Malaysian Journal of Library & Information Science, Vol. 26, no. 3, December, 2021: 97-103

Document type assignment by Web of Science, Scopus, PubMed, and publishers to "Top 100" papers

Andy Wai Kan Yeung
Oral and Maxillofacial Radiology,
Applied Oral Sciences and Community Dental Care,
Faculty of Dentistry, University of Hong Kong, Hong Kong, CHINA
e-mail: ndyeung@hku.hk

ABSTRACT

Document type (DT) assignment is an important feature from literature databases. This work evaluated how the literature databases and the publisher websites labeled "Top 100" (T100) papers, a recurring act for which researchers identify and analyze the 100 most cited entities (e.g. articles) within a pre-defined literature set. T100 papers concurrently indexed in the Web of Science (WoS), Scopus and PubMed databses were identified. Among the 248 T100 papers analyzed, no general consensus or consistent pattern was found for labeling T100 papers by the three data sources and the publishers' websites. All four sources labeled between 30–40% of the T100 papers as reviews. However, PubMed mostly did not give DT labels to the rest of the papers whereas WOS, Scopus, and publisher websites labeled them as articles. The inter-rater agreement was only fair; the decision seemed to be influenced by whether the authors mentioned the word "review" suggestive of the publication/document type in the title, abstract or keywords.

Keywords: Article; document type; review; Scopus; PubMed; Web of Science; Top 100; most cited

INTRODUCTION

Document type (DT) assignment is an important feature from literature databases. Bibliometricians may use DT as a filter to collect academic publications for various purposes, including scientometric research and evaluating institutional performance (Donner 2017). Sigogneau reported that, in the early 1990s, the Web of Science (WoS) gradually ceased to label publications as "notes" but "articles" instead (Sigogneau 2000).[1] She also reported an in-depth analysis of "proceedings papers" published in journals related to Physics based on the DT assignment by WoS in the same paper. Other bibliometric studies analyzed so-called original research only and thus filtered out proceedings papers and/or reviews (for selected examples, please see (Harzing 2013)). Besides affecting the inclusion of literature set, inaccurate DT assignment might over- or under-estimate the normalized citation scores during computation (Donner 2017). It is well-known that reviews tended to receive more citations than original research articles (for a comprehensive list of studies that investigated the relationship between DT and

[1] Consistent to the announcement released by WoS. Available at: https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Core-Collection-Document-Type-Note-merged-with-Article?language=en_US

citation, and showing citation bias towards review papers (Tahamtan, Afshar, and Ahamdzadeh 2016). Moreover, having highly cited "editorial materials" which are usually not peer-reviewed might theoretically, though uncommon, inflate the Journal Impact Factor (Campanario et al. 2011).

The author would like to know how the literature databases and the publisher websites labeled "Top 100" (T100) papers, a recurring act for which researchers identify and analyze the 100 most cited entities (e.g. papers) within a pre-defined literature set. In this study, T100 paper is defined as a publication that identified and analyzed the top 100 most cited publications for a specific research topic or field. Such papers have become more and more prevalent in the academic literature, and one even got published in *Nature* (Van Noorden, Maher, and Nuzzo 2014). Interestingly, it was labeled as a "news feature" by the publisher website, a "historical article" and "news" by PubMed, a "news item" by WoS, and a "review" by Scopus. Due to the different DT assignments, the citations received by this publication would contribute to both the numerator and denominator in computing the CiteScore, but only the numerator in computing the Impact Factor. T100 papers are essentially a type of bibliometric reports, and the latter were generally defined as "physical units of publications, bibliographic citations, and surrogates for them" (Broadus 1987, p. 377). As such. the objective of this study is to assess if different databases labelled T100 papers consistently, and if the conductance of statistical analysis would prompt databases to label them as original articles rather than reviews. Since many research evaluation reports limited their data sets to original articles only, the research question driving this study is: *Should T100 papers be treated as original research and hence included into research evaluation reports*?

It was largely unclear if various sources labeled T100 papers as articles or reviews. Though giving different definitions, all PubMed, WoS, and Scopus mentioned that a review should primarily be an examination or summary of existing literature without presenting new or novel findings (Yeung 2019). Would T100 papers with not only descriptive but analytical statistics be considered as presenting new or novel findings? Without much prior knowledge, it was hypothesized that T100 papers would be mainly assigned as articles or reviews in similar ratio. It was also hypothesized that T100 papers with p values listed in their abstract and full text would have a higher ratio of being labeled as articles. The reasoning behind was that papers reporting analytical statistics (hence with p values) would present some new or novel findings from the extracted bibliographic data, which would violate the definition of a review given by PubMed, WoS, and Scopus as described above.

