Selecting Researchers with a Not Very Long Career - The Role of Bibliometrics

Elizabeth S. Vieira[1] and José A. N.F. Gomes[2]

[1] *elizabeth.vieira@fc.up.pt,* [2] *jfgomes@fc.up.pt*
REQUIMTE/Departament of Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal

Introduction

The scientific community has developed many institutionalized forms of evaluation where peer review has an important role, but recently, bibliometric methods have been gaining some acceptability to assess the scientific performance.

The two techniques have been related to one another in different ways: 1) bibliometric methods have been used to analyze the peer review processes (Moed, 2005, chapters 19 and 20); 2) the peer review process uses bibliometric parameters as an auxiliary instrument (Moed, 2005 chapter 18, p. 233-234); and 3) peer reviewers are called in to validate and correct the results of some bibliometric process (e.g. Norris & Oppenheim, 2003; Rinia, van Leeuwen, van Vuren, & van Raan, 1998). There are some national scientific systems that use bibliometric techniques or a mix of bibliometric techniques and peer review to decide the allocation of funding (e.g. Excellence in Research for Australia (ERA); *Valutazione della Qualità della Ricerca* (VQR)). Taking into account the advantages and limitations of bibliometric techniques and the intensive use, recently, there is a growing interest in its potential in helping peers to prepare the final decisions and therefore several studies have been made on the subject (e.g. Vieira, Cabral, & Gomes, 2014a, 2014b, Bornmann & Leydesdorff, 2013). In this study, we exploit the usability of bibliometrics as support tool this time in selecting candidates that had been awarded their PhD's more than 6 and less than 12 years ago and had worked as independent researchers for less than 6 years. We deem this study important as: (1) there is a growing use of bibliometric indicators and it is important to know their caveats and strong points at the different levels; and (2) the use of bibliometric indicators is more controversial when applied to individual researchers, especially at initial steps of their careers.

Methodology

This study considers the applicants to the development grants of the opening *Investigador FCT* carried out in Portugal since 2012. The publications indexed in the Web of Science Core Collection of the 120 applicants from the Engineering and Technology (28), Natural Sciences (23), Exact Sciences (48) and Medical and Health Sciences (21) were used to calculate a set of bibliometric indicators that are intended to describe the scientific performance. Bibliometric techniques are not used in a formal way in the opening. However, we are looking for indicators that may be implicit in peer judgments. A set of 17 indicators was determined: *TD* (number of documents); *TDC* (number of cited documents); *NDF* (number of documents after fractionation by the total number of authors); *PA* (% of articles); *PP* (% of proceedings papers); *PR* (% of reviews); *PAP* (% of documents as articles and proceedings papers simultaneously); *PDAC* (% of documents as corresponding author); *h* index, h_{nf} index (Vieira & Gomes, 2011); $SNIP_m$ (median of all the *SNIP*s of the journals where the applicant has published, Moed, 2010); SJR_m (median value as in the $SNIP_m$, Gonzalez-Pereira, Guerrero-Bote, & Moya-Anegon, 2010); *PTDIF* (% of documents published in journals with Impact Factor- *IF*); *PQ1* (% of documents published in journals in the first quartile in its scientific domain, according to the *IF*); *HCD* (% of documents highly cited in the top 10%); *NI* (average number of citations per document after normalization); *DIC* (% of documents with international collaboration). There is a huge number of bibliometric indicators and we tried to select those that describe the several dimensions of the scientific production. Nevertheless other indicators could be used.

Using as dependent variable the decision of the peers panel (selected-1; not selected–0) and the bibliometric indicators as independent variables we applied binary logistic regression aimed at determining those indicators that can be used to predict the final decisions made by the peers.

