Journal of Informetrics 9 (2015) 954–961

Contents lists available at ScienceDirect

Journal of Informetrics

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

Funnel plots for visualizing uncertainty in the research performance of institutions

Giovanni Abramo[a,*], Ciriaco Andrea D'Angelo[b,1], Leonardo Grilli[c,2]

[a] Laboratory for Studies of Research and Technology Transfer at the Institute for System Analysis and Computer Science (IASI-CNR), National Research Council of Italy, Via dei Taurini 19, 00185 Roma, Italy
[b] Dipartimento di Ingegneria dell'Impresa, University of Rome "Tor Vergata", Via del Politecnico 1, 00133 Roma, Italy
[c] Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", University of Florence, Viale Morgagni, 59, 50134 Firenze, Italy

ARTICLE INFO

Article history:
Received 6 May 2015
Received in revised form 9 July 2015
Accepted 4 August 2015
Available online 19 November 2015

Keywords:
Funnel plots
Bibliometrics
Research evaluation
Universities
FSS

ABSTRACT

Research performance values are not certain. Performance indexes should therefore be accompanied by uncertainty measures, to establish whether the performance of a unit is truly outstanding and not the result of random fluctuations. In this work we focus on the evaluation of research institutions on the basis of average individual performance, where uncertainty is inversely related to the number of research staff. We utilize the funnel plot, a tool originally developed in meta-analysis, to measure and visualize the uncertainty in the performance values of research institutions. As an illustrative example, we apply the funnel plot to represent the uncertainty in the assessed research performance for Italian universities active in biochemistry.

1. Introduction

Data values almost always entail some degree of uncertainty. For example, survey values are usually provided with confidence intervals showing the likely range of the population values. Even when data are not obtained from surveys, as is the case in the evaluation of research performance, there is uncertainty due to a number of factors. For research performance, this is largely related to the assumptions and limits of the measurement instrument, and for aggregate measures, to the different sizes of the research units under consideration. Accounting for such uncertainty is crucial, to establish whether the performance of a unit is truly outstanding and not the result of random fluctuations. Indeed, the Royal Statistical Society recommends that performance reporting should always include measures of uncertainty, although in practice this is not always done (Bird et al., 2005). Indications of uncertainty are definitely not provided for the most popular yearly international university "league tables". This is true whether the rankings are produced by 'non-bibliometricians', such as the Shanghai Jiao Tong University Ranking (SJTU, 2014), QS World University Rankings (QS, 2014) and Times Higher Education World University Rankings (THE, 2014), or whether they are produced by bibliometricians themselves, such as the Scimago Institutions Ranking (Scimago, 2015). In our previous studies (Abramo, D'Angelo, & Di Costa, 2011a), the current authors are

* Corresponding author. Tel.: +39 06 771641; fax: +39 06 7716461.
E-mail addresses: giovanni.abramo@uniroma2.it (G. Abramo), dangelo@dii.uniroma2.it (C.A. D'Angelo), grilli@disia.unifi.it (L. Grilli).
[1] Tel.: +39 06 72597362; fax: +39 06 72597362.
[2] Tel.: +39 055 2751552; fax: +39 055 4223560.

http://dx.doi.org/10.1016/j.joi.2015.08.006
1751-1577/

like others in omitting the provision of the likely range of research performance values for Italian universities. The CWTS Leiden Rankings instead indicate stability intervals (Waltman et al., 2012). However, in the vast bibliometrics literature there are indeed relatively few works dealing with uncertainty in research performance measures. Colliander and Ahlgren (2011) distinguish between stability intervals and confidence intervals. To them, confidence intervals reflect uncertainty about a population parameter, whereas stability intervals reflect uncertainty about the indicators calculated for the dataset at hand. Schneider (2013) warns against the use of statistical significance tests (NHST) in research assessments. Instead, he advocates informed judgment, free of the 'NHST ritual', in decision-making processes. Bornmann, Mutz, and Daniel (2013) reformulated the 2011/2012 Leiden ranking by means of multilevel regression models, earlier introduced and applied by Bornmann, Mutz, Marx, Schier, and Daniel (2011) and Mutz and Daniel (2007). Williams and Bornmann (2014) propose guidelines for the consideration of percentile-rank classes of publications, when analyzed by citation impact. These authors, drawing on work by Cumming (2012), show how examination of the effect of sizes and confidence intervals can permit clearer understanding of citation impact differences.