MATERIALS AND METHODS

The PubMed, WoS, and Scopus databases were queried on 22 March 2021. The search looked for ("top 100" OR "100 most") AND (bibliometric*) within the titles, abstracts, and keywords of the indexed publications for WoS and Scopus, and within "all fields" for PubMed. The search yielded 367 papers from PubMed, 374 papers from WoS, and 428 papers from Scopus. During merging the records from PubMed and WoS, four commentaries (i.e. critique/comment on other T100 papers) were removed. Records from the databases were merged, and any duplicates were removed. A total of 257 papers remained. The abstracts of these 257 papers were screened and nine of them were subsequently excluded due to non-English or irrelevance (e.g. the paper was not analyzing the top 100 entities but something like "top 100 keywords were revealed from the data

set"). Finally, 248 papers entered the final analysis. It is cautioned that the analyzed literature set was confined to a portion of T100 papers only, as a quick search in WoS for ("top 100" OR "100 top" OR "100 most* cited" OR "100 top cited") without the word bibliometric* returned with 2397 papers. However, this number was beyond the author's capacity of manual inspection. The current sample size slightly exceeded 10 percent of this number, which should be reasonably representative.

The following parameters were recorded for each of the 248 papers:

a) Whether the word "review" appeared in the title, abstract and keywords. (Yes and DT-indicative; Yes but not DT-indicative (e.g. reporting the number of reviews found from the data set); or No). It was reasoned that the DT label selected by the authors during manuscript submission should be the gold standard. However, it was not possible to obtain such data in this study. Therefore, the second-best option was selected by recording whether the authors used the word "review" to indicate their intended DT (review vs non-review).
b) Whether p value appeared in the abstract. (Yes or No). It was assumed that having a p value would mean that statistical tests have been performed for some original data and hence suggesting that the paper should be an original article.
c) Whether p value appeared in the full text. (Yes or No)
d) DT assignment by PubMed, WoS, Scopus, and publisher websites, respectively. (Review including systematic review; article; editorial; or missing/miscellaneous)[2]. The data from the publishers' websites served as a reference to assess if the databases' data differ hugely from the reference. An article should present new concepts or findings, whereas a review should primarily be an examination or summary of existing literature without presenting new or novel findings.

Statistical analyses were performed with IBM SPSS Statistics 26.0. For statistical tests, rare DTs other than review and article would be recoded into missing/miscellaneous. Chi-squared tests were computed to evaluate if papers listing p values in their abstracts and full texts or papers mentioning the word "review" DT-indicatively would have a significantly higher ratio of being labeled as articles than papers without p values. Fleiss' kappa test were used to evaluate inter-rater agreement of the DT assignments between publisher websites, PubMed, WoS, and Scopus. Tests with $p < 0.05$ were considered statistically significant.

This study involves no animal or human subjects, so ethical approval was not needed.

RESULTS

Among the 248 T100 papers, 57.7 percent (n = 143) did not mention the word "review" in their title, abstract, or keywords, 17.3 percent (n = 43) mentioned "review" to indicate they were reviews, and 25.0 percent (n = 62) mentioned "review" that was non DT-indicative. The vast majority of the papers (89.1%; n = 221) did not list any p values in their abstract. When the full text was inspected, 58.9 percent (n = 146) still did not list any p values, implying that over half of the T100 papers were descriptive (one of the 146 papers claimed significant statistical results without reporting test statistics and p values).

[2] For PubMed, "article" here also included other DTs such as "comparative study", etc; whereas miscellaneous referred to "research support, non-US Gov't", etc. For publisher websites, miscellaneous referred to historical note, commentary, etc.

Interestingly, none of the 248 papers was labeled as "articles" by PubMed. Instead, over two-thirds (67.7%; n = 168) were having missing or miscellaneous DT (mostly missing), whereas 31.9 percent (n = 79) were reviews, and 0.4 percent (n = 1) was an editorial. For WoS, Scopus, and publisher websites, the ratios of review DT were 35.5 percent, 33.9 percent, and 37.1 percent, respectively (Table 1). In short, all four sources labeled between 30–40 percent of the T100 papers as reviews. However, PubMed mostly did not give DT labels to the rest of the papers whereas WoS, Scopus, and publisher websites labeled them as articles.

Table 1: Document Types Assigned to "Top 100" Papers by Various Sources.

Data sources Document Type	PubMed	Web of Science	Scopus	Publisher websites
Review	79 (31.9%)	88 (35.5%)	84 (33.9%)	92 (37.1%)
Article	0	157 (63.3%)	160 (64.5%)	142 (57.3%)
Missing or miscellaneous[a]	169 (68.1%)	3 (1.2%)	4 (1.6%)	14 (5.6%)

[a] Mostly missing for PubMed and mostly miscellaneous for publisher websites.

