



Bibliometric analysis of manuscript characteristics that influence citations: A comparison of six major psychiatry journals



Dawood M. Hafeez^{a,*}, Sabeena Jalal^b, Faisal Khosa^b

^a Capital Medical University, No. 10 Xitoutiao, You An Men Wai, Fengtai District, Beijing, 100069, PR China

^b Department of Radiology, Vancouver General Hospital, Vancouver, BC, Canada

ARTICLE INFO

Keywords:

Bibliometrics
Citation rate
Citation
Manuscript
Psychiatry

ABSTRACT

In this study we investigated the characteristics of psychiatry manuscript that influence its citation rate. We conducted a cross-sectional study of published articles ($n = 545$), from January to June 2007, from 6 major psychiatry journals with the highest 5-year impact-factor. Citation count for these articles was retrieved from Web Of Science (by Clarivate Analytics) and 22 article characteristics were tabulated manually. We then predicted the citation rate by performing univariate analysis, spearman rank-order correlation, and multiple regression model on the collected variables. Using spearman rank-order correlation, we found the following variables to have significant positive correlation with citations: abstract character count (r_s and p -value, 0.22 and 0.001 respectively), number of references (0.2, 0.01), abstract word count (0.17, 0.0005), number of pages (0.15, 0.003), open access (0.06, 0.05), study design reported in title (0.04, 0.0001), total number of words (0.03, 0.01) and structured abstract (0.03, 0.0009). In a multivariate linear regression model, the following variables predicted increased citation rates ($p < 0.001$, $R^2 = 0.38$): reporting of study design in title, structured abstract and open access. Editors and authors of psychiatry journals can improve the impact of their journals and articles by utilizing this bibliometric study when assembling their manuscript.

1. Introduction

There are approximately 2.5 million new scientific papers published each year (Ware and Mabe, 2015). From 2011 to 2015, 84,760 articles (approximately 17,000/year) were published in psychiatry journals worldwide (Zhang et al., 2017). With such a high output of publications, many indexing and abstracting services use citation count as an important statistical indicator to estimate article quality, importance and scientific influence (Agarwal et al., 2016). The influential scientific psychiatry articles are important to authors, institutions and journals, as well as funding agencies. As one example, It provides them information about the evolution and foci of research in the field (Wu and Duan, 2015; Naveed et al., 2017). Thus, it is no surprise that the topic of bibliometrics has gained significant interest in the scientific community (Agarwal et al., 2016). The quality of work itself could be the underlying strength in its impact on the scientific world. However, there can be situations where novel scientific works may not capture the attention due to less than optimal manuscript assembling. A manuscript without capturing the attention of other researchers will not be cited and has the potential to get lost in the pool of publications.

Studies across different fields have shown that, in addition to the

merit of the scientific work, there are several other factors that influence the citation count of articles (Onodera and Yoshikane, 2015; Shekhani et al., 2017; Antoniou et al., 2015; Khan et al., 2017; Falagas et al., 2013). These factors include characteristics related to authors, journals, and the article itself. Furthermore, because citation count depends on field-dependent citation practices, the characteristics influencing it also differ between each field (Moed et al., 1985). Therefore, the aim of our study was to determine the impact of various manuscript characteristics on the citation count of publications in the field of psychiatry. Our work has the potential to provide authors with knowledge on how to improve the visibility of their research and improve the chances of their publications being cited.

2. Material and methods

This retrospective study was exempt from Institutional Review Board since the study includes publically available data. To have an analysis with the greatest power, we selected high-impact-factor journals with the most visibility and citations. Journal Citation Report (by Clarivate Analytics) was used to select top-six high Impact Factor (IF) psychiatry journals; Journal of American Medical Association (JAMA)

* Corresponding author. Permanent address: 4724 Fort Knox Ct., Orlando, FL, 32822, USA.

E-mail addresses: daudm15@gmail.com (D.M. Hafeez), Sabeena.Jalalkhan@vch.ca (S. Jalal), fkhosa@hotmail.com (F. Khosa).

Table 1
Study characteristics and definitions.

