Journal of Informetrics 6 (2012) 11–18

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

# What factors determine citation counts of publications in chemistry besides their quality?

Lutz Bornmann[a,*], Hermann Schier[b], Werner Marx[b], Hans-Dieter Daniel[c,d]

[a] *Max Planck Society, Administrative Headquarters, Hofgartenstr. 8, D-80539 Munich, Germany*
[b] *Max Planck Institute for Solid State Research, Heisenbergstraße 1, D-70569 Stuttgart, Germany*
[c] *ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr. 24, CH-8092 Zurich, Switzerland*
[d] *University of Zurich, Evaluation Office, Mühlegasse 21, CH-8001 Zurich, Switzerland*

ARTICLE INFO

*Article history:*
Received 11 July 2011
Received in revised form 24 August 2011
Accepted 25 August 2011

*Keywords:*
Chemistry
Citation counts
Scientific quality
Single publication *h* index

ABSTRACT

A number of bibliometric studies point out that citation counts are a function of many variables besides scientific quality. In this paper our aim is to investigate these factors that usually impact the number of citation counts, using an extensive data set from the field of chemistry. The data set contains roughly 2000 manuscripts that were submitted to the journal *Angewandte Chemie International Edition* (AC-IE) as short communications, reviewed by external reviewers, and either published in AC-IE or, if not accepted for publication by AC-IE, published elsewhere. As the reviewers' ratings of the importance of the manuscripts' results are also available to us, we can examine the extent to which certain factors that previous studies demonstrated to be generally correlated with citation counts increase the impact of papers, controlling for the quality of the manuscripts (as measured by reviewers' ratings of the importance of the findings) in the statistical analysis. As the results show, besides being associated with quality, citation counts are correlated with the citation performance of the cited references, the language of the publishing journal, the chemical subfield, and the reputation of the authors. In this study no statistically significant correlation was found between citation counts and number of authors.



## 1. Introduction

Researchers publish papers not only to report the results of their own research but also so that other researchers use them as a reference point in their own work (Evidence Ltd., 2007). When researchers use others' research as a basis for their own work, they usually acknowledge this by citing that research in their publications: they cite the publication that reported it. If a publication was cited in other publications very frequently, it was evidently more useful for (more frequently the basis for) the research by other researchers in a field of science than a publication that was cited by other publications only seldom. This attention that a publication in the scientific community has received is interpreted "as an indicator of the importance, the visibility, or the impact of the researcher or the paper" (Hemlin, 1996, p. 221). This interpretation makes citation counts interesting for evaluation of research.

For appropriate assessment of research performance in science, evaluation faces the problem that "there is no mathematical formula that can quantify the 'quality' of an article" (Figueredo, 2006). For this reason, assessors have to depend on indicators that can adequately reflect the quality of a paper. Owing to the lack of alternatives that could be possible indictors, citation counts have become a normal part of the evaluation of papers (Garfield, 2002): the quality of a paper should be

* Corresponding author. Tel.: +49 89 2108 1265.
*E-mail address:* bornmann@gv.mpg.de (L. Bornmann).

1751-1577/$ – see front matter 
doi:10.1016/j.joi.2011.08.004

assessed as higher, the more frequently a paper is cited. According to van Raan (1996), citations yield “a good to even very good quantitative impression of at least one important aspect of quality, namely international impact” (p. 404). Citation analysis has long since become one of the most important instruments of research evaluation, so that for researchers it is no longer only important to publish as much as possible, but it also has to be demonstrable that the publications are “also heavily used, hence useful and important” (Harnad, 2007, p. 27).

The use of citation counts as indicators for scientific quality is justified mainly by the fact that citation counts for single publications correlate positively with quality indicators at a higher level of aggregation. Baird and Oppenheim (1994) wrote, for example: “Whatever measure you take for the eminence of an individual scientist or of a journal or of an institution, citation counts provide strong correlation with that result . . . high citation counts mean a statistical likelihood of high quality research” (p. 8). Cole (1992) put it this way:

> Extensive past research indicates that citations are a valid indicator of the subjective assessment of quality by the scientific community. The number of citations is highly correlated with all other measures of quality that sociologists of science employ. As long as we keep in mind that research of high quality is being defined as research that other scientists find useful in their current work, citations provide a satisfactory indicator. (p. 221)

However, Eugene Garfield, the founder of the Institute of Scientific Information (ISI, now Thomson Reuters, Philadelphia, PA, USA), already pointed out in the early 1970s that citation counts are a function of many variables besides scientific quality (Garfield, 1972). Since then, a number of variables that usually correlate with citation counts have emerged in bibliometric studies. For instance, Lawani (1986) and other researchers established that there is a positive relation between the number of authors of a publication and its citation count: a higher number of authors is usually associated with a higher number of citations.

