Journal of Informetrics 3 (2009) 348–352

Contents lists available at ScienceDirect

Journal of Informetrics

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

Extent of type I and type II errors in editorial decisions: A case study on *Angewandte Chemie International Edition*

Lutz Bornmann [a,*], Hans-Dieter Daniel [b,1]

[a] *ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr. 24, CH-8092 Zurich, Switzerland*
[b] *University of Zurich, Evaluation Office, Mühlegasse 21, CH-8001 Zurich, Switzerland*

ARTICLE INFO

Article history:
Received 4 March 2009
Received in revised form 30 April 2009
Accepted 6 May 2009

Keywords:
Editorial decision
Manuscript selection process
Predictive validity
Citation counts
Type I error
Type II error

ABSTRACT

Selection processes are never faultless. We investigate the predictive validity of the manuscript selection process at *Angewandte Chemie International Edition* (AC-IE), one of the prime chemistry journals worldwide, and conducted a citation analysis for manuscripts that were accepted by the journal or rejected but published elsewhere ($n = 1817$). With the bibliometric data, we were able to calculate the extent of type I and type II errors of the selection decisions. We found that the decisions regarding 15% of the manuscripts show a type I error (accepted manuscripts that did not perform as well as or worse than the average rejected manuscript). Moreover, the decisions regarding 15% of the manuscripts are affected by a type II error (rejected manuscripts that performed equal to or above the average accepted manuscript).

1. Introduction

We investigated the predictive validity of the manuscript selection process at *Angewandte Chemie International Edition* (AC-IE), one of the prime chemistry journals worldwide, and conducted a citation analysis for manuscripts that were accepted by the journal or rejected but published elsewhere. In the absence of other operationalizable criteria, a conventional approach is to use citation counts as a proxy for research quality, since they measure the international impact of the work by individuals or groups of scientists. Citation counts have been a controversial measure of both quality and scientific progress (Bornmann & Daniel, 2008d). Nevertheless, Lokker, McKibbon, McKinlay, Wilczynski, and Haynes (2008) succeeded in demonstrating for clinical articles that publications regarded shortly after their appearance as important by experts in the appropriate research field were cited much more frequently in subsequent years than publications that were less highly regarded. The Chemistry Division of the National Science Foundation (Arlington, VA) carried out a citation analysis with the goal "to explore the use of this relatively new tool for what it might tell about the discipline and its practitioners." The results of the study "generally support the idea that citations are meaningful" (Dewitt, Nicholson, & Wilson, 1980, p. 265). Furthermore, the results of a comprehensive citation content analysis conducted by Bornmann and Daniel (2008a) show that "an article with high citation counts had greater relevance for the citing author than an article with low citation counts" (p. 149).

Our analyses of the citation counts for the submissions at the AC-IE showed that the editorial decisions made by the AC-IE have high predictive validity (Bornmann & Daniel, 2008b, 2008c): on average (arithmetic mean and median), accepted manuscripts have clearly higher citation counts than rejected manuscripts that are published elsewhere. These results sug-

* Corresponding author. Tel.: +41 44 632 48 25; fax: +41 44 632 12 83.
E-mail addresses: bornmann@gess.ethz.ch (L. Bornmann), daniel@evaluation.uzh.ch (H.-D. Daniel).
[1] Tel.: +41 44 634 23 13; fax: +41 44 634 43 79.

1751-1577/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2009.05.002

gest for the AC-IE that the editorial decisions correspond on average to the manuscripts' future scientific impact. Another explanation for the impact differences might be that the higher citation counts of AC-IE publications result from the 'halo effect' (Merton, 1968; Shatz, 2004) produced by one of the prime chemistry journals worldwide. However, for Leimu and Koricheva (2005) it is a 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, see here also Seglen, 1997). Their study results "do not support this 'journal effect' hypothesis, because there was considerable variation in citation rates, especially for papers published in high-impact journals" (Leimu & Koricheva, 2005, p. 29).