Characteristics No.	Study Characteristic	Characteristic Definition
1	Psychiatry Journals	Top six major psychiatry journals with highest impact factor
2	Journal impact factor	The journal's impact factor according to JCR
3	Journal 5-year impact factor	The journal's 5-year impact factor according to JCR
4	Journal Origin	The continent the journal originated from
5	WOS citation number	No. of citations per article according to WOS
6	Year of Publication	Year the article was published in the journal
7	Words per title	No. of words in the article title
8	Characters per title	No. of characters in the article title
9	Study design stated in the title	Study design reported in the title of the article
10	Study results stated in the title	Study results reported in the title of the article
11	Punctuation in the title	Punctuation is present in the title
12	No. of authors	No. of authors in the article
13	Psychiatry authors	Are all authors from a psychiatry background?
14	Multi-institutional	Is the article a multi-institutional study?
15	Country of origin	The country that the article originated from, defined as the location of the institution of the first author
16	Design of study	Is the study design prospective, retrospective, review or crossover?
17	Sample size	Sample size (no.) included in the article
18	Abstract word count	Word count of the abstract
19	Abstract character count	Character count of the abstract
20	Structured abstract	Is the abstract one large paragraph, or is it divided into sections (i.e. Objective, Materials and Methods, Results, and Conclusion)?
21	Open access	Is the article open access?
22	Funding	Was there any funding for the study?
23	Total words	Total no. of words in the article
24	No. of references	Total no. of references in the article
25	No. of pages	Total no. of pages in the article
26	No. of figures	Total no. of figures in the article
27	No. of tables	Total no. of tables in the article

Table 2
Journal characteristics.

Journals	5 year Journal impact factor	1 year Journal impact factor	Median# of Citations	75th Percentile for Citations
JAMA	16.105	15.976	103	231.5
American Journal of Psychiatry	10.021	9.127	93.5	146
Molecular Psychiatry	9.671	10.9	80	128
Biological Psychiatry	8.668	8.456	55	89
Neuropsychopharmacology	6.347	6.157	43	65.5
British Journal of Psychiatry	6.059	5.446	53	92

(IF, 15.976), Molecular Psychiatry (10.9), American Journal of Psychiatry (9.127), Biological Psychiatry (8.456), Neuropsychopharmacology (6.157), and British Journal of Psychiatry (5.446). We derived the data from the list of all original research and review articles published in these journals between January 1, 2007 and June 30, 2007. Lee et al. reported in his study that peak citation rate occurs 4 years after publication in disciplines as a whole with a continued influence that gradually declines (Lee et al., 2005). Therefore, we selected the year 2007, as 10 years was enough time for a meaningful citation count to build up. A longer citation window would not only be unnecessary, it would also affect the relevance of our research to current practice. We employed Web Of Science (by Clarivate Analytics) to tabulate citation count for each article. The characteristics of the manuscript we analyzed and its definitions are shown in Table 1. The characteristics of journals included in this study are illustrated in Table 2. We organized the country of origin of article into continents.

For data analysis, we used STATA version 14.2 (StataCorp, 2015). Median and ranges were reported for quantitative variables. Frequency and percentages were reported for categorical variables.

2.1. Univariate analysis

At Univariate level, we applied simple linear regression. Each variable was regressed independently with Citations, their assumption

was checked and significance was noted. Regression coefficient with 95% confidence interval was noted. A significance level < 0.25 was considered as significant. The variables which were statistically insignificant (p value > 0.25) were excluded: results in the title (p -value = 0.31), number of authors (0.67), all authors belonging to the field of psychiatry (0.49), multi-institutional study (0.28), country of origin (0.37), sample size (0.29), abstract word count (0.53), funding (0.47), total number of words (0.38), number of pages (0.32), references (0.39), figures (0.56), tables (0.63), total words in paper and in title (0.83). Statistically significant variables were selected for inclusion into multivariate linear regression analysis.

2.2. Multivariable Analysis

A step-wise approach was used with assumption of p -value < 0.05 to be statistically significant. All factors that were considered significant at the univariate level were run through a multiple linear regression model, which includes assessment of Multi-collinearity of independent variable and were assessed by correlation coefficient. For two ordinal variables, Kendall Tau test was used, for two nominal variable, Cramer's V test was used, for one nominal and one ordinal variable, Eta test was used, for one continuous variable and one ordinal variable, Spearman test was used and for one continuous and one nominal variable, Eta test was used for assessing the multi-collinearity among independent variables. A correlation of more than 0.8 was treated as presence of Multi-collinearity. Presence of plausible interactions was assessed in the main model. The model was also assessed for presence of outliers and influential observations; corrections were made for influential observations.

