Which Factors Help Authors Produce the Highest Impact Research? Collaboration, Journal and Document Properties[1]

Fereshteh Didegah - Mike Thelwall
f.didegah@wlv.ac.uk - m.thelwall@wlv.ac.uk
Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY UK

ABSTRACT: This study assesses whether eleven factors associate with higher impact research: individual, institutional and international collaboration; journal and reference impacts; abstract readability; reference and keyword totals; paper, abstract and title lengths. Authors may have some control over these factors and hence this information may help them to conduct and publish higher impact research. These factors have been previously researched but with partially conflicting findings. A simultaneous assessment of these eleven factors for Biology & Biochemistry, Chemistry and Social Sciences used a single negative binomial-logit hurdle model estimating the percentage change in the mean citation counts per unit of increase or decrease in the predictor variables. The journal Impact Factor was found to significantly associate with increased citations in all three areas. The impact and the number of cited references and their average citation impact also significantly associate with higher article citation impact. Individual and international teamwork give a citation advantage in Biology & Biochemistry and Chemistry but inter-institutional teamwork is not important in any of the three subject areas. Abstract readability is also not significant or of no practical significance. Among the article size features, abstract length significantly associates with increased citations but the number of keywords, title length and paper length are insignificant or of no practical significance. In summary, at least some aspects of collaboration, journal and document properties significantly associate with higher citations. The results provide new and particularly strong statistical evidence that the authors should consider publishing in high impact journals, ensure that they do not omit relevant references, engage in the widest possible team working, when appropriate, and write extensive abstracts. A new finding is that whilst is seems to be useful to collaborate and to collaborate internationally, there seems to be no particular need to collaborate with other institutions within the same country.

Introduction

During an academic career, scholars make numerous choices about the type of research to conduct, how to present their research, and where to submit it for publication. It seems logical that researchers should aim to conduct the highest possible impact research in order to make the most of their talents and opportunities. Whilst the key decisions for this aim are likely to be specific to the topics researched, there are some more peripheral factors that are nevertheless relevant and that academics may also need to consider in order to maximise the impact of their efforts.

Citation counts are widely acknowledged as the main scientific research impact indicator and empirical studies have been carried out to seek associations between citation counts and various objective and easily measurable properties of research. These include the impact of the publishing journal (Boyack & Klavans, 2005), collaboration (Gazni & Didegah, 2010), the interdisciplinarity of the article references (Larivière & Gingras, 2010), the number and impact of references (Boyack & Klavans, 2005), and the size of the related field (Lovaglia, 1989). Thus, authors seeking to maximise the impact of their research may write more clear titles and abstracts and may also be particularly careful to ensure that their literature review does not miss any relevant highly cited papers. More generally, if they wish to conduct high

[1] Didegah, F., & Thelwall, M. (in press). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*.

impact research then they may also seek to engage in collaborations (hence generating more co-authors). Presumably, attempts to artificially manipulate these factors, such as by adding honorary international authors or irrelevant high impact interdisciplinary references, would not work since factors associating with higher citations presumably reflect underlying properties of research rather than surface features of an article. Nevertheless, knowledge of important factors may naturally push authors towards higher impact types of research, for example by looking to expand their collaboration network, by being open to interdisciplinary research influences, and by paying particular attention to relevant research in high impact international journals (e.g., rather than national research).

This study examines whether research collaboration, journal and reference impact, abstract readability, and article size attributes affect citation counts. These factors are at least to some extent under the control of the authors and so it would be useful to know whether researchers should pay attention to them to ensure that their research has the greatest possible impact. Research collaboration has been frequently analysed (Sooryamoorthy, 2009) and the other factors have also been examined (Zhao, 2010; Gazni, 2011) but they have not been examined simultaneously for multiple research fields using an optimal statistical model. This is an important omission because non-simultaneous tests may identify apparently important factors that have no effect when other factors are controlled. van Raan (1998) criticises the claim that a theory is needed for citation analysis and suggests replacing the theory with a feasible model that provides a possible approximation of reality. This study also helps to address this goal with its new, more integrated statistical model.

Literature review

As introduced above, research citation impact has been shown to be related to a number of objective factors, such as research collaboration, choice of journal, and properties of the article itself. This review does not consider article type as a factor, even though review articles are known to attract more citations (Aksnes, 2003), because it is concerned with primary research outputs. It also does not consider another factor, author reputation (Peters & van Raan, 1994), because this is presumably influenced by conducting high impact research and so is not a factor that authors can consider for individual articles.

Research collaboration

Multi-author research has become more common (Gazni, Sugimoto, & Didegah, 2012; Persson, Glänzel, & Danell, 2004) and receives more citations than does solo research (Gazni & Didegah, 2010; Sooryamoorthy, 2009; Leimu & Koricheva, 2005a&b). However, a few studies have found no correlation between more authors and increased citations (Bornmann, Schier, Marx, & Daniel, 2012; Haslam et al., 2008). These studies' findings are often not generalizable, however because they are limited to a single country (Sooryamoorthy, 2009), a single institution (Gazni & Didegah, 2010), a single field of study (Leimu & Koricheva, 2005a&b; Haslam et al., 2008) or a specific journal (Bornmann, Schier, Marx, & Daniel, 2012). Using correlation and regression tests, positive correlations between citation counts and the number of authors have been found (Gazni & Didegah, 2010; Sooryamoorthy, 2009; Leimu & Koricheva, 2005a&b; Haslam et al., 2008) but not the *extent* to which the number of authors contributes to increased citations. The differences between the results of previous studies might be due to the differing samples of publications used and disciplinary differences in particular. Whereas previous studies have conducted detailed micro-level analyses, macro level studies are also needed.

International collaboration can also lead to increased citations (Sooryamoorthy, 2009; Glänzel, 2001; Glänzel & Schubert, 2001; Katz & Hicks, 1997; Narin, Stevens, & Whitlow, 1991). Nevertheless, an investigation of Harvard University publications found no correlation between international collaboration and citation counts (Gazni & Didegah, 2010), but Harvard may be a special case as a world-leading institution. Most studies are geographically or institutionally limited and hence are difficult to generalise. Two studies (Glänzel, 2001; Glänzel & Schubert, 2001) avoid this issue by taking the full Science Citation Index (SCI) during a one or two-year period. However, they do not cover social science fields. To measure the impact of international collaboration on citation counts, the simple method of comparing the mean citation for domestic collaboration with that of international collaboration is often used. This has the limitation that the difference may be spurious in the sense of being caused by factors other than the ones investigated. International collaboration seems to be particularly beneficial for small institutions (Goldfinch, Dale, & DeRouen, 2003) rather than big institutions (Gazni & Didegah, 2010).

