# RESEARCH ARTICLE

Revised: 2 August 2021



# Classification and analysis of PubPeer comments: How a web journal club is used

# José Luis Ortega<sup>1,2</sup> 💿

<sup>1</sup>Institute for Advanced Social Studies (IESA-CSIC), Córdoba, Spain

<sup>2</sup>Joint Research Unit Knowledge Transfer and Innovation (UCO-CSIC), Córdoba, Spain

#### Correspondence

José Luis Ortega, Institute for Advanced Social Studies (IESA-CSIC), Córdoba, Spain. Email: jortega@iesa.csic.es

#### Abstract

This study explores the use of PubPeer by the scholarly community, to understand the issues discussed in an online journal club, the disciplines most commented on, and the characteristics of the most prolific users. A sample of 39,985 posts about 24,779 publications were extracted from PubPeer in 2019 and 2020. These comments were divided into seven categories according to their degree of seriousness (Positive review, Critical review, Lack of information, Honest errors, Methodological flaws, Publishing fraud, and Manipulation). The results show that more than two-thirds of comments are posted to report some type of misconduct, mainly about image manipulation. These comments generate most discussion and take longer to be posted. By discipline, *Health Sciences* and *Life Sciences* are the most discussed research areas. The results also reveal "super commenters," users who access the platform to systematically review publications. The study ends by discussing how various disciplines use the site for different purposes.

## **1** | INTRODUCTION

The peer-review system is an editorial practice that is recently being called into question by part of the scholarly community (Kelly et al., 2014; Mulligan, 2005). Many criticisms are levied against the subjectivity of the reviews, lack of transparency (anonymous referees and conflict of interests), and overexploitation of reviewers (Fox, 2017). The current transformations of the academic publishing world, with the consolidation of Open Access and the appearance of different ways of research dissemination (repositories, social networks, etc.), have favored the appearance of new proposals to reduce the limitations of this procedure or to contribute an alternative model. Many open access journals (*PeerJ, F1000, Biology Direct*) are instituting open peer-review processes in which reviewers' identities and reviews are publicly available on the site, providing complementary information about how the manuscript has been revised, corrected, and improved (Ross-Hellauer, 2017). Although this model increases transparency, it could influence the critical thinking skills of reviewers to avoid making enemies through the public exposure of their comments (Teixeira da Silva, 2019).

Another way to review research papers is through postpublication peer review (Knoepfler, 2015). This procedure places the review process after the article has been published and invites peers to publicly comment on the quality of the study. Although this method has only been embraced by a few journals (*Atmospheric Chemistry and Physics, MedEdPublish, Sci*), it has been most used in various web platforms associated with scientific publication (Publons, PubPeer, ResearchGate). These services

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

<sup>© 2021</sup> The Author. Journal of the Association for Information Science and Technology published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology.

<sup>2</sup> WILEY JASST

provide the technology and visibility to attract potential researchers interested in the public review of publications. The principal advantages of this system are that the entire process is transparent, it generates open and enriching scientific debate, and, in the case of web platforms, is independent of journal editorial policies. However, it is vulnerable to anonymous harassers, and the process could become endless (Teixeira da Silva & Dobránszki, 2015).

PubPeer was the first web site for postpublication peer review (Bik, 2019). This platform was created in 2012 as a journal club where research papers were discussed by scholars in an open environment (Torny, 2018). A request for protection by some students afraid of suffering reprisals led to anonymity being introduced in 2013. This change made the site very successful, as it began being used to pursue misconduct in research publications. PubPeer's significance not only lies in the reporting of bad practices, but also in the ongoing detection of failures in the peerreview system, with comments about conflicts of interests, metrics manipulation, or fake peer reviews. The massive disclosure of unethical practices is being wielded to demand more transparency in the scholarly publishing system and a higher commitment of publishing houses to detecting and disclosing these problems (Horbach & Halffman, 2018). This study analyzes the content of the comments posted to the platform, to explore general patterns in the use of this web space.

#### 2 **RELATED RESEARCH**

The literature on peer review has focused on appraising the various elements and agents involved in the process, to detect limitations and suggest improvements (Kassirer & Campion, 1994). Many studies have tested the consistency of reviewers' decisions, finding important disagreements between them (Pier et al., 2018; Rothwell & Martyn, 2000). In some cases, the subjectivity of the process has been reported when gender (Helmer et al., 2017), language (Herrera, 1999), and social biases (Walker et al., 2015) have been revealed. Other important limitations to peer review have been posited, including excessive time delay (Björk & Solomon, 2013), reviewers' conflicts of interests (Resnik & Elmore, 2018), and the inability to detect fraudulent misconduct (Horbach & Halffman, 2019; Smith, 2006).

These drawbacks of traditional peer review have encouraged improvements and alternative proposals that require testing. Reviewer anonymity is the most analyzed aspect (Ross-Hellauer, 2017). van Rooyen et al. (1999) observed that reviewer identification had no important effect on the process but significantly increased the likelihood of reviewers declining to review. Walsh et al. (2000) did find striking differences, observing that nonanonymous reviewers are more prone to being more

polite and to recommending the publication. Bolek et al. (2020) found similar results, particularly that women reviewers are more resistant to making their names public. However, Thelwall et al. (2020) detected an important country bias in non-anonymous reviewers, who favored the acceptance of publications from the same country. Studies about the impact of review reports being published openly are less frequent. Cosgrove and Cheifet (2018) found no significant differences in time delay or number of reluctant reviewers regarding open or non-open reviews. In a pilot study, Nature Communications (2016) confirmed that, on average, 60% of reviewers accepted the open publication of their reports.

Many of these changes in the peer-review system have been headed by recent web platforms that externally promote publication review-in particular, post-publication peer review (e.g., Publons, PubPeer, ResearchGate). However, studies about these platforms have been more descriptive than analytical. From this latter perspective, Ortega (2017) analyzed the relationship between reviewers' performance in Publons and their bibliometric outputs, finding a weak correlation between both activities. He also studied the association between Publons metrics, bibliometrics, and altmetrics, finding low correlations between these indicators (Ortega, 2019). Segado-Boj et al. (2018) found that respondents to an opinion survey who valued ResearchGate more highly were more critical of open review.

