Bibliometric Analysis of Manuscript Characteristics that Influence Citations: A Comparison of Six Major Family Medicine Journals

Hamza Paracha[1] MD; Amit Johal[2] MD; Maida Tiwana[3] MBBS, MPH; Dawood M Hafeez[4] MBBS; Sabeena Jalal[6] MBBS, MSc; Ateeq Rehman MD, FACP[5]; Syed Asim Hussain[7] MD CCFP; Faisal Khosa[6] MD, MBA

[1] International American University College of Medicine, Saint Lucia
[2] Resident Physician, Mercyhealth, Rockford, IL, USA
[3] University of Cincinnati, Cincinnati, Ohio, USA
[4] Cornell University, Ithaca, NY, USA
[5] Department of Internal Medicine, University of Wisconsin, Madison, Wisconsin
[6] Department of Radiology, Vancouver General Hospital, Vancouver, British Columbia, Canada
[7] Department of Family Medicine, University of Calgary, Calgary, Canada

Corresponding Author:
Hamza Paracha MD
International American University College of Medicine
PO Box 615, Gable Woods South
Vieux Fort, Saint Lucia, West Indies
Email: h.paracha@live.ca
Phone: +1 (587) 707-4269

Funding Statement/Disclosures:
Dr. Khosa is the recipient of the Canadian Association of Radiologists – Young Investigator Award (2019); Rising Star Exchange Scholarship Program of French Society of Radiology (2019) and Humanitarian Award of Association of Physicians of Pakistani Descent of North America (2019). The authors have no relevant disclosures.
Word count: 2459

ABSTRACT

Objective: The premise of our study was to investigate the characteristics of family medicine (FM) manuscripts that predicted to affect its citation rate.

Design: We conducted a cross-sectional study of published articles (n = 191), from January to June 2008, from 6 major FM journals with the highest impact factor. Annals of Family Medicine (IF = 1.864), British Journal of General Medicine (1.104), Journal of American Board of Family Medicine (1.015), Family Practice (0.976), BMC Family Practice (0.815) and Canadian Family Physicians (0.283). Citation counts for these articles were retrieved using Web of Science filter on SCImago and 26 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.

Results: Using spearman rank-order correlation, we found the following variables to have significant positive correlation with citations: number of references (r_s and p-value, 0.21529 and 0.0028 respectively), total words (0.22979, 0.0014), number of pages (0.22124, 0.0021), study design in title (0.03, 0.0001), structured abstract (0.06, 0.005), and open access (0.03, 0.003). 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 the title, structured abstract and open access.

Conclusion: Editors and authors of FM can enhance the impact of their journals and articles by utilizing this bibliometric study when assembling their manuscript.

Key Words: Bibliometrics; citation rate; citation; manuscript; Family Medicine

INTRODUCTION

There are about 2.5 million English and non-English scientific papers published per year. With such a high output of publications it is important for authors that their published papers be considered significant in the scientific community.[1] One measure of an article's impact on the scientific community is the number of citations the article receives.[2] Authors, institutions, and journals are interested in publishing articles that have scientific influence, therefore, making bibliometrics an emergent topic in the field of family medicine (FM). The scientific rigor and significance of the article is an important factor that determines how often an article is cited. However, there can be situations where scientific papers of great significance may not attract the recognition they merit due to the specifics of how a manuscript is assembled. Furthermore, because citation count is contingent on field-specific trend and citation practices, the characteristics influencing it also differ between those fields of medicine.[3] Hence, the purpose of our study was to determine various manuscript characteristics in FM and their impact on citation count. Our work has the potential to provide authors, journals, and institutions with knowledge on how to enhance their research's visibility and improve the chances of its citation.

