Journal of Informetrics 7 (2013) 89–90

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

Letter to the Editor

The advantage of the use of samples in evaluative bibliometric studies

Dear Sir,

In the natural sciences researchers publish their results mainly in international journals. As a rule these papers can be researched in literature databases appropriate for bibliometric analyses (especially Web of Science, Thomson Reuters, and Scopus, Elsevier). If it is the goal of an evaluative bibliometric study to compare several research institutions with a focus on natural sciences, the papers of the institutions included in the study have at least one author with the institution as his/her affiliation.

On the one hand, it is possible to include *all* the papers from the research institutions retrieved from the databases in the study. Using all the papers, that is the full survey, would have the advantage of including all the bibliometric information for an institution. The disadvantages are (1) that a full survey is associated with a high outlay. The larger the number of publications, the more expensive, as a rule, the purchase of advanced bibliometric indicators for the individual publications will be. (2) Furthermore, a full survey for an institution is generally speaking not possible as for the very recent publication years (the last two years) the citation windows are too small to allow a reliable statement about the citation impact of the publications. (3) Finally, the risk of errors increases with the quantity of bibliometric data, particularly when the data is obtained from more distant publication years (Marx, 2011).

Usually, a population, defined as the whole bibliometric data for an institution, is split up into natural, non-overlapping groups such as different publication years, journals, or authors. Such groups provide for clusters in a two-stage sampling design ("cluster sampling"), in which, firstly, one single cluster is randomly selected from a set of clusters (Levy & Lemeshow, 1999). For example, for an evaluation study, the clusters would consist of three consecutive publication years (e.g. cluster 1: 1990–1992, cluster 2: 1993–1995, etc.). Secondly, all the bibliometric data (publications and corresponding metrics) is gathered (census) for the selected cluster (e.g., cluster 2). The clusters should not differ significantly and should exhibit great heterogeneity in the metrics within the clusters. Cluster sampling is particularly efficient if each single cluster represents approximately the whole population. If bibliometric information is available for the whole institution, it is possible to test whether the selected cluster differs from the population in some important properties (e.g. distribution of publications over the journals). Weights can additionally help to adjust the sample for under or overrepresentation of certain properties in the sample in comparison with the population. Unfortunately, the practical advantages of the simple sampling procedure are at the expense of a lack of variance reduction as is the case in other sampling designs. For instance, in stratified sampling with homogeneous strata the resulting overall standard errors are smaller than in the case of a simple random sample from the whole population. Therefore, the standard errors for cluster-sampling are not smaller than in the case of drawing a random sample from the population.

With the cluster sample (of publication years) it should be ensured that a citation window of at least three years is possible for the papers in these years. Publication years make very suitable clusters because it is easier to create a subset of papers for an institution with them than with other possible clusters (such as authors) and no bias is expected in the data. Using authors (scientists) as the clusters would result in the following two problems, for example: (1) Because of homonyms it is a difficult task to compile the papers for a single scientist unambiguously. This problem does not arise with publication years. (2) As the differences in the scientific performance between the scientists are very large, the selection could result in a sample which cannot represent the population adequately. With publication years, we can assume that the performance between the years differs a little but within the years differs significantly.

1751-1577/$ – see front matter
http://dx.doi.org/10.1016/j.joi.2012.08.002

With statistical tests it is possible, based on a cluster sample, to verify the statistical significance of results (such as performance differences between two universities). If a statistical test which looks at the difference between two research institutions with regard to their performance turns out to be statistically significant (given a specific null hypothesis), it can be assumed that the difference has not arisen by chance, but can be interpreted beyond the data at hand (the results can be related to the population). This is a great advantage in evaluation studies, since as a rule the client of these studies wishes very recent publication years (e.g., the last three years). However, citation metrics for the last three years are as a rule unreliable because of the small citation window (Abramo, Cicero, & D'Angelo, 2011). If a (cluster) sample is used, the results from the selected earlier publication years, which enable reliable performance estimates, can be generalized to all the (very recent) years.

Yours sincerely,

References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2011). Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics*, *5*(4), 659–667. http://dx.doi.org/10.1016/j.joi.2011.06.004

Levy, P., & Lemeshow, S. (1999). *Sampling of populations – Methods and applications*. New York, NY, USA: Wiley.

Marx, W. (2011). Special features of historical papers from the viewpoint of bibliometrics. *Journal of the American Society for Information Science and Technology*, *62*(3), 433–439. http://dx.doi.org/10.1002/asi.21479

Lutz Bornmann*
Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstraße 8, 80539 Munich, Germany

Rüdiger Mutz
Professorship for Social Psychology and Research on Higher Education, ETH Zurich, Zurich, Switzerland

* Corresponding author.
E-mail address: bornmann@gv.mpg.de (L. Bornmann)

25 July 2012