"This article is interesting, however": exploring the language use in the peer review comment of articles published in the BMJ

Guangyao Zhang, Licheng Wang and Weixi Xie
WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology, Dalian, China
Furong Shang
College of Geography and Environment, Shandong Normal University, Jinan, China, and
Xinlu Xia, Chunlin Jiang and Xianwen Wang
WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology, Dalian, China

Received 27 June 2021
Revised 8 October 2021
2 November 2021
Accepted 8 November 2021

Abstract

Purpose – The purpose of this paper is to reveal a symbol – "however" that authors are very interested in, but few research studies pay attention to the existing literature. The authors aim to further insight its function.

Design/methodology/approach – In this research, the authors selected 3,329 valid comments on articles published in the British Medical Journal (BMJ) from 2015 to 2020 as the research objects. The authors showed the length distribution of reviewers' comments. In what follows, the authors analyzed the general distribution of words in comments and reviewer comments' position to understand reviewers' comments qualitatively in word dimension. Specially, the authors analyzed functions of "however" and "but", words that authors are most concerned with. In addition, the authors also discussed some factors, which may be related to "however," that reflect reviewers' praise through regression analysis.

Findings – The authors found that there are marked differences in the length of reviewers' comments under different review rounds. By mapping the reviewers' comments to different sections, the authors found that reviewers are deeply concerned to methods section. Adjectives and adverbs in comments written in different sections of the manuscripts also have different characteristics. The authors tried to interpret the turning function of "however" in scientific communication. Its frequency of use is related to reviewers' identities, specifically academic status. More precisely, junior researchers use "however" in praise more frequently than senior researchers do.

Research limitations/implications – The linguistic feature and function of "however" and "but" in the reviewers' comments of the rejected manuscripts may be different from accepted papers and also worth exploring. Regrettably, the authors cannot obtain the peer review comments of rejected manuscripts. This point may limit the conclusion of the investigation of this article.

Originality/value – Overall, the survey results revealed some language features of reviewers' comments, which could provide a basis of future endeavors for many reviewers in open peer review (OPR) field. Specially,

The authors would like to thank Yuqi Wang for the support provided in data analysis. The authors gratefully acknowledge the grant from the National Natural Science Foundation of China (71673038, 71974029). The authors thank two anonymous reviewers for their insightful comments and suggestions. The authors admit, to their interest, that this manuscript was received "however" in the review reports. And authors used the reviewer's comments "This article is interesting, however" to name this article. Thanks again to the reviewers.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Aslib Journal of Information Management

2050-3806
DOI 10.1108/AJIM-06-2021-0172

the authors also put forward an interesting symbol to examine the review comments, "however", for the first time.

Keywords Peer review, Reviewers' comments, Open peer review, Content analysis, However
Paper type Research paper

1. Introduction

Peer review, an effective method for the evaluation of scientific work, has a long history (Zuckerman and Merton, 1971). It has always been used as the "gold standard" to confirm the correctness of procedures and the rationality of results. At the same time, it provides the basis for the allocation of resources (Daryl and Edward, 1990). As the focus of scientists, peer review plays an important role in the quality control of research (Goodman *et al.*, 1994; Ware, 2008) and the communication of scholarly (Mulligan *et al.*, 2013).

Despite peer review being widely accepted, the academic community still has a lot to discuss regarding its downsides. On the one hand, the practice of peer review is controversial (Walker and da Silva, 2015) because of its limitations in stifling innovation (Horrobin, 2001; Spier, 2002; Brezis and Birukou, 2020), blocking publication (Armstrong, 1997; Benos *et al.*, 2007) and creating potential fairness and ethical risks in the process of peer review (Campanario, 1998; Budden *et al.*, 2008; Bornmann and Daniel, 2009; Emerson *et al.*, 2010). On the other hand, we still know very little about peer review (Jefferson *et al.*, 2002). Peer review is a black box for academia, for the content of peer review reports and generally remains known only to the authors and to the handling editors.

Under the circumstances of open science (Zhang *et al.*, 2021), we could get reliable access to peer review reports. We find *the British Medical Journal* (*BMJ*) which has narrowed the gap by providing a fully accurate open peer review database. So, in this article, we aim to understand reviewers' comments and extending previous research in depth by exploring referee reports on articles published in *BMJ* (Falk Delgado *et al.*, 2019). Although we can only obtain the review comments of accepted manuscripts, we hope to deconstruct the review reports step by step from the statistical description at the macro-level to the interpretation at the micro-level. This article is organized as follows. In the next section, we briefly review extant research studies which are related to the peer review process, open peer review and text analysis of peer review. A description of data sources and research methods follows additionally. Then we show the length distribution of reviewers' comments. In what follows, we analyze the general distribution of words in comments and reviewers' comments position to understand reviewers' comments qualitatively in word dimension. Specially, we discuss functions of "however" and "but", words that authors are most concerned with. As an important symbol in peer view, "however" is a word that attracts a big concern among authors who have received review comments. In this study, we aim to put forward this symbol which is very attractive and important but seldom pay attention to in the existing literature and further insight its function. This analysis helps us to understand the meaning of peer review, a special scientific communication situation. Furthermore, we find some factors which may be related to "however" reflect reviewers' praise. At the end of this article, we conclude with a discussion of suggestions and limitations.

2. Related works

2.1 Peer review process

Peer review is a process of quality control that involves the publications of science (Borgman and Furner, 2002; Rowland, 2002). It plays a significant role in academic communication (Mulligan *et al.*, 2013). Publishing science articles pays close attention to the peer review process (Rowland, 2002; Salem *et al.*, 2016), as it relates to the quality and credibility of

journals. In recent years, peer review as a social process has attracted research interests in the scientometrics community (Squazzoni *et al.*, 2017). Many researchers have emphasized effectiveness (Siler *et al.*, 2015; Jubb, 2016), reliability (Rennie, 2016) and bias (Seeber and Bacchelli, 2017) in the peer review process. Researchers can take various approaches or methods, such as building models (Ragone *et al.*, 2013; Brezis and Birukou, 2020) or conducting experiments (Dell'Anno *et al.*, 2020), to find problems in peer review process. The peer review process is a black box in former studies because peer review data are difficult to obtain for previous researchers. However, the opening of peer review reports provides a breakthrough point for drawing back the curtain. Based on open peer review reports on articles published in *BMJ*, we can better understand the process of peer review in science (Lee and Moher, 2017).

2.2 Open peer review

Traditional peer review is mainly based on closed peer review (CPR), which generally adopts a "single or double blind" review (Osterath, 2016). However, due to the attention of scholars and the efforts of journals, a large number of innovations have emerged in this field, especially with the advancement of the "open movement", an eye-catching form of OPR has appeared. Although OPR has not advanced to the point where we have a homogeneous theory (Ford, 2013), it can mitigate the aggressiveness and ideological bias of reviewers to a certain extent by increasing transparency. It can also help block "malicious" comments and control plagiarism (Mulligan *et al.*, 2013).

Controversies about merits and demerits of OPR still exist on this topic. For instance, supporters believe that OPR does not compromise the peer review process (Bravo *et al.*, 2019) and improve articles' citation counts (Zong *et al.*, 2020). However, some opponents hold that public intervention may increase the burden on reviewers (Callaway, 2016). Nevertheless, some journals which have been at the forefront of OPR have already begun to publish their referee reports. On May 22, 2019, *PloS* also announced that authors of all journals under its jurisdiction can choose to publish a peer review history together with the papers they have published. *PeerJ*, *the British Medical Journal, F1000Research*, *Nature Communications*, *Openreview.net* (Tran *et al.*, 2020) and many more pioneers have also started to open their peer review content in different implementations.