3. Results

3.1. Study characteristics

In our analysis, we included data from 545 articles, published between January 2007 and June 2007: 60 (11%) articles were published in JAMA, 49 (9%) in Molecular Psychiatry, 82 (15%) in American

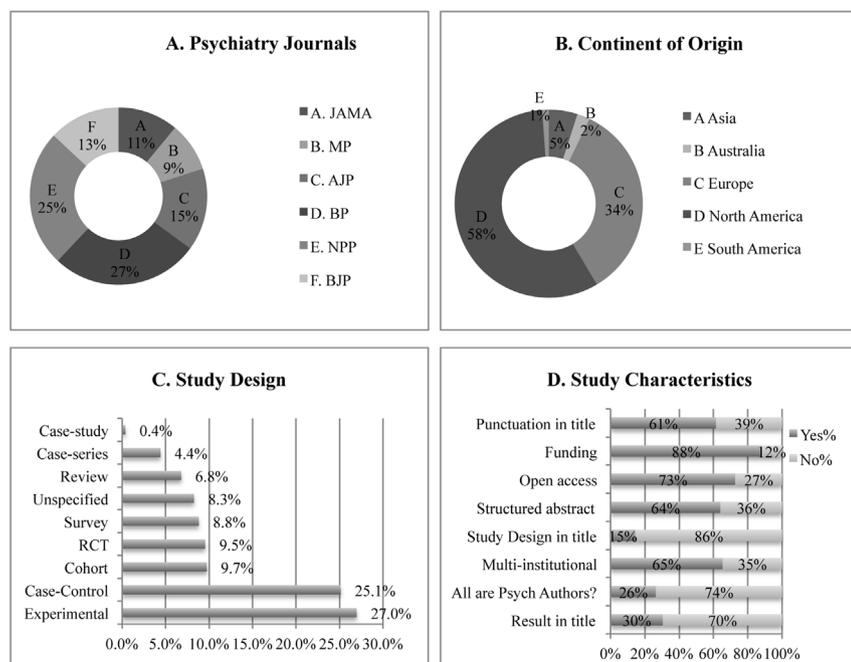


Fig. 1. Characteristics of 545 articles published in top six psychiatry journals between Jan 2007 to June 2007 that we evaluated. Parts A–D: (A) Shows percentage of articles by psychiatry journals (JAMA = Journal of American Medical Association, MP = Molecular Psychiatry, AJP = American Journal of Psychiatry, BP = Biological Psychiatry, NPP = Neuropsychopharmacology, BJP = British Journal of Psychiatry), (B) continent of origin, (C) study design, and (D) study characteristics. Percentages may not add to 100 due to rounding.

Table 3
Distribution of the median, range and 75th percentile of the Study Variables.

Variables	Median	Range	75th Percentile
Citations	61	0–1918	107
No. of authors per article	6	1–32	8
Sample size	104	0–1e7	535
Abstract word count	220	86–717	225
Abstract character count	1568.5	572–5374	1818
Total no. of words	6874	1762–27889	8172
References	47	6–383	63
No. of pages	9	3–31	10
No. of figures	3	0–15	5
No. of tables	2	0–10	4
No. of words in title	14	1–36	17
No. of characters in title	107	29–253	128

Journal of Psychiatry, 147 (27%) in Biological Psychiatry, 136 (25%) in Neuropsychopharmacology, and 71 (13%) in British Journal of Psychiatry. Post hoc analysis was done using the sample size to determine the power of the study. The G Power shows that a power of 1.00 was achieved. We organized the country of origin of article into continents, in which North America constituted 58% of the publications followed by Europe with 34%, Asia 5%, Australia 2% and South America 1%. 85% and 70% of the articles did not report study design and study results in the title, respectively. $n = 356$ studies (65%) were multi-institutional whereas $n = 189$ (35%) were single-center investigations. Approximately three-quarter, $n = 396$ (73%), of the articles reviewed were open access, and 480 studies (88%) acknowledged funding. The median number of authors was 6 (interquartile range [IQR], 4–8), and the median number of references was 47 (IQR, 36–63). Fig. 1 and Table 3 summarizes the study characteristics.

3.2. Citation analysis

The median number of citations received per article was 61 (IQR, 35–106). Spearman rank-order correlation was used to determine the relationship between various study variables with citations (Table 4). The following variables, arranged in descending strength, had statistically significant positive correlation with citations: Abstract Character Count (ACC) (r_s and p -value, 0.22 and 0.001 respectively), number of references (0.2, 0.01), Abstract Word Count (AWC) (0.17, 0.0005),

Table 4
Relationship between number of Citations and Study Variables.