Institutional collaboration, which involves researchers from different institutions, also associates with higher citation impact (Gazni & Didegah, 2010; Sooryamoorthy, 2009; Narin & Whitlow, 1990). These studies are also geographically and institutionally limited and use a simple correlation test for an association between institutional collaboration and citation counts, and so it may be that other factors explain the increased citations better than institutional collaboration.

Journal impact

High impact journals attract more attention from scholars and are therefore more visible (Haslam et al., 2008; Meadows, 1998). This popularity may influence the visibility and impact of their articles. The impact of the publishing journal, measured by the journal impact factor (JIF), has been shown to be important for gaining attention to research papers in Demography (Van Dalen & Henkens, 2005). In two micro-level studies of Nanotechnology & Nanoscience and Emergency Medicine, the JIF was the most significant determinant of the number of citations to papers (Didegah & Thelwall, 2013; Callaham, Wears, & Weber, 2002). A large scale study also found the JIF to be the most important determinant of citation impact in 17 disciplines (including Biology & Biochemistry, Chemistry and Social Sciences) out of 24 disciplines and there was a positive significant correlation between this factor and article citation counts (Boyack & Klavans, 2005). The *extent* to which this factor associates with increased citations was not determined in the above studies, however, but an investigation of Biomedicine articles found the JIF to contribute to an 11% increase in the number of citations to papers (Bornmann & Daniel, 2007).

While most studies have confirmed that the JIF significantly associates with citation counts for articles, there are some exceptions. For instance, the impact of ecological journals was not found significantly associate with the number of citations to individual articles (Leimu & Koricheva, 2005a). The considerable variation in citation rates of articles in high impact Ecological journals may have caused this result.

Reference impact

Articles with high impact references are cited more. To measure the impact of references, the average number of citations to the cited references (Didegah & Thelwall, 2013), the total number of citations to the references (Boyack & Klavans, 2005) and the h-index of the cited references (Bornmann, Schier, Marx, & Daniel, 2012) have been examined.

Abstract readability

Excellent writing skills are important for high impact research (Zimmerman, 1989). Readability refers to the level of difficulty of the language used to write a text. Using the Flesch difficulty score, Gazni (2011) found that papers with less readable abstracts were cited more than the papers with more readable abstracts in the five top institutions in the world. It may be that prestige in the world's top institutions ensures that their less readable abstracts seem more impressive, whereas unreadable abstracts may be taken as a sign of incompetence for researchers at other institutions. Alternatively, less readable abstracts may associate with higher citation areas of study, such as the more quantitative fields. However, medical articles with structured abstracts, using different sections in a way that is known to be more readable (Hartley & Benjamin, 1998), are, on average, more cited than articles with traditional unstructured abstracts (Hartley & Sydes, 1997).

It seems that there is not a strong relationship between article readability and citation impact in three sub-fields of Social Sciences: Marketing, Psychology and Education Science (Stremersch, Verniers, & Verhoef, 2007; Hartley, Sotto, & Pennebaker, 2002; Hartley & Trueman, 1992). Finally, three decades ago, Bottle, Rennie, Russ and Sardar (1983) claimed that the readability of articles had significantly decreased over time although the reasons for this were not clear and it is not known if this trend has continued.

Given that abstract readability and its association with research citation impact has been studied only to a limited degree, larger scale investigations are needed.

Size factors

Longer papers may likely be cited more if they have more content. A number of micro-studies in different subject areas have confirmed that the more pages, the higher the number of citations to a paper. In Social and Personality Psychology, longer papers with more figures and tables are cited more often (Haslam et al., 2008). Perhaps longer papers publish more original ideas and hence need more extensive and comprehensive explanations for different sections of the paper. The same result was found for publications in The Lancet, a leading journal in General Medicine. Longer medical papers receiving more citations also have many references and this may be another influence (Kostoff, 2007). Medical papers with longer abstracts have also been found to receive more citations (Kostoff, 2007) whereas papers with longer titles in Psychology receive fewer citations (Haslam et al., 2008).

Authors providing more (Haslam et al., 2008; Kostoff, 2007; Peters & van Raan, 1994) references attract more citations. In a comparison between four subject areas (Mathematics, Physics, Chemistry and Biology & Biochemistry), the number of references was found to positively and significantly correlate with the number of citations but the percentage increase in citations for each additional reference was not determined (Vieira & Gomes, 2010).

Research questions

Article citation impact factors have been widely scrutinized in the previous literature but have been considered separately (and mostly within a single field) whereas, in reality, citation impact results from interactions between different factors. A simultaneous assessment of these factors will fill this gap in the literature and represent a model closer to reality. Therefore, this study seeks to simultaneously analyse several factors in three different fields of research that are representatives of three broad areas of science (Life Sciences, Physical Sciences and Social Sciences). Finally, it goes further than the simple correlation between the factors and citation impact and provides evidence of the *extent* to which these factors associate with increased or decreased citations. This study seeks to answer the following research questions:

1. Do journal and reference characteristics (journal impact, reference impact and total references) associate with increased citation impact?
2. Do types of research collaboration (individual, institutional and international collaboration) associate with increased citation impact?
3. Do article size attributes (article, abstract, and title length and total keywords) associate with increased citation impact?
4. Do articles with more readable abstracts receive more citations?

5. *To what extent* do the above factors associate with increased citation counts?

Data and methods

Papers from Biology & Biochemistry (16,058 articles), Chemistry (16,378 articles) and Social Sciences (15,932 articles) covered by Thomson Reuters' Web of Science (WoS) from 2000-2009 were extracted. Using the *ScienceWatch.com* list of journals classifying each journal into one of the 22 ESI (Essential Science Indicators) fields, each paper in the sample was categorised into one ESI field. Only articles and conference proceedings were included because original research is mainly published in these two types of documents (Milojević & Leydesdorff, 2013).

Although the subject classification in WoS is journal-based, it is well-established and has frequently been used by scientometricians to classify individual papers. The three fields were picked up from a list of 22 different subject fields classified by ESI in WoS. Biology & Biochemistry was chosen as a representative for life sciences and Chemistry was chosen as a representative for physical sciences (see Nagaoka, Igami, Eto, & Ijichi (2011) for the categorization of subject fields), as they both are the largest fields (based on number of their papers) in their own category.

A limitation for the citation data in all three categories, and particularly for Social Sciences, is that different fields within each category will have different average citation levels. We chose not to normalise the citation counts (e.g., by dividing article citations by the average for their WoS subject area) in order to test the simplest model but future research could evaluate the impact of this choice.

Dependent and independent variables

The number of citations to papers is the dependent variable and the independent variables are research collaboration, abstract readability, journal and reference impact, and article size and metadata attributes. The JIF extracted from the Journal Citation Reports (JCR) was used as the indicator of journal impact. To measure reference impact, all references were matched against a dataset of all types of documents from 2000-2009. References not indexed in WoS were ignored. The average number of citations to its matched references was calculated for each paper in the dataset.