Due to various restrictions on the availability of peer review information in traditional peer review, studies on traditional peer review often qualitatively explain peer review behavior by journal editors from their own experience or rely on questionnaires to assess peer review behavior. However, with the rise of OPR, a large amount of data related to peer review has grabbed the attention of scholars (Hopewell *et al.*, 2014), enabling scientific researchers to understand more about review behaviors from more aspects. Analyzing referee reports is one of the most important observational perspectives.

2.3 Text analysis of peer review

The movement of opening referee reports, which helps to open the black box of peer review processes in science, is of great significance in interpreting the behavior of reviewers. With this background, open peer review texts have spawned a large number of studies. As a special kind of corpus, peer review reports, which contain abundant information about review behaviors, have attracted the attention of researchers in computer science. Some researchers have revealed the structure of peer reviews based on open peer review data. For example, Hua studied the content and structure of peer reviews under the argument mining framework (Hua *et al.*, 2019). Qin and Zhang explored the distribution of peer review comments in different structural functions of academic articles (Qin and Zhang, 2021). Meng proposed a method for building multi-level aspects of peer reviews for academic

articles, which can provide a well-rounded optimization direction for submitters (Meng *et al.*, 2021). Natural language processing and sentiment analysis are becoming more common in research of peer review texts than decades ago. Wang and Wan proposed a multiple instance learning network with a novel abstract-based memory mechanism (MILAM) to identify the sentences with positive and negative sentiment polarities from a peer review text (Wang and Wan, 2018). Chakraborty performed an exploratory analysis on a peer review dataset and investigated how the aspect sentiment distributions vary between accepted and rejected papers (Chakraborty *et al.*, 2020). There has even been a sentiment analysis program, PeerJudge, to detect praise and criticism in peer evaluations (Thelwall *et al.*, 2020). However, the latest research shows that manual analysis can provide more reliable results than algorithms (TextBlob and VADER) in the analysis of peer review reports (Luo *et al.*, 2021).

The peer review reports can be used as a supplement to articles that are being published. Researchers find the correlation between review texts and their mother articles by analyzing peer review content from F1000Research and reviewing reports using natural language processing (NLP) techniques (Rashidi *et al.*, 2020). Peer review reports can be used to predict the acceptance or rejection of manuscripts from missing reviewers (Ghosal *et al.*, 2019). Several studies have shown that the quality of reviews can be quantified or evaluated through Zipf's law (Ausloos *et al.*, 2016) or other metrics, such as the number of distinct words, tone and content (Yadav and Gehringer, 2016). What is more, there are also several related works about quantitative linguistic analysis of research grant evaluation. Kaatz carried out a quantitative linguistic analysis of National Institutes of Health R01 application critiques from investigators at one institution. They found that critiques of funded R01s contain more positive evaluation and (Kaatz *et al.*, 2015) fewer negative evaluation words than critiques of unfunded applications. In later studies, researchers have identified gender bias in peer review by showing that reviewers may score male investigators more competitively than females by text mining (Magua *et al.*, 2017).

For another, researchers can find interesting phenomena and get several enlightenments from reviewing texts from a bibliometric perspective. As for review length, some studies have indicated that comments in public peer review are much longer than those in CPR (Bornmann *et al.*, 2012). Casnici considered the length of the review text to measure the quality of reviews (Casnici *et al.*, 2017). The *Global State of Peer Review Report* (Publons, 2018), which mentioned the length distribution of reviewers' referee reports in different countries, suggested that the length of comments should be regarded as a weak variable of review quality. As to comment styles and language features, reviewer recommendation has the greatest influence on linguistic characteristics (and length) of referee reports (Buljan *et al.*, 2020). Besselaar analyzed the linguistic characteristics of review comments for European Research Council Starting Grants, and they found that negative linguistic categories in review comments have a strong effect on the grant scores (van den Besselaar *et al.*, 2018). This may result in low predictive validity and in bias. The features of peer review text may be related to the identity of reviewers. For instance, reviewers from established economies with a more comprehensive peer review mechanism may experience review training (Xu *et al.*, 2019). Gender effects in peer review have aroused great interest of researchers (Marsh *et al.*, 2009). Female reviewers are often more cooperative, focusing on the establishment and maintenance of good relationships (Maltz and Borker, 1983; Tannen, 1992). Differences are also found in the relationship between the academic status of reviewers and the recommendations they made. Younger reviewers provide more positive and objective recommendations, and male reviewers tend to write more constructive reports (Bravo *et al.*, 2019). Senior researchers are harsher in their comments than junior researchers (Casnici *et al.*, 2017).

3. Method and data

Compared with other journals, *BMJ* implements a full OPR strategy after a series of experiments in 1999. We can get various peer review information from articles, which have been accepted from BMJ.com, including reviewers' comments, reviewers' names, job titles and institutions. As Figure 1 shows, research papers submitted to *BMJ* after September 2014 usually have their prepublication history posted on the BMJ.com after being accepted (actually, papers published after March 2015 have published their prepublication history), including authors' manuscripts, reviewers' comments and authors' responses. All of these records are presented in PDFs, which can be downloaded at BMJ.com (Groves and Loder, 2014). We collected 3,917 PDFs from March 2015 to April 2020 at BMJ.com. After sorting out these documents, we got a total of 690 articles, with the number of review rounds from 1 to 6. Their distribution is shown in Table 1. The vast majority of hired manuscripts were reviewed within one or two rounds (accounted for 86.4%). Only a few manuscripts were counted over four rounds (accounted for 3.6%). We supplemented the document types of papers from Web of Science, including 563 articles, 92 reviews and 35 editorial materials. In addition, the paper whose DOI is 10.1136/bmj.h3147 has no document types in Web of Science while PubMed marked it as "Randomized Controlled Trial." So, we classified it as "Article."

We then used Python to extract what we need from referee reports, including reviewers' comments, their names, job titles and institutions. In this study, we mainly used the reviewers' comments of 690 papers (a total of 3,329 valid comments are extracted) as the research test bed. We employed a gender inference system named AMiner Gender Prediction

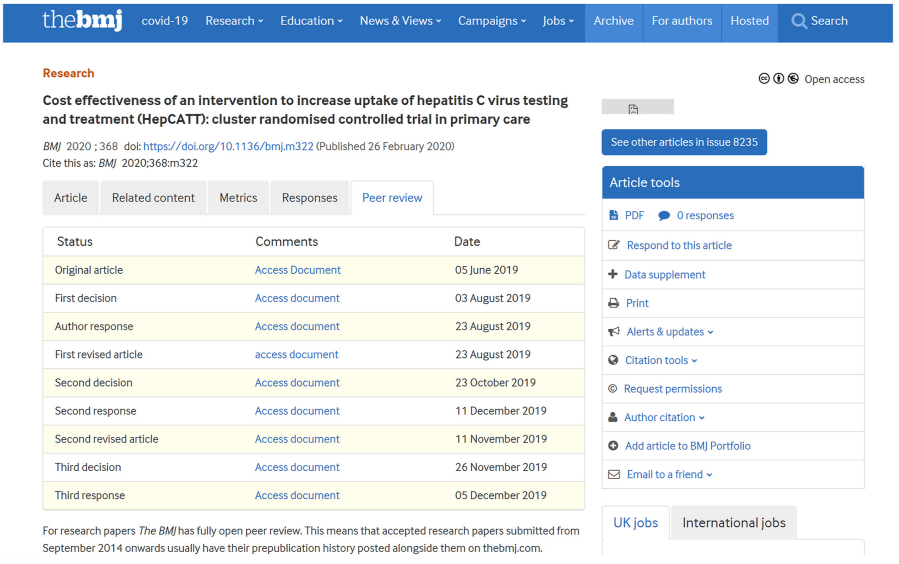

Figure 1. Prepublication history of an article from BMJ.com (https://www.BMJ.com/content/368/bmj.m322/peer-review)