In this paper we want to use an extensive data set from the field of chemistry to investigate a number of these factors that usually correlate with citation counts. The data set contains roughly 2000 manuscripts that were submitted to the journal *Angewandte Chemie International Edition* (AC-IE) as short communications, reviewed by external reviewers, and either published in AC-IE or, if rejected for publication by AC-IE, published elsewhere. For each manuscript in the data set we have the reviewers' ratings of the importance of the obtained results. We can therefore examine to what extent certain factors that other studies demonstrated to be associated with citation counts increase the impact of papers, controlling in the statistical analysis for the quality of the manuscript (here: its importance) via the ratings of the reviewers.

If the correlation of these factors with citation counts is examined empirically and it is found, for example, that a higher number of authors is associated with a higher citation count, it cannot be ruled out that the higher number of authors led to an increase in the quality of a paper, which is what led to the higher citation count. In this study, to ensure that the correlation of the factors (such as, for example, number of authors) with citation counts is measured *independently* of the quality of the paper, the factors and in addition the reviewers' ratings were included in the analysis.

## 2. Factors that usually correlate with citation counts

The research activity of researchers, publication of their findings, and citation of the publications by colleagues in the field are all social activities. This means that citation counts for the publications are not only an indicator of the impact of the researchers' scientific work (as one part of the quality of their work) on the advancement of scientific knowledge. Citation counts also reflect (social) factors that do not have to do with the accepted conventions of scholarly publishing (Bornmann & Daniel, 2008c). As Martin and Irvine (1983) described it:

> There are ‘imperfections’ in the scientific communications system, the result of which is that the importance of a paper may not be identical with its impact. The ‘impact’ of a publication describes its actual influence on surrounding research activities at a given time. While this will depend partly on its importance, it may also be affected by such factors as the location of the author, and the prestige, language, and availability of the publishing journal. (p. 70)

Bibliometric studies published in recent years have generally revealed an association of a number of factors with citation counts. For the factors presented in the following (such as the number of authors of a paper, or the performance of papers cited in a paper), in this study we examine the association with citation counts when *controlling for* quality (as measured by the rated importance of the research paper).

### *2.1. Number of authors*

A number of studies found that “multi-authorship increases above all the probability to be cited by others” (Glänzel, Debackere, Thijs, & Schubert, 2006, p. 274): the more authors a paper has, the higher the number of citations of this paper that can be expected. Four reasons for this association can be found in the literature (see here Leimu & Koricheva, 2005): (1) each additional author increases the probability of self-citations, (2) papers with many authors are most probably multidisciplinary papers, so that citations in various disciplines can be expected, (3) the more authors a paper has, the larger the network in which the paper will become known through personal contacts, and (4) not only informal but also formal communication in the scientific community can contribute to the greater visibility (and thus to a higher citation count) of a multi-authorship paper: “The longer the author list is, the greater the probability of the paper being presented to several conferences is, especially if the team is multidisciplinary" (Valderas, 2007). (5) Scientific communities with very expensive experiments, e.g., high energy physics or astrophysics, can afford only very few experiments, sometimes only one, and developed worldwide cooperation for doing them documented by an exhaustive list of authors.

### 2.2. Single publication h index for the cited references

"It is not possible to publish new material whatever its quality without demonstrating a minimal overlap with the status quo by including relevant references to reach this aim" (Opthof & Wilde, 2009, p. 146). As Webster, Jonason, and Schember (2009) and Vieira and Gomes (2010) showed, there is a positive correlation between citation counts and the number of cited references: the more cited references a paper contains, the higher the citation count a paper will be expected to have. Webster et al. (2009) found, for example, that "reference counts explained 19% of the variance in citation counts" (p. 356). Webster et al. concluded that one of the reasons for this connection was that "the tit-for-tat nature of 'I cite you, you cite me,' may be at work: the more people you cite in your paper, the more people are likely to cite your paper (the paper they were cited in) in the future" (p. 349). Two further studies (Boyack & Klavans, 2005; Lancho-Barrantes, Guerrero-Bote, & Moya-Anegon, 2010) found that not only the number but also the citation impact of the cited references is correlated with the citation counts for a paper: the higher the impact of the cited references, the higher the later impact of the citing paper.