Since in *every* selection process some good manuscripts are rejected and some bad manuscripts are accepted due to random error or systematic bias, it is instructive to calculate the extent of "erroneous" decisions. In type I errors, the editors concluded that a manuscript had the scientific potential for publication and was accepted, when it in fact did not, as reflected in a manuscript's low scientific impact subsequent to publication. Type I errors lead to the *overestimation* of the manuscript's future performance, i.e. the accepted manuscript will perform on the same level as or below the average of the rejected group. In type II errors, the editor concluded that a manuscript did *not* have the scientific potential for publication and was rejected, when it actually did—as reflected in a high scientific impact subsequent to publication. Type II errors lead to the *underestimation* of the manuscript's future performance, i.e. the rejected manuscript will perform on the same level as or above the average of the accepted group. This approach of calculating the extent of "erroneous" decisions in fellowship selection processes was developed by Bornmann and Daniel (2007).

In this study, the extent of "erroneous" decisions are calculated for the first time for editorial decisions using the example of AC-IE. The concept of type I and type II errors was recently discussed by Straub (2008a, 2008b) in two editorials of the journal *MIS Quarterly* to reframe the thinking around the type II problem in editorial decision-making.

2. Methods

2.1. Manuscript reviewing at AC-IE

AC-IE is one of the prime chemistry journals in the world, with a higher annual Journal Impact Factor (JIF, provided by Thomson Reuters, Philadelphia, PA) than the JIFs of comparable journals (10.031 in the 2007 Journal Citation Reports, Science Edition). AC-IE is a journal of the German Chemical Society (Gesellschaft Deutscher Chemiker (GDCh), Frankfurt am Main, Germany) and is published by Wiley-VCH (Weinheim, Germany).

A manuscript submitted to AC-IE is usually subject to 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 editors find this to be the case, the submitted manuscript is sent to several independent referees (external review), who review it using an evaluation form and a comment sheet. The referees know the authors' identities, but reviews are not signed (single blinding). The journal editors then make the decision to accept or reject a manuscript for publication on the basis of these reviews and on their own evaluations.

AC-IE introduced peer review in 1982, primarily in conjunction with one of the document types published in the journal, "Communications," which are short reports on works in progress or recently concluded experimental or theoretical investigations. What the editors of AC-IE look for most of all is excellence in chemical research. Submissions that referees deem to be of high quality are selected for publication: For most submissions, a manuscript is published only if two external referees rate the results of the study reported in the manuscript as (very) important and also recommend publication in the journal (Bornmann & Daniel, 2009).

2.2. Database for the present study and conducting of citation analysis

For the investigation of manuscript selection at AC-IE, we used information on all 1899 manuscripts that were reviewed internal and external in the year 2000. Of the 1899 manuscripts, 46% ($n = 878$) were accepted for publication in AC-IE, and 54% ($n = 1021$) were rejected. A search in the literature databases Science Citation Index (SCI, Thomson Reuters) and Chemical Abstracts (CA, Chemical Abstracts Services, CAS, Columbus, OH) revealed that of the 1021 rejected manuscripts, 959 (94%) were later published in 136 other (different) journals. As early as 1980 Abelson (1980) reported a similar finding – that almost all of the manuscripts rejected by a journal were published later in other journals – for the journal *Science*. Other studies on the fate of manuscripts rejected by a journal report percentages ranging from 28% to 85% for manuscripts later published elsewhere (Weller, 2002). For manuscripts rejected by *Angewandte Chemie* in the year 1984, Daniel (1993/2004) determined a percentage of 71% (see here also Cronin & McKenzie, 1992; Ray, Berkwits, & Davidoff, 2000).

For accepted manuscripts and manuscripts that were rejected (but published elsewhere), we determined the number of citations for a fixed time window of three years after the publication year. "Fixed citation windows are a standard method in bibliometric analysis, in order to give equal time spans for citation to articles published in different years, or at different times in the same year" (Craig, Plume, McVeigh, Pringle, & Amin, 2007, p. 243). The citation analyses for the present study were conducted in the year 2007 based on CA. CA is a comprehensive database of publicly disclosed research in chemistry and related sciences (see http://www.cas.org/).

