSP scholars publishing in a multilingual journal. *Journal of English for Academic Purposes*, 13, 65–77. <https://doi.org/10.1016/j.jeap.2013.10.010>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–

101. <https://doi.org/10.1191/1478088706qp063oa>
- Chapelle, F. H. (2014). The history and practice of peer review. *Groundwater*, 52(1), 1. <https://doi.org/10.1111/gwat.12139>
- Delgado, F. D., Garretson, G., & Delgado, F. A. (2019). The language of peer review reports on articles published in the BMJ, 2014-2017: An observational study. *Scientometrics*, 120, 1225–1235. <https://doi.org/10.1007/s11192-019-03160-6>.
- Di Ferrante, L., Bernstein, K. A., & Gironzetti, E. (2019). Towards decentering English: Practices and challenges of a multilingual academic journal. *Critical Multilingual Studies*, 7(1), 105–123. <https://doi.org/10.2458/cms.6569>
- Fiammenghi, C., & Hartle, S. (2025). Computer-assisted personalized instruction in Academic Medical English writing: A socio-material view. In *Advancing CALL: New research agendas-EUROCALL 2025 Short Papers* (pp. 551–558). <https://doi.org/10.4995/EuroCALL2025.2025.21190>
- Flowerdew, J., & Habibie, P. (2022). *Introducing English for Research Publication Purposes*. New York: Routledge. <https://doi.org/10.4324/9780429317798>
- Gosden, H. (2003). “Why not give us the full story?”: Functions of referees’ comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes*, 2(2), 87–101. [https://doi.org/10.1016/S1475-1585\(02\)00037-1](https://doi.org/10.1016/S1475-1585(02)00037-1)
- Groves, T. (2010). Is open peer review the fairest system? Yes. *BMJ*, 341, c6424. doi:<https://doi.org/10.1136/bmj.c6424>.
- Groves, T., & Loder, E. (2014). Prepublication histories and open peer review at the BMJ. *BMJ*, 349, g5394. <https://doi.org/10.1136/bmj.g5394>.
- Hartle, S. (2023). From learner corpus to data-driven learning (DDL) in EAP writing. Improving lexical usage in academic writing. *EuroAmerican Journal of Applied Linguistics and Languages*, 10(2), 9–31. <https://doi.org/10.21283/2376905X.1.10.2.2748>
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12(1), 17–29. [https://doi.org/10.1016/S1060-3743\(02\)00124-8](https://doi.org/10.1016/S1060-3743(02)00124-8)
- Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press. <https://doi.org/10.3998/mpub.23927>
- Jenkins, J., & Morán Panero, S. (2024). *Global Englishes: A resource book for students*. Routledge. <https://doi.org/10.4324/9781003027805>
- Kachru, B. (1985). Institutionalized second-language varieties. In S. Greenbaum (Ed.), *The English language today* (pp. 211–226). Oxford: Pergamon.
- Keserlioglu, K., Kilicoglu, H., & ter Riet, G. (2019). Impact of peer review on discussion of study limitations and strength of claims in randomized trial reports: A before and after study. *Research Integrity and Peer Review*, 4(19). <https://doi.org/10.1186/s41073-019-0078-2>
- Lee, K. P., Boyd, E. A., Holroyd-Leduc, J. M., Bacchetti, P., & Bero, L. A. (2006). Predictors of publication: Characteristics of submitted manuscripts associated with acceptance at major biomedical journals. *Medical Journal of Australia*, 184(12), 621–626. <https://doi.org/10.5694/j.1326-5377.2006.tb00418.x>
- Li, L.-J., & Ge, G.-C. (2009). Genre analysis: Structural and linguistic evolution of the English-medium medical research article (1985–2004). *English for Specific Purposes*, 28(2), 93–104. <https://doi.org/10.1016/j.esp.2008.12.004>
- Mungra, P., & Webber, P. (2010). Peer review process in medical research publications: Language and content comments. *English for Specific Purposes*, 29, 43–53. <https://doi.org/10.1016/j.esp.2009.07.002>
- Paltridge, B. (2017). *The discourse of peer review: Reviewing submissions to academic journals*. Sidney: Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-48736-0>
- Polka, J. K., Kiley, R., Konforti, B., Stern, B., & Vale, R. D. (2018). Publish peer reviews. *Nature*, 560, 545–547. <https://doi.org/10.1038/d41586-018-06032-w>.
- Räsänen, C. (1999). *The conference forum as a system of genres*. Göyrbpht: Acta Universitatis Gothoburgensis.
- Ravinetto, R., & Amir Singh, J. (2023). Responsible dissemination of health and medical research: some guidance