Results

The model

The application of the binary logistic regression lead to the following model:

$$P_i = \frac{e^{-1.88+1.116 SJR_m+0.064 HCD}}{1+e^{-1.88+1.116 SJR_m+0.064 HCD}}$$

where P_i is the probability of the applicant i to be selected by the peers for funding. The SJR_m and the HCD are the indicators that were found to be able to represent the decisions made by the peers panel. The sensitivity determined for this model was 73.2%, the percentage of false positives obtained was 35% and 70% of the cases are predicted correctly by the model. The probability of the forecasted probability by the model for a selected applicant to be higher than that of a non-selected one is 75.3% (ROC curve).

Forecasts

The predictions given by the model are useful in preparing the decisions to be taken by the peers, but the use can be increased if complemented with some type of uncertainty measure. Here, this is shown using the margins concept. Margins are being used in bibliometrics at the individual level for the first time as far as we know.

In Figure 1 is shown the probability of a given applicant to be selected for funding as we increase the value of the HCD and SJR_m, respectively, and maintaining the average value of the other variable. For each predicted value is also shown the confidence interval at 95%, working as the uncertainty measure. All this information can be used by the peers to improve the decision making process.
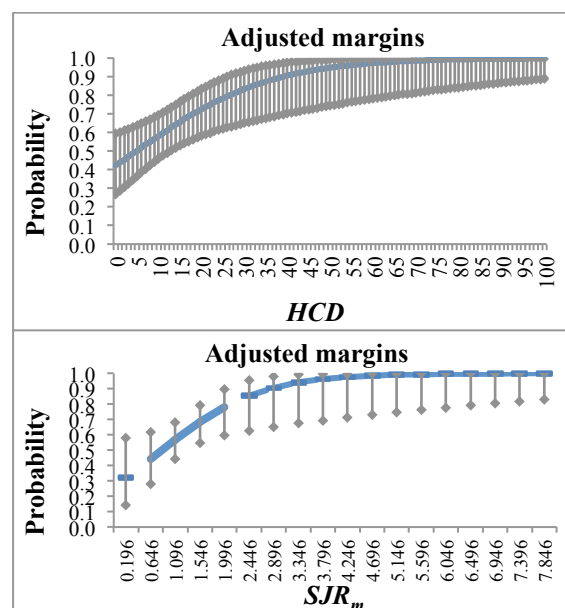

Figure 1. Predicted probabilities complemented with confidence intervals (95%). The dashed zone represents values with a few observations.

Conclusions

From this study some findings can be drawn:

✓ The bibliometric indicators are useful in describing the performance of applicants with PhD's earned 6 to 12 years ago.

✓ A composite indicator (*HCD* and *SJRm*) when used by the peers will have a positive impact on the final decision.
✓ Bibliometric indicators can be used, for example, as input tool helping peers panel in their decision making process as the indicators can give consistent and objective information.
✓ The *HCD* is a serious candidate as tool in support decisions of peer evaluations as it was also found to be useful in describing the final decisions in other types of openings (Vieira et al., 2014a, 2014b).

References

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from incites and f1000. *Journal of Informetrics, 7*(2), 286-291.

Gonzalez-Pereira, B., Guerrero-Bote, V.P., & Moya-Anegon, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics, 4*(3), 379-391.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. The Netherlands: Springer

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265-277.

Norris, M., & Oppenheim, C. (2003). Citation counts and the research assessment exercise V - archaeology and the 2001 RAE. *Journal of Documentation, 59*(6), 709-730.

Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria - evaluation of condensed matter physics in the Netherlands. *Research Policy, 27*(1), 95-107.

Vieira, E. S., Cabral, J. A. S., & Gomes, J.A.N.F. (2014a). Definition of a model based on bibliometric indicators for assessing applicants to academic positions. *Journal of the Association for Information Science and Technology, 65*(3), 560-577.

Vieira, E. S., Cabral, J. A. S., & Gomes, J.A.N.F. (2014b). How good is a model based on bibliometric indicators in predicting the final decisions made by peers? *Journal of Informetrics, 8*(2), 390-405.

Vieira, E. S., & Gomes, J.A.N.F. (2011). An impact indicator for researchers. *Scientometrics, 89*(2), 607-629.