According to PubPeer, Wager and Veitch (2017) categorized comments about 150 biomedicine research articles to measure the proportion of misconduct detected by the platform. Their findings revealed that only 7% deserved a journal action. More recently, Bordignon (2020) used PubPeer comments to test the influence of negative comments in the retraction/correction of papers, concluding that PubPeer contributes more to the correction of science than negative citations. Ortega (2020) studied the incidence of editorial notices in PubPeer, observing that editorial notices are more common in multidisciplinary journals and those specializing in biochemistry and medicine. Other publications only focused on theoretical discussions about the problem of anonymity and the implications of post-publication review (Blatt, 2015; Teixeira da Silva, 2018; Torny, 2018; Townsend, 2013). However, there is an important gap regarding the overall content of this site and how it has been used by the scholarly community.

#### 3 **OBJECTIVES**

The aim of this study is to explore the use of PubPeer by the scholarly community through the content categorization of comments posted to the platform. It analyzes the characteristics of each type of comment (delay, number of posts),

disciplinary differences, and the most relevant profiles in the network. Three research questions are formulated:

- What are the most frequent types of comment? And what is the difference between them according to time delay and number of posts?
- What is the disciplinary distribution of each type of comment? And in which disciplines is each type of comment more prevalent?
- How do users use this platform? Specifically, what are the differences between genders and disciplines?

# 4 | METHODS

#### 4.1 | Source

PubPeer is a scientific forum or journal club where scientific publications are discussed after publication. A special feature of this postpublication peer-review site, created in October 2012, is that comments can be anonymous. As a result, the forum has become a specialized site for reporting bad publishing practices. Considerable controversy has ensued because many authors feel defenseless in the face of unknown complainants (Torny, 2018). However, many users are grateful for this format because bad practices can be flagged up with no reprisals.

Until 2018, a generic "Unregistered Submission" or "Peer" was used to post anonymous comments. However, this impeded identifying who posted which comment when several anonymous users participated in the same thread. To solve this problem, registration was required for posters, and the scientific name of an organism was assigned to ensure anonymity (Teixeira da Silva, 2018). Recently, PubPeer has integrated several platforms to indicate when a paper is being discussed: Zotero and a navigator extension signal when comments are made about a publication. Moreover, tweets about research publications can be read in PubPeer as a way of capturing discussion outside the platform (PubPeer, 2021).

#### 4.2 | Data

Two samples were extracted from PubPeer: an initial sample of 32,097 threads and 65,179 posts was obtained in March 2019 and was then updated with 7,659 threads and 21,200 posts in January 2020. These samples were obtained by searching for the first letters of the alphabet—a, b, and c—in the standard search box. PubPeer ranks results by date, and therefore entries after 2018 were retrieved. Each entry is a comment linked to a thread with an identifier. This ID was extracted, and

a crawler was designed to sequentially extract the bibliographic data about publications and the metadata (user, text, date, etc.) of each comment linked to each thread. WebQL Studio (www.ql2.com) was used for this task. In total, 86,379 posts from 39,757 threads associated with 24,779 publications were retrieved. From this set of posts, 11,469 (13.3%) were automatically generated by a robot (statcheck), and 6,328 (7.3%) were retrieved without user comments. These posts were removed. Finally, 68,595 (79.4%) posts produced by 26,456 users were selected for the study.

Each publication included in the sample was thematically classified according to the journal in which it was published. This information was obtained from the SCImago Journal & Country Rank portal (scimagojr. com), which uses All Science Journal Classification (ASJC) to categorize and rank journals. Each publication was assigned to more than one discipline according to the number of categories in which a journal is classified. Documents published at other venues (repositories, conferences, books, etc.) were classified manually.

To identify the type of user (anonymous or nonanonymous), the system applies the at (@) symbol to most nonanonymous users. In any case, a manual inspection was performed to distinguish nicknames and Latin species names from first name/surname structures. Genre identification was carried out manually by searching for first names in Gender Checker (genderchecker.com). Unisex names such as Chris, Alex or Chang were not considered.

Posts or comments are interchangeable and refer to each comment or opinion posted by a user about a publication. Thread refers to the collection of comments discussing a specific issue in one publication. A publication may have different threads, though this is uncommon. Only 539 (2.2%) publications were discussed in more than one thread.

## 4.3 | Classification

An important limitation in the study of web posts and comments is that the text must be processed for content to be categorized and the principal meaning of each post defined. Unlike Twitter, PubPeer does not include hashtags or other types of words that allow content classification. Accordingly, highly informative words (keywords) that describe the principal statement of each comment were extracted. This procedure can be carried out in various ways. A common method is a panel of several coders that contrast their subjective interpretation of content (Hooge et al., 2018; Malički et al., 2021)—a team of collaborators is required for this purpose. However, more often, categorization through keywords has been ▲ WILEY JASST

used to classify posts on social networks (Carreño & Winbladh, 2013; Salehan & Kim, 2016; Thelwall et al., 2012) when the volume of entries is considerable and an automatic approach is necessary.

The most frequent nouns and verbs in the entire corpus (more than 1%) were selected to search comments that included those terms. These keywords were stemmed (plural/singular, verb forms, etc.) and normalized (Methods = Methodology, Methodological) to become concepts. For example, Figure (27%), Image (24%), Retraction (4%), and Duplication (3%) were the most frequent concepts. Each comment is thus described by several terms that describe the general content. Therefore, an entry with the terms Figure and Splicing might suggests a warning about a possible manipulation of images, while Methods, Discussion, and Clarifications might allude to doubts about how the results were obtained.

Next, a classification scheme was designed to assign each comment to a category. Several sources and publications were used as a reference. For example, the three types of misconduct defined by the National Science Foundation (Electronic Codes of Federal Regulations, 2019); the types of misconduct in biomedicine identified by Kumar (2008); differences between misconduct and honest error (Resnik & Stewart, 2012); and the classification of misconduct proposed by Kuroki (2018). However, all these examples focus on scientific fraud, and PubPeer is a journal club that includes comments about any type. Therefore, an ad hoc classification scheme was designed according to the degree of misconduct and seriousness of the comments, and 24,016 publications (97%) were classified following this scheme:

- Positive review: Comments that praise and highlight publications according to the reach and importance of the results. Associated concepts are, for instead, Interesting, Excellent, and Useful.
- · Critical review: Comments that discuss the methods and results and their interpretations. This group includes discussions about theoretical implications and scientific disagreements (Resnik & Stewart, 2012). Associated concepts are, for example, Methods and Discussion. Within this category:
  - Lack of information: Inside Critical review, this is a subcategory that addresses the problematic absence of information about how the study was performed, the availability of raw data, and lack of relevant bibliographic references. Associated concepts are, for example, Data available and Missing information.
- · Errors: These comments reveal scientific studies that use wrong methods or data but with no intention of committing deliberate fraud. Two groups of comments:
  - Honest errors (Resnik & Stewart, 2012): They could be rectifiable mistakes (e.g., erratum) due to

confusion and oversight in the writing of the paper. Associated concepts are, for example, Error, Correction, and Typo error.

