Retracted Papers with Incorrect Document Type Indexing in PubMed, Scopus, and Web of Science

Randi Proescholdt and Jodi Schneider
randiep2@illinois.edu, jodi@illinois.edu
School of Information Sciences, University of Illinois at Urbana-Champaign

Introduction

Finding all retracted publications in a database is important for bibliometric researchers; meanwhile domain scientists may need to eliminate all retracted publications from search results, or to determine whether a particular publication is retracted. Yet the quality of retraction indexing has received little attention. Donner's study of document type classification in Web of Science (WoS) found that article, review, and letter document types were reasonably accurate yet had shortcomings for bibliometric analysis (2017)—though it does not mention "retraction" nor "retracted publication" document types, which were not introduced to WoS until October 2016 (Clarivate Analytics, 2016). Schmidt (2018) analyzed retraction indexing – both document type and interlinking of retracted publications and retraction notices – using a sample of biomedical articles from 1980-2013 in PubMed and WoS; much of the data for that study was collected in 2015, and a search as of March 2017 found only a small fraction (less than 100) of retraction notices from 2015-2016 indexed as such, and no articles indexed as retracted publications (Schmidt, 2018). We aim to raise awareness of the continuing challenge of quality control issues relating to retracted publications in PubMed, Scopus, and WoS.

Methods

To find possible retracted publications or retraction notices not indexed as such, we searched for "retracted article", a phrase commonly found in retracted publication titles, and refined our search by eliminating retraction-related document types. Search strings are shown in Table 1. To identify results likely to be retracted publications or retraction notices, we first sorted by title. Then we read each title and classified it as a likely retracted publication (or retraction notice), unclear, or about retraction. For WoS we added two additional categories: items with no title exported and items with titles that did not include "retracted article."

Results

Table 1 shows the number and percentage of results as of August 17 (PubMed and Scopus) and August 19 (WoS), 2020, that were likely to be retracted. Despite WoS's plans to begin using the new retraction and retracted publication document types in 2016 and to re-process previously indexed items back to 2008 (Clarivate Analytics, 2016), our search of WoS found missing document types after 2008, as well as after 2016. We also found common patterns in the titles of our search results, as shown in Figure 1. The most common was "RETRACTED ARTICLE:", as well as variations and translations of this phrase in multiple languages.

Table 1: Identifying possible retracted items or retraction notices

	Document types	**Search 1: "Retracted article" in title, but not indexed with retracted publication document type**	**Number of results**	**Likely retracted (percentage)**
PubMed (New PubMed)	Retracted Publication Retraction of Publication Published Erratum Duplicate Publication Corrected and Republished Article	"Retracted Article"[Title] NOT "Retracted Publication"[PT] NOT "Retraction of Publication"[PT] NOT "Published Erratum" [PT] NOT "Corrected and Republished Article"[PT] NOT "Duplicate Publication"[PT]	61	58 (95%)
Scopus	Erratum Retracted	TITLE ("RETRACTED ARTICLE") AND (EXCLUDE (DOCTYPE,"tb") OR EXCLUDE (DOCTYPE,"er"))[1]	8683	8654 (99%)
Web of Science (All databases, all years)	Retracted Publication Correction Retraction Correction, Addition	TI="Retracted article" Refined by: [excluding] DOCUMENT TYPES:(RETRACTED PUBLICATION OR RETRACTION OR CORRECTION)	80	56 (70%)

[1]OR EXCLUDE is the default syntax in this case. Scopus changes any use of AND EXCLUDE to OR EXCLUDE.