Three different patterns of research collaboration were used: individual collaboration (number of authors in each paper), institutional collaboration (number of institutions in the author affiliations of each paper) and international collaboration (number of countries in the author affiliations of each paper). The number of authors per paper was automatically counted from the WoS author names field. To identify and count institutional and international collaborations, the number of distinct institutions and countries contributing to the WoS affiliation field of each paper was automatically counted.

There are numerous formulae to measure the readability of a text but their validity is still a matter of debate. To prevent readability formula limitations affecting the results of our study, seven different readability formulae were used: Kincaid formula, Automated Readability Index (ARI), Coleman-Liau formula, Flesch Reading Ease formula, Fog Index, Lix formula, and SMOG Grading. The STYLE program was used to automatically calculate these scores (Cherry & Vesterman, 1981). There was a significant correlation between the seven readability scores in the three fields (Tables 1-3). The Flesch Reading Ease Score was used since it seems to be the most popular and also has a high correlation with the other six scores ($r \sim 0.8$). The Flesch Score ranges between 0 and 100 where 0 indicates a text that is the most difficult to read and 100 represents the easiest text to read.

The length of a paper was measured by its number of pages and the length of an abstract and title was measured by the number of words.

Statistical procedures

Count models provide a structural framework for analysing count data. Given that the study dependent variable is count data (citations), these types of regression models are the most appropriate. The research data set is overdispersed (i.e., the variance of the data is greater than its mean). A Poisson regression model, the basic count model, assumes mean and variance equality (Cameron & Trivedi, 2001) and cannot adequately deal with overdispersed data so this option was rejected.

Initially, standard, zero-inflated and hurdle negative binomial models were considered. A *standard* negative binomial model is frequently used to model overdispersed data. *Hurdle models* seek first to determine the probability of an observation being positive or zero, and then estimate the parameters of the count distribution for positive observations. *Zero-inflated* models assume two types of zeros in the data: zeros which arise from a count distribution and zeros which arise from a "perfect-zero" distribution (Hilbe, 2011). We fitted these three models on the dataset and hurdle models were found to give the best fit to the data. The hurdle model is also intuitively a good choice because it seems reasonable to assume that it is a significant hurdle for a paper to receive its first citation but after this it is more likely to be cited in the future. More citations may occur because a cited paper is listed higher in information retrieval systems (e.g., Google Scholar) or because of the endorsement of a citation reported in such systems.

There are different types of hurdle model. Logit and complementary log-log (cloglog) hurdle models were fitted on the data set and found to have identical AIC values. AIC (Akaike Information Criterion) is an indicator of the statistical goodness of fit and helps to choose between two models. The logit and cloglog models are the binary models for modelling the zero counts and specify the relationship between the predictors and the dependent variable. As the results from the logit model are easier to interpret, it was used (Hilbe, 2011). In the negative binomial-logit hurdle model, two parameters are predicted with the negative binomial model: The overdispersion parameter and the mean of the negative binomial model. With the logit model, an odds ratio in the form of Log [P(citations>1)/P(citations=0)] is predicted.

Since the citation counts are not normalized by year of publication, we entered the publication year into both the logit and negative binomial models to control for the effect of the publication year.

Results

The results of the negative binomial-logit hurdle model provide coefficients for both the negative binomial (non-zero citation counts) and the logit (proportion of uncited papers) components of the model (Tables 4 to 6).

Journal Impact

With respect to the negative binomial model, the JIF significantly associates with increased citations in the three fields and most strongly in Chemistry. A unit increase in the JIF increases the mean citation count by a substantial 31.9%, 27.8%, and 15.6% in Chemistry, Social Sciences and Biology & Biochemistry, respectively. With respect to the logit model, this change significantly contributes to 88.9%, 59.1% and 38.9% decreases in the mean number of zero citations in Chemistry, Biology & Biochemistry and Social Sciences, respectively. In summary, the evidence consistently shows that higher JIFs associate with increased mean citations.

The number of references and their impact

The two article reference features, impact and number, associate with increased citation counts in the three fields. A unit increase in the average impact of an article's references associates with 0.8%, 0.5% and 0.4% increases in the mean citations to articles in Social Sciences, Biology & Biochemistry and Chemistry, respectively. Whilst this change seems to be too small to be significant, the references impact scores can have quite a wide range (e.g., 95% are in the range 0-100 in Biology & Biochemistry) and so an increase by about 50 average citations to references seems possible, in theory for a typical article. This would lead to mean citation increases for the article of 50 times larger than the above figures (i.e., 40%, 25% and 20%). In practical terms, this might mean forgetting to cite two key extremely cited articles could be very costly for the eventual impact of an article. Each additional reference also associates with 0.9%, 0.8% and 0.7% increases in the mean citations to Chemistry, Social Sciences and Biology & Biochemistry articles, respectively. These differences are potentially substantial since an author could reasonably easily add ten references to a paper through a more extensive literature review. These variables significantly associate with decreased zero citations in Social Sciences and Biology & Biochemistry. In summary, using an appropriate number of impact references will increase the likelihood of a greater citation impact for the citing article.

Research collaboration

The coefficients of the negative binomial model show that among the patterns of research collaboration, international and individual collaborations significantly associate with increased citations in Biology & Biochemistry and Chemistry. The number of countries is not a significant determinant of citation counts in Social Sciences but additional authors associate with increased citations in this field. One additional country increases the mean citation count by 8.6% and 5.5% in Chemistry and Biology & Biochemistry, respectively: for papers that are cited at least once in Chemistry, each extra country attracts, on average, 8.6% more citations. International collaboration is not significant for zero citations in Biology & Biochemistry and Social Sciences while it significantly associates with decreased zero citations in Chemistry. Each additional country associates with a 34% decrease in the mean number of zero citations in Chemistry. In summary, individual and international collaboration are the two types of research collaboration contributing to increased citation impact.

According to the negative binomial model, each additional author increases the mean number of citations by 2.9%, 1.3% and 0.9% in Social Sciences, Chemistry and Biology & Biochemistry, respectively. The results of logit models show that this variable is not significant for zero citations in Biology & Biochemistry while it significantly associates with decreased zero citations in Social Sciences and Chemistry.

Spearman correlations are moderate between the number of institutions with the number of authors and the number of countries (Tables 7, 8 & 9). Perhaps because of collinearity, the results of the simultaneous hurdle model and the separate hurdle model for the number of institutions differ. In other words, in the analysis of the number of institutions together with the other variables, this variable associates with decreased citation counts while in a separate hurdle model for the number of institutions only, it significantly associates with increased citation counts in the three fields. The effect of this variable on citation counts was scrutinized separately in more detail. Keeping the number of authors and the number of countries constant at different values, extra hurdle models were run. In the majority of cases, the coefficient of the number of citations was not significant and the results were not consistent and varied from one number of countries to another. So the overall evidence of the impact of the number of institutions in Biology & Biochemistry is unclear (Table 10), but it seems that this is not an important factor.