From Scopus data, one full-length paper was labeled as "proceedings paper" which correctly described that it was published in a special issue related to an academic conference (notwithstanding it could not indicate the content type of this paper). Scopus also labeled another paper published in 2018 as "article in press" which was inaccurate and could not indicate the content type. From publisher websites, two papers were directly labeled as bibliometric research, both of which were published in *International Journal of Ophthalmology* (ISSN 2222-3959 / 2227-4898). Editorial was also a rare DT, with one paper in PubMed, three in WoS, two in Scopus, and one in publisher websites. These rare DTs were recoded into missing/miscellaneous in Table 1 and SPSS during statistical analysis.

Chi-squared test showed that there was no significant relationship between the listing of p value and the DT assignment, regardless of the p value location and DT source (all $p > 0.05$, Table 2). For all databases, mentioning the word "review" DT-indicatively in the paper's title, abstract, or keywords would significantly associate with a higher ratio of being actually assigned to a "review" DT label, relative to counterparts mentioning the word "review" non DT-indicatively or without mentioning it (all $p < 0.05$, Table 3). Meanwhile, Fleiss' kappa showed that there was fair agreement ($\kappa = 0.346$, $p < 0.001$) between the DT assignments by PubMed, WoS, Scopus, and publisher website. They had higher agreement for "article" DT ($\kappa = 0.463$) than "review" ($\kappa = 0.346$) than missing/miscellaneous ($\kappa = 0.192$) (all $p < 0.001$). In overall, 21.9 percent of the papers had identical DT assignment by all four sources (all four agreed), 67.6 percent had three agreed, 10.5 percent had two agreed, and 0 percent had none agreed.

The extracted data in Excel format and the coded SPSS data were uploaded as Supplementary Tables 1 and 2 respectively, and available for download from Figshare (doi: 10.6084/m9.figshare.17104322).

Table 2. Relationship Between Listing of P Value and Document Type Assignment.

Data Sources	Listing of p value															
	PubMed				Web of Science				Scopus				Publisher websites			
	Abstract		Full text		Abstract		Full text		Abstract		Full text		Abstract		Full text	
Document Type	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Review	72	7	45	34	82	6	56	32	77	7	54	30	85	7	58	34
Article	0	0	0	0	136	21	88	69	140	20	89	71	123	19	79	63
Missing/miscellaneous	149	20	101	68	3	0	2	1	4	0	3	1	13	1	9	5
P value		0.484		0.676		0.238		0.493		0.477		0.342		0.345		0.485

Table 3. Relationship Between the Mentioning of "Review" Indicating the Document Type (DT) and the Actual DT Assignment.

Data sources	Mentioning of "review" in the title, abstract, or keywords											
	PubMed			Web of Science			Scopus			Publisher websites		
Actual Document Type	Yes (DT-indicative)	Yes (Not DT-indicative)	No	Yes (DT-indicative)	Yes (Not DT-indicative)	No	Yes (DT-indicative)	Yes (Not DT-indicative)	No	Yes (DT-indicative)	Yes (Not DT-indicative)	No
Review	26	19	34	26	19	43	25	18	41	26	20	46
Article	0	0	0	17	42	98	17	43	100	15	40	87
Missing/miscellaneous	43	43	109	0	1	2	1	1	2	2	2	10
P value			< 0.001			0.006			0.007			0.009

DISCUSSION AND CONCLUSIONS

To the best of the author's knowledge, no report has been published on DT assignment of bibliometric reports, including T100 papers. T100 papers were relatively homogeneous in terms of the article structure as the author queried certain literature databases, identified the 100 most cited relevant papers according to some selection criteria, and evaluated the bibliographic data with or without performing statistical tests. Here it was revealed that DT assignment of T100 papers was mixed: PubMed basically labeled them as reviews, whereas WoS, Scopus, and publisher websites labeled around 60 percent of them as articles and the rest mainly as reviews. Interestingly, the presence of p values did not seem to affect the DT assignment. With p values and thus statistical tests conducted, should the report be considered as a merely summary of existing literature without presenting new or novel findings? This seemed to be philosophical. Meanwhile, it was not possible to obtain the data about the DT label selected by the corresponding authors during manuscript submission. Therefore, it was assessed whether the corresponding authors used the word "review" to indicate their intended DT (review vs non-review). Should this be the gold standard then? Unfortunately, previous literature analyses found that around 20 percent of observational studies reported incorrect study design (not consistent to what was really done) (LeBrun et al. 2020), and >70 percent of articles using the phrase "case series" in their titles were actually mislabeled (Esene et al. 2014; Sargeant et al. 2017). Some authors suggested that DT assignment should be done by the data collectors themselves (Di Girolamo and Reynders 2020), but project-specific DT labels would then be limited to a predefined literature set and not readily accessible by others.