Today, the *h* index (Hirsch, 2005) is a widely used measure of scientific performance: "The automatic calculation of *h*-indices has even become a built-in feature of major bibliographic databases such as Web of Science and Scopus" (van Eck & Waltman, 2008, p. 263). Customarily the *h* index is used as a single number that provides information on the publication output of a scientist or a journal and the citation impact of these publications (Bornmann & Daniel, 2007, 2009c). Recently, Schubert (2009) suggested calculating the *h* index for a single publication (see here also Thor & Bornmann, 2011):

> A Hirsch-type index can be used for assessing single highly cited publications by calculating the *h*-index of the set of papers citing the work in question. This index measures not only the direct impact of a publication but also its indirect influence through the citing papers. (p. 559)

In this study, to examine whether both the number and impact of the cited references correlate with the citation counts for a paper, we calculated the *h* index for the cited references for each paper and included it in the analysis.

### 2.3. Language of the journal in which the publication appeared

A further factor that is assumed to have an influence on citation counts is the language of the journal in which a publication appears (see here, for example, the study of Lansingh & Carter, 2009). With regard to language the main thing is whether a journal publishes papers in English or papers in another language. Since English is the lingua franca in practically all science disciplines (including chemistry), for reception in the scientific community papers published in other languages generally have a disadvantage compared to papers published in English.

### 2.4. Chemical subfield

The chance of a paper being cited is especially related to the number of papers published in the different disciplines and fields. For this reason, far fewer citations are to be expected in small fields than in more general fields. A number of studies showed that there are very great differences in average citation counts between disciplines and between the subfields of a discipline. For instance, Neuhaus and Daniel (2009) examined the biochemistry sections of Chemical Abstracts (CA). Chemical Abstracts Services (CAS, Columbus, OH, USA) categorizes publications in 80 different sections, each section covering only one broad area of scientific inquiry. Each abstract in CA appears in only one CA section, categorized according to the most important aspects of the publication. The 80 individual sections are listed according to five broad headings for five main areas of chemical research: (1) organic, (2) physical, inorganic, and analytical, (3) macromolecular, (4) applied, and (5) biochemistry. As Neuhaus and Daniel found, the average numbers of citations for the 20 sections in the area of biochemistry ranged from 2.37 for "Agrochemical bioregulators" to 19.67 for "General biochemistry."

Because the papers used in this study are chemistry-related publications, in the statistical analysis we used the subfield-specific assignments from CA. The papers' assignments to the five main areas mentioned above were entered into the statistical analysis, and the association with citation counts was examined.

### 2.5. Number of authors of a paper who are listed in ISIHighlyCited.com

Highly cited scientists may attract more citations, because they are well-known within their field (Peters & van Raan, 1994). Consistent with Robert K. Merton's interpretation of the Matthew effect in science (Merton, 1968) and Cozzens (1985) "success–breeds–success" phenomenon, publications by authors whose works have been very frequently cited in the past can be expected to be cited more often than publications by authors who have not published highly cited works in the past. This means that publications of the same intrinsic worth will be cited differently depending on the status of the author (see also Garfield, 2002).

The status of the authors of a publication as most highly influential scientists in an area of research was entered into the statistical analysis of this study as authors who are listed in ISIHighlyCited.com, which is a part of the ISI Web of Knowledge platform provided by Thomson Reuters. ISIHighlyCited.com provides "a tool to identify individuals ... that have made fundamental contributions to the advancement of science and technology in recent decades... These individuals are the most highly cited within each category for the period 1981 to 1999" (ISIHighlyCited.com). For each paper in the data set, we determined the number (percentage) of a paper's authors who are listed in ISIHighlyCited.com.

Further (possible) factors which correlate with citation counts but were not taken into account in this study are study design and study topic (Willis, Bahler, Neuberger, & Dahm, 2011), the presence of industry funding and an industry-favoring result (Kulkarni, Busse, & Shams, 2007) and the impact factor of the original publishing journal (Callaham, Wears, & Weber, 2002).

## 3. Methods

### 3.1. Manuscript review at AC-IE

A manuscript submitted to AC-IE usually undergoes internal and external review. First, editors at the journal evaluate whether the manuscript contributes to the development of an important area of research (internal review). If the editorial office finds that this is the case, the submitted manuscript is sent to several independent referees (external review), who review it using an evaluation form and a comment sheet (Bornmann, Weymuth, & Daniel, 2010). The evaluation form for reviewers contained among other things the following question concerning the importance of a manuscript: "How important do you consider the results?" (four response categories: very important, important, less important, unimportant). The journal editors make the decision to accept or reject a manuscript for publication on the basis of these reviews and their own evaluations (Bornmann & Daniel, 2009b, 2010).

