The Impact of Article Length on the Number of Future Citations: A Bibliometric Analysis of General Medicine Journals

Matthew E. Falagas1,2,3*, Angeliki Zarkali1, Drosos E. Karageorgopoulos1,4,5, Vangelis Bardakas1,6, Michael N. Mavros1

1 Alfa Institute of Biomedical Sciences (AIBS), Marousi, Athens, Greece, 2 Department of Medicine and Infectious Diseases, Mitera General Hospital, Hygeia Group, Athens, Greece, 3 Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, United States of America, 4 Department of Medicine, Hygeia Hospital, Marousi, Athens, Greece, 5 Hellenic Center of Disease Control and Prevention, Marousi, Athens, Greece, 6 Department of Applied Mathematical and Physical Science, National Technical University of Athens, Athens, Greece

Abstract

Background: The number of citations received is considered an index of study quality and impact. We aimed to examine the factors associated with the number of citations of published articles, focusing on the article length.

Methods: Original human studies published in the first trimester of 2006 in 5 major General Medicine journals were analyzed with regard to the number of authors and of author-affiliated institutions, title and abstract word count, article length (number of print pages), number of bibliographic references, study design, and 2006 journal impact factor (JIF). A multiple linear regression model was employed to identify the variables independently associated with the number of article citations received through January 2012.

Results: On univariate analysis the JIF, number of authors, article length, study design (interventional/observational and prospective/retrospective), title and abstract word count, number of author-affiliated institutions, and number of references were all associated with the number of citations received. On multivariate analysis with the logarithm of citations as the dependent variable, only article length [regression coefficient: 14.64 (95% confidence intervals: (5.76–23.50)] and JIF [3.37 (1.80–4.948)] independently predicted the number of citations. The variance of citations explained by these parameters was 51.2%.

Conclusion: In a sample of articles published in major General Medicine journals, in addition to journal impact factors, article length and number of authors independently predicted the number of citations. This may reflect a higher complexity level and quality of longer and multi-authored studies.

Introduction

An article’s citations are considered a measure of the scientific recognition the study has received, and thus an indicator of its value and impact on the scientific field [1]. The citations are also the main factor determining the scientific impact of a journal, as expressed by the journal impact factor [2]. This indicator represents the mean number of citations received in an index calendar year, by all the citable articles published in a journal during the previous two years [3,4]. Researchers commonly aim to publish articles that will attract citations and will thus be regarded to have a high scientific impact, as this may be associated with their career advancement.

Several studies have been conducted to explore the factors associated with the citation count of scientific articles. While the effect of journal impact factor [5–10] and study design [11–16] on citations received has been established by different studies, the published evidence on other potentially relevant variables, such as open access to the full text of the article, [17–20] or article length, [13] seems conflicting.

In this context, we aimed to examine the factors associated with the number of citations received by published articles, focusing on the article’s length.

Methods

Data sources

Original human research articles published in the first trimester of 2006 in the 5 highest impact factor journals in the field of general and internal medicine were analyzed (the New England Journal of Medicine, the Lancet, the Journal of the American Medical Association, the Annals of Internal Medicine, and the British
Medical Journal. Experimental studies, review articles, and meta-analyses were excluded. The 2006 journal impact factors were retrieved from the Thomson Reuters Journal Citation Reports. The number of citations to each article was last assessed in January 2012, according to the Thomson Reuters Web of Knowledge.

Data extraction

The abstract and/or full-text manuscript of each article was accessed to collect information regarding article length and characteristics that were reported to affect the number of citations in previous studies. Specifically, we documented variables comprised the number of authors and affiliated institutions, title and abstract word count, article length (as the number of pages), number of bibliographic references, study design (human or experimental studies; prospective or retrospective; interventional or observational), access to the article (open access or requiring subscription), and 2006 journal impact factor (JIF).