- Methodological flaws: They are motivated by a lack of awareness of statistical or other scientific techniques (e.g., western blots, spectroscopy) that throw up wrong results (e.g., correlation fishing, bar errors, loading controls). This category could be bordering on fraud, because this confusion could be intended to obtain the desired results. However, such intentionality is not always evident, and these issues are given the benefit of doubt and considered as Errors. Associated concepts are, for example, Sampling issues, Lack of references, Error, Wrong method, and Ethical issues.
- · Fraud: These are the most serious posts. They report intentional misconduct that aims to highlight findings or take advantage of the research evaluation system. Two subsections were defined:
  - Publishing fraud: Interference with the publishing system to increase production and impact. Associated concepts are, for example, Salami publishing, Reusing, Republishing, Plagiarism, Duplicated publication, and Citation farm.
  - Manipulation: Intentional edition and manipulation of data and images to obtain better results than those expected, to corroborate the desired hypothesis. Associated concepts are, for example, Splicing, Duplication, and Manipulation.

#### 4.4 | Classification process and validation

A heuristic was used to proceed with classification. To reduce time and effort, only the first comment in a thread was analyzed-the first post is often the most informative as it presents the initial argument about the publication. If the starting post did not include sufficient information (i.e., only an external link, a figure with embedded text), the remaining posts were read manually. Because the classification scheme is gradual, when a thread addressed several issues, the publication was assigned to the most serious category. For example, if a paper was reported for Methodological flaws and Plagiarism, then that paper was classified as Publishing fraud. This fact is rather unusual since only 1.5% of threads include comments of different types.

To test the accuracy of the classification process, a random subsample of 4,000 (16.7%) classified posts (predicted) were also manually classified (actual). The percentage of cases in the sub-sample is proportional to the original sample to avoid biases in the validation. Using a simple rule of three, the number of cases for each

TABLE 1 Confusion matrix with the proportion of predicted assignments according to actual classification

Predicted Actual	Manipulation (%)	Publishing fraud (%)	Methodological flaws (%)	Honest errors (%)	Lack of information (%)	Critical review (%)	Positive review (%)
Manipulation	93.9	29.5	15.1	1.0	1.1	5.3	0.0
Publishing fraud	4.3	66.5	5.1	0.0	0.0	0.0	0.0
Methodological flaws	1.5	0.8	72.4	4.1	1.1	1.7	0.0
Honest errors	0.0	0.3	0.0	80.6	0.0	0.6	0.0
Lack of information	0.0	0.5	0.6	1.0	75.5	1.3	0.0
Critical review	0.3	2.4	6.7	13.3	22.3	90.6	28.3
Positive review	0.0	0.0	0.0	0.0	0.0	0.5	71.7

FIGURE 1 Distribution of comments by publication (log–log) [Color figure can be viewed at wileyonlinelibrary.com]



category were randomly selected sorting PubPeer IDs alphabetically.

Table 1 gives the precision between keyword and manual classification. The test shows a general precision of 88.1%; Publishing fraud (66.5%), Positive review (71.7%), and Methodological flaws (72.4%) were the categories with the most incorrect assignments. In the case of Publishing fraud, misclassification is due to the ambiguity of the term "duplication," because it could suggest duplication of elements in a figure (Manipulation), or duplication of text or figures in different publications (Publishing fraud). Twenty-eight percent of Positive reviews are in fact Critical reviews because many Critical reviews also include positive comments. Misclassification in Methodological flaws is due to the difficulty of specifying when an error could be intentional or not. This lead to the wrong assignment of 15.1% of Methodological flaws posts to Manipulation. However, the overall result indicates that the classification process works fairly well,

because close to nine out of 10 posts (precision = 88.1%) were correctly assigned.

The total list of classified posts (24,016) and the subsample of reclassified posts (4,000) are available at the following site: https://osf.io/2zfrw/.

#### 5 | RESULTS

#### 5.1 | Distribution of comments

Of the comments, 63,588 (92.7%) were associated with a research publication. Distribution of comments by article is skewed, following a power-law trend ( $\alpha = 3.06$ ) (Figure 1). Hence, most papers are only commented upon once (50.25%) or twice (34.56%), while just seven articles have more than 100 comments (Table 2). This result shows that very few publications provoke clear discussion with numerous comments and replies.

ASST -WILEY

TABLE 2 Distribution of the number of comments by publication and percentage of contribution

Number of comments	Publications	Publications (%)	Comments	Contribution (%)
1	12,069	50.25	12,069	18.98
2–4	8,299	34.56	21,883	34.41
5–9	2,595	10.81	16,516	25.97
10–99	848	3.53	11,690	18.38
<100	7	0.03	1,430	2.25
Total	24,016		63,588	

TABLE 3 Distribution of publications, author responses, and anonymous posts by type of comments in PubPeer

Group of comments	Type of comments	Publications	Publications (%)	Author response	Author response (%)	Anonymous	Anonymous (%)
Positive review	Positive review	190	0.8	13	6.9	76	49.4
Critical review	Critical review	3,791	15.8	222	5.9	2,694	73.8
	Lack of information	398	1.7	26	6.5	335	86.6
Errors	Honest errors	577	2.4	40	6.8	474	83.6
	Methodological flaws	1,809	7.5	76	4.2	1,413	82.0
Fraud	Publishing fraud	2,316	9.6	88	3.8	1,969	91.6
	Manipulation	14,935	62.2	1,196	8.0	12,865	93.8
Total		24,016		1,661	6.9	19,826	85.6

Table 3 depicts the number and proportion of publications classified according to the type of comment they receive. Appendix A includes all information about the estimation and correction of these results according to the accuracy test. Differences between predicted and actual cases are small and hardly significant, indicating that both classification methods report similar results. Publications reported for Manipulation of data and figures are the majority (62.2%), followed by Critical reviews (15.8%), and Publishing fraud (9.6%). By group of comments, 71.8% of publications are reported for fraud of any kind, while 17.5% are just critical reviews of methods or discussions about the theoretical implications of the results. These results show that the main use of PubPeer is to report fraudulent studies-more than two thirds of the comments. The remaining comments are about the research quality of studies and flag up methodological errors.