### 3.2. Dataset for the study

For investigation of the AC-IE peer review process, information on a total of 1899 manuscripts reviewed in the year 2000 was used. The information was drawn from material in the archives of the journal's publishing house, Wiley-VCH. In addition to internal review by the publisher, there were a total of 4593 external reviews of the 1899 manuscripts using an evaluation form and/or a comment sheet. For the statistical analysis in this study, for each manuscript we determined the median of the independent ratings of the importance of a submission. According to Thorngate, Dawes, and Foddy (2009) the average error in ratings decreases with an increasing number of raters. Based on the external reviews, 46% ($n = 878$) of the 1899 manuscripts were accepted for publication in AC-IE, and 54% ($n = 1021$) were rejected. Research in the literature databases Web of Science (WoS, Thomson Reuters) and CA revealed that 959 (94%) of the 1021 rejected manuscripts were published later in other journals in a more or less revised form (Bornmann & Daniel, 2008a, 2008b, 2009a; Bornmann et al., 2009; Bornmann, Mutz, Marx, Schier, & Daniel, 2011).

In this study 1765 manuscripts could be included that were published in AC-IE (as short communications) or in one of 125 other journals (as short communications, articles, etc.). Of the total 1899 manuscripts, only those could be used in this study that had no missing values for the variables included in the statistical analysis. For each of the 1765 accepted or rejected (but published elsewhere) manuscripts, we determined in WoS the number of citations from time of publication up to March 2010. For the calculation of the $h$ index for the cited references for each paper, we retrieved all references from the reference list of the papers in WoS, including the times-cited information for each reference (which is published in a WoS source journal). The single publication $h$ index for the references was computed from the list sorted by times-cited. The time window for the citation data covers the period from the year of publication to the date of searching (March 2010).

### 3.3. Statistical procedure

We performed a multiple regression analysis, which reveals the factors that exert a primary effect on a certain outcome. The coefficients in the regression model, called "partial" regression coefficients (Rabe-Hesketh & Everitt, 2004), represent the effects of each factor, controlling for all other factors in the model. Since the skewness of citation counts (Seglen, 1992) suggests the use of a negative binomial specification (Glänzel & Schubert, 1993), we calculated a Negative Binomial Regression Model (NBRM, Long & Freese, 2006, section 8.3) (see also Bornmann & Daniel, 2006). The citation counts for the accepted and rejected manuscripts that were published in AC-IE or elsewhere enter into the NBRM as a dependent variable. The independent variables considered were the factors that were found in previous research to be associated with citation counts (see Section 2) and the reviewers' ratings (the indicator of scientific quality).

In the NBRM the publication year of each accepted and rejected (but published elsewhere) manuscript was included as exposure time (Long & Freese, 2006, pp. 370–372). By using the exposure option provided in the statistical package Stata (StataCorp., 2009), the amount of time that a paper is "at risk" of being cited is considered. The violation of the assumption of independent observations by including citation counts of more than one paper per journal is considered in the NBRM by using the cluster option in Stata (StataCorp., 2009). This option specifies that the citation counts are independent across papers published in different journals but are not necessarily independent within papers of the same journal (see Hosmer &

**Table 1**
Description of the dependent and independent variables.

| Variable | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|
| Dependent variable | | | | |
| Citation counts | 42.79 | 49.76 | 0 | 501 |
| Independent variables | | | | |
| Median reviewers' ratings of importance (1 = very important, 4 = unimportant) | 2.28 | 0.59 | 1 | 4 |
| Single publication *h* index for the cited references | 21.67 | 12.72 | 3 | 167 |
| Number of authors | 4.26 | 1.95 | 1 | 16 |
| Language of the journal in which the publication appeared (0 = English, 1 = multiple or non-English language) | 0.07 | | 0 | 1 |
| Chemical subfield | | | | |
| Organic chemistry | 0.52 | | 0 | 1 |
| Physical, inorganic, and analytical chemistry | 0.29 | | 0 | 1 |
| Macromolecular chemistry | 0.06 | | 0 | 1 |
| Applied chemistry | 0.03 | | 0 | 1 |
| Biochemistry (reference category) | 0.10 | | 0 | 1 |
| Number (percentage) of authors of a paper who are listed in ISIHighlyCited.com | 0.03 | 0.09 | 0 | 0.5 |

Lemeshow, 2000, section 8.3). According to Meadows (1998) "the more highly regarded a journal, the more likely it is that researchers will want to make use of its contents" (p. 165).