MATERIAL AND METHODS

We selected high-impact-factor journals with the most visibility and citations. Using the Web of Science filter on SCImago we selected six high Impact Factor (IF) FM journals; Annals of Family Medicine (IF = 1.864), British Journal of General Medicine (1.104), Journal of American Board of Family Medicine (1.015), Family Practice (0.976), BMC Family Practice (0.815) and

Canadian Family Physicians (0.283). We derived the data from the list of all original research and review articles published in these journals between January 1, 2008 and June 30, 2008. As reported in a previous study by Lee et al., 90.7% of papers are not cited within the first year following publication.[4] He also mentions that the peak citation occurs 4 years after publication with continued influence and then a gradual decline that could take up to 20 years.[4] Thus, the data selected from 2008 would allow enough time for a meaningful citation count to build up and provide a substantial data for our analysis. We employed Web of Science (Clarivate Analytics) to tabulate citation count for each article. The characteristics of the manuscripts 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. We excluded editorials, case reports, letters to the editor, and commentaries. We compared 26 characteristics among the six FM journals with the highest impact factors.

Analyses Description

We applied simple linear regression. Each variable was regressed independently with citations. Regression coefficient with 95% confidence interval was noted. A significance level < 0.25 was considered as significant. The variables that were not significant (p value >0.25) were excluded: number of authors (p-value = 0.85), figures (0.39), multi-institutional study (0.8), sample size (0.68), and abstract word count (0.26). The significant variables would be total number of words (0.0014), number of pages (0.02), references (0.0028), tables (0.09) and total words in title (0.9). Statistically significant variables were selected for inclusion into multivariate linear regression analysis.

For multivariate analysis we used an assumption of *p*-value < 0.05 as statistically significant. All factors that were considered significant at the univariate level were run through a multiple linear regression model. Spearman rank order correlation was used to see the relation between the study variable and citation. A correlation of more than 0.8 was treated as presence of multi-collinearity. Positive interactions were assessed along with the presence of outliers.

RESULTS:

Study Characteristics:

In our analysis, we included data from 191 articles. Altogether, there were six journals in which these articles were published. The data was collected from January 1, 2008 till June 3, 2008. 23 articles were taken from Journal of the American Board of Family Practice (12.04% of total), 33 from Annals of Family Medicine (17.28%), 36 from British Journal of General Practice (18.85%), 40 from BMC Family Practice (20.94%), 20 from Canadian Family Physician (10.47%) and 39 from Family Practice (Oxford Press) (20.42%). The G Power shows that a power of 1.00 for the study was obtained.

A total of 72.77% articles did not report study design in the title. Approximately, 86.91% of the studies were multi institutional studies and the rest of the 13.09% were single center studies. With respects to continent of origin of articles, 53.92% are from Europe, 40.31% are from North America, 3.66% are from Australia and 1.04% are from Asia. In a country-wise breakdown for

the number of articles, 52 were from UK, 47 from US and 21 from Canada with Australia at 7 and Asia at 2. Approximately 75.91% (145) of the articles noted funding.

The median number of authors is approximately 5, whereby the IQR {Q3-Q1=9.5-3.5}, interquartile range for number of authors is 6. Out of 191 articles, 30 (15.71%) had no FM authors; 161(84.29%) had reported "yes" as FM authors. Fig 1 and Table 3 summarizes the study characteristics.

According to the results of our study, the statistically significant positive correlation with citations was seen in variables pertaining to the length of the articles. These variables were, **total words** (r_s = 0.22979, *p*-value = 0.0014), **number of references** (0.21529, 0.0028), and **number of pages** (0.22124, 0.0021). Moreover, **study design in title** (0.03, 0.0001), **structured abstract** (0.06, 0.005), and **open access** (0.03, 0.003) also revealed statistically significant positive correlation.

Following variables showed positive correlation with citations but have no statistical significance: **words per title** (r_s = 0.12152, *p*-value = 0.0940), **characters per title** (0.12757, 0.0786), **number of authors (**-0.01371, 0.8507), **sample size** (0.03032, 0.6870), **results stated in title** (0.13, 0.06), **abstract word count** (0.08247, 0.2567), **abstract character count** (0.10596, 0.1446), **number of figures (**0.06240, 0.3911), **number of tables** (0.12297, 0.0901), **multi-institutional** (0.003, 0.8), **funding** (0.0011, 0.4), and **punctuation in title** (0.007, 0.5). The median number of citations received per article was 19.

Simple Linear Regression,

The simple linear regression was performed and found that citation count was statistically significant with two other variables: total words with ***p*-value = 0.0014** and total number of references with ***p*-value = 0.0028**

A multiple regression was run to predict citation numbers from multiple variables mentioned, including number of references, total words per title, and characters per title. These variables added statistical significance to the prediction.