According to the response rate of authors, in general, the reply average is quite low (6.9%), which suggests that many authors are not aware of the comments expressed about their papers. The highest response rates are in Manipulations (8.0%) and Positive reviews (6.9%). For Manipulations, these could be motivated by authors' needs to defend themselves against serious accusations. Conversely, for Positive reviews, it could be that some authors use PubPeer as a network for dissemination. More than half the replies are single posts (54%) reporting the release of new and positive advances.

Nevertheless, with an average percentage of 85.6%, anonymous comments are highly frequent. By type of comment, the degree of anonymity drops as the comments become less problematic. Comments with the highest percentage of anonymous posters are therefore Manipulations (93.8%) and Publishing frauds (91.6%), while Positive (49.4%) and Critical reviews (73.8%) are the type of comments published by fewer anonymous users.

Figure 2 displays the distribution of the number of posts per publication according to type of comment. Publications categorized as Manipulation  $(3.3 \pm 0.08)$  and Publishing fraud  $(3.15 \pm 0.47)$  gather more posts on average than the other categories, while Positive review  $(1.4 \pm 0.19)$  is the type of comment that accumulates the fewest posts. By group of comments, Fraud publications  $(3.25 \pm 0.094)$  clearly produce more posts than Critical review  $(2.55 \pm 0.32)$  and Errors  $(2.48 \pm 0.65)$ . These differences are statistically significant (Kruskal–Wallis chi-square = 1152.1, p < .0001), and these results clearly show that publications reported for fraud receive many more posts than other types of publications. Bilateral comparisons (Dunn's test) show that these differences are not significant between Critical review and Honest errors

FIGURE 2 Boxplot of the distribution of the number of posts (log) by type of comment [Color figure can be viewed at wileyonlinelibrary.com]



FIGURE 3 Boxplot of the distribution of time delay in posting (log) according to type of comment [Color figure can be viewed at wileyonlinelibrary.com]



FIGURE 4 Bar plot of the proportion of posts by type of comment and grouped by research area [Color figure can be viewed at wileyonlinelibrary.com]

(p = .015) and Lack of information (p = .854); Positive review and Honest errors (p = .735); and Publishing fraud and Lack of information (p = .032).

Figure 3 displays the time delay in days between when a paper is published and when comments are posted on PubPeer, to understand how long it takes for certain types of comments to be posted. Only articles published from 2015 were selected, to avoid outliners from very old articles published before the platform existed. The findings appear to show that the more problematic the comments, the longer they take to be posted. Thus, comments about Manipulation and Publishing fraud take, on average,  $593 \pm 15.4$  days  $(1.6 \pm 0.04 \text{ years})$  and  $533 \pm 34.2 \text{ days}$   $(1.5 \pm 0.09 \text{ years})$ , respectively, while Positive and Critical reviews happen 239  $\pm$  60.9 days (0.65  $\pm$  0.17 years) and 241  $\pm$  14.6 days (0.66  $\pm$  0.04 years) later. Differences in the delay between Manipulation and Publishing fraud and the other groups are statistically significant (Kruskal-Wallis chi-square = 1079.3, p < .0001). Bilateral comparisons (Dunn's test) show that these differences are not significant between Methodological flaws and Positive review (p = .283) and Honest errors (p = .395); Lack of information and Honest errors (p = .561); and Publishing fraud and Manipulation (p = .152)

#### **Disciplinary analysis** 5.2

The most commented research areas in PubPeer are Life Sciences (50.6%) and Health Sciences (27.6%). Physical

Positive review Critical review Lack of information Honest errors Methodological flaws Publishing fraud

Sciences contributes 18.2% of publications, while the presence of Social Sciences & Humanities is anecdotal, with only 3.7% of comments.

Figure 4 shows the proportion of the type of comment according to the four main research areas. The picture clearly depicts a significant difference between disciplines: Health Sciences (77%) and Life Sciences (79%) are the research areas with a higher proportion of fraud reports, especially regarding Manipulation (68% in Health Sciences and 70.5% in Life Sciences), whereas Social Science & Humanities publications include more Critical reviews (43.5%) and Errors (33.2%). In this same research area, the most frequent comment type is Methodological flaws (29.7%). This clear difference between Social Sciences & Humanities and other research areas could be because Social Sciences & Humanities include speculative disciplines (e.g., History, Philosophy) based on qualitative methodologies that favor critical discussion and disciplines (e.g., Psychology, Economics) with a strong statistical component that facilitates the proliferation of methodological errors in data analysis and treatment

Figure 5 presents a more detailed view at the disciplinary level, clearly showing that Life Sciences and Health Sciences are the most problematic research areas. Concretely, Immunology and Microbiology (83.1%), Biochemistry, Genetics and Molecular Biology (83%), and Pharmacology, Toxicology and Pharmaceutics (82.5%) receive the most comments on fraud, with posts on data and image manipulation (more than 70%) being the most



FIGURE 5 Barplot of the proportion of comments by type and grouped by discipline [Color figure can be viewed at wileyonlinelibrary.com]



TABLE 4 Distribution of the number of comments by user, proportion of anonymity, and percentage of contribution

Number of comments	Users	Users (%)	Anonymous	Anonymous (%)	Comments	Contribution (%)
1	15,857	59.94	14,489	91.37	15,857	23.12
2-4	8,344	31.54	7,675	91.98	21,087	30.74
5–9	1,678	6.34	1,557	92.79	10,584	15.43
10–99	552	2.09	498	90.22	9,310	13.57
<100	25	0.09	19	76.00	11,757	17.14
Total	26,456		24,238	91.62	68,595	

TABLE 5 Distribution of posts by research area and type of comment of the five most active commenters