Normally, the Journal Impact Factor (Garfield, 2006) is considered to be a numerical indicator of the prestige of a journal. But by using the cluster option in the statistical analysis, we did without the use of the Journal Impact Factor as a predictive factor in citation counts. According to Seglen (1997) "article citation rates determine the journal impact factor, not vice versa." Leimu and Koricheva (2005) found that contrary to the widespread belief "that publication in a high-impact journal might by itself enhance the citation rate of an article by increasing its visibility or persuasiveness of the arguments presented" (p. 29), their study results did "not support this 'journal effect' hypothesis, because there was considerable variation in citation rates, especially for papers published in high-impact journals" (p. 29).

## 4. Results

Table 1 shows a description of the dependent and independent variables that were included in the NBRM. In addition to the citation counts, the models take into account: the median reviewers' ratings of the importance of the paper, the single publication *h* index for the cited references, the number of authors, the language (English, or non-English or multiple) of the journal in which the publication appeared, the chemical subfield, and the number (percentage) of authors of a paper listed in ISIHighlyCited.com. As the table shows, the total 1765 papers that were included in the study were cited on average 42.79 times (arithmetic average), with a range from 0 citations to 501 citations. The single publication *h* index for the cited references was on average 21.67, with a range from a minimum of 3 to a maximum of 167. With an *h* index of 167, at least 167 papers in a cited reference list were cited at least 167 times. As the assignments of the papers to the chemical subfields in Table 1 shows, approximately half of the papers were assigned by CAS to organic chemistry; 29% of the papers belong to the physical, inorganic, and analytical chemistry categories.

Table 2 presents the results of the NBRM for predicting citation counts for accepted and rejected (but published elsewhere) manuscripts. Statistically significant effects (independently of the importance of individual papers) could be found in the expected direction for four factors that bibliometric studies have demonstrated to generally correlate with citation counts: (a) the higher the single publication *h* index for the cited references in a publication, the higher the expected citation count for that paper; (b) more citations are to be expected for a publication in an English-language journal than for a publication in a multiple-language or non-English language journal; (c) the results reveal an association between citations and the research area. Publications in physical, inorganic, and analytical chemistry were more frequently cited than publications in biochemistry (the reference category; AC-IE papers in biochemistry are mostly in the field of synthetic biochemistry and not in genetics or in life sciences); (d) a higher number (percentage) of the authors of a paper who are listed in ISIHighlyCited.com is associated with a greater number of citations of a publication. However, contrary to expectations, the NBRM reveals no statistically significant effect for number of authors of a paper (although the (positive) sign of the coefficient is in the expected direction).

To gain an indication of the size of the effect of those factors found to be statistically significantly correlated with citation counts, following the NBRM we calculated percent changes in expected count for a unit increase in an independent variable (see Long & Freese, 2006). For example, for the variable "language of the journal," this calculation shows that being published in a multiple language or non-English language journal decreases the expected number of citations by about 48% – holding all

**Table 2**
Negative binomial regression model predicting citation counts.

| Independent variable | ML-point estimates | Percent change in expected count for a unit increase |
|---|---|---|
| Median reviewers' ratings of importance (1 = very important, 4 = unimportant) | −0.238*** (−4.09) | |
| Single publication *h* index for the cited references | 0.0264*** (14.53) | 2.7 |
| Number of authors | 0.0227 (1.58) | |
| Language of the journal (1 = multiple or non-English language) | −0.650*** (−3.74) | −47.8 |
| Organic chemistry | −0.0130 (−0.20) | |
| Physical, inorganic, and analytical chemistry | 0.262** (2.98) | 30 |
| Macromolecular chemistry | 0.101 (1.41) | |
| Applied chemistry | 0.0573 (0.35) | |
| Number (percentage) of authors of a paper who are listed in ISIHighlyCited.com | 0.548*** (4.01) | 73 |
| Intercept | −4.111*** (−33.99) | |
| $n_{papers}$ | 1765 | |
| $n_{journals\,(clusters)}$ | 125 | |
| Papers per journal (cluster) | Minimum = 1<br>Mean = 14.1<br>Maximum = 877 | |

*Notes*: *t* statistic in parentheses.
** $p < 0.01$.
*** $p < 0.001$.

other variables constant. For each additional increase in the single publication *h* index for the cited references of a publication by a value of 1, the odds of being cited increases by nearly 3% – holding all other variables constant.