Round				Year				Document types	
Round	Num	Round	Num	Year	Num	Year	Num	Document types	Num
1	301	4	18	2015	104	2018	140	Article	563
2	294	5	6	2016	140	2019	145	Review	92
3	70	6	1	2017	126	2020	35	Editorial material	35

Table 1. The quantity distribution of manuscripts in round/year/document types

API (https://www.aminer.cn/gender) that uses three gender inference methods (*Face recognition (FR)*, *Facebook-generated name list (FGNL)* and *web-based gender predictor (WebGP)*) with a specific *voting model (Final)* to identify reviewers' gender (Tang *et al.*, 2011; Gu *et al.*, 2016). The accuracy of gender identifying achieved in *FGNL* and *WebGP* are around 95.2 and 93.44%. Finally, this API used *voting model (Final)* selects the gender value with more votes as the final result. We divided the reviewers' academic into two categories: senior researchers (mainly including professors and associate professors) and intermediate/junior researchers (mainly including assistant professor, postdoc and research fellow). We determined countries where the reviewers work according to their institutions and divided the countries' economic backgrounds into two categories: established and emerging economies, according to *the Human Development Report* (UNDP, 2010).

A Poisson regression was considered for analyzing the factors that may be related to "however" that reflected reviewers' praise because the dependent variable was measured as a count variable. Although the dependent variable did not indicate overdispersion, as its variance (0.20) was approximately equal to its mean (0.19). We also supplemented results of negative binomial regression to enhance the robustness of the results.

4. Results

4.1 An overview of reviewers' comments

We regarded words as the basic unit of length and calculating the length of each comment. The average length of reviewers' comments was 542 (words). The median length was 445, which was higher than the average length of Publons, 391 (Publons, 2018). On the whole, reviewers' comments presented relatively long, and there was a strong tailing phenomenon. So, we deleted outliers with excessive length (length >2098, 99th percentile), and comments' length distribution is shown in Figure 2. Reviewers' comments had different characteristics under different rounds. We counted up the length of reviewers' comments under different review rounds and then calculated the median in groups. We excluded one paper with six

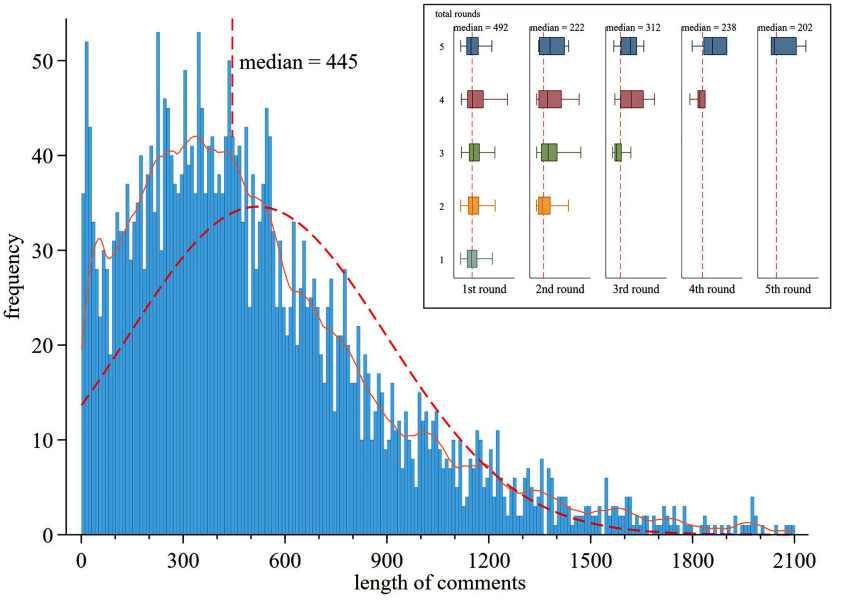

Figure 2.
Length distribution of reviewers' comments

rounds of review in the dataset because this sample size was too small. It can be seen from Figure 2 that reviewers wrote comments of different lengths in different rounds, and the length of the first round of peer reviewers' comments was often longer than the median. The average length of the first round of reviewers' comments was 595, and the median was 492, which was higher than the average level of reviewers' comments (542). Reviewers may prefer to put forward questions or suggestions in the first round of peer review. But in the final review process, reviewers may express their views briefly.

Commenting is a kind of social practice. Reviewers' views and attitudes can be reflected in various words and grammatical means. It is therefore worth analyzing words that reviewers frequently use and their distribution. After calculating the length of comments, we made an in-depth analysis of the word dimension. Because great heterogeneity (such as great differences in length and content) may be existing between different rounds in peer review comments, we analyzed the whole thing with only one round of comments (301 manuscripts, including 1074 comments). It is worth noting that we deleted a comment in the subsequent analysis because the content of this comment is "Reviewer 1 declared a disqualifying competing interest." We split sentences and identified parts of words using Python NLTK module (Bird, 2006). Since the evaluation function of comments is mainly carried out with adjectives (Biber *et al.*, 1999; Martin and White, 2007) and adverbs (Quirk and Greenbaum, 1973; Quirk *et al.*, 1985), we mainly analyzed these two kinds of words. After further cleaning segmentation results, we deleted "table," which is also an adjective but often refers to noun in comments from results. Overall, 85,273 words were obtained, and 4,219 words were unique.

Zipf pointed out that in any language, the influence of high frequency words is relatively limited. Among the top 20 words, the vast majority did not have practical meaning (including emotional and subjective tendency), which also included words "also," "clinical," etc. These words conformed with the "least effort" rule, so we delete them in Table 2. In addition, positive words, such as "important," "well," etc., often appeared in the first five commonly used words so does "however." This was consistent with traditional cognition, that is, reviewers often praise articles and then put forward questions or suggestions.