Name	Research area	Posts	Post (%)	Group	Posts	Post (%)
Actinopolyspora Biskrensis	Health Sciences	318	34.6	Fraud	662	88.4
	Life Sciences	404	43.9	Errors	41	5.5
	Multidisciplinary	96	10.4	Critical review	46	6.1
	Physical Sciences	97	10.5			
	Social Sciences & Humanities	5	0.5			
	Total	920		Total	749	
Elisabeth M. Bik	Health Sciences	520	35.1	Fraud	979	90.6
	Life Sciences	666	45.0	Errors	83	7.7
	Multidisciplinary	97	6.6	Critical review	18	1.7
	Physical Sciences	188	12.7			
	Social Sciences & Humanities	9	0.6			
	Total	1,480		Total	1,080	
Hoya Camphorifolia	Health Sciences	647	30.6	Fraud	1,483	93.9
	Life Sciences	835	39.4	Errors	43	2.7
	Multidisciplinary	100	4.7	Critical review	53	3.4
	Physical Sciences	523	24.7			
	Social Sciences & Humanities	12	0.6			
	Total	2,117		Total	1,579	
Indigofera Tanganyikensis	Health Sciences	346	47.1	Fraud	497	90.9
	Life Sciences	289	39.3	Errors	37	6.8
	Multidisciplinary	38	5.2	Critical review	12	2.2
	Physical Sciences	60	8.2	Positive review	1	0.2
	Social Sciences & Humanities	2	0.3			
	Total	735		Total	547	
Lydia M. Maniatis	Health Sciences	209	24.7	Fraud	63	11.5
	Life Sciences	404	47.8	Errors	93	17.0
	Multidisciplinary	102	12.1	Critical review	387	70.9
	Physical Sciences	43	5.1	Positive review	3	0.5
	Social Sciences & Humanities	88	10.4			
	Total	846		Total	546	

frequent. However, comments about Publishing fraud are more common in Health Sciences and Physical Sciences, particularly Dentistry (31%) and Veterinary (26.9%), Earth and Planetary Sciences (27.3%), and Engineering (21.4%). The figure also shows that Social Sciences & Humanities publications have more Errors. Business, Management and

<b>TABLE 6</b> Distribution of users   with more than two posts by gender	Gender	Users	Users (%)	Comments	Comments (%)	Comments/user
with more than two posts by gender	Men	231	76.0	2,465	44.2	10.7
	Women	73	24.0	3,116	55.8	42.7
	Total	304		5,581		18.4

Accounting (53.5%) and Economics, Econometrics and Finance (44.8%) receive the most comments about Methodological flaws, which could be explained by the fact that they are mainly based on quantitative methods prone to mistakes in statistical data. Yet Decision Sciences (58.3%), Mathematics (53%), and Computer Science (50.6%) receive the most critical reviews. However, these disciplines have a low number of publications in PubPeer, and these percentages are distorted by one specific user's criticism about fuzzy logic methods. If these comments are removed, then Arts and Humanities (42%) and Social Sciences (49.2%) have the largest number of critical reviews. These results now fit better with speculative disciplines in which theoretical discussions between different schools and traditions are more commonplace

# 5.3 | Distribution of users

Figure 6 depicts the cumulative distribution of the number of commented publications by user, and Table 4 the number of users, anonymity and contribution by number of comments. As for comment distribution, the trend fits with a power law ( $\alpha = 2.52$ ), suggesting that few users are responsible for most comments (25 users post 17.1% of the comments), and most users just comment on one (59.9%) or two (31.5%) publications. Interestingly, the tailend of the distribution shows more posts per user than expected. This deviation means that a group of "super commenters" uses the journal club as an active medium for reporting specific problems, rather than for discussing one-time articles. The proportion of anonymous users (91.62%) is slightly higher than the percentage of anonymous posts (85.6%), which could indicate that the nonanonymous users could be more active than the anonymous ones. The fact that the anonymity in users with more than 10 posts is under the mean, could confirm this assumption

A close look at the five most active commenters reveals some distinctive features (Table 5). Most focus almost entirely on reporting fraudulent activities, mainly image manipulation. The only exception is Lydia Maniatis, who spotlighted the discussion of certain methods and theories (Critical review) in Neuroscience. Another characteristic is that these prolific commenters specialize in discussing *Life Sciences* and *Health Sciences* publications, particularly *Biochemistry, Genetics and Molecular Biology, and Medicine, except for Hoya Camphorifolia, who focused on Physical Sciences as well.* 

Interestingly, only two of these five "super commenters" are non-anonymous-and they are women. This led us to explore gender differences in how users engage with the platform. Table 6 shows the number of users and comments by gender, limited to non-anonymous users with more than two posts. The results indicate that many more men (76.0%) than women (24.0%) comment on papers, but that women contribute a slightly higher proportion of posts (men = 44.2%, women = 55.8%) and more posts on average (men = 10.7, women = 42.7). These differences are statistically significant (Wilcox test = 5,153,819, p < 2.2e-16), perhaps due to the high participation of Elisabeth M. Bik and Lydia Maniatis, which drags the mean up to more than three times more than for men. This could also explain why the percentage of comments about misconduct is higher in women (91.2%) than in men (88.1%), as is reporting more data manipulations (men = 76.0%; women = 83.2%), while women post fewer positive reviews (men = 3.0%; women = 0.5%).

## 6 | DISCUSSION

Exploratory analysis of the comments posted on PubPeer has provided several insights into how this popular journal club is used. Most posts generate few replies, and close to 70% of the comments receive fewer than two responses. This suggests that the platform is used for reporting problems about specific papers rather than for generating critical discussion about new research fronts or the potential of some methods, a perception confirmed by the content of the comments. That 71% of the posts are about fraud, especially image manipulation, reveals that PubPeer is mainly being used to report fraudulent research. In addition, comments about dishonest studies attract on average more attention than other comments (3.5 posts), indicating that the audience of this site could be especially interested in detecting misconduct in publications. Interestingly, comments about fraudulent publications take longer to get published (1.5 years), which emphasizes the considerable effort and prudence involved in the formalization of these accusations.

Another sign of poor engagement is the low response rate per author (6.9%), although this percentage climbs to WILEY\_ JASIST

8.0% for comments about manipulations. Even then, the proportion of authors that do not respond to these accusations is very high, which could indicate that many authors are not aware of the platform, or that they refuse to publicly discuss misconduct in their publications. However, it is also true that many authors do not respond as such but as regular users, which could distort the response rate.

But perhaps the most outstanding feature are the anonymous posts: around 86% of comments do not have a recognizable user. This percentage reaches 94% for fraudulent publications and highlights the need for anonymous whistleblowing channels to avoid reprisals in science (Bouter & Hendrix, 2017; Yong et al., 2013). However, the high percentage of anonymity in posts about critical reviews (76%), even in positive reviews (49%), signals a warning about an unfair use of this feature, which should only be used for fraud cases to protect the informant (Blatt, 2015; Teixeira da Silva & Blatt, 2016).