## 5. Discussion

Citation analysis is one of the most important instruments in the evaluation of research. Citation analyses have been conducted for assessment of national science policies and disciplinary development (Lewison, 1998; Oppenheim, 1997; Tijssen, van Leeuwen, & van Raan, 2002), departments and research laboratories (e.g., Bayer & Folger, 1966; Narin, 1976), books and journals (e.g., Garfield, 1972; Nicolaisen, 2002), and individual scientists (e.g., Cole & Cole, 1973; Garfield, 1970). In these studies the number of citations of papers was used to measure the impact of the work of scientists on the scientific community, as high quality work by a scientist will trigger more responses (citations) from scientific colleagues than low quality work (Raan, Visser, van Leeuwen, & van Wijk, 2003). However, for many years now, the usefulness of citation counts for measuring research impact has been questioned (see here MacRoberts & MacRoberts, 2010). It is thought that the probability of being cited depends on many factors that do not have to do with the accepted conventions of scholarly publishing.

In this study we examined the correlation between citation counts and a number of the factors that previous studies had demonstrated to be associated with citation counts. In contrast to many other studies, however, the distinguishing characteristic of this study is that the correlation of these factors with the citation counts of individual papers was examined independently of the quality of the papers (as measured by their rated importance). This means that the *adjusted* covariation between one factor and citation impact was determined. By using this particular design of analysis, our aim was to make it largely impossible for higher or lower citation counts for a group of papers to *not* be attributable to their higher or lower importance but rather attributable to membership in a certain group (for example, published in a non-English journal). We examined a total of five factors: (1) the single publication *h* index for the cited references, (2) the number of authors, (3) the language (English, or non-English or multiple) of the journal in which the publication appeared, (4) the chemical subfield, and (5) the number (percentage) of authors of a paper who are listed in ISIHighlyCited.com.

This study found a statistically significant correlation with citation counts for four of five factors. Hence, the citation counts of a paper in chemistry are not only connected with scientific quality (as measured here by reviewers' ratings) but also with the performance of the references cited in the paper, the language of the publishing journal, the chemical subfields, and the reputation of the authors. As other studies have pointed out the importance of these factors as well (see Section 2), based on the findings we recommend that these factors be taken into consideration in evaluations of a citation analysis in chemistry. To control for these factors, they could – whenever possible – be considered in the statistical analysis (a regression model) of bibliometric data as independent variables in addition to the variable of actual interest (see here as an example Bornmann & Daniel, 2006).

As this study looked at only chemistry-related papers, it would be necessary to conduct a similar study also in other fields. It would be important to examine whether in other fields the factors investigated here have an impact on citation counts and whether further factors also play a role. For these studies we find it imperative that, just as we did here, the scientific quality of the individual papers is controlled for using adequate indicators. It will also be important for further studies to examine the association between number of authors and citations. Contrary to the findings of previous studies, this study did not find a statistically significant correlation between this factor and citation counts.

## Acknowledgements

The entire research study, which is also investigating quality assurance of open access journals, is supported by a grant from the Max Planck Society. The authors would like to thank Dr. Christophe Weymuth (formerly at the Organic Chemistry Institute of the University of Zurich and now at BIOSYNTH AG, Switzerland) for investigation of the manuscripts rejected by *Angewandte Chemie International Edition* and published elsewhere.

We thank Dr. Peter Gölitz, Editor-in-Chief of *Angewandte Chemie*, the Editorial Board of *Angewandte Chemie*, and the German Chemical Society (GDCh, Frankfurt am Main, Germany) for permission to conduct the evaluation of the selection process of the journal, and to the members of the editorial office for their generous support during the carrying out of the study.

## References

Baird, L. M. & Oppenheim, C. (1994). Do citations matter? *Journal of Information Science*, *20*(1), 2–15.

Bayer, A. E. & Folger, J. (1966). Some correlates of a citation measure of productivity in science. *Sociology of Education*, *39*(4), 381–390.

Bornmann, L. & Daniel, H.-D. (2006). Selecting scientific excellence through committee peer review – A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics*, *68*(3), 427–440.