All in all, if we take 0.05 as the benchmark *p*-value anything lesser than 0.05 would be considered significant. In terms of correlation, the closer the value of 'r_s' to 1, the better is the correlation. Anything closer to "0" can be considered as weak association. Anything closer to -1 is negative association.

The median number of authors was 5, the median number of references was 27 and the median number of characters in title is 96 as shown in Table 3. The data set is not evenly distributed but the median still gives us a good measure of the distribution of data. The range of all the variables tell us about the spread of the data and it includes the outliers. Our data is widely distributed so outliers are there. The median ranges of all the variables predict how these will affect the number of citations. As shown in the table. The statistically significant variables are those that pertain to the length as shown by the results earlier. The total number of words and the word count are positively related to number of citations.

The median number of citations and the category of being open access or not had a positive influence on number of citations. The percentage of open access is 81.15%. The regression model accounted for the major variability in the model. The study design has a positive effect on number of citations. In order of decreasing percentage, survey constituted (15%) of the total number of study designs in our data, followed by randomized control trial (14.14%), cross sectional (14%), cohort (8.38%), review (6%), case series (5.24%) and case studies (0.02%).

DISCUSSION

Our study results show that having a FM article published with open access had a significant positive correlation with higher citation count. Journal articles with open access are readily available and accessible, which allows greater visibility and audience, therefore, higher chances for citations.[5] Davis PM reported that, although open access publishing may reach more readers, there was no evidence found to support a citation advantage for open access articles cited within the first year after publication, when compared with subscription-access control articles cited within first 3 years of publication.[6] Based on this hypothesis, we decided to collect data from articles published over 10 years ago. This allowed us to overcome a potential bias.

Characteristics relating to the length of the article showed a statistically significant positive correlation with an increased number of citations. These characteristics were; total word count, number of pages and number of references. Antoniou et al. also reported a positive correlation between the length of the article, number of references and the mentioning of study design in the

title which increased the citation count in their analysis, even after adjusting for several potentially confounding variables.[7] This may signify that longer articles may contain more ideas, information, and greater diversity of results attracting more opportunity for higher citation rates.

Similarly, the number of figures and the number of tables showed a positive correlation only. While the number of tables and figures also pertains to the length of the article; in our study their statistical significance was not observed. Shekhani et al. suggested that having a greater number of contributors or authors on a radiology scientific article may increase its quality, length, references, and citations.[8] However, our data for FM journal articles only correlates to the length and references of the article.

Aksnes and Falagas et al. reported that, generally, team-authored studies and multi-institutional collaborations got cited more.[9,10] However, our study showed a slight positive correlation with no statistical significance in multi-institutional studies. We also found a negative association in an increased number of authors, and this did not correlate to increasing the number of citations for FM journal articles. Since the articles were pertaining to different subspecialties, we can hypothesize that an increase number of authors or multi-institutional collaboration does not significantly influence FM journal articles.

Words and characters per title also had a positive correlation and are comparable to a similar publication in psychiatry.[11] These results are supported by another previous studies that stated the length of the title, presence of a colon in the title and presence of an acronym showed a positive correlation.[12] The punctuation in the title showed a positive correlation, which

contrasted the previously stated study by Jacques et al.[12] However, other studies have shown that shorter titles may potentially be easier to understand and results in an increased citation. [13,14] Some authors maintain that many of the literature searches are performed electronically on sites such as PubMed or Google. Naturally, a longer title can lead to an increased chance of the article appearing in a search result and therefore can increase the chances of a researcher finding the work and citing it. On the other hand, misleading titles may skew results the other way.[12,15]

The length of the abstract, i.e. abstract word count and abstract character count had a positive correlation but no statistical significance with citation count. However, structured abstracts significantly correlated with an increased citation count. Using a structured abstract may allow readers to have a quick overview of the article. Previous studies have shown that the format of the abstracts is influenced by the publishing journal, the instructions for authors and processes of review and editing play important roles in promoting appropriate abstract formats. There are two possible outcomes for those abstracts that do not follow instructions in place for authors. The content in the abstract may be inappropriate for the format suggested by the journal, or the abstract could be submitted, reviewed, and edited before the instructions for authors had been released.[16] The limitations and instructions employed by the journal are essential to maintain the quality of the abstracts. Editors may also benefit their journal and authors by supporting and promoting the use of structured abstracts.