Abstract readability

Abstract readability is a significant determinant of decreased citations in Biology & Biochemistry. A unit increase in the readability score decreases the mean citation count to 99.7% which statistically has no practical significance. Around 60% of readability scores range between 10 and 30 which is a change of 20 units. Hence, a twenty-unit increase in the readability score (i.e., from the bottom to the top of the normal range) decreases the mean citation count by only 6.1% (coef.=0.06), which is probably too small to be worth considering. No significant association was found between this variable and citation counts in Social Sciences and Chemistry. Moreover, with respect to the logit model, abstract readability is a significant determinant of zero citations in none of the fields.

Article size attributes

With respect to the negative binomial model, among the article size attributes, abstract length significantly associates with increased citations in all three fields, although its association is minor (0.2% in Social Sciences and Biology & Biochemistry and 0.1% in Chemistry). The number of keywords statistically significantly associates with decreased citations in the three fields but its association is of no practical significance. The number of keywords in the articles is 4 or 5 in about 45% of articles in all three fields. Therefore, the main unit of change in the number of keywords is one. A unit increase in the number of keywords associates with 99% decrease in the mean citation count, which is of no practical significance. Paper length is not a significant determinant of citations in Biology & Biochemistry and Social Sciences but each additional page contributes to a 2.8% increase in the mean citations for Chemistry. Title length statistically significantly associates with decreased citations in Biology & Biochemistry and Social Sciences, although its association is of no practical significance (Exponential Coefficient=99%). The number of words in the title ranges between 10 and 20 for around 70% of articles in Biology & Biochemistry and between 8 and 13 for around 50% of articles in Social Sciences. A ten-unit increase in the title length associates with a 9.4% increase in the mean citation counts in Biology & Biochemistry and a five-unit increase in this variable associates with a 7.7% increase in Social Sciences. No significant association was found between this factor and citations in Chemistry.

Spearman correlation results

Using a model to simultaneously assess the factors is one of the goals of this study that previous research has not addressed. Using Spearman correlations, the relationships between the factors and citation counts were also measured individually. Approximately the same results were found to those for the simultaneous assessment, showing that the advanced model has not uncovered any surprising relationships that were hidden by, or caused by, associations with other variables. The three research collaboration factors have a significant positive but weak correlation with citation counts. However, with respect to the negative binomial model a contradictory result was found for the number of institutions because this factor associates with decreased citations in Biology & Biochemistry and Chemistry and is not a significant determinant in Social Sciences. Multi-collinearity presumably causes this contradiction. Moreover, abstract length positively correlates with citations while there is a negative correlation between the number of keywords and citations in the three fields examined. However, with respect to the Spearman results the impact and number of references have similar correlation coefficients to the JIF and positively and significantly correlate with the number of citations.

The overdispersion parameters are significant in all three models, further justifying the negative binomial model (p for alpha<0.001).

Discussion

Journal Impact

The analysis of the factors affecting citation counts of the papers that are cited at least once indicates that the JIF is the main determinant of article citation impact in Biology & Biochemistry, Chemistry, and Social Sciences. The same result has also been found in the previous literature (Didegah & Thelwall, 2013; Boyack & Klavans, 2005; Callaham, Wears, & Weber, 2002). Based on the results, if the JIF increases by one unit, the mean citation counts of articles in Chemistry, Social Sciences and Biology & Biochemistry increase by 31.9%, 27.8%, and 15.6%, respectively. The JIF is measured based on the current number of citations to the journal articles published over the last two years divided by the number of articles in the two years considered.

The impact and the number of references

The impact and the number of cited references are also significant determinants of increased citation impact in the three fields. The results are consistent with previous studies (Didegah & Thelwall, 2013; Bornmann, Schier, Marx, & Daniel, 2012; Haslam et al., 2008; Kostoff, 2007; Boyack & Klavans, 2005; Peters & van Raan, 1994).

Research collaboration

Individual collaboration associates with increased citations in the three fields. Conversely, however, a study of a specific journal in Chemistry found no correlation between the number of authors and increased citation counts (Bornmann, Schier, Marx, & Daniel, 2012). This difference may result from the difference between the micro-level and macro-level analyses or the smaller sample size for the single journal studied giving insufficient statistical power to identify the association. In addition, the number of authors has not been found to be a significant determinant of citations in social and personality psychology (Haslam, et al., 2008). The authors believed that team-working is not necessarily a true reflection of research collaboration in this field.

International collaboration associates with increased citations in Biology & Biochemistry and Chemistry whereas it is neither a significant determinant of citation counts nor zero citations in Social Sciences. The number of countries has been significant for increased citations in the majority of previous studies except for an institutionally-limited investigation of Harvard University. This university is one of the world's top universities and it seems logical in this context that its researchers benefit more from institutional collaboration than from international collaboration (Gazni & Didegah, 2010).With regard to the negative binomial model, no clear evidence of the number of institutions was found in the three fields examined. This variable significantly associates with increased citation counts when it is individually modelled. The multi-collinearity between the number of institutions and the two other research collaboration variables may have caused the contradictory results of the simultaneous and non-simultaneous models for this variable. Moreover, the results reveal that inter-institutional collaboration does not have the influence value of the individual and international collaboration on the article citation impact.

The contradiction between the results of this study and some previous studies of international and institutional collaboration may result from the limited geographical and institutional coverage of previous research whereas the current study has a global coverage and seeks results at a macro-level. This study goes beyond a simple correlation between a predictor variable and citation counts. A co-analysis of predictors is considered here and the results are therefore more reliable although factors such as impact of authors, countries or institutions not considered in the analysis may also influence the results. Furthermore, the influence of research collaboration on research citation impact is not uniform and varies across domains particularly for the institutional and international types of collaboration (Gazni & Didegah, 2010; Sooryamoorthy, 2009). However, the positive impact of the number of authors on the citation counts in all fields of science is generally acknowledged (Franceschet & Costanini, 2010).