Table 2. Distribution of adjectives and adverbs of reviewers' comments

Order	Word	Num	Coverage	Intensity	Word 1	Word 2	Word 3
1/4	Important	1176	58.38%	1.88	Paper	Patients	Risk
2/8	Well	960	62.66%	1.43	Written	Paper	Results
3/9	However	788	43.76%	1.68	Risk	Patients	Results
4/11	Different	767	42.18%	1.69	*But*	Risk	Studies
5/14	Clear	675	36.87%	1.70	*But*	Data	Results
6/18	General	584	35.38%	1.54	Work	Population	Patients
7/21	Primary	493	24.49%	1.87	Outcome	Analysis	Patients
8/22	Major	492	29.61%	1.55	Risk	Paper	Data
9/24	Large	477	35.57%	1.25	Data	Number	Cohort
10/25	Overall	469	31.75%	1.38	Design	*But*	Well
11/26	Likely	456	26.16%	1.62	Patients	Risk	*But*
12/27	Main	435	36.96%	1.10	Results	*But*	Patients
13/28	Possible	434	28.31%	1.43	*But*	Data	Risk
14/31	Clearly	420	28.49%	1.37	Research	Question	Results
15/32	Potential	409	26.63%	1.43	Data	Bias	Risk
16/33	Specific	408	28.86%	1.32	Comments	*But*	Data
17/34	Further	407	26.16%	1.45	*But*	Discussion	Patients
18/35	Relevant	405	27.09%	1.39	References	Patients	Paper
19/36	Significant	402	25.88%	1.45	Statistically	Results	Risk
20/38	Better	392	24.77%	1.47	Patients	*But*	Results

We calculated the coverage and intensity of each word. Order is the order of the number of a word, where the number before "/" is the order of a word in adjectives and adverbs, and the number after "/" is the order of a word in all words. Coverage is measured by the proportion of the number of comments containing the word to the total, and intensity is measured by the frequency of the word divided by the number of comments containing the word. The formula is defined as (4.1) and (4.2). In terms of coverage, words including "well," "important" and "however" had the highest coverage rates, while in terms of intensity, words including "important," "clear" and "however" all exceeded 1.6, which means they were used nearly twice in each comment. We calculated the number of other words in the same sentence and kept the first three, which were called word 1, word 2 and word 3. Since the words "authors" and "study" appeared 19 times and 17 times in word 1–word 3, they were deleted from Table 2 to avoid redundancy. As shown in Table 2, reviewers seem to pay more attention to research methodology and results, such as "results," "risk," "data" and "patients." Interestingly, we also found nine"but" in this section, so we also listed them in italics.

$$Coverage = \frac{The\ \ number\ \ of\ \ comments\ \ containing\ \ the\ \ word}{The\ \ number\ \ of\ \ all\ \ comments} \tag{4.1}$$

$$Intensity = \frac{The\ \ number\ \ of\ \ the\ \ word\ \ in\ \ all\ \ comments}{The\ \ number\ \ of\ \ comments\ \ containing\ \ the\ \ word} \tag{4.2}$$

4.2 Analysis of reviewers' comments position

It is important for us to understand the function of comments and how an author reviews a manuscript. Especially, the language style across different sections of articles is one of the concerns of researchers (Hartley *et al.*, 2003). Similarly, we hope to explore these language traits qualitatively across different sections in reviewers' comments. We mapped 1073 reviewers' comments manually to different sections of the manuscript. These comments were divided into seven sections: overall, abstract, introduction, methods, results, discussion and others. Overall is a total evaluation of an article, and others refers to unclassified content. We complied with the following rules in section classification. First, several reviewers clearly pointed out the corresponding position of comments in the manuscript in 217 comments. In this case, we classified sections according to these positions' information. Second, reviewers repeated original paragraphs or sentences in manuscript or gave line numbers in manuscript. So, we can determine the corresponding position of the original text according to reviewers' comments. Third, we classified sections according to feature words. For example, we recognized introduction by "background," "literature review," recognized methods by "method(s)," "data," "experiment," recognized results by "result(s)," "measurement(s)," and recognized discussion by "discussion," "conclusion." Besides, we put the comments that hard-to-divide into "other". Judging from the proportion of different sections in comments, almost all comments evaluated articles as a whole (accounting for about 98.88%); additionally, our classification results were consistent with Qin's, that is, reviewers highly value materials, methods and results (Qin and Zhang, 2021). More comments evaluated methods, results and discussion, accounting for about 67.13, 60.34 and 52.05%, while a few comments included an abstract section and an introduction section, which only appeared in about 20% of the comments. In addition, 66 comments supplemented the references of articles, accounting for 6.15% of all manuscripts.

Reviewers listed their suggestions or questions in comments one by one. We defined each suggestion or question as an "opinion" and calculated the length of comments, number of opinions in various sections and average length of opinions (the ratio of comment length to number of opinions in each article). The calculated mean and median are presented in Table 3. The methods section was found to the maximum regardless of length, number of opinions

and average length of opinions. This result shows that the methods section is the most concerning section of reviewers and the most conspicuous section to reveal some problems. Although the length of results and discussion was lower than overall, the number of contained opinions were second only to methods. For the overall section, reviewers give a long overall evaluation of each article no matter what attitude they hold, so its length was second only to methods. But the number and average length of contained opinions were relatively short and overall did not undertake specific evaluation functions. The length and average length of opinions of abstract section, introduction section and other section were relatively short.

We list high-frequency adjectives and adverbs contained in the four sections of overall (in blue), methods (in green), results (in yellow) and discussion (in pink) in Table 4. By doing this, we found that high-frequency words in different sections varied distinctly. "Important" and "however" were in the first five words of the four sections. "Important," "well" and "interesting" often appeared in the overall and evaluated articles as a whole, while "different," "however" and "clear" ranked higher in methods and results. "Possible," "further" and

		Length		Number		Average length	
Section	Ratio	Med	Mean	Med	Mean	Med	Mean
Overall	98.98%	128	185.65	2	2.22	73	86.04
Abstract	20.20%	40	60.32	1	1.55	19	35.24
Introduction	22.35%	67	90.80	1	1.84	37	59.87
Methods	67.13%	143	192.77	2	3.13	88	134.00
Results	60.34%	116	164.72	2	2.73	70	110.79
Discussion	52.05%	108	146.09	2	2.27	65	95.60
Other	44.60%	66	98.44	2	2.62	35	60.01

Table 3. Summary statistics of different comments sections

order	word	num	coverage	intensity	word	num	coverage	intensity
1	important	689	42.27%	1.52	different	175	12.20%	1.34
2	well	598	50.09%	1.11	however	170	13.04%	1.21
3	general	351	22.07%	1.48	clear	163	10.71%	1.42
4	however	313	24.30%	1.20	important	144	10.43%	1.29
5	overall	254	20.11%	1.18	primary	132	8.38%	1.47
6	major	234	16.95%	1.29	statistical	130	9.68%	1.25
7	interesting	231	23.28%	0.92	likely	130	8.38%	1.44
8	clear	212	14.99%	1.32	well	124	10.99%	1.05
9	relevant	207	15.08%	1.28	possible	122	9.03%	1.26
10	clearly	198	15.74%	1.17	main	97	9.96%	0.91
1	different	187	12.66%	1.38	important	118	9.03%	1.22
2	however	145	9.50%	1.42	different	104	8.29%	1.17
3	important	138	9.87%	1.30	possible	94	7.26%	1.21
4	clear	118	9.03%	1.22	however	92	7.26%	1.18
5	primary	103	5.77%	1.66	clear	85	6.98%	1.13
6	well	102	8.75%	1.09	likely	83	6.70%	1.15
7	likely	100	6.61%	1.41	further	82	5.96%	1.28
8	main	100	9.59%	0.97	potential	79	6.33%	1.16
9	significant	97	7.26%	1.24	well	71	7.17%	0.92
10	useful	80	6.42%	1.16	current	67	5.68%	1.10

Table 4. Distribution of adjectives and adverbs in different sections of reviewers' comments (overall/methods/results/discussion)

"potential" ranked higher in the discussion. We also put the distribution of verbs and nouns in appendix.