Table 1. Characteristics of the analyzed studies.

| Study characteristics                              | Hypothesis (increase citations) |
|----------------------------------------------------|---------------------------------|
| No of authors, median (range)                      | 9.88 (1–48)                     | More authors                      |
| No of author-affiliated institutions, median (range)| 5.33 (1–43)                     | More institutions                 |
| Title word count, median (range)                   | 13.75 (6–29)                    | Longer title                      |
| Abstract word count, median (range)                | 294.88 (105–589)                | Longer abstract                    |
| Article length (print pages), median (range)       | 7.88 (2–15)                     | Lengthier article                 |
| No of bibliographic references, median (range)     | 29.31 (3–61)                    | More references                   |
| Nature of study                                    | Prospective vs Retrospective    | IF Prospective study              |
| Study design                                       | Interventional vs Observational | IF Interventional study           |
| Open access versus restricted access               | 90.2% vs 9.8%                   | If Open access                    |
| 2006 journal impact factor (JIF), median (range)   | 26.88 (9.25–51.3)               | Higher JIF                        |
| Citation count, median (range)                     | 166.2 (5–1314)                  |                                   |

doi:10.1371/journal.pone.0049476.t001

Table 2. Results of statistical analysis.

| Variable                                           | Univariate | Multivariate |
|----------------------------------------------------|------------|--------------|
|                                                   | Correlation coefficient | p-value* | Regression Coefficient (95% Confidence Interval) | p-value |
| No of Authors                                      | 0.50       | <0.001       | 0.003 (−0.007, 0.014) | 0.555 |
| No of author-affiliated institutions                | 0.35       | <0.001       | 0.006 (−0.001, 0.014) | 0.103 |
| Title word count                                   | −0.26      | <0.001       | −0.03 (−0.016, 0.009) | 0.587 |
| Abstract word count                                | −0.22      | <0.001       | 0.001 (0.000, 0.002) | 0.107 |
| Article length (number of print pages)             | 0.70       | <0.001       | 0.079 (0.055, 0.102) | <0.001 |
| No of References                                   | 0.33       | <0.001       | 0.003 (−0.002, 0.008) | 0.284 |
| Retrospective study                                | -          | <0.001       | −0.045 (−0.158, 0.069) | 0.438 |
| Observational study                                | -          | <0.001       | −0.020 (−0.146, 0.107) | 0.758 |
| Multi-center study                                 | -          | 0.066        | Not included            | -     |
| Open access                                        | -          | 0.704        | Not included            | -     |
| 2006 journal impact factor                         | 0.63       | <0.001       | 0.008 (0.004, 0.013) | <0.001 |

*Refers to Spearman’s correlation or to Mann-Whitney U test.
doi:10.1371/journal.pone.0049476.t002

Data analysis and statistical methods

Statistical analyses were performed using SPSS Version 20.0. Initially, the association of each independent variable with the dependent variable (citation count) was assessed with univariate analyses (Mann-Whitney for categorical and Spearman’s correlation for continuous variables); we used non-parametric methods, because citations of articles published in General Medicine journals are known to have a non-parametric distribution [21]. Variables significantly associated with the citation count in univariate analysis (p<0.10) were then entered in a backward multiple linear regression model to identify independent predictors of higher number of citations. The multiple linear regression model was also run with logarithmic transformation of the dependent variable (number of citations) to assess for a logarithmic, rather than linear relationship between the dependent and independent variables. Since the logarithmic transformed model performed better, only the results of this model were presented. To exclude the possibility of a false positive association between the article length and the number of authors and the number of citations, we repeated the multiple regression analysis separately.
for each of the journals, as the journal impact factor has been well established to be a major factor affecting citations.

All assumptions of linear regression were met by this model, including lack of error term correlation (Durbin-Watson = 2.013). Graphical examination of residuals did not suggest a violation of the linearity and normality assumption. Multicollinearity was deemed not important (VIF <5) for every independent variable. Homoscedasticity was checked by examination of the scatterplot of residuals and predicted values, and was met when outliers were excluded from the model. We also tested for outliers using added value and residual plots. Three outliers were identified with citations 1314, 1185 and 793, and were excluded. A variable was considered statistically significant if it had a p-value <0.05 in the final multivariable model.

**Results**

A total of 196 articles were analyzed. Experimental studies were excluded, leading to a total of 192 articles. The citation count varied from 5 to 1314 with a median of 96.5 (mean = 166). The majority of studies were prospective (67.2%), open-access (90.2%) and multi-center (67.2%). The most common type of study in our sample was that of a trial (39.6%, both randomized control trials and non-randomized trials). The study characteristics are presented in Table 1.