By discipline, PupPeer's comments center on Health and Life Sciences, which are also the most troublesome disciplines, with the highest percentages of fraudulent publications. Specifically, Immunology and Microbiology, Biochemistry, Genetics and Molecular Biology, and Pharmacology, Toxicology and Pharmaceutics receive the most comments about publishing fraud and data manipulation. This result coincides with previous studies about the disciplinary distribution of retractions (Ortega, 2020; Tripathi et al., 2019) and highlights the great problem of image manipulation in Biomedicine today (Bik et al., 2016; Oksvold, 2015). Otherwise, social science and humanity disciplines indicate a high proportion of comments about critical reviews and methodological flaws, which exposes the importance of theoretical discussion in the humanities and the misuse of statistical methods in the social sciences (Brown & Hedges, 2009; Lamiell, 2019). These results suggest that this platform could be used differently according to discipline. Thus, while biochemists access the site to report misconduct (Manipulation, Publishing fraud), social scientists and humanists use it to discuss conclusions and detect methodological errors (Critical review, Methodological flaws).

The distribution of posts by user pinpoints the different uses of the platform. The results show that while most users post about one or two articles (95%), a small group of "super commenters" access the forum to systematically report misconduct. That the most prolific users specialize in *Biochemistry, Genetics and Molecular Biology*, and *Medicine*, is unsurprising and would explain why disciplines with more troublesome publications favor the emergence of users who specialize in detecting misconduct. This result suggests, to some extent, a certain degree of professionalization. The appearance of "Image forensics" is an example of researchers that develop competence in the detection of misconduct practices and focus their research career on reporting these issues (Shen, 2020). The fact that the anonymity is less frequent in this group could indicate that some "professionals" are less afraid of reprisals and opting for being identified.

#### 6.1 | Limitations

The main difficulty of this study is that a specific classification scheme has been designed to gather every type of comment and provide a complete overview of the use of PubPeer. Although the definition of these seven categories is based on previous studies (Electronic Codes of Federal Regulations, 2019; Kumar, 2008; Kuroki, 2018; Resnik & Stewart, 2012), the assignation of comments to categories always involves the risk of misclassification. The accuracy test revealed that 88.1% of the posts were correctly classified, which can be considered a suitable rate. Yet this precision increased to 94% at group level. However, the significant misclassification rate in some categories should be examined: the results underestimated the proportion of Positive reviews (28%) and Honest errors (17%), while overestimating Critical review (3.6%) and Manipulation (1.7%).

However, this accuracy test is only based on the opinion of one classifier which could introduce a subjective perception about the category of each post, mainly according to posts whose content is unclear and imprecise and they are hard to assign to a category. Thus, although the assignation has followed the defined criteria in the classification scheme, it is possible that the interpretation of some cases might be debatable by different classifiers. To relieve this problem, the result of the accuracy test is publicly available to be discussed by the research community. Even then, this important limitation encourages more research on the content of this platform that make possible to confirm these results.

Another limitation could be the representativeness of the sample. PubPeer provides no information about the number of posts, publications, or users covered in the platform, making it extremely difficult to check whether the sample is proportional to the population, or whether any bias has been introduced. Although the strategy for extracting data was designed to be comprehensive, possible distortions in the sample must be considered.

#### 7 | CONCLUSIONS

This exploratory analysis of the content of comments uploaded to PubPeer has allowed us to draw several conclusions about how this platform is used. First, over two thirds of the comments are posted to report some type of misconduct, mainly concerning image manipulation. Consequently, PubPeer cannot be the place for many discussions: only 31.5% of publications received more than three comments, and the response rate of authors is very low (7.5%). When discussion did take place, it centered on publications accused of misconduct. Therefore, comments about fraud are the most frequent, attract more posts, and take longer to be uploaded. Anonymity is widely used (85.6%) for all types of comment, even for those that do not require this protection.

According to distribution by research area and discipline, *Health* and *Life Sciences* are the most problematic research areas. Specifically, *Immunology and Microbiology, Biochemistry, Genetics and Molecular Biology,* and *Pharmacology, Toxicology and Pharmaceutics* have more issues with image manipulation. Otherwise, *Social Sciences & Humanities* attract more comments related to critical discussions about methodological and theoretical aspects.

The study about the participation in PubPeer revealed a small group of "super commenters" that use the platform to systematically review publications, mainly regarding image manipulation in biology. The distribution of users shows there are more men than women, although this imbalance does not influence the number of posts by genre.

#### ORCID

José Luis Ortega D https://orcid.org/0000-0001-9857-1511

#### REFERENCES

- Bik, E. M. (2019). PubPeeer—A website to comment on scientific papers. Science Integrity Digest. Retrieved from https:// scienceintegritydigest.com/2019/07/16/pubpeer-a-website-tocomment-on-scientific-papers/
- Bik, E. M., Casadevall, A., & Fang, F. C. (2016). The prevalence of inappropriate image duplication in biomedical research publications. *mBio*, 7(3), e00809-16. https://doi.org/10.1128/mBio. 00809-16
- Björk, B. C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, 7(4), 914–923. https://doi.org/10.1016/j.joi.2013.09.001
- Blatt, M. R. (2015). Vigilante science. *Plant Physiology*, 169(2), 907–909. https://doi.org/10.1104/pp.15.01443
- Bolek, C., Marolov, D., Bolek, M., & Shopovski, J. (2020). Revealing reviewers' identities as part of open peer review and analysis of the review reports. *LIBER Quarterly*, 30(1), 1–25. https://doi. org/10.18352/lq.10347
- Bordignon, F. (2020). Self-correction of science: A comparative study of negative citations and post-publication peer review. *Scientometrics*, 124, 1225–1239. https://doi.org/10.1007/s11192-020-03536-z