Every scientific study has limitations and our study is no exception. Our study is subject to selection bias as the journals selected were those with high-impact factor. Results may not be representative of all family medicine journal articles. Our sample size, n=191 from six journals,

and the time frame of six months may seem appropriate but inclusion of more journals, a larger sample size and a longer time frame could potentially yield more powerful results. The correlation between manuscript characteristics and citations may become stronger due to increased variability in the data and lesser effect of outliers. Human error is an item that must be addressed due to our manual data extraction.

Open access status also presents limitations in that journal impact varies among subspecialties. In a previous study, Davis PM et al. stated that a citation advantage from open access may be an artifact of self-selection.[6] Other databases such as Web of Science (WOS), Scopus (Elsevier), and Google Scholar (Google) also differ in citation counts. We chose WOS articles through SCImago as our database because it offers a strict evaluation process. It also provided the most reliable and integrated citation metrics from multiple sources in a single interface.[17] Our research is limited in the context that family medicine research articles are scattered among family medicine journal and even more extensively through non-family medicine journals making it difficult to fully incorporate all articles with vast numbers of citations.[18] Lastly, the statistical significance of the data does not inevitably interpret to a correlation being the representative of causation.

Based on the result of our bibliometric analysis, authors of articles publishing in family medicine journals can improve their citation count by paying close attention to certain article characteristics. Our study findings suggested mentioning study design in the title, publishing open access, having structured abstract, along with the higher total words, number of pages and references showed a statistical significance in the multi regression model. The results also

showed a negative correlation with an increased number of authors. The readership of family medicine journal articles is different from other subspecialties. As compared to bibliometric studies in Psychiatry[11] or Radiology[8], family medicine is a diverse field of medicine which encompasses articles from many different subspecialties. By employing some changes to the structure of the abstract and manuscript, authors can improve the presentation of their manuscript for the Family Medicine journal readership. Our results suggest that bibliometric knowledge is beneficial for authors and journals to enhance the visibility and citation count of their articles.

REFERENCES

1. Ware M, Mabe M. *The STM Report An Overview of Scientific and Scholarly Journal Publishing Mark Ware Consulting International Association of Scientifi Fic, Technical and Medical Publishers*.; 2015. www.markwareconsulting.com.
2. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci*. 2005. doi:10.1073/pnas.0507655102
3. Moed HF, Burger WJM, Frankfort JG, Van Raan AFJ. The application of bibliometric indicators: Important field- and time-dependent factors to be considered. *Scientometrics*. 1985. doi:10.1007/BF02016935
4. Lee JD, Cassano-Pinché A, Vicente KJ. Bibliometric analysis of Human Factors (1970-2000): A quantitative description of scientific impact. *Hum Factors*. 2005;47(4):753-766. doi:10.1518/001872005775570970
5. Davis PM. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *FASEB J*. 2011. doi:10.1096/fj.11-183988
6. Davis PM, Lewenstein B V, Simon DH, Booth JG, Connolly MJL. Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ*. 2008;337:a568. doi:10.1136/bmj.a568
7. Antoniou GA, Antoniou SA, Georgakarakos EI, Sfyroeras GS, Georgiadis GS. Bibliometric analysis of factors predicting increased citations in the vascular and endovascular literature. *Ann Vasc Surg*. 2015;29(2):286-292. doi:10.1016/j.avsg.2014.09.017
8. Shekhani HN, Shariff S, Bhulani N, Khosa F, Hanna TN. Bibliometric Analysis of Manuscript Characteristics That Influence Citations: A Comparison of Six Major Radiology Journals. *AJR Am J Roentgenol*. 2017;209(6):1191-1196. doi:10.2214/AJR.17.18077
9. Aksnes DW. Characteristics of highly cited papers. *Res Eval*. 2003;12(3):159-170. doi:10.3152/147154403781776645
10. Falagas ME, Zarkali A, Karageorgopoulos DE, Bardakas V, Mavros MN. The Impact of Article Length on the Number of Future Citations: A Bibliometric Analysis of General Medicine Journals. *PLoS One*. 2013. doi:10.1371/journal.pone.0049476
11. Hafeez DM, Jalal S, Khosa F. Bibliometric analysis of manuscript characteristics that

influence citations: A comparison of six major psychiatry journals. *J Psychiatr Res*. 2019. doi:10.1016/j.jpsychires.2018.07.010