Abstract readability

Abstract readability was found to be a statistically significant determinant of decreased mean citations in Biology & Biochemistry. This variable is neither a significant determinant of citation counts nor zero citations in Chemistry and Social Sciences. The Spearman correlation between this variable and citation counts is close to zero in the three fields, although it is a negative correlation coefficient in Biology & Biochemistry and Chemistry but positive in Social Sciences. However, previous research confirmed a negative correlation between the abstract readability and citation impact of publications in the top institutions of the world (Gazni, 2011). There are numerous readability measures available but each of them has its own limitations (Gazni, 2011). To prevent the limitations of a single measure to negatively affect the results of the current study, a range of readability formulae was chosen to examine their associations with citation counts. But given that the different readability formulae significantly correlate with each other, the hypothesis that different formulae may differently influence the citation counts is not confirmed. However, all readability measures have two common limitations: first, they do not consider the characteristics of readers. The readers of scientific papers are experts in their own fields and have prior knowledge and interest in them; second, they fail to consider the characteristics of the text affecting text comprehension such as content familiarity, text structure, and author style (Armbruster, Osborn, & Davison, 1985). Hence an abstract graded as difficult based on its Flesch score may not be difficult for the scholars of the field (Gazni, 2011). On the other hand, scholars may scan the abstracts for keywords to find if a paper is relevant rather than reading the entire abstract. Therefore, this limitation may have affected the results and particularly the negative association between the readability score and article citation impact in Biology & Biochemistry may be due to this limitation.

Abstract length

The abstract length is another variable that significantly associates with increased citation counts in the three fields revealing that the longer the abstract, the higher the article citation impact, although the extent of its association is not considerable. The same result was found in Medicine: the longer the abstract, the higher the number of citations to the medical articles (Kostoff, 2007). Perhaps an extensive abstract is a more complete representation of the paper, providing readers with more details and enabling them to decide about the paper’s usefulness and this explains why an article with a longer abstract may receive more citations.

Number of keywords

The number of keywords statistically but not practically associates with decreased citations in the three fields. Keywords mainly aim at easing information retrieval that may also lead to a higher visibility. But they are mostly picked up from the title and abstract and that is probably why they are not important determinants of article citation impact.

Title length

The statistical association between the title length and citation counts reveals that the shorter the title, the higher the article citation impact in Biology & Biochemistry and Social Sciences, although the association is again of no practical significance. No significant association between this variable and citation counts was found in Chemistry. Whereas a negative correlation was found between the title length and citation counts to psychological articles (Haslam et al., 2008), the results of Spearman correlation in the current study show a positive, although weak, correlation between this variable and citation counts in the three fields. Given that an article's title is the first point of contact with the target readers, it should be informative enough and reflective of the article's content. An informative title can be of any size and that is why the title length is not an important factor of citations based on the combined statistical model.

Paper length

In a co-analysis of the paper length together with all the other variables, the paper length is not a significant determinant of citation counts in Biology & Biochemistry and Social Sciences but it significantly associates with increased citation counts when it is individually modelled. Since this variable significantly correlates with the number of references ($r{\sim}0.5$; p-value<0.001 in the three fields), it could be assumed that the paper length does not associate with increased citations unless a considerable number of references are cited. In other words, long articles with few cited references are not necessarily receiving higher number of citations in Biology & Biochemistry and Social Sciences. Conversely in Chemistry the number of pages is a significant determinant of increased citations together with the number of cited references. This suggests that long articles do not necessarily need to have a long list of cited references to receive higher number of citations in this area of science. This finding is contrary to previous studies that claim a higher citation impact for longer papers. The reason is probably that previous studies were not based on a simultaneous assessment of this variable together with some other factors (Haslam et al., 2008; Kostoff, 2007; Leimu & Koricheva, 2005a; Van Dalen & Henkens, 2005).

Conclusion

To answer the first research question, the journal and reference impact and the number of references are all significant determinants of increased citations to articles in the three fields.

For the second question, two types of research collaboration, the number of authors and the number of countries, significantly associate with increased citations in the three fields, except that there is no significant association between the number of countries and increased citation impact in Social Sciences. Probably due to multi-collinearity, there is surprisingly a negative association with the number of institutions and the number of citations in Biology & Biochemistry and Chemistry and an insignificant association in Social Sciences. More detailed analyses with the number of authors and the number of countries held constant shows that the number of institutions is not an important citation factor.

To answer the third question, among the article size attributes, abstract length significantly associates with increased citation impact in all fields. The number of keywords and the title length statistically associate with decreased citations. Article length associates with increased citation impact only in Chemistry. In Biology & Biochemistry and Social Sciences, article length also strongly associates with on the number of references. Therefore, longer papers per se do not associate with the increased citation impact unless they include more references.

For the fourth question, the abstract readability statistically but not practically associates with decreased citations in Biology & Biochemistry and it is a significant determinant of citations neither in Chemistry nor in Social Sciences.

Finally, the JIF increases the mean citation count more in Chemistry and Social Sciences than in Biology & Biochemistry. A unit increase in the impact factor increases the mean citation count by 31.9%, 27.8%, and 15.6% in Chemistry, Social Sciences and Biology & Biochemistry, respectively. The impact and the number of references both associate with increased citations. A unit increase in the reference impact associates with 0.8%, 0.5% and 0.4% increases in the mean citations to articles in Social Sciences, Biology & Biochemistry and Chemistry, respectively. Each extra reference also associates with 0.9%, 0.8% and 0.7% increases in the mean citations to Chemistry, Social Sciences and Biology & Biochemistry articles, respectively. Each additional author increases the mean number of citations by 2.9%, 1.3% and 0.9% in Social Sciences, Chemistry and Biology & Biochemistry and each additional country increases the mean by 8.6% and 5.5% in Chemistry and Biology & Biochemistry, respectively. Among article size attributes, abstract length associates with a 0.2% increase in the mean citation count in Social Sciences and Biology & Biochemistry and 0.1% increase in Chemistry.

Acknowledgements

This research is part of the FP7 EU-funded project ACUMEN on assessing Web indicators in research evaluation and it is an extended version of a paper accepted for presentation at the 14th International Conference on Scientometrics and Informetrics, with six more citation factors.

References

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159-170.

Armbruster, B.B. Osborn, J. H. & Davison, A. L. (1985). Readability Formulas May Be Dangerous to Your Textbooks. *Educational Leadership*, 42, 7, 18-20.

Bottle, R. S., Rennie, J. S., Russ, S. & Sardar, Z. (1983). Changes in the communication of chemical information I: some effects of growth. *Journal of Information Science,* 6, 103-108.

Bornmann, L. & Daniel, H. D. (2007). Multiple publication on a single research study: does it pay? The influence of number of research articles on total citation counts in biomedicine. *Journal of American Society for Information Science*, 58, 1100-1117.

Bornmann, L., Schier, H., Marx, W. & Daniel, H-D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics,* 6, 11-18.

Boyack, K. W. & Klavans, R. (2005). Predicting the Importance of Current Papers. *Proceedings of ISSI 2005*, 335–342. Edited by P. Ingwersen and B. Larsen. July 24-28, Stockholm, Sweden.

Callaham, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Journal of the American Medical Association*, 287, 2847-2850.

Cameron, C. A. & Trivedi, P. K. (2001). *Essentials of Count Data Regression, A companion to theoretical Econometrics*. Oxford: Blackwell.