- points. *BMJ Evidence-Based Medicine*, 28(3), 144–147. <https://doi.org/10.1136/bmjebm-2022-111967>
- Rooyen, V., et al. (2010). Effect on peer review of telling reviewers that their signed reviews might be posted on the web: Randomised controlled trial. *BMJ*, 341, c5729. <https://doi.org/10.1136/bmj.c5729>.
- Ross-Hellauer, T. (2017). What is Open Peer Review? A Systematic Review. *F1000Research*, 6, 588. <https://doi.org/10.12688/f1000research.11369.2>
- Ross-Hellauer, T., Deppe, A., & Schmidt, B. (2017). Survey on Open Peer Review: Attitudes and Experience Amongst Editors, Authors and Reviewers. *Plos One*, 12, e0189311. <https://doi.org/10.1371/journal.pone.0189311>
- Ross-Hellauer, T., & Horbach, S. P. (2024). Additional experiments required: A scoping review of recent evidence on key aspects of Open Peer Review. *Research Evaluation*, 33, rvae004. <https://doi.org/10.1093/reseval/rvae004>
- Schriger, D. L., Sinha, R., Schroter, S., Liu, P. Y., & Altman, D. G. (2006). From submission to publication: A retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. *Annals of Emergency Medicine*, 48(6), 750–756. <https://doi.org/10.1016/j.annemergmed.2006.06.017>
- Squazzoni, F., Ahrweiler, P., Barros, T., Bianchi, F., Birukou, A., Blom, H. J., & Willis, M. (2020). Unlock ways to share data on peer review. *Nature*, 578(7796), 512–514. <https://doi.org/10.1038/d41586-020-00500-y>
- Wolfram, D., Wang, P., Hembree, A., & Park, H. (2020). Open peer review: promoting transparency in open science. *Scientometrics*, 125, 1033–1051. <https://doi.org/10.1007/s11192-020-03488-4>

Software

ATLAS.ti Scientific Software Development GmbH. (2023). ATLAS.ti Windows (version 25) [Qualitative data analysis software]. <https://atlasti.com>

Notes

Note 1. Retrievable at: <https://bmjopen.bmj.com/pages/reviewer-guidelines>. Last accessed 12th June 2026.

Author Statements

Author(s) declared the statement below:

Acknowledgments

Not applicable.

Author Contributions

Not applicable.

Consent for Publication

All authors have read and approved the published version of the manuscript.

Funding

This research received no external funding.

Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed Consent

Not applicable.

Ethics Approval

This article does not require ethical approval.

Data Availability Statement

Available upon request: The datasets generated and/or analyzed during the current study are available from the

corresponding author upon reasonable request.

Use of Artificial Intelligence

During the preparation of this work the author used ChatGPT Version 5.1 (EDU license) for language editing, for clarity, and for stylistic refinement. This is the typical prompt used: “read this paragraph/sentence and (1) check it for grammatical accuracy (2) if necessary, highlight any unclear words/phrases/sentences and suggest possible revisions”. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Supplementary Materials

No supplementary materials are available.

Journal Information

The Editorial Board declared the below statement on behalf of the journal and publisher.

Ethics Statement

The journal’s policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Preservation Policy

This journal is archived by Library and Archives Canada (LAC).

Open Access

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Please find details at <https://creativecommons.org/licenses/by/4.0>.

Copyright © The Author(s)

Copyright remains with the author(s). The author(s) grant the journal the right of first publication, and the article is published under the terms of the Creative Commons Attribution (CC BY 4.0) License.