In this work we introduce the funnel plot as a tool to visualize the uncertainty in the research performance values of institutions. The funnel plot was originally developed in meta-analysis (Egger, Smith, Schneider, & Minder, 1997), and to the best of our knowledge has not been applied to bibliometric rankings of research performance. A funnel plot shows the uncertainty in data values by adding confidence bands, indicating the range where the true research performance value is likely to lie. The visualization of uncertainty is useful in both analyzing the data and communicating the results. The visualizations help: (i) signpost that there is some uncertainty about the true values, and that the data values are subject to random fluctuations; (ii) identify truly outstanding units, namely units whose difference from the overall mean is statistically significant; and (iii) highlight which datasets are more reliable for decision-making, to permit for example giving more weight to datasets with smaller confidence bands (such as those based on larger samples). Finally, we provide an example of funnel plot application, utilizing it to visualize the uncertainty in the evaluation of research performance of Italian universities active in Biochemistry.

In the next section we describe the factors that are likely to cause distortion and uncertainty in research performance values. In Section 3 we illustrate the general use of the funnel plot as a tool for visualizing uncertainty. In Section 4 we present the dataset and the research performance indicator used in the analysis. In Section 5 we show the results from applying the funnel plot to our selected field of observation. Section 6 offers the conclusions.

2. Uncertainty in the assessment of research performance

According to Bougnol and Dulá (2015) all assessment processes are critically affected by 'subjective' aspects (arising mainly from assumptions about the operationalization of the performance indicator), as well as by purely technical issues (related to data handling). The literature on university rankings for educational performance (Guarino, Ridgeway, Chun, & Buddin, 2005) proposes probabilistic approaches to treat the uncertainty and assess the statistical significance of differences across universities. Turning to measures of research performance, the inherent assumptions and limits again prompt the adoption of probabilistic approaches rather than deterministic ones, in reporting and interpreting the assessment results. A very important consideration is that any policy, administrative or operational decisions based on performance assessments should consider the uncertainty levels embedded in the results.

Bibliometric measures of performance are based on countable research output/impact, and in some cases also on inputs. Although the outputs captured by bibliometric indicators have well known limitations[1], bibliometricians agree that in certain fields they are reliable proxies of overall output. Similarly, notwithstanding their limitations[2], citations are considered a reliable proxy of the impact of scientific research, as long as there is a sufficient time lapse from the publication date to the observation of citations (Abramo, Cicero, & D'Angelo, 2011b).

To begin, we first examine the output/impact-side factors that introduce uncertainty in bibliometric indicators. We expect the factors of uncertainty to affect the indicators randomly, meaning they will generate fluctuations without systematic effects in favor of or against particular groups of researchers[3]. We argue that the main sources of uncertainty are the following:

(1) Variability in the intensity of production due to personal events (e.g. the researcher has occasional periods of additional teaching or managerial duties, family problems, etc.);
(2) Variability in the intensity of production due to patterns in research projects (e.g. the researcher has variable access to funding, or is cyclically unproductive due to engaging in the early stages of long-term projects or frontier projects);

[1] The limitations of bibliometric indicators include the following: not all new knowledge produced can be codified into countable documents, and a part of it remaining intangible; not all documents are indexed in the bibliometric databases typically used in assessment exercises, such as WoS or Scopus.

[2] Measuring the impact of a publication through citations suffers from problems such as negative citations and 'network' citations (Glänzel, 2008).

[3] On the contrary, bias factors generate fluctuations with systematic effects. For example, in comparing professors belonging to different fields, gross aggregations may favor those belonging to a particular field. The analyst should make efforts to eliminate the effects of such bias factors, for example by making comparisons within homogeneous fields (Abramo, D'Angelo, & Di Costa, 2008) or by field-normalizing citations. Bias can be reduced, even if no method can accomplish the perfect fine-grained classification of scientist or the perfect overlapping of citation distributions when assessing a large number of fields (Zhang, Cheng, & Liu, 2014Abramo, Cicero, & D'Angelo, 2012aRadicchi, Fortunato, & Castellano, 2008).