



Assessing the varying level of impact measurement accuracy as a function of the citation window length

Giovanni Abramo^{a,b,*}, Tindaro Cicero^b, Ciriaco Andrea D'Angelo^b

^a Institute for System Analysis and Computer Science (IASI-CNR), National Research Council of Italy, Italy

^b Laboratory for Studies of Research and Technology Transfer, School of Engineering, Department of Management, University of Rome "Tor Vergata", Italy

ARTICLE INFO

Article history:

Received 21 April 2011
Received in revised form 27 May 2011
Accepted 15 June 2011

Keywords:

Research evaluation
Scientific impact
Citation window
Peer review
Bibliometrics
Italy

ABSTRACT

With the passage of more time from the original date of publication, the measure of the impact of scientific works using subsequent citation counts becomes more accurate. However the measurement of individual and organizational research productivity should ideally refer to a period with closing date just prior to the evaluation exercise. Therefore it is necessary to compromise between accuracy and timeliness. This work attempts to provide an order of magnitude for the error in measurement that occurs with decreasing the time lapse between date of publication and citation count. The analysis is conducted by scientific discipline on the basis of publications indexed in the Thomson Reuters Italian National Citation Report.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The use of national exercises to evaluate research systems is becoming ever more diffuse. One of the major objectives is to support efficient allocation of public resources to the various actors in the national systems. Traditionally the assessment exercises relied on peer review approaches, but advances in bibliometric techniques have led to many governments adopting bibliometric indicators to inform or even entirely substitute peer review, at least for the hard sciences. The penetration of bibliometrics can be appreciated by examining the typologies of three assessment frameworks: the Research Excellence Framework (REF) in the UK, the Quinquennial Research Evaluation (VQR) in Italy, and the Excellence in Research for Australia initiative (ERA). For the ERA, preparations for submissions began in June 2010. For the VQR, detailed guidance on submissions and assessment criteria is expected in 2011. For the United Kingdom REF, guidelines will be published during 2011, with institutions invited to make submissions during 2013 and actual assessment taking place in 2014. The REF is a typical example of a so called "informed peer-review" exercise, where the assessment outcomes will be a product of expert review informed by citation information and other quantitative indicators. It will substitute the previous Research Assessment Exercise series which were pure peer-review. The Italian VQR, substituting the previous pure peer-review Triennial Evaluation Exercise (VTR), can be considered a hybrid: a varying mix of pure peer-review, informed peer-review and the bibliometric approach. To prepare judgments of research output quality, the panels of experts appointed in each of fourteen disciplines, can choose one or both of two methodologies for evaluating any particular output: (i) citation analysis; and/or (ii) peer-review by external experts, selected by a collegial decision of the panel. The Australian ERA assessment in the hard sciences is conducted through a pure bibliometric approach. Single research outputs are evaluated by a citation index referring to world and Australian

* Corresponding author at: Dipartimento di Ingegneria dell'Impresa, Università degli Studi di Roma "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy. Tel.: +39 06 72597362.

E-mail address: giovanni.abramo@uniroma2.it (G. Abramo).

benchmarks. Because the entire research staff of the institutions must submit their full research product, indicators of research volume are also used to evaluate overall research performance.

Studies have demonstrated that there is indeed a positive relationship between citations of a work and the opinions of experts concerning its quality (Aksnes & Taxt, 2004; Franceschet & Costantini, 2011; Oppenheim, 1997; Reale, Barbara, & Costantini, 2006; Rinia, Van Leeuwen, Van Vuren, & Van Raan, 1998), however there are numerous differences between peer review and bibliometrics, including in the limitations of the two approaches.