On univariate analysis, all tested independent variables except access (free versus restricted) and multicenter or single-center study, were found to have a statistically significant correlation to citations (Table 2). Therefore, the following variables were entered in the multivariate model: JIF, number of authors, article length, prospective or retrospective design, type of study (interventional or observational), abstract and title word count, number of affiliated institutions, and number of references, with the logarithm of the number of citations as the dependent variable.

A backward linear regression analysis was performed, removing insignificant independent variables one by one. Two variables were found to independently predict the number of citations: article length (number of pages) [regression coefficient (95% confidence interval): 0.079 (0.055–0.102), p<0.001; Figure 1 and JIF [0.008 (0.004–0.013), p<0.001; Figure 2]. The variance of citations explained by these factors is 51.2% (adjusted R² = 50.7%), p<0.001. The findings of the univariate and multivariate analyses are presented in table 2.

**Subgroup analyses**

For the subgroup of articles published in two of the five included journals, article length was found to be the only factor independently associated with citations, with a parameter estimate of 0.080 [(0.032–0.127), p = 0.002] and 0.058 [(0.013–0.104), p = 0.013], respectively. For articles published in the third journal, statistically significant factors included the number of institutions.
[0.050 (0.001–0.098), p = 0.04] and the number of references [0.014 (0.003–0.024), p = 0.01], while in the fourth journal significant were the number of authors [0.029 (0.006–0.051), p = 0.015] and the number of references [0.025 (0.011–0.040), p = 0.001]. In the remaining journal, no variable was found to be significantly associated with citations, although that may reflect the smaller sample size (n = 23). Last, article length was significantly associated with the number of citations in the single-center studies subgroup [0.109 (0.075–0.143), p<0.001].

Discussion

The main finding of this study is that the article length and journal impact factor are independently associated with the number of citations received by each article. Although several previous studies have reported that the journal impact factor is associated with the article citations, this is the first study, to the best of our knowledge, to report a positive association between the article length and the article citations after adjustment for several potentially confounding variables, such as the study design, prospective or retrospective nature of the study, abstract and title word count, number of author–affiliated institutions and number of bibliographic references. Specifically, we found an increase by an average of 0.079 in the logarithm of citations per article for each additional page, 0.008 for every unit of increase in the journal impact factor. The greater article length could reflect increased greater scientific complexity and higher methodological quality of a study; in addition, lengthier articles are expected to contain more information, thus increasing the possibilities that part of it will be appropriate to be cited by other researchers. Furthermore, in lengthier compared with shorter articles, the study methodology and findings could be more clearly and elaborately presented and discussed, and can therefore have a greater impact. It should be highlighted that our findings probably do not apply to long articles where the results have been improperly “inflated”; after all, some of the greatest discoveries in science have been described only briefly [22].

A few studies have assessed, albeit not comprehensively, the impact of the article length on future citations. In the field of Astronomy and Astrophysics, lengthier articles were cited more often in some journals [23]. In the fields of Infectious Diseases, Clinical Microbiology and Antimicrobial Agents, brief reports were cited less often than full articles, even after adjustment for the journal impact factor [24]. This was not the case in another study assessing 504 articles and adjusting for several confounding factors [13]. In contrast to our study, in which we assessed only original study articles, the authors included in their analysis numerous Cochrane reviews and reports from the Technology Assessment database (n = 108), that are typically lengthy; in addition, they excluded articles not meeting specific methodological and clinical relevance criteria. That study reported a slightly negative
Table 3. Published studies examining factors that affect citations.

