The problem of citation impact assessments for recent publication years in institutional evaluations

Lutz Bornmann*

Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany

A R T I C L E   I N F O

A B S T R A C T

Bibliometrics has become an indispensable tool in the evaluation of institutions (in the natural and life sciences). An evaluation report without bibliometric data has become a rarity. However, evaluations are often required to measure the citation impact of publications in very recent years in particular. As a citation analysis is only meaningful for publications for which a citation window of at least three years is guaranteed, very recent years cannot (should not) be included in the analysis. This study presents various options for dealing with this problem in statistical analysis. The publications from two universities from 2000 to 2011 are used as a sample dataset ($n = 2652$, univ 1 = 1484 and univ 2 = 1168). One option is to show the citation impact data (percentiles) in a graphic and to use a line for percentiles regressed on 'distant' publication years (with confidence interval) showing the trend for the 'very recent' publication years. Another way of dealing with the problem is to work with the concept of samples and populations. The third option (very related to the second) is the application of the counterfactual concept of causality.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Modern science evaluates and is also subject to evaluation. Without research assessments, it is impossible to ensure the quality of research. That is why, according to the founder of the modern sociology of science Robert K. Merton (1973), one of its norms is "organised scepticism". From the 17th century, peer review was used almost exclusively to evaluate research until the 1980s and 1990s when indicator-based evaluation and multi-stage evaluation procedures were introduced (Daniel, Mittag, & Bornmann, 2007). It is now standard for an evaluation report of an institution to include bibliometric indicators on the number of publications and the citation impact of these publications (for the natural and life sciences). Appropriate standards such as those formulated by Bornmann et al. (in press) can be used to conduct a bibliometric study.

However, institutional evaluations frequently present the problem that it is precisely the research performance over very recent years that needs to be measured as interest is focussed on these years. It is only possible to measure the citation impact of a publication reliably around three years after it has appeared. The most recent 1 to 2 publication years of an institution cannot be included in the evaluation, even if methods of field normalization are used (Wang, 2013). According to the Council of Canadian Academies (2012) "past research suggested that, for the natural sciences and engineering, an appropriate citation window is typically between three and five years . . . More recent evidence, however, has proposed that a citation window as short as two years may be appropriate in some cases . . . This evidence implies that citation-based indicators should be limited to assessing research published at least two years previously. Any attempt to use citation-based

* Tel.: +49 89 2108 1265.
 E-mail address: bornmann@gv.mpg.de

indicators for more recent research may result in spurious or misleading findings" (p. 68). This study therefore describes options for statistical procedures which allow a statement to be made about very recent publication years on the basis of those publication years which can be included in the evaluation (that is, earlier publication years). This study follows up on activities which Bornmann and Mutz (2013) initiated with their publication on the use of samples in institutional evaluations.

A number of advanced indicators are used in bibliometrics with which it is possible to measure the citation impact of publications from a research institution. They are used to show the citation impact achieved by a publication relative to the impact which other publications from the same year and in the same field have made (Rehn, Kronman, & Wadskog, 2007). Up to now, these indicators have been calculated by determining the average citation impact over the publications in a year and a field, but recently percentiles have been proposed as an important alternative (Bornmann & Mutz, 2011). A percentile is a value below which a certain proportion of publications fall: The higher the percentile for a publication, the more citations it has received compared to publications in the same field and publication year (Bornmann, Mutz, Marx, Schier, & Daniel, 2011). Although there is still some uncertainty concerning the exact method of calculating percentiles (Bornmann, in press; Bornmann, Leydesdorff, & Mutz, 2013), compared to earlier indicators they have the advantage that they do not require a (arithmetic) mean to be established. As distributions of citations are skewed to the right, the mean is not suitable as a measure of the central tendency. Percentiles are therefore used as an indicator of citation impact in this study.

2. Methods

In this study, the publications from two universities from 2000 to 2011 are used as a sample dataset. For each publication, the citation window extends from the publication to the end of 2011. There are total of 2652 publications (articles and reviews) for the universities (univ 1 = 1484, univ 2 = 1168); they published an average of 221 publications per year (univ 1 = 124, univ 2 = 97). The percentiles for the publications are researched in InCites. InCites (http://incites.thomsonreuters.com/) is a web-based research evaluation tool allowing the assessment of the productivity and citation impact of institutions. Percentiles are defined by Thomson Reuters as follows: "The percentile in which the paper ranks in its category and database year [that means, in its reference set], based on total citations received by the paper. The higher the number [of] citations, the smaller the percentile number. The maximum percentile value is 100, indicating 0 cites received. Only article types *article, note, and review* are used to determine the percentile distribution, and only those same article types receive a percentile value. If a journal is classified into more than one subject area, the percentile is based on the subject area in which the paper performs best, i.e. the lowest value" (http://incites.isiknowledge.com/common/help/h_glossary.html). InCites defines percentiles in the inverse direction than the standards in the literature (Bornmann & Marx, 2013).

In general, three steps are needed in order to calculate the percentiles for a reference set and all these steps can be differently conducted (Bornmann et al., 2013).

First, the rank-frequency function (see Egghe & Rousseau, 2006) is calculated. All publications in the set are ranked in decreasing order by their number of citations, and the number of publications in the (reference) set is determined.

Secondly, the minimum or maximum, respectively, of the percentile scale must be determined. InCites assign publications with 0 citations a percentile of 100. Furthermore, publications with a high citation impact are assigned a low percentile and publications with a low citation impact are assigned a high percentile in InCites. By assigning the value 100 to the publications with 0 citations it is ensured that the missing citation impact of publications is reflected in the percentiles in the same way in every case. Different values for publications with 0 citations would arise if percentiles are calculated without using a constant value of zero.

Thirdly, each publication is assigned a percentile based on the citation distribution (sorted in decreasing order). However, percentiles can be calculated in different ways (Cox, 2005). InCites and, for example, Rousseau (2012) calculate the quantiles – that is, the continuous variable from which percentiles can be derived by rounding – using the ranks (i) and the number of publications (n) ($i/n \times 100$). The formula ($(i - 0.5)/n \times 100$) derived by Hazen (1914) is used very frequently nowadays for the calculation of percentiles (for example by StataCorp, 2011).

The analyses for this study were performed with the statistical software Stata (StataCorp, 2011).

3. Results

Fig. 1 uses box plots to show the universities' distributions of the percentiles in each publication year. The recent publication years are also included in this figure. It is clearly visible for both universities that 2011 (on average) resulted in a significantly lower citation impact for the publications, compared to other years. Including the final year in statistical bibliometric analyses for an evaluation study or considering it in isolation would result in an erroneous representation of the performance of the two universities in terms of their citation impact. As a percentiles distribution such as that shown in Fig. 1 is not unusual, but can be seen generally in publication sets, recent years should not be included in an evaluation study and ways should be sought with which to achieve a generalising statement about the citation impact of a university based on the other years (which then relates to the recent years).