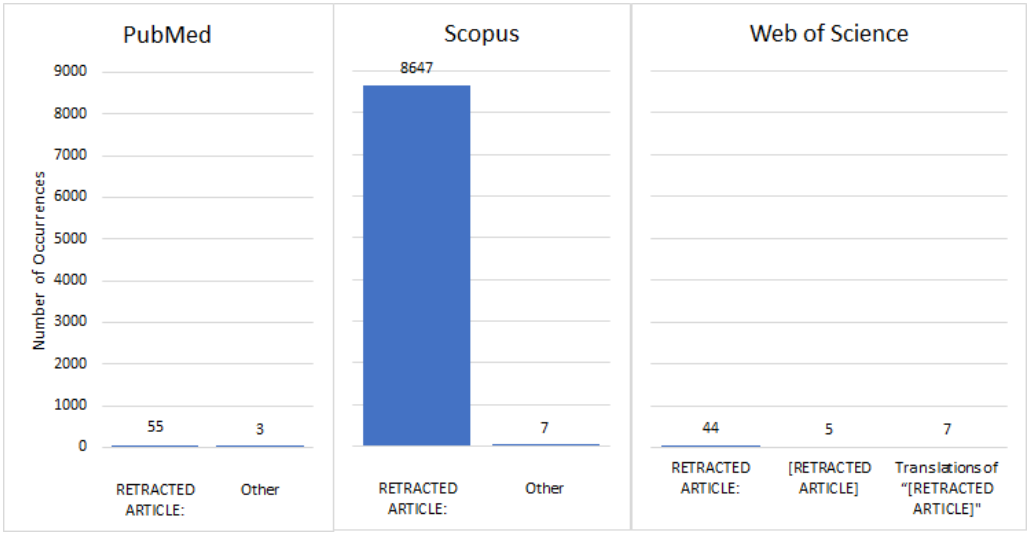

Figure 1: Commonly found phrases in titles of PubMed, Scopus, and WoS results for the searches shown in Table 1

We dug deeper into Scopus errors. 8654 items from Scopus appeared to be retracted publications or retraction notices that were not indexed with the appropriate document types. We were able to confirm 6328 (73%) directly against the Retraction Watch database (2020); spot-checking suggests that the remaining items were also retracted publications or retraction notices.

Most of Scopus' errors (96%; 8308/8654) were in records for IEEE publications; note that IEEE had a large-scale retraction, primarily consisting of conference papers published between 2009 and 2011 (McCook, 2018). In the Retraction Watch matched data, dates were available for 6052 IEEE published items: they were retracted from 2007 to 2017 (i.e., three or more years ago). Scopus is not keeping current with IEEE retractions.

Conclusions

Quality control of document type classification is essential for identifying retracted publications. The poor indexing of retracted publications has broader implications: it is difficult for bibliometric researchers to find all retracted publications; it is difficult for domain scientists to eliminate retracted research from their searches; and it is difficult for domain scientists to reliably identify the retraction status of a single item using PubMed, Scopus, and WoS. Database providers should devote greater attention to quality control; even just correcting document types of those items beginning with "RETRACTED ARTICLE:" or those retracted by IEEE would be an improvement. Meanwhile, because of the reality of imperfections in these databases, we recommend researchers consult multiple sources to confirm whether or not an article is retracted.

Data availability

Data (except the Retraction Watch-matched data) is publicly available at https://osf.io/epd2s/.

Acknowledgements

Alfred P. Sloan Foundation G-2020-12623. Retraction Watch for use of their data.

References

Clarivate Analytics. (2016, October 31). *Web of Science Release Notes v5.23*. Wokinfo.com. http://wokinfo.com/media/pdf/WOS_5-23-external-release-notes.pdf

Clarivate Analytics. (2020). *Searching the Document Type Field*. Web of Science Core Collection Help. https://images.webofknowledge.com/images/help/WOS/hs_document_type.html

Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. *Scientometrics*, 113(1), 219–236. https://doi.org/10.1007/s11192-017-2483-y

Elsevier. (2020). *How do I search for a document?* Scopus: Access and use Support Center. https://service.elsevier.com/app/answers/detail/a_id/11213/c/10545/supporthub/scopus/

McCook, A. (2018). One publisher, more than 7000 retractions. *Science*. 362 (6413), 393. https://doi.org/10.1126/science.362.6413.393

National Library of Medicine. (2020, March 11). Publication Characteristics (Publication Types) with Scope Notes. Nih.gov. https://www.nlm.nih.gov/mesh/pubtypes.html

Retraction Watch Database. (2020, Sept 21). CSV file version. http://retractiondatabase.org/

Schmidt, M. (2018). An Analysis of the Validity of Retraction Annotation in PubMed and the Web of Science. *JASIST*. 69(2):318–328. https://doi.org/10.1002/asi.23913