Variables	rs	Correlation	p-value
Authors	0.074	Positive	0.13
References	0.2	Positive	0.01
Sample Size	0.038	Positive	0.4
AWC	0.17	Positive	0.0005
ACC	0.22	Positive	0.0001
Words	0.03	Positive	0.01
Pages	0.15	Positive	0.003
Figures	-0.0155	Negative	0.75
Tables	0.08	Positive	0.07
Words in Title	0.02	Positive	0.7
Characters in Title	0.003	Positive	0.9
Result in title	0.0092	Positive	0.08
Multi-institutional	0.002	Positive	0.7
Study Design in title	0.04	Positive	0.0001
Structured abstract	0.03	Positive	0.0009
Open access	0.06	Positive	0.05
Funding	0.0011	Positive	0.4
Punctuation present in title	0.007	Positive	0.5

Note– AWC = Abstract Word Count. ACC = Abstract Character Count. The r_s value is the Spearman rank-order correlation coefficient. Boldface indicates a statistically significant difference, which was defined by $p < 0.05$.

number of pages (0.15, 0.003), total number of words (0.3, 0.01), open access (0.06, 0.05), study design reported in title (0.04, 0.0001), and structured abstract (0.03, 0.0009).

The multivariable analysis was conducted and best model fit supported the inclusion of open access (p -value = 0.03), study design stated in title (0.001), structured abstract (0.002), and study design (0.046) in the model.

3.3. The model

3.3.1. Regression equation

Y (Citations) = $\beta_0 + \beta_1$ (study design stated in title) + β_2 (structured abstract) + β_3 (open access) + β_{41} (Experimental) + β_{42} (Case Study/series) + β_{43} (unspecified) + β_{44} (Case Control) + β_{45} (RCT) + β_{46} (Cohort) + β_{47} (Review and Meta Analysis).

This prediction equation accounted for major variability in the model as adjusted R square = 0.38, F test was 91.04, and p -value was

≤ 0.001.

We also studied correlation of study design with the citation rate. By applying Kruskal-Wallis test we found that the following study designs had a significant difference in the number of citations, in decreasing order: Review & Meta-analysis (Median no. of citations = 105, Range = 11–1918), Cohort (81, 15–441), Cross-sectional study & Survey (77.5, 21–1704), RCT (76.5, 3–595), Case-Control (68, 6–606), Unspecified (48, 8–301), Case-Study/Series (44.5, 7–356), and Experimental (44, 401).

4. Discussion

According to the results of our study, the statistically significant positive correlation with citations was frequently seen in variables pertaining to the length of the article. These variables were ACC, references, AWC, and number of pages and words. It may signify that longer articles may contain more ideas, information, and greater diversity of results attracting more opportunity for higher citation rate. We observed that the top psychiatry journals with impact factor ranking first and second, JAMA and American Journal of Psychiatry, allow up to 350 and 400 words, respectively, in the abstract as compared to 250 word limit in third, fourth, fifth and sixth ranking journal. Editors of lower ranked journals may perhaps update their guidelines for authors by offering flexibility in regards to the length of abstract.

While number of tables and figures also pertain to length of article but in our study their statistical significance was not observed. Similarly, although weak positive correlation, no statistical significance was seen in our study for an increased number of authors and multi-institutional studies. Whereas, Aksnes has reported that, generally, team-authored studies and multi-institutional collaborations get cited more (Aksnes, 2003).

Looking beyond the length of article, structured abstract significantly correlated with high citation count. Structured abstract potentially makes it easier and quicker for readers, authors and editors to navigate through, find and select appropriate articles. Because abstract formats are influenced by publishing journal, authors can choose to publish their studies in Journals that have adopted this quality standard. In 1993, the International Committee of Medical Journal Editors recommended the use of structured abstracts in the “Uniform Requirements for Journals Submitted to Biomedical Journals” (“Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors,” 1993). Editors may also benefit their journal and authors by supporting and promoting the use of structured abstracts.

We also found a significant positive correlation of study designs and its reporting in the title with citations. Most study designs reported in the title were meta-analysis, randomized controlled trials, surveys, cohorts or reviews. Such study designs are higher up in hierarchy of evidence (Gordis, 2014). Our study has demonstrated just as Patsopoulos et al. concluded in their study that “the citation impact of various study designs is commensurate with most proposed hierarchies of evidence” (Patsopoulos et al., 2005). Reporting study design in the title informs “readers about the contents and merits of articles, and the ability of titles to attract potential readers (“marquee appeal”) (Goodman et al., 2001). This may explain that these study designs and their reporting in title attracted scientific community seeking higher level of evidence and hence increased the number of citations. Psychiatry authors of manuscripts working with such study designs may consider reporting it in their title. This may attract readers, increase visibility of the work and hence the potential for higher citation.