Of all 1837 manuscripts published in the AC-IE (accepted manuscripts) or another journal (rejected manuscripts, see above), 1817 could be included in the analysis of this study. For 20 rejected manuscripts that were published elsewhere

between 2004 and 2006, a citation window of three years (one year after publication up to the end of 2006) was not available.

3. Results

The box plots in Fig. 1 show the distributions of the citation counts for the accepted manuscripts and the manuscripts that were rejected but published elsewhere by using medians, quartiles, and outliers. In the box plots, it can be clearly seen that the citation counts of both accepted and rejected manuscripts vary around the median (see the boxes and the outliers in the figure), by which the citation counts of both groups partly overlap. Among the rejected manuscripts are manuscripts that have citation counts that are substantially higher than those of accepted manuscripts, even above the median of the citation counts for accepted manuscripts. Conversely, among accepted manuscripts, we find manuscripts that have citation counts equal to or lower than median citation counts for rejected manuscripts.

The overlaps in the distributions point to two different error types in decision-making by the AC-IE (see Table 1): In type I errors, the AC-IE editors concluded that a manuscript had the scientific potential for publication (and was accepted), when it actually did not (as reflected in a manuscript's low citation counts). In type II errors, the AC-IE editors concluded that a manuscript did not have the scientific potential for publication (and was rejected), when it actually did (as reflected in high citation counts). To determine the extent of type I and type II errors in the selection process (see here Bornmann & Daniel, 2007), we categorized the editorial decision to accept manuscripts with citation counts equal to or less than the median value for rejected manuscripts as a type I error. Type II errors were defined as the rejection of manuscripts with citation counts equal to or higher than the median of accepted manuscripts (see the median citation counts in Fig. 1). Based on these definitions, we calculated the extent of type I and type II errors in the AC-IE manuscript selection process.

Seventy percent of the editorial decisions can be called correct according to our definition (see Table 2). The further percentages in the table reveal that the editors made both error types equally frequently: 15% of the manuscripts were rejected but later went on to demonstrate the same or greater scientific impact than manuscripts that were accepted, and 15% of the manuscripts were accepted but were subsequently not as successful as or on the same level as an "average" rejected manuscript. In Table 2, the proportion of type I errors within the accepted group and the proportion of type II errors within the rejected group are also reported. These proportions show that the error rates within accepted and rejected groups are approximately one-third, whereby the extent of type I errors slightly exceeds the extent of type II errors.

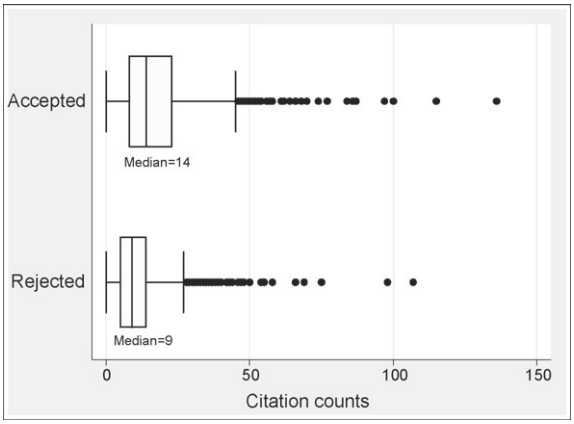

Fig. 1. Box plots for citation counts for accepted manuscripts, and rejected manuscripts that were published elsewhere ($n = 1817$). *Notes*. The vertical line in the middle of each box indicates the median, while the left and right borders of the box mark the 25th and 75th percentiles, respectively. The whiskers to the left and right of the box mark the 10th and 90th percentiles. The points beyond the whiskers are outliers beyond the 90th percentile.

Table 1
Type I and type II errors in editorial decisions.

Manuscript's scientific impact	Editorial decision	
	Acceptance	Rejection
Manuscript's scientific impact is *high*	*Correct*: citation counts are higher than the median citation counts for rejected manuscripts	*Type II error*: citation counts are equal to or higher than the median citation counts for accepted manuscripts
Manuscript's scientific impact is *low*	*Type I error*: citation counts are equal to or lower than the median citation counts for rejected manuscripts	*Correct*: citation counts are lower than the median citation counts for accepted manuscripts

Table 2
Proportion of type I and type II errors in the editorial decisions of AC-IE.