12. Jacques TS, Sebire NJ. The impact of article titles on citation hits: an analysis of general and specialist medical journals. *JRSM Short Rep*. 2010;1(1):2. doi:10.1258/shorts.2009.100020
13. Paiva C, Lima J, Paiva B. Articles with short titles describing the results are cited more often. *Clinics*. 2012. doi:10.6061/clinics/2012(05)17
14. Habibzadeh F, Yadollahie M. Are Shorter Article Titles More Attractive for Citations? Cross-sectional Study of 22 Scientific Journals. *Croat Med J*. 2010. doi:10.3325/cmj.2010.51.165
15. Chokshi FH, Kang J, Kundu S, Castillo M. Bibliometric Analysis of Manuscript Title Characteristics Associated With Higher Citation Numbers: A Comparison of Three Major Radiology Journals, AJNR, AJR, and Radiology. *Curr Probl Diagn Radiol*. 45(6):356-360. doi:10.1067/j.cpradiol.2016.03.002
16. Nakayama T, Hirai N, Yamazaki S, Naito M. Adoption of structured abstracts by general medical journals and format for a structured abstract. *J Med Libr Assoc*. 2005.
17. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J*. 2008;22(2):338-342. doi:10.1096/fj.07-9492LSF
18. Dunikowski LG, Freeman TR. Impact of family medicine research: Bibliometrics and beyond. *Can Fam Physician*. 2016;62(3):266-268. http://www.ncbi.nlm.nih.gov/pubmed/26975920. Accessed July 29, 2019.

Figure 1: Characteristics of 191 articles published in top six family medicine journals between Jan 2008 to June 2008 that we evaluated. Parts A–D: (A) Shows percentage of articles by family medicine journals (B) continent of origin, (C) study design, and (D) study characteristics. Percentages may not add to 100 due to rounding.