In our study, open access journal also had a significant positive correlation with higher citation count. Manuscripts with this characteristic are readily searchable, available and easily accessible, allowing the article greater visibility and audience and hence higher chances for citations. In contrast, open access status did not affect the citation count of ophthalmology articles (Lansingh and Carter, 2009).

This could be suggesting, again, the field-dependent citation practices and difference in readership of psychiatry journals.

In conclusion, both editors of psychiatry journals and authors may potentially improve the impact of their journals and articles by utilizing the results of our bibliometric study when assembling their manuscript. Journals may benefit by recommending to the authors to provide structured abstracts, and allowing flexible length in abstract word count. Psychiatry authors may consider writing detailed manuscripts with more references, reporting study designs in title, writing structured and longer abstracts, and opting to publish in open access and/or high-impact factor journals to have increased potential for higher citation rate. In the era of digital research and massive publication productions, bibliometric characteristics should be considered when editing and assembling a manuscript. Lastly, we are not suggesting that the article assembly trumps science, but rather the method of presentation can complement and supplement science and has the potential to increase citation.

4.1. Limitations

Our study may be subjected to selection bias because we selected only high impact factor journals. Results, therefore, may not be representing all psychiatry journals and articles. We analyzed studies published in the year 2007, giving citations 10 years to build up for this analysis. The statistically positive spearman correlations in our study were weak to very weak. Stronger correlations with citation would be better explained by scientometric analysis, which is a science of measuring and analyzing science itself. Factors such as research area, topics and time period may influence citations strongly. This can be visualized in Naveed et al.'s scientometric study in which he demonstrated publication trends and citation output in 4 consecutive decades. His study showed that some topics in certain time periods attracted more interest than other subjects. This was explained through the evolution and change in foci of research in the field with new discoveries in epidemiology, assessment techniques, genetics, neurosciences and therapeutics (Naveed et al., 2017).

If we had included more journals in our bibliometric study, the correlation between manuscript characteristics and citations may have become stronger due to increased variability in the data and lesser effect of outliers. There was twice as much difference between the median number of citations of first ranking journal as compared with 3rd, 4th and 5th ranking journals, a potential for skewed data results. The effect of skewed data on correlation will be greater in smaller dataset than in the larger one (Goodwin and Leech, n.d.). Although our sample size, $n = 545$ from 6 journals, and the time frame, 6 months, from which data was extracted, may seem appropriate but compared to approximately 17,000 articles published annually in psychiatry, inclusion of more journals, a larger sample size and a longer time frame would have the potential to yield more powerful results. Because the data was extracted manually for each individual article, a potential for human error must be addressed. Finally, having a statistically significant data does not necessarily mean that the correlation is representative of causation.

Contributors

- Dawood Hafeez MBBS: Conducted data collection and authored the manuscript.
- Sabeena Jalal MBBS, MSc: Performed statistical analysis of the data and its interpretation.
- Faisal Khosa MD, MBA: conceptualized the idea for this project and oversaw from the start to the finish including several revisions of the manuscript.

Funding

This research did not receive any specific grant from funding

agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

None.

Declaration of Interest

None.

Disclosure

All authors have approved the final article.

Dr Khosa is the recipient of Vancouver Coastal Health – Healthcare Hero Award (2018).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jpsychires.2018.07.010>.