Cherry, L. L. & Vesterman, W. (1981). *Writing tools: the STYLE and DICTION programs*. Bill Cronin Laboratories, Murray Hill, New Jersey.

Didegah, F. & Thelwall, M. (2013). Determinants of Research Citation Impact in Nanoscience and Nanotechnology. *Journal of the American Society for Information Science & Technology*, 64(55), 1055-1064.

Franceschet, M. & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics*, 4, 540-553.

Gazni, A. (2011). Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the world. *Journal of Information Science*, 37 (3), 273-281.

Gazni, A. & Didegah, F. (2010). Investigating Different Types of Research Collaboration and Citation Impact: A Case Study of Harvard University's Publications. *Scientometrics,* 87(2), 251-265.

Gazni, A., Sugimoto, C. R., & Didegah, F. (2012). Mapping world scientific collaboration: authors, institutions, and countries. *Journal of the American Society for Information Science & Technology,* 63(2), 323-335.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. Scientometrics, 51(1), 69-115.

Glänzel, W. & Schubert, A. (2001). Double effort=Double impact? A critical view at international co-authorship in chemistry. *Scientometrics,* 50(2), 199-214.

Goldfinch, S., Dale, T., & DeRouen, K. (2003). Science from the periphery: Collaboration, networks and “periphery effects” in the citation of New Zealand Crown Research Institutes articles, 1995–2000. *Scientometrics*, 57(3), 321–337.

Hartley, J., & Benjamin, M. (1998). An evaluation of structured abstracts in journals published by the British Psychological Society. *British Journal of Educational Psychology*, 68(3), 443-456.

Hartley, J., Sotto, E., Pennebaker, J. (2002). Style and substance in psychology: Are influential articles more readable than less influential ones? *Social Studies of Science*, 32, 321-334.

Hartley, J. & Sydes, M. (1997). Are structured abstracts easier to read than traditional ones? *Journal of Research in Reading*, 20(2), 122-136.

Hartley, J. & Trueman, M. (1992). Some observations on using journal articles in the teaching of psychology. *Psychology Teaching Review*, 1(1), 46-51.

Haslam, N. et al. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, 76(1), 169-185.

Hilbe, J. M. (2011). *Negative Binomial Regression, second edition* (5th print, 2012). Cambridge, UK: Cambridge University Press.

Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541-554.

Kostoff, R. (2007). The difference between highly and poorly cited medical articles in the journal Lancet. *Scientometrics*, 72, 513-520.

Larivière, V. & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology*, 61, 126-131.

Leimu, R., & Koricheva, J. (2005a). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, 20, 28-32.

Leimu, R., & Koricheva, J. (2005b). Does scientific collaboration increase the impact of ecological articles? *BioScience*, 55, 438-443.

Lovaglia, M. J. (1989). Status characteristics of journal articles for editor’s decisions and citations. *The Society for Social Studies of Science Annual Meeting*, November 15- 18, University of California at Irvine, Irvine, CA.

Meadows, A. J. (1998). *Communicating Research*. London: Academic Press.

Milojević, S. & Leydesdorff, L. (2013). Information Metrics (iMetrics): A Research Specialty with a Socio-Cognitive Identity? *Scientometrics*, 95(1), 141-157.

Nagaoka, S., Igami, M., Eto, M., & Ijichi, T. (2011). *Knowledge Creation Process in Science: Basic findings from a large-scale survey of researchers in Japan*. Report by National Institute of Science and Technology Policy (NISTEP), Institute of Innovation Research (IIR), Hitotsubashi University.

Narin, F. & Whitlow, E. S. (1990). *Measurement of scientific co-operation and co-authorship in CEC related areas of science*. Report EUR 12900. Office for Official Publications in the European Communities, Luxembourg.

Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific cooperation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313-323.

Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421-432.

Peters, H. P. F. & van Raan, A. F. J. (1994). On Determinants of Citation Scores: A Case Study in Chemical Engineering. *Journal of the American Society for Information Science*, 45(1), 39-49.

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, 81(1), 177-193.

Stremersch, S. Verniers, I. & Verhoef, P. C. (2007). The quest for citations: drivers of article impact. *Journal of Marketing*, 71, 171-193.

Van Dalen, H. P. & Henkens, K. E. (2005). Signals in science- on the importance of signalling in gaining attention in science. *Scientometrics*, 64, 209-233.

van Raan, A. F. J. (1998). In matters of quantitative studies of science: The fault of theorists is offering too little and asking too much. *Scientometrics*, 43(1), 129-139.

Vieira, E. S., & Gomes, J. A. N. F. (2010). Citation to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, 4(1), 1-13.

Zhao, D. Z. (2010). Characteristics and impact of grant-funded research: a case study of the library and information science field. *Scientometrics*, 84(2), 293-306.
Zimmerman, J. L. (1989). Improving a manuscript's readability and likelihood of publication. *Issues in Accounting Education*, 4, 458-466.

Appendix- Tables 1-10

Table 1. The correlation between the seven abstract readability scores in Biology & Biochemistry

Spearman Correlation	Kincaid	ARI	Coleman-Liau	Flesch Score	Fog Index	Lix	SMOG-Grading
Kincaid	1						
ARI	0.961	1					
Coleman-Liau	0.464	0.522	1				
Flesch Score	-0.868	-0.819	-0.772	1			
Fog Index	0.954	0.909	0.462	-0.849	1		
Lix	0.898	0.92	0.571	-0.827	0.88	1	
SMOG-Grading	0.948	0.905	0.457	-0.842	0.99	0.874	1

Table 2. The correlation between the seven abstract readability scores in Chemistry

Spearman Correlation	Kincaid	ARI	Coleman-Liau	Flesch Score	Fog Index	Lix	SMOG-Grading
Kincaid	1						
ARI	0.96	1					
Coleman-Liau	0.454	0.548	1				
Flesch Score	-0.85	-0.832	-0.773	1			
Fog Index	0.95	0.907	0.424	-0.808	1		
Lix	0.89	0.924	0.567	-0.812	0.87	1	
SMOG-Grading	0.927	0.89	0.446	-0.812	0.979	0.857	1

Table 3. The correlation between the seven abstract readability scores in Social Sciences

Spearman Correlation	Kincaid	ARI	Coleman-Liau	Flesch Score	Fog Index	Lix	SMOG-Grading
Kincaid	1						
ARI	0.97	1					
Coleman-Liau	0.408	0.429	1				
Flesch Score	-0.881	-0.821	-0.72	1			
Fog Index	0.957	0.917	0.419	-0.87	1		
Lix	0.895	0.917	0.541	-0.836	0.867	1	
SMOG-Grading	0.951	0.912	0.411	-0.861	0.991	0.858	1