Bornmann, L. & Daniel, H.-D. (2007). What do we know about the *h* index? *Journal of the American Society for Information Science and Technology*, *58*(9), 1381–1385. doi:10.1002/asi.20609

Bornmann, L. & Daniel, H.-D. (2008a). The effectiveness of the peer review process: Inter-referee agreement and predictive validity of manuscript refereeing at *Angewandte Chemie*. *Angewandte Chemie International Edition*, *47*(38), 7173–7178. doi:10.1002/anie.200800513

Bornmann, L. & Daniel, H.-D. (2008b). Selecting manuscripts for a high impact journal through peer review: A citation analysis of Communications that were accepted by *Angewandte Chemie International Edition*, or rejected but published elsewhere. *Journal of the American Society for Information Science and Technology*, *59*(11), 1841–1852. doi:10.1002/asi.20901

Bornmann, L. & Daniel, H.-D. (2008c). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. doi:10.1108/00220410810844150

Bornmann, L. & Daniel, H.-D. (2009a). Extent of type I and type II errors in editorial decisions: A case study on *Angewandte Chemie International Edition*. *Journal of Informetrics*, *3*(4), 348–352.

Bornmann, L. & Daniel, H.-D. (2009b). The luck of the referee draw: The effect of exchanging reviews. *Learned Publishing*, *22*(2), 117–125. doi:10.1087/2009207

Bornmann, L. & Daniel, H.-D. (2009c). The state of *h* index research. Is the *h* index the ideal way to measure research performance? *EMBO Reports*, *10*(1), 2–6. doi:10.1038/embor.2008.233

Bornmann, L. & Daniel, H.-D. (2010). The manuscript reviewing process: Empirical research on review requests, review sequences and decision rules in peer review. *Library & Information Science Research*, *32*(1), 5–12. doi:10.1016/j.lisr.2009.07.010

Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A. & Daniel, H. D. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry. Citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of Informetrics*, *3*(1), 27–35. doi:10.1016/j.joi.2008.11.001

Bornmann, L., Mutz, R., Marx, W., Schier, H. & Daniel, H.-D. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: Do the editors of a high-profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society – Series A (Statistics in Society)*, *174*(part 4) doi:10.1111/j.1467-985X.2011.00689.x

Bornmann, L., Weymuth, C. & Daniel, H.-D. (2010). A content analysis of referees' comments: How do comments on manuscripts rejected by a high-impact journal and later published in either a low- or high-impact journal differ? *Scientometrics*, *83*(2), 493–506.

Boyack, K. W. & Klavans, R. (2005). Predicting the importance of current papers. In P. Ingwersen, & B. Larsen (Eds.), *Proceedings of the 10th international conference of the international society for scientometrics and informetrics* (pp. 335–342). Stockholm, Sweden: Karolinska University Press.

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*(21), 2847–2850.

Cole, J. R. & Cole, S. (1973). *Social stratification in science*. Chicago, MA, USA: The University of Chicago Press.

Cole, S. (1992). *Making science. Between nature and society*. Harvard University Press: Cambridge, MA, USA.

Cozzens, S. E. (1985). Comparing the sciences – Citation context analysis of papers from neuropharmacology and the sociology of science. *Social Studies of Science*, *15*(1), 127–153.

Evidence Ltd. (2007). *The use of bibliometrics to measure research quality in UK higher education institutions*. London, UK: Universities UK.

Figueredo, E. (2006). The numerical equivalence between the impact factor of journals and the quality of the articles. *Journal of the American Society for Information Science and Technology*, *57*(11), 1561.

Garfield, E. (1970). Citation indexing for studying science. *Nature*, *227*(5259), 669–671.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, *178*(4060), 471–479.

Garfield, E. (2002). Highly cited authors. *Scientist*, *16*(7), 10.

Garfield, E. (2006). The history and meaning of the Journal Impact Factor. *Journal of the American Medical Association*, *295*(1), 90–93.

Glänzel, W., Debackere, K., Thijs, B. & Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics*, *67*(2), 263–277.

Glänzel, W. & Schubert, A. (1993). A characterization of scientometric distributions based on harmonic means. *Scientometrics*, *26*(1), 81–96.

Harnad, S. (2007). Open Access scientometrics and the UK Research Assessment Exercise. In D. Torres-Salinas, & H. F. Moed (Eds.), *Proceedings of the 11th conference of the international society for scientometrics and informetrics* (pp. 27–33). Madrid, Spain: Spanish Research Council (CSIC).

Hemlin, S. (1996). Research on research evaluations. *Social Epistemology*, *10*(2), 209–250.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572. doi:10.1073/pnas.0507655102

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Chichester, UK: John Wiley & Sons, Inc.