Peer review presents a series of well documented and much discussed limitations regarding each of its three fundamental steps: (i) the choice of products for submission to evaluation; (ii) the choice of experts entrusted with evaluation of the products; (iii) the inherent subjectivity in the judgments given by the reviewer, as offered for each product (Bornmann, 2008; Horrobin, 1990; Moxham & Anderson, 1992). However bibliometrics also has its own limitations. The most notable is the fact that it can only be applied to disciplines where publication in journals is considered a reliable proxy of research output, meaning only the hard sciences (Moed, 2005). For the hard sciences, Abramo and D'Angelo (2011) have compared the results of the Italian VTR with those from a bibliometric simulation and have shown that bibliometric approach is greatly preferable to peer review for accuracy, robustness and functionality of measurement, and for the costs and times involved. However the 2011 study by Abramo and D'Angelo did not deal with the critical concern of the time that elapses between date of publication and the date of actually counting the citations, which is necessary to obtain citation counts that can give an accurate measure of the true publication impact. In theory the peer-review approach would permit an evaluation of quality immediately on release of the publication, but with bibliometrics the citations of a work can only be a good proxy of true impact when there has been a sufficient lapse from the date of publication. A minimum "citation window" is necessary. The potential problem is that whatever the intentions for the evaluation exercise (selective funding, informing research policies and management decisions, reducing information asymmetry between suppliers of knowledge and users), it is highly desirable that the evaluation results be available in close reference to the period being evaluated. This factor could affect the applicability of bibliometric methods. We note that in spite of the questions raised, the time necessary to implement peer-review exercises is longer (two years or more for the entirety of steps) than for the mechanisms of bibliometric exercise. Also, peer-review exercises typically occur over cycles of 5–6 years (the latest RAE covered an eight year period), which is slower and less frequent than is desirable for evaluation aims. Evaluations based on bibliometric techniques can be more frequent and thus more effective in stimulating continuous improvement in the research system.

Given the concerns, it is of great interest to understand the number of years necessary before citations of a publication can be considered an accurate and robust proxy of real scientific impact, and if this window of time differs from one discipline to another. The present work intends to provide answers to these questions and to define a methodology (in terms of citation windows) for conduct of bibliometric exercises that will offer the necessary robust ratings and rankings.

There are not many works that have dealt with this question. In general, we can say that citations have increased gradually over time, as shown by the growing value of journals' impact factors; moreover, impact factors vary widely across fields (Althouse, West, Bergstrom, & Bergstrom, 2008). Glänzel, Schlemmer, and Thijs (2003), analyzing a set of works published in 1980 and indexed in the Science Citation Index (SCI), demonstrate that the probability of publications that are not cited or poorly cited over an initial period of 3–5 years should then become highly cited beyond the standard bibliometric time horizon (i.e. in a time window of 21 years after publication) is very remote and limited to rare exceptions. However, like Rousseau (1988), they note that in certain fields (e.g. mathematics-related), the standard bibliometric time horizon is greater than in others: for correct evaluation of impact of a work in mathematics the citation window should be more than three years. A subsequent study by Adams (2005) includes a conclusion that "initial citation counts" (i.e. citations received 1 and 2 years after publication) "might be useful as a forward indicator of the long-term quality of research publications". This author's findings are based on observation of publications for 1993 in the life and physical sciences, extracted from the UK National Citation Report licensed from Thomson Reuters by the UK Office of Science and Technology (OST). Considering a window from 1993 to 2002, Adams detects a strong correlation between the ranking lists for publications per number of citations in the first two years and continuing over subsequent years. Stringer, Sales-Pardo, and Nunes Amaral (2008), investigate the time scale for the full impact of papers published in a given journal to become apparent, and find that it varies from less than 1 year to 26 years, depending on the journal. Continuing from this previous literature, the intention of our current work is to study how accuracy in the measurement of publication impact varies, in each hard science discipline, in function of the length of the citation window between date of publication and citation count. All the limits of citation counts as proxy of impact, amply discussed in the literature (Glänzel, 2008; MacRoberts & MacRoberts, 1989; Moed, 2005), remain. Furthermore, we also report on the issues of the citation patterns seen in the various subject categories, the first-citation speed, and the error in evaluating a publication as having nil impact when it has not matured any citations within a given date. The study is based on publications indexed in the Thomson Reuters 2001–2008 Italian National Citation Report, extracted from Web of Science (WoS).

The following section describes the dataset used for the analysis, the elaborations, and the results concerning the accuracy of measure of impact in function of the length of the citation window. The final section provides a summary of the main findings, discusses their implications, and indicates opportunities for further consideration and examination.

Download English Version:

<https://daneshyari.com/en/article/523496>

Download Persian Version:

<https://daneshyari.com/article/523496>

[Daneshyari.com](https://daneshyari.com)