	Absolute	%
Correct decision	1275	70
Type I	278	15
Type II	264	15
Total	1817	100
Errors among acceptances		
Type I (n = 878)	278	32
Errors among rejections		
Type II (n = 939)	264	28

Table 3
Proportion of type I and type II errors in the funding decisions of EMBO and BIF.

Error type	Long-Term Fellowship program (EMBO)		Young Investigator program (EMBO)		Post-doctoral fellowship program (BIF)	
	Absolute	%	Absolute	%	Absolute	%
Correct decision	362	54	204	69	277	67
Type I	48	7	10	3	22	5
Type II	258	39	83	28	115	28
Total	668	100	297	100	414	100
Errors among approvals						
Type I	48	37 (n = 130)	10	26 (n = 39)	22	34 (n = 64)
Errors among rejections						
Type II	258	48 (n = 538)	83	32 (n = 258)	115	33 (n = 350)

In the research on funding and publication decisions to date, two other studies have been performed in which the extent of type I and type II errors was determined. Bornmann and Daniel (2007) investigated the validity of decisions for awarding long-term fellowships to post-doctoral researchers as practiced by the Boehringer Ingelheim Fonds (BIF). Bornmann, Wallon, and Ledin (2008) analyzed the Long-Term Fellowship (LTF) and the Young Investigator (YI) programs of the European Molecular Biology Organization (EMBO). The results of both studies on the extent of type I and type II errors in funding decisions are shown in Table 3. It can be clearly seen for all three programs that the type II error is considerably greater than the type I error. In this regard, the results for the three research funding programs differ from the findings for AC-IE, in which the extent of both error types in the publication decisions turned out to be equally large (see above).

4. Discussion

Selection processes are never faultless. With the bibliometric data for manuscripts accepted by the AC-IE and manuscripts that are rejected but published elsewhere, we were able to calculate the extent of over- and underestimation (type I and type II errors) of the future success of the manuscripts. We found that 15% of all manuscripts were overestimated (accepted manuscripts that did not perform as well as or worse than the average rejected manuscript), and 15% were underestimated (rejected manuscripts that performed equal to or above the average selected manuscript). This result regarding publication decisions differs from the results regarding funding decisions presented to date in that in funding decisions, the type II error turns out to be considerably greater than the type I error.

When interpreting the distributions of correct decisions, type I errors, and type II errors in different selection processes, it must be taken into account that the extent of type I and type II errors is generally dependent on the approval and rejection rates of a selection process. If the approval rate is low, only a few publication or funding decisions (acceptances or approvals) are at risk of a type I error. On the other hand, if the rejection rate is low, there is less risk of a type II error in the publication or funding decisions (rejections). Since funding rates in the 1990s were less than 20% at BIF and EMBO, it is to be expected that these funding decisions are considerably more affected by a type II error than by a type I error. Given an approval rate of about 50% (and a rejection rate of about 50%) for the AC-IE at the beginning of 2000, a result other than an equally large extent for both error types would have been surprising.

Our review of the literature revealed that, in addition to the studies on the funding decisions of BIF and EMBO, some other studies also contain information on the occurrence of errors in publication decisions. Rousseeuw (1991), for example, comment as follows: "It is commonly known and a constant course of frustration that even well-known refereed journals contain a large fraction of bad articles which are boring, repetitive, incorrect, redundant, and harmful to science in general. What is perhaps even worse, the same journals also stubbornly reject some brilliant and insightful articles (i.e., your own) for no good reason" (p. 41, see also Starbuck, 2005). For Cole (1992), "physics journals prefer to make 'Type I' errors of accepting unimportant work rather than 'Type II' errors of rejecting potentially important work. This policy often leads to

the publication of trivial articles with little or no theoretical significance, deficits which are frequently cited by referees in social science fields in rejecting articles. Other fields, such as sociology in the United States, follow a norm of rejecting an article unless it represents a significant contribution to knowledge. Sociologists prefer to make Type II errors" (p. 114). According to our study on chemistry, the extent of both error types turned out to be equally large. Straub (2008a) has pointed out that the publication of weak papers (type I error) does not actually hurt a journal as much as a type II error: In the case of a type II error "the original journal has lost the chance to publish what could be a highly cited paper" (Straub, 2008b, p. v).