| Author; Years; | Analysis | Conclusions |
|---------------|----------|-------------|
| Year          | Databases/journals studied; Sample size; (Specialty) | Dependent variable; Independent variables |
| Perneger [5]; | NR; Pubmed, Scopus Citation comparison of consensus articles in different journals | 1.0 log unit of citations increase per unit of JIF (95% CI: 0.7–1.3, \( P = 0.001 \)) |
| 2010          | 4 consensus statements Citations | 33 articles JIF (NR) |
| Etter [6];    | NR Univariate and multivariate linear regression | More citations if: Statistically significant results (median 541 vs. 17, \( P = 0.001 \)), higher JIF (10.2 citations/JIF point, \( P = 0.001 \)) |
| 2009          | Cochrane Citations | 150 RCTs JIF, Favorable outcome, Year Funding, Country, Product type (Nicotine replacement RCTs) |
| Fillion [7];  | 1998–2004 NR Journal and country most strongly associated with citations | Citation rate (high impact factor, US) |
| 2008          | ISI Web of Science Citation rate | 72 articles Authors, JIF, Topic, Institution, Country, Year (Epidemiology articles on child injuries and coronary disease) |
| Nieminen [8]; | 1996 Mann-Whitney tests, Kruskal-Wallis, ANOVA, and negative binomial regression | No citation advantage for reporting quality and statistical analysis; JIF as important as quality |
| 2006          | Am J Psych, Arch Gen Psych, BJPsych and NUPsych Citations | 448 articles Reporting quality, Sample size, JIF (Psychiatry) |
| Montori [9];* | 2000 Multiple linear regression | Twice as many citations for systematic vs narrative reviews (95% CI: 1.5–2.7); JIF = weaker predictor than quality |
| 2003          | Hand search of 170 journals Citations | 271 reviews JIF, Type of review (NR) |
| Callaham [10];| 1991 Multivariate regression | JIF = the strongest predictor (100%); Newsworthiness score (89.9% as strong); Subjective quality score (61.5%). Positive outcome bias not significant. |
| 2002          | Emergency medicine specialty meeting articles Citations | 204 articles JIF, Subject, Quality, Study design, Positive result, Newsworthiness (Emergency medicine) |
| Okike [11];   | 2002–2003 Multiple linear regression, log-transformation | More citations at 5 years in: high level of evidence, large sample size, multiple institutions, self-reported conflict of interest, sports medicine and arthroscopy. Less citations for: Pediatric orthopedic articles |
| 2011          | JBJS Am vol, JBJS Brit vol, CORR; Citations | 661 articles JIF, Level of evidence, Sample size, Self-reported conflict of interest, Subject, Location, Control/blinding, No of authors/ institutions, Prospective study (Orthopedics) |
| Willis [12];  | 2004 Binary logistic regression | More citation rates: RCTs [OR=115.5 (9.4–1419.6, \( p<0.001 \)), topic of oncology [OR=2.5 (1.4–4.7, \( p = 0.004 \))]. |
Table 3. Cont.

| Author; Years; Journal | Analysis | Conclusions |
|------------------------|----------|-------------|
| Lokker [13]; 2005      | Multiple regression | More citations if: More authors, higher clinical relevance scores, more references |
| Kulkarni [14]; 1999–2000 | Univariate and multivariate linear regression | Increased citation rates: larger sample, journal of publication, funding and industry-favoring result |
| Bhandari [15]; 2000    | Regression analysis | Citations: Meta-analyses (mean = 15.5), Randomized trials (9.3), Basic science papers (7.6), Observational (retrospective 5.3, prospective 4.2), Case reports (1.5) |
| Kim [17]; 2009–2010    | NR | 43% decrease in citations per month from 2009 to 2010 (p = 0.00064). No difference in 2009 vs 2010 simulated |
| Davis [19];** 2007     | Logistic and negative binomial regression | OA: 89% more full text downloads (76–103%), 42% more PDF downloads (32–52%), 23% more unique visitors (16–30%), 24% less abstract downloads (~29–19%). No evidence of citation advantage. |
| Eyesenbach [20]; 2004–2005 | Logistic and linear regression | OA articles more recognized and cited [OR 2.1 (1.5–2.9)] |
correlation between the article length and the number of citations received \([-0.11 \text{ to } -0.01}\); however, when Cochrane reviews and reports from the Technology Assessment database were excluded, no association between the article length and citations was identified. Although the difference between these findings and those of our study is probably attributed to the difference in the type of articles assessed (inclusion/exclusion of review articles), it remains to be proven whether our findings can be generalized to a larger part of the biomedical literature than just the 5 highest impact factor journals in General & Internal Medicine.