- Bouter, L. M., & Hendrix, S. (2017). Both whistleblowers and the scientists they accuse are vulnerable and deserve protection. *Accountability in Research*, 24(6), 359–366. https://doi.org/10. 1080/08989621.2017.1327814
- Brown, B. L., & Hedges, D. (2009). Use and misuse of quantitative methods: Data collection, calculation, and presentation. In D. M. Mertens & P. E. Ginsberg (Eds.), *Handbook of social research ethics* (pp. 373–386). Sage.
- Carreño, L. V. G., & Winbladh, K. (2013). Analysis of user comments: An approach for software requirements evolution. In 35th international conference on software engineering (ICSE), San Francisco, CA, USA (pp. 582–591). IEEE. https://doi.org/ 10.1109/ICSE.2013.6606604
- Cosgrove, A., & Cheifet, B. (2018). Transparent peer review trial: The results. *Genome Biology*, *19*, 206. https://doi.org/10.1186/ s13059-018-1584-0
- Electronic Codes of Federal Regulations (e-CFR). (2019). Public Welfare, National Science Foundation § 689 Research Misconduct et seq. Retrieved from https://www.nsf.gov/oig/\_pdf/cfr/ 45-CFR-689.pdf
- Fox, C. W. (2017). Difficulty of recruiting reviewers predicts review scores and editorial decisions at six journals of ecology and evolution. *Scientometrics*, 113(1), 465–477. https://doi.org/10.1007/ s11192-017-2489-5
- Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer review. *eLife*, 6, e21718. https://doi.org/ 10.7554/eLife.21718
- Herrera, A. J. (1999). Language bias discredits the peer-review system. *Nature*, *397*(6719), 467–467. https://doi.org/10.1038/17194
- Hooge, I. T., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2018). Is human classification by experienced untrained observers a gold standard in fixation detection? *Behavior Research Methods*, 50(5), 1864–1881. https://doi.org/ 10.3758/s13428-017-0955-x
- Horbach, S. P., & Halffman, W. (2018). The changing forms and expectations of peer review. *Research Integrity and Peer Review*, 3, 8. https://doi.org/10.1186/s41073-018-0051-5
- Horbach, S. P., & Halffman, W. (2019). The ability of different peer review procedures to flag problematic publications. *Scientometrics*, *118*(1), 339–373. https://doi.org/10.1007/s11192-018-2969-2
- Kassirer, J. P., & Campion, E. W. (1994). Peer review: Crude and understudied, but indispensable. *Jama*, 272(2), 96–97. https:// doi.org/10.1001/jama.1994.03520020022005
- Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer review in scientific publications: Benefits, critiques, & a survival guide. *Ejifcc*, 25(3), 227–243.
- Knoepfler, P. (2015). Reviewing post-publication peer review. *Trends in Genetics*, *31*(5), 221–223.
- Kumar, M. N. (2008). A review of the types of scientific misconduct in biomedical research. *Journal of Academic Ethics*, 6(3), 211–228. https://doi.org/10.1007/s10805-008-9068-6
- Kuroki, T. (2018). New classification of research misconduct from the viewpoint of truth, trust, and risk. Accountability in Research, 25(7–8), 404–408. https://doi.org/10.1080/08989621. 2018.1548283
- Lamiell, J. T. (2019). Psychology's misuse of statistics and persistent dismissal of its critics. Palgrave Macmillan.
- Malički, M., Costello, J., Alperin, J. P., & Maggio, L. A. (2021). Analysis of single comments left for bioRxiv preprints till

September 2019. Biochemia Medica, 31, 177-184. https://doi. org/10.11613/BM.2021.020201

- Mulligan, A. (2005). Is peer review in crisis. Oral Oncology, 41(2), 135-141. https://doi.org/10.1016/j.oraloncology.2004.11.001
- Nature Communications. (2016). Transparent peer review one year on. Nature Communications, 7, 13626. https://doi.org/10.1038/ ncomms13626
- Oksvold, M. P. (2015). Incidence of data duplications in a randomly selected pool of life science publications. Science and Engineering Ethics, 22(2), 487-496. https://doi.org/10.1007/s11948-015-9668-7
- Ortega, J. L. (2017). Are peer-review activities related to reviewer bibliometric performance? A scientometric analysis of Publons. Scientometrics, 112(2), 947-962. https://doi.org/10.1007/s11192-017-2399-6
- Ortega, J. L. (2019). Exploratory analysis of Publons metrics and their relationship with bibliometric and altmetric impact. Aslib Journal of Information Management, 71(1), 124-136. https:// doi.org/10.1108/AJIM-06-2018-0153
- Ortega, J. L. (2020). The relationship and incidence of three editorial notices in PubPeer: Errata, expressions of concern, and retractions. Learned Publishing, 34(2), 164-174. https://doi.org/ 10.1002/leap.1339
- Pier, E. L., Brauer, M., Filut, A., Kaatz, A., Raclaw, J., Nathan, M. J., & Carnes, M. (2018). Low agreement among reviewers evaluating the same NIH grant applications. Proceedings of the National Academy of Sciences of the United States of America, 115(12), 2952-2957. https://doi.org/10.1073/pnas.1714379115
- PubPeer. (2021). Frequently asked questions. Retrieved from https:// pubpeer.com/static/faq
- Resnik, D. B., & Elmore, S. A. (2018). Conflict of interest in journal peer review. Toxicologic Pathology, 46(2), 112-114. https://doi. org/10.1177/0192623318754792
- Resnik, D. B., & Stewart, C. N. (2012). Misconduct versus honest error and scientific disagreement. Accountability in Research, 19(1), 56-63. https://doi.org/10.1080/08989621.2012.650948
- Ross-Hellauer, T. (2017). What is open peer review? A systematic review. F1000Research, 6, 588. https://doi.org/10.12688/ f1000research.11369.2
- Rothwell, P. M., & Martyn, C. N. (2000). Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone? Brain, 123(9), 1964-1969. https://doi.org/10.1093/brain/123.9.1964
- Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. Decision Support Systems, 81, 30-40. https://doi. org/10.1016/j.dss.2015.10.006
- Segado-Boj, F., Martín-Quevedo, J., & Prieto-Gutiérrez, J. J. (2018). Attitudes toward open access, open peer review, and altmetrics among contributors to Spanish scholarly journals. Journal of Scholarly Publishing, 50(1), 48-70. https://doi.org/10.3138/jsp.50.1.08
- Shen, H. (2020). Meet this super-spotter of duplicated images in science papers. Nature, 581, 132-136. https://doi.org/10.1038/ d41586-020-01363-z
- Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. Journal of the Royal Society of Medicine, 99(4), 178-182.
- Teixeira da Silva, J. A. (2018). The issue of comment ownership and copyright at PubPeer. Journal of Educational Media & Library Sciences, 55(2), 181-191.