The current study had some limitations. The citation data of the T100 papers were not collected, as many of them were published within these few years so that they did not have enough time to accumulate citations. Therefore, it was not tested if bibliometric papers with "review" DT label were generally more cited than their counterparts with "article" label. Certainly, the research topics covered by the papers might also affect the citations received. These issues should be addressed by future studies that perhaps incorporate a larger sample size that covers beyond the T100 papers.

In conclusion, there was no general consensus or consistent pattern for labeling T100 papers. All four sources labeled between 30–40 percent of the T100 papers as reviews. However, PubMed mostly did not give DT labels to the rest of the papers whereas WoS, Scopus, and publisher websites labeled them as articles. Their inter-rater agreement was only fair. Their decision seemed to be influenced by whether the authors mentioned the word "review" suggestive of the publication/document type in the title, abstract or keywords, but not by the listing of p values in the paper abstract or full text. Various stakeholders who assign or utilize such labels should be aware of this discrepancy between different literature databases. It is arguable that the T100 papers often synthesize a summary of the bibliographic and semantic context of the identified papers, but that summary is different from the one synthesized by conventional review papers that focused on the scientific content. Therefore, the author would like to advocate for a new document type mutually exclusive from articles and reviews, solely for bibliometric papers.

ACKNOWLEDGMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The author declares that he has no conflict of interest.

REFERENCES

Broadus, R.N. 1987. Toward a definition of "bibliometrics". *Scientometrics,* Vol. 12, no. 5-6: 373-379. Available at: https://doi.org/10.1007/BF02016680.

Campanario, J.M., Carretero, J., Marangon, V., Molina, A. and Ros, G. 2011. Effect on the journal impact factor of the number and document type of citing records: a wide-scale study. *Scientometrics,* Vol. 87, no. 1: 75-84. Available at: https://doi.org/10.1007/s11192-010-0333-2.

Di Girolamo, N., and Reynders, R.M. 2020. Characteristics of scientific articles on COVID-19 published during the initial 3 months of the pandemic. *Scientometrics,* Vol. 125, no. 1: 795-812. Available at: https://doi.org/10.1007/s11192-020-03632-0.

Donner, P. 2017. Document type assignment accuracy in the journal citation index data of Web of Science. *Scientometrics,* Vol. 113, no. 1: 219-236. Available at: https://doi.org/10.1007/s11192-017-2483-y.

Esene, I.N., Ngu, J., El Zoghby, M., Solaroglu, I., Sikod, A.M., Kotb, A., Dechambenoit, G. and El Husseiny, H. 2014. Case series and descriptive cohort studies in neurosurgery: The confusion and solution. *Child's Nervous System,* Vol. 30, no. 8: 1321-1332. Available at: https://doi.org/10.1007/s00381-014-2460-1.

Harzing, A.-W. 2013. Document categories in the ISI Web of Knowledge: Misunderstanding the social sciences? *Scientometrics,* Vol. 94, no. 1: 23-34. Available at: https://doi.org/10.1007/s11192-012-0738-1.

LeBrun, D.G., Kocher, M.S., Baldwin, K.D. and Patel, N.M. 2020. How often are study design and level of evidence misreported in the pediatric orthopaedic literature? *Orthopaedic Journal of Sports Medicine,* Vol. 40, no. 5: e385-e389. Available at: https://doi.org/10.1097/BPO.0000000000001470.

Sargeant, J., O'Connor, A., Cullen, J., Makielski, K. and Jones‐Bitton, A. 2017. What's in a name? The incorrect use of case series as a study design label in studies involving dogs and cats. *Journal of Veterinary Internal Medicine,* Vol. 31, no. 4: 1035-1042. Available at: https://doi.org/10.1111/jvim.14741.

Sigogneau, A. 2000. An analysis of document types published in journals related to physics: Proceeding papers recorded in the Science Citation Index database. *Scientometrics,* Vol. 47, no. 3: 589-604. Available at: https://doi.org/10.1023/a:1005628218890.

Tahamtan, I., Afshar, A.S. and Ahamdzadeh, K. 2016. Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics,* Vol. 107, no. 3: 1195-1225. Available at: https://doi.org/10.1007/s11192-016-1889-2.

Van Noorden, R., Maher, B. and Nuzzo, R. 2014. The top 100 papers. *Nature,* Vol. 514, no. 7524: 550-553. Available at: https://doi.org/10.1038/514550a.

Yeung, A.W.K. 2019. Comparison between Scopus, Web of Science, PubMed and publishers for mislabelled review papers. *Current Science,* Vol. 116, no. 11: 1909-1914. Available at: https://doi.org/10.18520/cs/v116/i11/1909-1914.