4.3 Analysis of "however" and "but"

As we all know, authors are worried about seeing "however" in reviewers' comments because reviewers often praise manuscripts out of politeness before "however" and then put forward a lot of problems after this word. Some reviewers even reject manuscripts by using "however." In both Table 2 and Table 4, we can point out that "however" appears frequently in reviewers' comments. It seems that few researchers have paid attention to the meaning of this word in reviewers' comments. So, what do reviewers talk about when they talk about "however"? Of 470 comments, specifically, 774 sentences had the word "however." First, we confirmed the position of "however" in comments. Second, we took the previous sentence and the next sentence containing the word "however" as context and classified them manually in terms of context. We show the labels and definitions in Table 5 and the turning function of "however" in Figure 3. As shown in Figure 3, of the 774 sentences, 697 "however" (accounting for 90.05%) expressed the meaning of turn. After stating manuscript content, reviewers often express their doubts about the research content or put forward suggestions for improvement, accounting for 59.82%, or, after praising researchers, they put forward different views (26.10%), and a few reviewers (0.52%) make decisions to refuse this manuscript after praising. After pointing out the problems or questioning the results, some reviewers think that problems have little impact on the research and may even think that the manuscript can be published. "However" is also used to compare the views of reviewers and authors or to

	Label	Definition
Before however	Reviewers' praise	Reviewers expressed their praise for the manuscript and the authors
	Manuscript content	Reviewers stated the contents of the manuscript
	Questions or mistakes	Reviewers put forward suggestions or questions regarding the manuscript
	Suggestions for improvement	Reviewers put forward suggestions for improving the manuscript
	Reviewers' views	Reviewers express different views than authors
	Expected results	Reviewers put forward their expected research results
After however	Questions or mistakes	Reviewer pointed out mistakes in the manuscript or asked questions
	Suggestions for improvement	Reviewers put forward suggestions for improving the manuscript
	Questions (or mistakes) and suggestions	Reviewers put forward questions and suggestions regarding the manuscript
	Reviewers' disagreements	Reviewers put forward views different from those of the authors
	Questions do not harm the value of research	Although there are some minor problems with the manuscript, reviewers believe that these problems will have no negative impact on the manuscript
	The value of research	Reviewers highlighted the value of the manuscript
	Actual results	Reviewers pointed out actual research results in the manuscript
	Rejection	Reviewers think that this article should be rejected
	Authors' views	Reviewers quoted the author's views briefly
	Can be published	Reviewers agreed that the manuscript could be published
	Author decide whether or not to accept	Although reviewers expressed their suggestions, authors can decide whether to accept these suggestions or not

Table 5.
The labels and definitions of the sentence before and after "however"

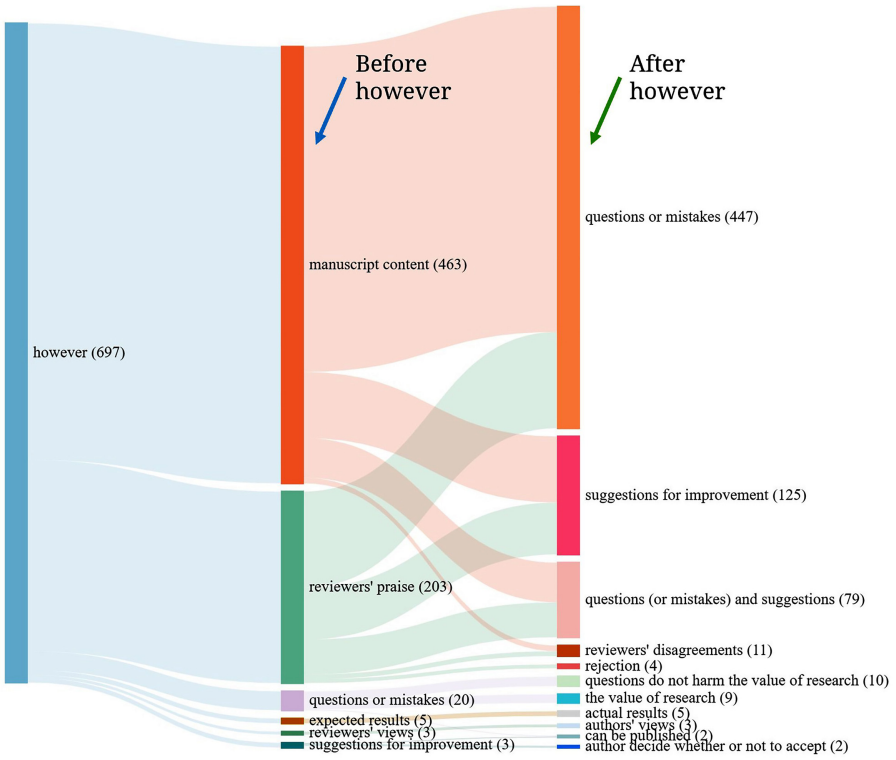

Figure 3.
Turning function distribution of "however"

compare the expected results with the actual results. Usually, "however" means "but" or "no matter how," but a subsequent sentence has different explanations, for example, turning function in authors' manuscripts (38), linking function (linking context) (33), asking questions (3), research background (1), irrelevant to research content (1) and reviewers' views (1).

We also observed the word "but" in Table 2. What is the difference between the functions of "but" and "however"? Of 800 comments, specifically, 1947 sentences had "but." We treated "but" the same way as "however." As shown in Table 6, we sorted out 1947 "but" in reviewers' comments and found that more than half of the "but" did not have reviewers' turning views. In total, 889 "but" (accounting for 45.66%) were turning points in the research content, which did not involve the reviewers' views. "However" is usually used to connect two sentences; "but" often appears in one sentence, and the function of "but", which is usually used as a

Function	Number	Ratio	Function	Number	Ratio
Manuscript content	889	45.66%	Comparison of two results	27	1.39%
No turning function	125	6.42%	Good job – poor job	23	1.18%
Content – questions	567	29.12%	Emphasize the key points	17	0.87%
Content – suggestions	119	6.11%	The change of viewpoints	3	0.15%
Content – praise	13	0.67%	Can be published – rejection	2	0.10%
Praise – questions	120	6.16%	Questions – praise	1	0.05%
Praise – suggestions	41	2.11%			

Table 6.
Function distribution of "but"

conjunction, is similar to "and." A total of 125 "but" (accounting for 6.42%) did not indicate turn. Among sentences which mentioned manuscript content before "but", 567 "but" (29.12%) pointed out problems in manuscripts, 119 "but" (6.11%) made suggestions after "but" and 13 "but" (0.67%) praised manuscripts after "but." In total, 161 sentences praised manuscripts before the "but", among which 120 "but" (6.16%) pointed out problems in manuscripts, and 41 "but" (2.11%) put forward suggestions for improvement. In addition, the "but" also appeared to be the difference between results and the emphasis of the research focus, but its proportion was only 3.75%.

4.4 What factors are related to "however" that reflect reviewers' praise?

We hope to better understand the meaning of "however" in scientific communication, specifically those that expresses praise before propounding suggestions or questions. Could these "fine comments" be related to the identity of reviewers? Is the frequency of "however" used by reviewers related to countries' economic backgrounds where reviewers work, gender and academic status? We added economic background, gender, academic status and length of review to the regression and controlled for the length of comments. The regression model is as follows:

$Y_i = X_i + \varepsilon_i$ where Y represents the frequency of "however" in praise, X_i ($X_i = [economic\ background, gender, academic\ status, length]^T$) is a vector of characteristics related to Y_i and ε_i is an error term. Table 7 presents variables in this section.

In our definition, male = 0, female = 1; emerging economies = 0, established economies = 1; senior = 0 and intermediate/junior = 1.