Table 4. The results of hurdle model in Biology & Biochemistry

Logit Model	**Coef.**	**Exp(Coef.)**	**Std. Err.**	**z**	**P>z**	**[95% Conf.**	**Interval]**
No. of Authors	0.004	1.004	0.014	0.31	0.758	-0.024	0.033
No. of Institutions	-0.042	0.959	0.038	-1.1	0.272	-0.116	0.033
No. of Countries	0.055	1.057	0.072	0.77	0.443	-0.085	0.195
Abs. Readability	0.001	1.001	0.003	0.41	0.678	-0.004	0.006
Impact Factor	0.465	1.591	0.025	18.89	0.000	0.416	0.513
Ref. Impact	0.001	1.001	0.000	2.25	0.024	0.000	0.002
No. of Refs	0.017	1.017	0.003	6.54	0.000	0.012	0.022
Art. Length	-0.003	0.997	0.01	-0.25	0.803	-0.022	0.017
Abs. Length	0.001	1.001	0.000	1.76	0.079	0.000	0.002
No. of Keys	0.014	1.014	0.013	1.1	0.272	-0.011	0.039
Title Length	0.001	1.001	0.007	0.09	0.93	-0.013	0.014
Constant	0.374	1.454	0.162	2.32	0.02	0.058	0.691
NB Model	**Coef.**	**Exp(Coef.)**	**Std. Err.**	**z**	**P>z**	**[95% Conf.**	**Interval]**
No. of Authors	0.009	1.009	0.004	2.26	0.024	0.001	0.016
No. of Institutions	-0.032	0.968	0.011	-2.95	0.003	-0.054	-0.011
No. of Countries	0.054	1.055	0.02	2.72	0.006	0.015	0.093
Abs. Readability	-0.003	0.997	0.001	-3.35	0.001	-0.004	-0.001
Impact Factor	0.145	1.156	0.005	28.98	0.000	0.135	0.155
Ref. Impact	0.005	1.005	0.000	29.1	0.000	0.005	0.006
No. of Refs	0.007	1.007	0.001	11.1	0.000	0.006	0.008
Art. Length	0.001	1.001	0.003	0.45	0.653	-0.005	0.008
Abs. Length	0.002	1.002	0.000	11.44	0.000	0.001	0.002
No. of Keys	-0.007	0.993	0.004	-1.99	0.047	-0.014	0.000
Title Length	-0.009	0.991	0.002	-4.74	0.000	-0.012	-0.005
Constant	1.19	3.286	0.053	22.45	0.000	1.086	1.293
Alpha	**0.089**	**1.093**	**0.018**	**4.9**	**0.000**	**0.053**	**0.125**

Table 5. The results of hurdle model in Chemistry

Logit Model	**Coef.**	**Exp(Coef.)**	**Std. Err.**	**z**	**P>z**	**[95% Conf.**	**Interval]**
No. of Authors	0.055	1.056	0.015	3.75	0.000	0.026	0.083
No. of Institutions	-0.094	0.91	0.039	-2.44	0.015	-0.17	-0.019
No. of Countries	0.293	1.34	0.073	4.01	0.000	0.15	0.436
Abs. Readability	0.001	1.001	0.002	0.75	0.453	-0.002	0.005
Impact Factor	0.636	1.889	0.026	24.04	0.000	0.584	0.688
Ref. Impact	-0.007	0.993	0.003	-2.48	0.013	-0.013	-0.002
No. of Refs	0.017	1.017	0.002	8.13	0.000	0.013	0.021
Art. Length	0.000	1.000	0.008	0.01	0.994	-0.015	0.015
Abs. Length	0.001	1.001	0.000	2.47	0.014	0.000	0.002
No. of Keys	0.026	1.026	0.01	2.51	0.012	0.006	0.046
Title Length	0.03	1.03	0.005	5.66	0.000	0.019	0.04
Constant	-0.722	0.486	0.112	-6.47	0.000	-0.941	-0.504
NB Model	**Coef.**	**Exp(Coef.)**	**Std. Err.**	**z**	**P>z**	**[95% Conf.**	**Interval]**
No. of Authors	0.012	1.013	0.006	2.22	0.026	0.001	0.023
No. of Institutions	-0.073	0.929	0.018	-4.13	0.000	-0.108	-0.039
No. of Countries	0.082	1.086	0.03	2.78	0.005	0.024	0.14
Abs. Readability	-0.002	0.998	0.001	-1.85	0.064	-0.003	0.000
Impact Factor	0.277	1.319	0.01	27.49	0.000	0.257	0.296
Ref. Impact	0.004	1.004	0.001	3.07	0.002	0.001	0.007
No. of Refs	0.009	1.009	0.001	10.88	0.000	0.007	0.01
Art. Length	0.027	1.028	0.004	6.86	0.000	0.019	0.035
Abs. Length	0.001	1.001	0.000	3.57	0.000	0.000	0.001
No. of Keys	-0.024	0.977	0.005	-5.01	0.000	-0.033	-0.014
Title Length	0.000	1.000	0.002	0.15	0.883	-0.004	0.005
Constant	0.993	2.699	0.056	17.73	0.000	0.883	1.102
Alpha	**0.483**	**1.62**	**0.025**	**19.32**	**0.000**	**0.434**	**0.532**

Table 6. The results of hurdle model in Social Sciences

Logit Model	Coef.	Exp(Coef.)	Std. Err.	z	P>z	[95% Conf.	Interval]
No. of Authors	0.06	1.062	0.02	3	0.003	0.021	0.099
No. of Institutions	0.029	1.03	0.037	0.79	0.427	-0.043	0.101
No. of Countries	-0.035	0.965	0.071	-0.5	0.619	-0.175	0.104
Abs. Readability	0.004	1.004	0.002	1.76	0.078	0.000	0.008
Impact Factor	0.329	1.389	0.041	8.04	0.000	0.249	0.409
Ref. Impact	0.005	1.005	0.001	6.19	0.000	0.003	0.006
No. of Refs	0.008	1.008	0.001	5.39	0.000	0.005	0.01
Art. Length	0.006	1.006	0.004	1.61	0.108	-0.001	0.013
Abs. Length	0.002	1.002	0.000	4.78	0.000	0.001	0.003
No. of Keys	-0.029	0.971	0.009	-3.16	0.002	-0.048	-0.011
Title Length	-0.007	0.993	0.006	-1.24	0.214	-0.019	0.004
Constant	0.057	1.058	0.135	0.42	0.675	-0.208	0.322
NB Model	**Coef.**	**Exp(Coef.)**	**Std. Err.**	**z**	**P>z**	**[95% Conf.**	**Interval]**
No. of Authors	0.028	1.029	0.012	2.35	0.019	0.005	0.052
No. of Institutions	0.01	1.01	0.021	0.47	0.636	-0.031	0.052
No. of Countries	0.069	1.071	0.04	1.71	0.087	-0.01	0.148
Abs. Readability	0.002	1.002	0.001	1.53	0.126	-0.001	0.005
Impact Factor	0.245	1.278	0.025	9.85	0.000	0.196	0.294
Ref. Impact	0.008	1.008	0.001	13.99	0.000	0.007	0.009
No. of Refs	0.008	1.008	0.001	9.13	0.000	0.007	0.01
Art. Length	0.004	1.004	0.002	1.57	0.116	-0.001	0.009
Abs. Length	0.002	1.002	0.000	5.99	0.000	0.001	0.002
No. of Keys	-0.034	0.966	0.006	-5.53	0.000	-0.047	-0.022
Title Length	-0.015	0.985	0.004	-3.91	0.000	-0.022	-0.007
Constant	0.437	1.548	0.093	4.72	0.000	0.255	0.618
Alpha	0.822	2.276	0.044	18.58	0.000	0.736	0.909