- Teixeira da Silva, J. A. (2019). Challenges to open peer review. Online Information Review, 32(2), 197-200. https://doi.org/ 10.1108/OIR-04-2018-0139
- Teixeira da Silva, J. A., & Blatt, M. R. (2016). Does the anonymous voice have a place in scholarly publishing? Plant Physiology, 170(4), 1899-1902. https://doi.org/10.1104/pp.15.01939
- Teixeira da Silva, J. A., & Dobránszki, J. (2015). Problems with traditional science publishing and finding a wider niche for post-publication peer review. Accountability in Research, 22(1), 22-40. https://doi.org/10.1080/08989621.2014.899909
- Thelwall, M., Allen, L., Papas, E. R., Nyakoojo, Z., & Weigert, V. (2020). Does the use of open, non-anonymous peer review in scholarly publishing introduce bias? Evidence from the F1000Research postpublication open peer review publishing model. Journal of Information Science. https://doi.org/10.1177/0165551520938678
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163-173. https://doi.org/10.1002/asi.21662
- Torny, D. (2018). Pubpeer: Vigilante science, journal club or alarm raiser? The controversies over anonymity in post-publication peer review. In PEERE international conference on peer review, PEERECOST network, Mar 2018, Rome, Italy. CCSD.
- Townsend, F. (2013). Post-publication peer review: PubPeer. Editors' Bulletin, 9(3), 45-46.
- Tripathi, M., Sonkar, S., & Kumar, S. (2019). A cross sectional study of retraction notices of scholarly journals of science. DESIDOC Journal of Library & Information Technology, 39(2), 74-81. https://doi.org/10.14429/djlit.39.2.14000
- van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomised trial. BMJ, 318(7175), 23-27. https://doi.org/10.1136/bmj.318.7175.23
- Wager, E., & Veitch, E. (2017). The role of PubPeer comments in alerting editors to serious problems with clinical research publications. In Eighth international congress on peer review and scientific publication, September 2017, Chicago, USA. American Medical Association. Retrieved from https://peerreviewcongress.org/ abstract/the-role-of-pubpeer-comments-in-alerting-editors-toserious-problems-with-clinical-research-publications/
- Walker, R., Barros, B., Conejo, R., Neumann, K., & Telefont, M. (2015). Personal attributes of authors and reviewers, social bias and the outcomes of peer review: A case study. F1000Research, 4, 21. https://doi.org/10.12688/f1000research.6012.2
- Walsh, E., Rooney, M., Appleby, L., & Wilkinson, G. (2000). Open peer review: A randomised controlled trial. British Journal of Psychiatry, 176(1), 47-51. https://doi.org/10.1192/bjp.176.1.47
- Yong, E., Ledford, H., & Van Noorden, R. (2013). Research ethics: 3 ways to blow the whistle. Nature News, 503(7477), 454. https://doi.org/10.1038/503454a

How to cite this article: Ortega, J. L. (2021). Classification and analysis of PubPeer comments: How a web journal club is used. Journal of the Association for Information Science and Technology, 1-16. https://doi.org/10.1002/asi.24568

### APPENDIX A: ESTIMATION OF CASES ACCORDING TO THE ACCURACY TEST

This section includes information to estimate the amount of cases in each category according to the result of the accuracy test. Table A1 displays the distribution of threads, publications, author responses and anonymous posts of the selected subsample classified according to the heuristic and previous to the accuracy test (Predicted); Table A2 includes the same information but now with the reassigned cases after the accuracy test (Actual); and finally, Table A3 shows the estimation/correction of cases from the result of the accuracy test (Estimated). The following formula was used to estimate the results:

SST WILEY

$$\begin{aligned} \text{Estimated} = \left( \left( \frac{\text{Predicted} - \text{Actual}}{\text{Predicted}} \right) \text{Observed} \right) \\ & - \text{Observed}, \end{aligned}$$

where Predicted is the number of cases assigned by the heuristic that were selected to the manual classification (Table A1), Actual is the corrected cases in the manual classification (Table A2) and Observed are the cases observed in the study (Table 3). The estimation was not calculated for categories with 0 cases.

**TABLE A1** Distribution of threads, publications, author responses, and anonymous posts by type of comments in PubPeer based on the initial heuristic (N = 4,000)

Group of comments	Type of comments	Threads	Threads (%)	Publications	Publications (%)	Author response	Author response (%)	Anonymous	Anonymous (%)
Positive review	Positive review	36	1.0	36	0.9	5	13.9	25	69.4
Critical review	Critical review	662	15.2	644	16.4	44	6.8	493	74.5
	Lack of information	84	1.9	81	2.1	10	12.3	70	83.3
Error	Honest error	84	2.7	84	2.1	0	0	0	0.0
	Statistical flaws	281	8.4	277	7.0	15	5.4	239	85.1
Fraud	Publishing fraud	370	9.7	358	9.1	11	3.1	0	0.0
	Manipulation	2,483	61.1	2,450	62.3	267	10.9	2,345	94.4
	Total	4,000		3,930		352	9.0	3,172	79.3

**TABLE A2** Distribution of threads, publications, author responses, and anonymous posts by type of comments in PubPeer after the accuracy test (N = 4,000)

Group of comments	Type of comments	Threads	Threads (%)	Publications	Publications (%)	Author response	Author response (%)	Anonymous	Anonymous (%)
Positive review	Positive review	46	1.2	47	1.2	6	12.8	27	58.7
Critical review	Critical review	638	16.0	620	15.8	45	7.3	462	72.4
	Lack of information	94	2.4	91	2.3	11	12.1	80	85.1
Error	Honest error	98	2.5	96	2.4	10	10.4	83	84.7
	Statistical flaws	312	7.8	304	7.7	16	5.3	273	87.5
Fraud	Publishing fraud	373	9.3	364	9.3	19	5.2	347	93.0
	Manipulation	2,439	61.0	2,409	61.3	252	10.5	2,318	95.0
	Total	4,000		3,931		359	9.1	3,590	89.8

**TABLE A3** Distribution of threads, publications, author responses, and anonymous posts by type of comments in PubPeer, estimated from the accuracy test (N = 24,001)

Group of comments	Type of comments	Threads	Threads (%)	Publications	Publications (%)	Author response	Author response (%)	Anonymous	Anonymous (%)
Positive review	Positive review	242.8	1.0	231.3	1.0	15.6	6.7	82.1	33.8
Critical review	Critical review	3653.6	15.2	3644.3	15.2	227.0	6.2	2508.2	68.7
	Lack of information	445.4	1.9	441.7	1.8	28.6	6.5	382.9	86.0
Error	Honest error	645.7	2.7	637.1	2.7	40.0	6.3	474.0	73.4
	Statistical flaws	2008.6	8.4	1993.5	8.3	81.1	4.1	1614.0	80.4
Fraud	Publishing fraud	2334.8	9.7	2335.1	9.8	152.0	6.5	1969.0	84.3
	Manipulation	14670.3	61.1	14662.2	61.2	1128.8	7.7	12716.9	86.7
	Total	24001.1		23945.2		1673.1	7.0	19747.0	82.5