Table 8 provides the results of regression model. The benchmark models included Poisson model and Nbreg model. The log likelihood value of Nbreg model improved more slightly

Table 7. Summary of descriptive statistics

Variable	Economic background Categories	No	Gender Categories	No	Academic status Categories	No
	Developed	998	Male	383	Senior	507
	Developing	57	Female	667	Intermediate/junior	472
	Unknown	18	Unknown	23	Others	94

Note(s): Others including patients and public reviewers and the reviewers who do not have titles. We excluded a comment that involved a disqualifying competing interest

Table 8. Regression results

	Benchmark models		Excluding physicians	
	Poisson	Nbreg	Poisson	Nbreg
Economic background	0.204 (0.38)	0.202 (0.38)	0.114 (0.40)	0.122 (0.40)
Gender	0.017 (0.16)	0.021 (0.16)	0.025 (0.17)	0.032 (0.17)
Academic status	0.342** (0.15)	0.340** (0.15)	0.398** (0.16)	0.393** (0.16)
Length	0.001*** (0.00)	0.001*** (0.00)	0.001*** (0.00)	0.001*** (0.00)
Cons	−2.458*** (0.40)	−2.465*** (0.41)	−2.401*** (0.42)	−2.420*** (0.42)
Ln alpha		−1.494 (1.14)		−1.230 (0.95)
N	962	962	841	841
Log likelihood	−492.19	−491.76	−431.40	−430.76
Akaike information criterion (AIC)	994.38	995.51	872.80	873.52
Bayesian information criterion (BIC)	1018.73	1024.73	896.48	901.93

Note(s): Standard errors are in parentheses. $^{**}p < 0.05$; $^{***}p < 0.01$

than the value of Poisson model. Academic status was significantly associated with the frequency of "however" ($p < 0.05$). That means intermediate or junior researchers use more "however" than senior researchers. Although effects of gender and reviewers' economic background were not significant, the direction of their coefficient was positive. Some reviews, such as physicians, cannot be classified in these senior or intermediate/junior reviewers. Therefore, we excluded physicians, doctors and senior physicians from our regression model. The results were similar to benchmark models. In addition, the zero-inflated Poisson regression and zero-inflated negative binomial regression were consistent with benchmark models. Due to layout limitations, the specific inspection processes are not listed one by one.

5. Discussion

This study, based on peer review comments from articles which have been accepted, contributes to revealing the meaning of the term "however" that authors are very interested in, but few research studies pay attention to the existing literature and further revealing its function. "However" plays an interesting role in comments, even peer review. It has high frequencies in different sections. What is more, the information it conveys is often more concerning to the author. After a comparative analysis of "however" and "but," we found that there are differences in function and sentence structure between the two groups. "However" often connects two sentences, while "but" is often used in one sentence. The word "however" in sentences often has a turning meaning, but the "but" does not. We also tried to interpret the significance of "however" in scientific communication. Its frequency of use is related to reviewers' identities, specifically academic status. More precisely, junior researchers use "however" in praise more frequently than senior researchers do. The politeness and goodwill embodied in "however" exudes humanistic care in science to a certain extent. Admittedly, rejection is inevitable and dispiriting during the process of peer review. But the kindness of reviewers may save authors from the abyss, especially among young researchers who have just embarked on the academic road. Moreover, the manner of reviewers is also related to the credibility of the journal (Rockwell, 2005). The respect of reviewers is also a valuable part of comments (Ward *et al.*, 2015).

We found that there are marked differences in the length of reviewers' comments under different review rounds. With the advancement of peer review rounds, the length of reviewers' comments has shown a downward trend. So, the differences of different rounds should be considered when studying reviewers' comments. Considering that the peer review comments we used were selected from articles published in the BMJ, the possible explanation was that authors accepted proposed changes in manuscripts and increase agreement with reviewers in communications. In addition, we analyzed the general distribution of words in comments and reviewers' comments position to understand reviewers' comments qualitatively in word dimension. We found that reviewers attached importance to methods and results in manuscripts. It is important for authors to ensure the reliability of methods and the accuracy of results in their research studies.

Publication should be regarded as the beginning, not the end, of the peer review process, and the reviewers' comments will become a part of scientists' contributions to the field, which will be recognized and disseminated by academic circles. In many countries, such as the USA, reviewing a manuscript is regarded as a significant contribution, even comparable to publishing a paper; however, in some countries, such as China, the significance of peer review is seriously underestimated. As a kind of unpaid contribution, it is of great significance to recognize the contribution of reviewers from the perspective of reviewers' comments to solve the "reviewer crisis" (Arns, 2014; Xu *et al.*, 2019). For journals, the analysis of reviewers' comments can also help inexperienced researchers to learn how to review manuscripts more effectively.

Finally, we acknowledge that our research has several limitations. The conclusion of this study was that preliminary exploration has certain value, but the open peer review data are still limited to the articles that have been accepted, which may limit the extrapolation of the conclusions. We used manual classification to ensure as much accuracy as possible from the qualitative perspective for the lack of suitable algorithms to analyze some interesting points on the micro-level. However, machine learning methods are needed when mining large-scale peer review reports. In addition, the referee reports received by rejected manuscripts may have more research value. Especially, the linguistic feature and function of "however" and "but" in the reviewers' comments of the rejected manuscripts may be different from accepted papers and also worth exploring. Sometime, authors concerned the "however" and "but" word in the reviewers' comments for the rejected manuscript. But these analyses need further data opening. According to our survey, in 2021, *eLife* has started to make peer review reports available to the public, whether or not a manuscript is accepted for publication. We hope more and more journal follow this initiative. The reviewer information and authors' reactions in referee reports is also worth analyzing, so we will continue to explore features of reviewers in the future.

6. Conclusion

The increasingly abundant review data provided by OPR are of great value for understanding the peer review process and broadening the horizon of quantitative social science. In this study, we analyzed the reviewers' comments on articles published in *BMJ*. By mapping the reviewers' comments to different sections, reviewers find more problems in the methods section, and it is easier to write longer comments on methods. Adjectives and adverbs in comments written in different sections of the manuscripts also have different characteristics. Specially, we revealed an interesting symbol - "however" and discussed some factors which may be related to "however" reflect reviewers' praise through regression analysis.

References

Armstrong, J.S. (1997), "Peer review for journals: evidence on quality control, fairness, and innovation", *Science and Engineering Ethics*, Vol. 3 No. 1, pp. 63-84.

Arns, M. (2014), "Open access is tiring out peer reviewers", *Nature*, Vol. 515 No. 7528, p. 467.

Ausloos, M., Nedic, O., Fronczak, A. and Fronczak, P. (2016), "Quantifying the quality of peer reviewers through Zipf's law", *Scientometrics*, Vol. 106 No. 1, pp. 347-368.

Benos, D.J., Bashari, E., Chaves, J.M., Gaggar, A., Kapoor, N., LaFrance, M., Mans, R., Mayhew, D., McGowan, S., Polter, A., Qadri, Y., Sarfare, S., Schultz, K., Splittgerber, R., Stephenson, J., Tower, C., Walton, R.G. and Zotov, A. (2007), "The ups and downs of peer review", *Advances in Physiology Education*, Vol. 31 No. 2, pp. 145-152.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999), *Longman Grammar of Spoken and Written English*, Longman, London.

Bird, S. (2006), "NLTK: the natural language toolkit", *Proceedings of the COLING/ACL on Interactive Presentation Sessions*, Association for Computational Linguistics (COLING-ACL '06), Sydney, pp. 69-72.