In addition to the number of print pages, we found that the impact factor of the journal and the number of authors were associated with the citation count. Although we limited our analysis only to articles from high impact factor journals, the articles published in the highest impact factor journals were cited significantly more often. It should be noted that we used the 2006 journal impact factor (that refers to articles published in 2004 and 2005) for our analysis (that referred to articles published in 2006) to avoid a potential bias. In this regard, our findings are in concordance with previous studies that found the journal impact factor to be a major predictor of the article citation count [5–10].

Several other variables assessed in previous studies were incorporated in our analysis, but failed to show a statistically significant association with the number of citations. The characteristics and findings of all relevant studies are briefly presented in Table 3. Some authors have described an association between the type of the study and the future citations, with more citations received by meta-analyses and randomized control trials and less citations received by observational studies [11,12,14,16]; their findings are have been limited by selection bias (articles of a specific specialty) [11,12,14,16] and inappropriate adjustment of confounding factors [16]. Such findings were not verified in our analysis, as we found no citation advantage neither for interventional over observational studies, nor for any specific type of study (trial, cohort, cross-sectional or case-control); however, this could also be attributed to the relatively small sample size of each subset of articles of different study type. It has been debated whether

### Table 3. Cont.

| Author; Years; Analysis Conclusions |  |
|------------------------------------|---|
| **2003** | **PNAS** | Citations |
| 1492 original research articles | OA, No of authors, Country, Funding, Authors’ lifetime publication count | (NR) |
| Main factor studied: No of authors | **2006** | Citations |
| Science, Cell, Nature, NEJM, Lancet, JAMA | No of authors and institutions | (NR) |
| Main factor studied: Title length | **2010** | Citations |
| Hazibzadeh [26]; 2005 | Linear regression model | Increased citation rates in longer titles (more in high JIF) |
| 22 English journals via Scopus | Title characteristics | (NR) |
| Jacques [27]; 2005 | NR | Increased citation rates in: Longer titles (rho = 0.62, 2-sided P<0.0001), presence of a colon or acronym in title |
| 50 articles | (NR) |
| Main factor studied: Hit count online | **2004** | Citations |
| Perneger [28]; 1999 | NA | More citations for papers with most hits on BMJ website the first week: (extra 3.7 citations/100 hits, P<0.001) |
| BMJ | (NR) |
| All studies were cohort studies of published articles except: * Cross-sectional studies, ** Randomized control trial. Abbreviations: Am J Psych: American Journal of Psychiatry, Arch Gen Psych: Archives of General Psychiatry, BJPsych: British Journal of Psychiatry, BMJ: British Medical Journal, CORR: Clinical Orthopedics and Related Research, JAMA: Journal of American Medical Association, JBJS Am vol.: Journal of Bone and Joint Surgery American Volume, JBJS Brit vol.: Journal of Bone and Joint Surgery British Volume JIF: Journal’s Impact factor, NEJM: New England Journal of Medicine, NIPych: Nordic Journal of Psychiatry, OA: Open Access, OR: Odds ratio, PNAS: Proceedings of the National Academy of Sciences, RCT: Randomized control trial. doi:10.1371/journal.pone.0049476.t003 |
open access distribution of articles leads to more citations [18–20,25,26] or that scientific collaboration positively influence citation count [11,15,16], paper quality [9,10], funding [15,18] or country of origin of the authors [7,18], which are factors that have been found to affect citations by other authors. Last, in our assessment of article length, we only analyzed page count (not word count) and inter-journal variance in the number of words per page cannot be excluded.

In conclusion, for original research articles published in the major General Medicine journals, in addition to journal impact factor, the article length independently predicts the number of future citations. This probably reflects a higher complexity level and quality of longer studies and does not apply to inappropriately inflated articles. Additional studies are warranted to verify the generalizability of our findings to a largest part of the biomedical literature.

**Author Contributions**

Conceived and designed the experiments: MEF DEK. Performed the experiments: AZ DEK. Analyzed the data: VB MNM. Contributed reagents/materials/analysis tools: MEF AZ DEK MNM. Wrote the paper: MEF AZ DEK VB MNM.