Finally, we would like to stress again that although the results of our study point to type I and type II errors in the AC-IE publication decisions, accepted manuscripts, as compared to manuscripts that are rejected but published elsewhere, had on average greater impact. While it would certainly be desirable to completely eliminate both error types in decision-making, it simply cannot be done. In fact, reducing one cause for one error type (e.g., by increasing the publication rate) automatically increases the risk for the other error type.

Acknowledgments

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 also thank Dr. Werner Marx and Dr. Hermann Schier of the Central Information Service for the institutes of the Chemical Physical Technical (CPT) Section of the Max Planck Society (located at the Max Planck Institute for Solid State Research in Stuttgart, Germany) for conducting the citation search for citations of the accepted and rejected (but published elsewhere) manuscripts in the literature database Chemical Abstracts (CA).

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.

The authors wish to express their gratitude to two anonymous referees for their helpful comments.

References

Abelson, P. H. (1980). Scientific communication. *Science*, *209*(4452), 60–62.
Bornmann, L., & Daniel, H.-D. (2007). Convergent validation of peer review decisions using the *h* index: extent of and reasons for type I and type II errors. *Journal of Informetrics*, *1*(3), 204–213.
Bornmann, L., & Daniel, H.-D. (2008a). Functional use of frequently and infrequently cited articles in citing publications. A content analysis of citations to articles with low and high citation counts. *European Science Editing*, *34*(2), 35–38.
Bornmann, L., & Daniel, H.-D. (2008b). 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.
Bornmann, L., & Daniel, H.-D. (2008c). 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.
Bornmann, L., & Daniel, H.-D. (2008d). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80.
Bornmann, L., & Daniel, H.-D. (2009). The luck of the referee draw: The effect of exchanging reviews. *Learned Publishing*, *22*(2), 117–125.
Bornmann, L., Wallon, G., & Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European Molecular Biology Organization programmes. *PLoS One*, *3*(10), e3480.
Cole, S. (1992). *Making science. Between nature and society*. Cambridge, MA, USA: Harvard University Press.
Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*, *1*(3), 239–248.
Cronin, B., & McKenzie, G. (1992). The trajectory of rejection. *Journal of Documentation*, *48*(3), 310–317.
Daniel, H.-D. (1993/2004). *Guardians of science. Fairness and reliability of peer review*. Weinheim, Germany: Wiley-VCH. Published online 16 July 2004, Wiley Interscience, doi:10.1002/3527602208.
Dewitt, T. W., Nicholson, R. S., & Wilson, M. K. (1980). Science citation index and chemistry. *Scientometrics*, *2*(4), 265–275.
Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, *20*(1), 28–32.
Lokker, C., McKibbon, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: retrospective cohort study. *British Medical Journal*, *336*(7645), 655–657.
Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56–63.
Ray, J., Berkwits, M., & Davidoff, F. (2000). The fate of manuscripts rejected by a general medical journal. *American Journal of Medicine*, *109*(2), 131–135.
Rousseeuw, P. J. (1991). Why the wrong papers get published. *Chance*, *4*(1), 41–43.
Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, *314*(7079), 498–502.
Shatz, D. (2004). *Peer review: A critical inquiry*. Lanham, MD, USA: Rowman & Littlefield.
Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science*, *16*(2), 180–200.
Straub, D. W. (2008a). Thirty years of service to the IS profession: time for renewal at MISQ? *MIS Quarterly*, *32*(1), iii–viii.
Straub, D. W. (2008b). Type II reviewing errors and the search for exciting papers. *MIS Quarterly*, *32*(2), V–X.
Weller, A. C. (2002). *Editorial peer review: Its strengths and weaknesses*. Medford, NJ, USA: Information Today, Inc.