Borgman, C.L. and Furner, J. (2002), "Scholarly communication and bibliometrics", *Annual Review of Information Science and Technology*, Vol. 36 No. 1, pp. 2-72.

Bornmann, L. and Daniel, H.-D. (2009), "Reviewer and editor biases in journal peer review: an investigation of manuscript refereeing at Angewandte Chemie International Edition", *Research Evaluation*, Vol. 18 No. 4, pp. 262-272.

Bornmann, L., Wolf, M. and Daniel, H.-D. (2012), "Closed versus open reviewing of journal manuscripts: how far do comments differ in language use?", *Scientometrics*, Vol. 91 No. 3, pp. 843-856.

Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B. and Squazzoni, F. (2019), "The effect of publishing peer review reports on referee behavior in five scholarly journals", *Nature Communications*, Vol. 10 No. 1, p. 322.

Brezis, E.S. and Birukou, A. (2020), "Arbitrariness in the peer review process", *Scientometrics*, Vol. 123 No. 1, pp. 393-411.

Budden, A.E., Tregenza, T., Aarssen, L.W., Koricheva, J., Leimu, R. and Lortie, C.J. (2008), "Double-blind review favours increased representation of female authors", *Trends in Ecology and Evolution*, Vol. 23 No. 1, pp. 4-6.

Buljan, I., Garcia-Costa, D., Grimaldo, F., Squazzoni, F. and Marušić, A. (2020), "Large-scale language analysis of peer review reports", *ELife*, Vol. 9, p. e53249.

Callaway, E. (2016), "Open peer review finds more takers", *Nature*, Vol. 539 No. 7629, p. 343.

Campanario, J.M. (1998), "Peer review for journals as it stands today - Part 1", *Science Communication*, Vol. 19 No. 3, pp. 181-211.

Casnici, N., Grimaldo, F., Gilbert, N. and Squazzoni, F. (2017), "Attitudes of referees in a multidisciplinary journal: an empirical analysis", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 7, pp. 1763-1771.

Chakraborty, S., Goyal, P. and Mukherjee, A. (2020), "Aspect-based sentiment analysis of scientific reviews", *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, Association for Computing Machinery, New York, NY, pp. 207-216.

Daryl, E.C. and Edward, J.H. (1990), *Peerless Science: Peer Review and U.S. Science Policy*, State University of New York Press, New York.

Dell'Anno, R., Caferra, R. and Morone, A. (2020), "A 'Trojan Horse' in the peer-review process of fee-charging economic journals", *Journal of Informetrics*, Vol. 14 No. 3, p. 101052.

Emerson, G.B., Warme, W.J., Wolf, F.M., Heckman, J.D., Brand, R.A. and Leopold, S.S. (2010), "Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial", *Archives of Internal Medicine*, Vol. 170 No. 21, pp. 1934-1939.

Falk Delgado, A., Garretson, G. and Falk Delgado, A. (2019), "The language of peer review reports on articles published in the BMJ, 2014-2017: an observational study", *Scientometrics*, Vol. 120 No. 3, pp. 1225-1235.

Ford, E. (2013), "Defining and characterizing open peer review: a review of the literature", *Journal of Scholarly Publishing*, Vol. 44 No. 4, pp. 311-326.

Ghosal, T., Verma, R., Ekbal, A. and Bhattacharyya, P. (2019), "DeepSentiPeer: harnessing sentiment in review texts to recommend peer review decisions", *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, pp. 1120-1130.

Goodman, S.N., Berlin, J., Fletcher, S.W. and Fletcher, R.H. (1994), "Manuscript quality before and after peer review and editing at annals of internal medicine", *Annals of Internal Medicine*, Vol. 121 No. 1, pp. 11-21.

Groves, T. and Loder, E. (2014), "Prepublication histories and open peer review at the BMJ", *BMJ*, Vol. 349, doi: 10.1136/bmj.g5394.

Gu, X., Yang, H., Tang, J. and Zhang, J. (2016), "Web user profiling using data redundancy", *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, IEEE Press, Davis, CA, pp. 358-365.

Hartley, J., Pennebaker, J.W. and Fox, C. (2003), "Abstracts, introductions and discussions: how far do they differ in style?", *Scientometrics*, Vol. 57 No. 3, pp. 389-398.

Hopewell, S., Collins, G.S., Boutron, I., Yu, L.-M., Cook, J., Shanyinde, M., Wharton, R., Shamseer, L. and Altman, D.G. (2014), "Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study", *BMJ*, Vol. 349, doi: 10.1136/bmj.g4145.

Horrobin, D.F. (2001), "Something rotten at the core of science?", *Trends in Pharmacological Sciences*, Vol. 22 No. 2, pp. 51-52.

Hua, X., Nikolov, M., Badugu, N. and Wang, L. (2019), "Argument mining for understanding peer reviews", *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2019*, June 2, 2019 - June 7, 2019, Vol. 1, Association for Computational Linguistics (ACL), Minneapolis, MN, pp. 2131-2137.

Jefferson, T., Wager, E. and Davidoff, F. (2002), "Measuring the quality of editorial peer review", *JAMA: The Journal of the American Medical Association*, Vol. 287 No. 21, pp. 2786-2790.

Jubb, M. (2016), "Peer review: the current landscape and future trends", *Learned Publishing*, Vol. 29 No. 1, pp. 13-21.

Kaatz, A., Magua, W., Zimmerman, D.R. and Carnes, M. (2015), "A quantitative linguistic analysis of National Institutes of Health R01 application critiques from investigators at one institution", *Academic Medicine*, Vol. 90 No. 1, pp. 69-75.

Lee, C.J. and Moher, D. (2017), "Promote scientific integrity via journal peer review data", *Science*, Vol. 357 No. 6348, pp. 256-257.

Luo, J., Feliciani, T., Reinhart, M., Hartstein, J., Das, V., Alabi, O. and Shankar, K. (2021), "Analysing sentiments in peer review reports: evidence from two science funding agencies", *Quantitative Science Studies*, pp. 1-52.

Magua, W., Zhu, X., Bhattacharya, A., Filut, A., Potvien, A., Leatherberry, R., Lee, Y.-G., Jens, M., Malikireddy, D., Carnes, M. and Kaatz, A. (2017), "Are female applicants disadvantaged in National Institutes of Health peer review? Combining algorithmic text mining and qualitative methods to detect evaluative differences in R01 reviewers' critiques", *Journal of Womens Health*, Vol. 26 No. 5, pp. 560-570.

Maltz, D.N. and Borker, R.A. (1983), "A cultural approach to male–female miscommunication", in *Language and Social Identity*, Cambridge University Press, Cambridge, pp. 196-216.

Marsh, H.W., Bornmann, L., Mutz, R., Daniel, H.-D. and O'Mara, A. (2009), "Gender effects in the peer reviews of grant proposals: a comprehensive meta-analysis comparing traditional and multilevel approaches", *Review of Educational Research*, Vol. 79 No. 3, pp. 1290-1326.

Martin, J.R. and White, P.R.R. (2007), *The Language of Evaluation*, Palgrave Macmillan, London.

Meng, M., Wang, Y. and Zhang, C. (2021), "Building multi-level aspects of peer reviews for academic articles", *18th International Conference on Scientometrics and Informetrics Conference, ISSI 2021*, July 12, 2021 - July 15, 2021, International Society for Scientometrics and Informetrics, Leuven, pp. 1519-1520.