Table 7. Spearman correlations in Biology & Biochemistry

Spearman	Citation Count	No. of Authors	No. of Institutions	No. of Countries	Abs. Readability	Impact Factor	Ref. Impact	No. of Refs	Art. Length	Abs. Length	No. of Keys	Title Length
Citation Count	1											
No. of Authors	0.076	1										
No. of Institutions	0.037	0.465	1									
No. of Countries	0.051	0.241	0.549	1								
Abs. Readability	-0.073	-0.042	-0.025	-0.015	1							
Impact Factor	0.455	0.142	0.07	0.083	-0.095	1						
Ref. Impact	0.416	0.098	0.019	0.000	-0.057	0.405	1					
No. of Refs	0.265	0.045	0.062	0.095	-0.111	0.372	0.17	1				
Art. Length	0.12	0.02	0.061	0.089	-0.077	0.146	0.017	0.554	1			
Abs. Length	0.153	0.1	0.058	0.053	0.052	0.193	0.03	0.334	0.348	1		
No. of Keys	-0.223	-0.079	-0.017	-0.024	0.063	-0.421	-0.204	-0.13	-0.018	-0.097	1	
Title Length	0.021	0.101	0.05	0.035	0.028	0.045	-0.028	0.113	0.13	0.221	-0.009	1

Table 8. Spearman correlations in Chemistry

Spearman	Citation Count	No. of Authors	No. of Institutions	No. of Countries	Abs. Readability	Impact Factor	Ref. Impact	No. of Refs	Art. Length	Abs. Length	No. of Keys	Title Length
Citation Count	1											
No. of Authors	0.055	1										
No. of Institutions	0.016	0.420	1									
No. of Countries	0.065	0.232	0.560	1								
Abs. Readability	-0.046	-0.037	0.016	0.016	1							
Impact Factor	0.459	0.095	0.034	0.077	-0.094	1						
Ref. Impact	0.359	0.071	0.015	0.010	-0.044	0.279	1					
No. of Refs	0.304	0.020	0.045	0.081	-0.061	0.454	0.187	1				
Art. Length	0.129	-0.016	0.075	0.092	0.049	0.059	-0.006	0.448	1			
Abs. Length	0.148	0.021	0.088	0.074	0.239	0.173	0.035	0.324	0.522	1		
No. of Keys	-0.112	0.016	0.006	-0.027	0.098	-0.255	-0.144	-0.034	0.188	0.142	1	
Title Length	0.107	0.048	0.049	0.035	0.022	0.135	0.008	0.177	0.152	0.253	0.085	1

Table 9. Spearman correlations in Social Sciences

Spearman	Citation Count	No. of Authors	No. of Institutions	No. of Countries	Abs. Readability	Impact Factor	Ref. Impact	No. of Refs	Art. Length	Abs. Length	No. of Keys	Title Length
Citation Count	1											
No. of Authors	0.129	1										
No. of Institutions	0.093	0.617	1									
No. of Countries	0.03	0.246	0.44	1								
Abs. Readability	0.014	0.033	0.011	0.026	1							
Impact Factor	0.186	0.275	0.217	0.107	0.008	1						
Ref. Impact	0.302	0.236	0.158	0.025	-0.032	0.234	1					
No. of Refs	0.104	-0.157	-0.063	-0.009	-0.167	0.035	0.023	1				
Art. Length	0.023	-0.309	-0.173	-0.043	-0.155	-0.193	-0.08	0.489	1			
Abs. Length	0.122	0.26	0.178	0.096	0.096	0.26	0.123	0.012	-0.129	1		
No. of Keys	-0.024	0.069	0.053	0.064	0.046	0.104	0.006	0.039	-0.079	0.155	1	
Title Length	0.014	0.16	0.103	0.039	-0.011	0.096	0.071	0.029	-0.026	0.176	0.072	1

Table 10. The results of extra hurdle models (only the negative binomial part) for the effect of the number of institutions on citation counts using a range of different fixed numbers of authors and countries (e.g., 3au_2cnty means 3 authors from 3 different countries)

Biology & Biochemistry					Chemistry				
Status	**Coef.**	**Exp (coef.)**	**P>|z|**	**Sample Size**	**Status**	**Coef.**	**Exp (coef.)**	**P>|z|**	**Sample Size**
2au_1cnty	-0.044	0.96	0.52	1935	2au_1cnty	-0.27	0.76	0.00	2562
3au_1cnty	-0.054	0.95	0.05	2307	3au_1cnty	-0.168	0.85	0.00	3090
4au_1cnty	-0.098	0.91	0.01	2144	4au_1cnty	-0.11	0.9	0.02	2686
5au_1cnty	-0.0003	0.99	0.9	1772	5au_1cnty	-0.065	0.94	0.18	1713
6au_1cnty	0.055	1.06	0.01	1315	6au_1cnty	-0.102	0.9	0.05	1008
7au_1cnty	-0.017	0.98	0.7	864	7au_1cnty	-0.1	0.9	0.14	505
8au_1cnty	-0.102	0.90	0.04	499	8au_1cnty	0.08	1.08	0.48	188
9au_1cnty	0.0054	1.01	0.9	325	9au_1cnty	0.03	1.03	0.74	135
10au_1cnty	0.125	1.13	0.1	199	10au_1cnty	0.028	1.03	0.8	67
3au_2cnty	0.02	1.02	0.85	377	3au_2cnty	0.03	1.03	0.84	424
4au_2cnty	-0.125	0.88	0.2	452	4au_2cnty	-0.069	0.93	0.53	513
5au_2cnty	0.02	1.02	0.68	448	5au_2cnty	-0.056	0.95	0.5	448
6au_2cnty	-0.11	0.9	0.04	423	6au_2cnty	-0.25	0.78	0.01	289