Kulkarni, A. V., Busse, J. W. & Shams, I. (2007). Characteristics associated with citation rate of the medical literature. *PLoS One*, *2*(5) doi:10.1371/journal.pone.0000403

Lancho-Barrantes, B. S., Guerrero-Bote, V. P. & Moya-Anegon, F. (2010). What lies behind the averages and significance of citation indicators in different disciplines? [Article]. *Journal of Information Science*, *36*(3), 371–382. doi:10.1177/0165551510366077

Lansingh, V. C. & Carter, M. J. (2009). Does Open Access in ophthalmology affect how articles are subsequently cited in research? [Article]. *Ophthalmology*, *116*(8), 1425–1431. doi:10.1016/j.ophtha.2008.12.052

Lawani, S. M. (1986). Some bibliometric correlates of quality in scientific research. *Scientometrics*, *9*(1–2), 13–25.

Leimu, R. & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, *20*(1), 28–32.

Lewison, G. (1998). Gastroenterology research in the United Kingdom: Funding sources and impact. *Gut*, *43*(2), 288–293.

Long, J. S. & Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (2nd ed.). College Station, TX, USA: Stata Press, Stata Corporation.

MacRoberts, M. H. & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, *61*(1), 1–12.

Martin, B. R. & Irvine, J. (1983). Assessing basic research – some partial indicators of scientific progress in radio astronomy. *Research Policy*, *12*(2), 61–90.

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

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56–63.

Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ, USA: Computer Horizons.

Neuhaus, C. & Daniel, H.-D. (2009). A new reference standard for citation analysis in chemistry and related fields based on the sections of Chemical Abstracts. *Scientometrics*, *78*(2), 219–229.

Nicolaisen, J. (2002). The J-shaped distribution of citedness. *Journal of Documentation*, *58*(4), 383–395.

Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation*, *53*(5), 477–487.

Opthof, T. & Wilde, A. A. M. (2009). The Hirsch-index: A simple, new tool for the assessment of scientific output of individual scientists. The case of Dutch professors in clinical cardiology [Article]. *Netherlands Heart Journal*, *17*(4), 145–154.

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.

Rabe-Hesketh, S. & Everitt, B. (2004). *A handbook of statistical analyses using Stata*. Boca Raton, FL, USA: Chapman & Hall/CRC.

Schubert, A. (2009). Using the *h*-index for assessing single publications. *Scientometrics*, *78*(3), 559–565.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628–638.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, *314*(7079), 498–502.

StataCorp. (2009). *Stata statistical software: Release 11*. College Station, TX, USA: Stata Corporation.

Thor, A. & Bornmann, L. (2011). The calculation of the single publication h index and related performances measures: A Web application based on Google Scholar data. *Online Information Review*, *35*(2), 291–300.

Thorngate, W., Dawes, R. M. & Foddy, M. (2009). *Judging merit*. New York, NY, USA: Psychology Press.

Tijssen, R. J. W., van Leeuwen, T. N. & van Raan, A. F. J. (2002). *Mapping the scientific performance of German medical research. An international comparative bibliometric study*. Schattauer: Stuttgart, Germany.

Valderas, J. M. (2007). Why do team-authored papers get cited more? *Science*, *317*(5844), 1496. doi:10.1126/science.317.5844.1496b

van Eck, N. J. & Waltman, L. (2008). Generalizing the *h*- and *g*-indices. *Journal of Informetrics*, *2*(4), 263–271.

van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, *36*(3), 397–420.

van Raan, A. F. J., Visser, M. S., van Leeuwen, T. N. & van Wijk, E. (2003). Bibliometric analysis of psychotherapy research: Performance assessment and position in the journal landscape. *Psychotherapy Research*, *13*(4), 511–528.

Vieira, E. S. & Gomes, J. A. N. F. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, *4*(1), 1–13. doi:10.1016/j.joi.2009.06.002

Webster, G. D., Jonason, P. K. & Schember, T. O. (2009). Hot topics and popular papers in evolutionary psychology: Analyses of title words and citation counts in *Evolution and Human Behavior*, 1979–2008. *Evolutionary Psychology*, *7*(3), 348–362.

Willis, D. L., Bahler, C. D., Neuberger, M. M. & Dahm, P. (2011). Predictors of citations in the urological literature [Review]. *BJU International*, *107*(12), 1876–1880. doi:10.1111/j.1464-410X.2010.10028.x