References

- Agarwal, A., Durairajanayagam, D., Tatagari, S., Esteves, S.C., Harlev, A., Henkel, R., Roychoudhury, S., Homa, S., Puchalt, N.G., Ramasamy, R., Majzoub, A., Ly, K.D., Tvrda, E., Assidi, M., Kesari, K., Sharma, R., Banihani, S., Ko, E., Abu-Elmagd, M., Gosalvez, J., Bashiri, A., 2016. Bibliometrics: tracking research impact by selecting the appropriate metrics. *Asian J. Androl.* 18, 296–309. <https://doi.org/10.4103/1008-682X.171582>.
- Aksnes, D.W., 2003. Characteristics of highly cited papers. *Res. Eval.* 12, 159–170. <https://doi.org/10.3152/147154403781776645>.
- Antoniou, G.A., Antoniou, S.A., Georgakarakos, E.I., Sfyroeras, G.S., Georgiadis, G.S., 2015. Bibliometric analysis of factors predicting increased citations in the vascular and endovascular literature. *Ann. Vasc. Surg.* 29, 286–292. <https://doi.org/10.1016/j.avsg.2014.09.017>.
- Falagas, M.E., Zarkali, A., Karageorgopoulos, D.E., Bardakas, V., Mavros, M.N., 2013. The impact of article length on the number of future citations: a bibliometric analysis of general medicine journals. *PLoS One* 8, e49476. <https://doi.org/10.1371/journal.pone.0049476>.
- Goodman, R.A., Thacker, S.B., Siegel, P.Z., 2001. What's in a Title? A descriptive study of article titles in peer-reviewed medical. *J. Sci. Educ.* 24, 77.
- Goodwin, L.D., Leech, N.L., n.d. Understanding correlation: factors that affect the size of r. *J. Exp. Educ.* <https://doi.org/10.2307/20157427>.
- Gordis, L., 2014. *Epidemiology, fifth ed.* Elsevier/Saunders, Philadelphia, PA.
- Khan, M.S., Usman, M.S., Fatima, K., Hashmani, N., Siddiqi, T.J., Riaz, H., Khan, A.R., Khosa, F., 2017. Characteristics of highly cited articles in interventional cardiology. *Am. J. Cardiol.* 120, 2100–2109. <https://doi.org/10.1016/j.amjcard.2017.08.030>.
- Lansingh, V.C., Carter, M.J., 2009. Does open access in ophthalmology affect how articles are subsequently cited in research? *Ophthalmology* 116, 1425–1431. <https://doi.org/10.1016/j.ophtha.2008.12.052>.
- Lee, J.D., Cassano-Pinche, A., Vicente, K.J., 2005. Bibliometric analysis of Human Factors (1970-2000): a quantitative description of scientific impact. *Hum. Factors* 47, 753–766. <https://doi.org/10.1518/001872005775570970>.
- Moed, H.F., Burger, W.J.M., Frankfort, J.G., Raan, A.F.J., Van, 1985. The application of bibliometric indicators: important field- and time-dependent factors to be considered. *Scientometrics* 8, 177–203. <https://doi.org/10.1007/BF02016935>.
- Naveed, S., Waqas, A., Majeed, S., Zeshan, M., Jahan, N., Haaris Sheikh, M., 2017. Child psychiatry: a scientometric analysis 1980-2016. *F1000Res.* 6 (1293). <https://doi.org/10.12688/f1000research.12069.1>.
- Onodera, N., Yoshikane, F., 2015. Factors affecting citation rates of research articles. *J. Assoc. Inf. Technol.* 66, 739–764. <https://doi.org/10.1002/asi.23209>.
- Patsopoulos, N.A., Analatos, A.A., Ioannidis, J.P.A., 2005. Relative citation impact of various study designs in the health sciences. *Jama-Journal Am. Med. Assoc.* 293, 2362–2366. <https://doi.org/10.1001/jama.293.19.2362>.
- Shekhani, H.N., Shariff, S., Bhulani, N., Khosa, F., Hanna, T.N., 2017. Bibliometric analysis of manuscript characteristics that influence citations: a comparison of six major radiology journals. *Am. J. Roentgenol.* 209, 1191–1196. <https://doi.org/10.2214/AJR.17.18077>.
- StataCorp, 2015. *STATA Statistical Software* [14.2].
- Uniform requirements for manuscripts submitted to biomedical journals, 1993. International committee of medical journal editors. *J. Am. Med. Assoc.* 269, 2282–2286.
- Ware, M., Mabe, M., 2015. *The STM Report an Overview of Scientific and Scholarly Journal Publishing.* (The Hague, Netherlands).
- Wu, Y., Duan, Z., 2015. Analysis on evolution and research focus in psychiatry field. *BMC Psychiatr.* 15 (105). <https://doi.org/10.1186/s12888-015-0482-1>.
- Zhang, J., Chen, X., Gao, X., Yang, H., Zhen, Z., Li, Q., Lin, Y., Zhao, X., 2017. Worldwide research productivity in the field of psychiatry. *Int. J. Ment. Health Syst.* 11 (20). <https://doi.org/10.1186/s13033-017-0127-5>.