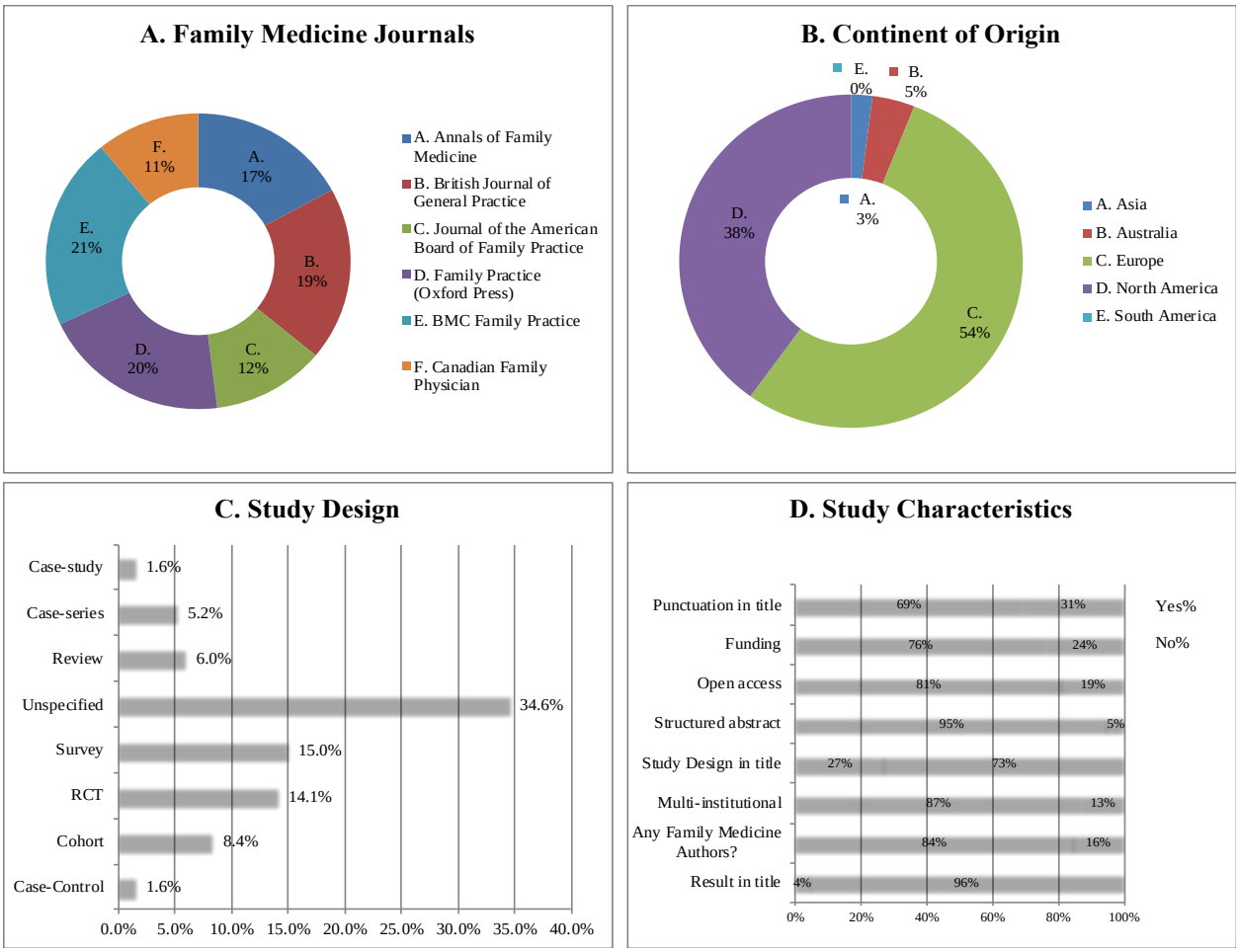

Table 1. Study Characteristics and Definitions

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

Table 2. Journal Characteristics

Journal	Impact factor	Observations	Mean	Standard Deviation	Minimum	Maximum
Journal of the American Board of Family Practice	1.015	23	22.13	18.19	1	66
Annals of Family Medicine	1.864	33	48.54	43.93	0	165
British Journal of General Practice	1.104	36	24.06	28.25	4	171
BMC Family Practice	0.815	40	21.15	17.71	4	83
Canadian Family Physician	0.283	20	26.15	19.68	2	78
Family Practice (Oxford Press)	0.976	39	22.74	21.13	0	82

Table 3. Study Variables With Associated Mean and SDs

Variables	Median	Range
Citations	19	0-171
Number of authors	5	1-11
Sample size	210	0-3630296
Abstract Word Count	255	0-499
Abstract Character Count	1705	0-3106
Total number of words	4146	744-7200
References	27	5-91
Number of pages	7	2-16
Number of figures	0	0-5
Number of tables	2	0-10
Number of words in title	13	3-247
Number of Characters in title	96	29-1769

Table 4. Relationship between number of Citations and Study Variables

Variables	r_s	Correlation	P value
Number of authors	-0.013	Negative	0.85
Sample size	0.03	Positive	0.687
Abstract word count	0.08	Positive	0.25
Abstract character count	0.1	Positive	0.144
Total words	**0.229**	**Positive**	**0.0014**
Number of references	**0.215**	**Positive**	**0.0028**
Number of pages	**0.225**	**Positive**	**0.0021**
Number of figures	0.06	Positive	0.391
Number of tables	0.12	Positive	0.0901
Words per title	0.12	Positive	0.994
Characters per title	0.12	Positive	0.078
Result in title	0.13	Positive	0.06
Multi-institutional	0.003	Positive	0.8
Study design in title	0.03	Positive	0.0001
Structured Abstract	**0.06**	**Positive**	**0.005**
Open access	**0.03**	**Positive**	**0.003**
Funding	0.0011	Positive	0.4
Punctuation in title	0.007	Positive	0.5

Note--The rs value is the Spearman rank-order correlation coefficient. Boldface indicates a statistically significant difference, which was defined by $p<0.05$.