Mulligan, A., Hall, L. and Raphael, E. (2013), "Peer review in a changing world: an international study measuring the attitudes of researchers", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 1, pp. 132-161.

Osterath, B. (2016), "Peer-review 'heroes' do lion's share of the work", *Nature*. doi: 10.1038/nature.2016.21031.

Publons (2018), *Publons' Global State of Peer Review 2018*, Publons, London, doi: 10.14322/publons.GSPR2018.

Qin, C. and Zhang, C. (2021), "Exploring the distribution regularities of referees' comments in IMRAD structure of academic articles", *18th International Conference on Scientometrics and Informetrics Conference, ISSI 2021*, July 12, 2021 - July 15, 2021, International Society for Scientometrics and Informetrics, Leuven, pp. 1527-1528.

Quirk, R. and Greenbaum, S. (1973), *A University Grammar of English*, Longman, London.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985), *A Comprehensive Grammar of the English Language*, Longman, London.

Ragone, A., Mirylenka, K., Casati, F. and Marchese, M. (2013), "On peer review in computer science: analysis of its effectiveness and suggestions for improvement", *Scientometrics*, Vol. 97 No. 2, pp. 317-356.

Rashidi, K., Sotudeh, H., Mirzabeigi, M. and Nikseresht, A. (2020), "Determining the informativeness of comments: a natural language study of F1000Research open peer review reports", *Online Information Review*, Vol. 44 No. 7, pp. 1327-1345.

Rennie, D. (2016), "Let's make peer review scientific", *Nature*, Vol. 535 No. 7610, pp. 31-33.

Rockwell, S. (2005), "Ethics of peer review: a guide for manuscript reviewers", available at: https://ori.hhs.gov/sites/default/files/prethics.pdf (accessed 29 April 2021).

Rowland, F. (2002), "The peer-review process", *Learned Publishing*, Vol. 15 No. 4, pp. 247-258.

Salem, R.M., Culbertson, N.M. and O'Connell, A. (2016), "Process for selecting and implementing a manuscript management system: experiences of a new peer-reviewed journal", *Learned Publishing*, Vol. 29 No. 1, pp. 55-59.

Seeber, M. and Bacchelli, A. (2017), "Does single blind peer review hinder newcomers?", *Scientometrics*, Vol. 113 No. 1, pp. 567-585.

Siler, K., Lee, K. and Bero, L. (2015), "Measuring the effectiveness of scientific gatekeeping", *Proceedings of the National Academy of Sciences*, Vol. 112 No. 2, pp. 360-365.

Spier, R.E. (2002), "Peer review and innovation", *Science and Engineering Ethics*, Vol. 8 No. 1, pp. 99-108.

Squazzoni, F., Brezis, E. and Marušić, A. (2017), "Scientometrics of peer review", *Scientometrics*, Vol. 113 No. 1, pp. 501-502.

Tang, C., Ross, K., Saxena, N. and Chen, R. (2011), "What's in a name: a study of names, gender inference, and gender behavior in Facebook", in *Database Systems for Adanced Applications*, Springer, Berlin and Heidelberg, pp. 344-356.

Tannen, D. (1992), *You Just Don't Understand*, William Morrow Paperbacks, New York.

Thelwall, M., Papas, E.-R., Nyakoojo, Z., Allen, L. and Weigert, V. (2020), "Automatically detecting open academic review praise and criticism", *Online Information Review*, Vol. 44 No. 5, pp. 1057-1076.

Tran, D., Valtchanov, A., Ganapathy, K., Feng, R., Slud, E., Goldblum, M. and Goldstein, T. (2020), "An open review of OpenReview: a critical analysis of the machine learning conference review process", ArXiv:2010.05137 [Cs], available at: http://arxiv.org/abs/2010.05137 (accessed 31 July 2021).

UNDP (2010), "Human development report 2010: the real wealth of Nations - pathways to human development", New York, available at: http://hdr.undp.org/en/content/human-development-report-2010.

van den Besselaar, P., Sandström, U. and Schiffbaenker, H. (2018), "Studying grant decision-making: a linguistic analysis of review reports", *Scientometrics*, Vol. 117 No. 1, pp. 313-329.

Walker, R. and da Silva, P.R. (2015), "Emerging trends in peer review - a survey", *Frontiers in Neuroscience*, Vol. 9, p. 169.

Wang, K. and Wan, X. (2018), "Sentiment analysis of peer review texts for scholarly papers", *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, New York, NY, pp. 175-184.

Ward, P., Graber, K.C. and Mars, H.V.D. (2015), "Writing quality peer reviews of research manuscripts", *Journal of Teaching in Physical Education*, Vol. 34 No. 4, pp. 700-715.

Ware, M. (2008), *Peer Review: Benefits, Perceptions and Alternatives*, Publishing Research Consortium, London.

Xu, S., Zhang, G., Sun, Y. and Wang, X. (2019), "Understanding the peer review endeavor", *Proceedings of the Association for Information Science and Technology*, Vol. 56 No. 1, pp. 316-325.

Yadav, R.K. and Gehringer, E.F. (2016), "Metrics for automated review classification: what review data show", in *State-of-the-Art and Future Directions of Smart Learning*, Springer, Singapore, pp. 333-340.

Zhang, G., Wang, Y., Xie, W., Du, H., Jiang, C. and Wang, X. (2021), "The open access usage advantage: a temporal and spatial analysis", *Scientometrics*, Vol. 126 No. 7, pp. 6187-6199.

Zong, Q., Xie, Y. and Liang, J. (2020), "Does open peer review improve citation count? Evidence from a propensity score matching analysis of PeerJ", *Scientometrics*, Vol. 125 No. 1, pp. 607-623.

Zuckerman, H. and Merton, R. (1971), "Patterns of evaluation in science - institutionalisation, structure and functions of referee system", *Minerva*, Vol. 9 No. 1, pp. 66-100.

Appendix

Table A1.
Distribution of nouns in reviewers' comments

	Overall		Methods		Results		Discussions	
Order	Word	Num	Word	Num	Word	Num	Word	Num
1	paper	832	data	601	results	748	discussion	598
2	patients	793	methods	514	data	365	data	259
3	data	765	analysis	505	patients	333	page	255
4	risk	662	patients	484	risk	320	patients	253
5	manuscript	506	risk	359	analysis	273	risk	237
6	results	498	page	333	line	229	line	216
7	analysis	420	line	326	page	225	results	200
8	comments	387	time	267	figure	220	findings	160
9	use	354	results	245	outcomes	202	analysis	142
10	health	353	information	225	number	198	limitations	121

Note(s): We deleted "authors", "study" and "studies" from Table A1 to avoid redundancy

Table A2.
Distribution of verbs in reviewers' comments

	Overall		Methods		Results		Discussions	
Order	Word	Num	Word	Num	Word	Num	Word	Num
1	used	300	used	395	included	168	think	108
2	written	231	included	280	presented	153	used	89
3	using	224	using	170	reported	143	discussed	85
4	think	212	given	164	see	119	suggest	84
5	based	184	provide	147	used	114	given	80
6	associated	170	based	141	given	111	based	80
7	given	163	described	138	think	105	included	75
8	included	154	see	134	please	93	using	72
9	found	149	think	120	compared	93	include	69
10	make	148	excluded	120	seems	88	see	65

Note(s): We deleted the different forms of "be", "have" and "do" from Table A2 to avoid redundancy

Corresponding author
Xianwen Wang can be contacted at: xianwenwang@dlut.edu.cn