Response to 'How good is research really?'

Bornmann and Marx [1] analyse the use of bibliometric parameters to determine "how good the research is", advising against the use of certain scientometric indicators, such as the impact factor and the *h*-index, and proposing percentiles as an alternative. However, they do so by using debatable arguments and ignoring validation, which is the basis for any meaningful indicator. The need to validate scientometric predictors is based on general principles [2] and is especially necessary in the case of science owing to the manner in which it advances: a large volume of "normal" research supports the small fraction of "revolutionary" research that pushes scientific progress [3]. Thus, Bornmann and Marx's "good research" could refer to either "normal" or "revolutionary" research, or both [4]. If this uncertainty is not resolved, any further discussion about a reliable indicator of good research is futile.

The reliability of research indicators can be assessed in the light of one of their major applications: "to measure the effectiveness of research expenditures" [5]. To compare two institutions, we should divide the values for their research indicator by their corresponding research expenditures, and the one with the higher ratio can be said to be more cost-effective. However, for this calculation to be meaningful, the indicator must be sensitive to research that is relevant for the progress of science and that provides benefits for society, and insensitive to the large volume of "normal" results that are important for researchers but not for society at large [4]. One possibility to assess the results of a comparison is to assume that institutions obtaining Nobel Prizes are more successful in producing research that is relevant to society. Surprisingly, when applying this simple test to most conventional research indicators, the finding is that the most competitive elite universities show a low effectiveness in their use of research expenditures [4], which suggests that in these indicators the weight of "normal" research is high.

Bornmann and Marx [1] propose the proportion of the top 10% most cited papers as research indicators. Because validation is not reported, it is unknown why the top 10% is the correct percentile and not, for example, the top 5% or 20%. Yet, the reliability of this indicator can be tested by comparing universities, which is possible because it is used in the Leiden (PP_{top10%} indicator) and SCIMAGO (Exc indicator) rankings [6]. Massachusetts Institute of Technology (MIT), USA, and Complutense University of Madrid, Spain are good examples because there are previous comparisons [4]. The values for these universities are: PP_{top10%} 25.2% and 8.1% (<http://www.leidenranking.com/ranking.aspx>), and Exc 29.4% and 12.4% (<http://www.scimagoir.com/>), respectively, which gives MIT and Complutense indicator ratios of 3.1 and 2.4, respectively. Introducing the number of papers, the ratios of the transformed indicators increase to 7.1 and 5.4, respectively. Taking into account that research expenditure is ten times higher at MIT, the conclusion is that Complutense University is more effective than MIT in the use of research funding, which is not reasonable. In fact, it has been proposed that ratios for reliable research indicators should be at least 50 [4], and much higher ratios have been reported for indicators validated in terms of Nobel Prizes [7]. Similar comparisons can be made with other universities.

The research indicators proposed by Bornmann and Marx [1] join a large number of published, non-validated bibliometric indicators of research performance. Their expansion and prevalence is probably a

cause for the extended feeling among scientists that there is a lack of reliable research indicators, a feeling that is summarized in a 2011 report by the Royal Society, UK: "Better indicators are required in order to properly evaluate global science" [8]. A question recently raised by Francis Narin—an expert on science and technology analysis—relates to the same problem: "for both old and new indicators, basic validity and relevance issues remain, such as by what standard can we validate our results, and what external use can appropriately be made of them?" [9].

CONFLICT OF INTEREST

The author declares that he has no conflict of interest.

REFERENCES

1. Bornmann L, Marx W (2013) *EMBO Rep* **14**: 226–230
2. Harnad S (2009) *Scientometrics* **79**: 147–156
3. Kuhn TS (1962) *The Structure of Scientific Revolutions*. University of Chicago Press
4. Rodríguez-Navarro A (2012) *PLoS ONE* **7**: e47210
5. Garfield E, Welljams-Dorof A (1992) *Sci Public Policy* **19**: 321–327
6. Leydesdorff L, Bornmann L (2012) *Scientometrics* **92**: 781–783
7. Rodríguez-Navarro A (2011) *PLoS ONE* **6**: e20510
8. The Royal Society (2011) *Knowledge, Networks and Nations*. The Royal Society
9. Narin F (2012) *Scientometrics* **92**: 391–393

**Alonso Rodríguez-Navarro is at the Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Spain.
E-mail: alonso.rodriguez@upm.es**

EMBO reports (2013) **14**, 493; published online 14 May 2013; doi:10.1038/embor.2013.60

Comments to the response of Rodríguez-Navarro

In his critical response [1] to our article 'How good is research really?', Alonso Rodríguez-Navarro lists several points that we address here. First, he criticizes that we use debatable arguments for

our preference for percentiles and that we ignore the validity of the approach. Unfortunately, he does not specify which of the arguments he finds debatable. We believe that there are many good reasons

in favour of percentile-based indicators: they are normalized for subject area and time period, they are independent of the skewed distribution of citations and they offer the option of focusing on specific percentile rank classes. Rodríguez-Navarro himself proposes the x-index, a “percentile-based index of the high-citation tail” [2]. He must, therefore, be persuaded by the percentile approach.

Lutz Bornmann has already undertaken several studies into the validity of bibliometric indicators, which are in line with the recommendation from Harnad [3]: “Scientometric predictors of research performance need to be validated by showing that they have a high correlation with the external criterion they are trying to predict.” For example, in references [4,5] the *h*-index is validated with the assessments of experts as an external criterion for quality. Bornmann and Leydesdorff use data from F1000 to validate the percentiles of citation counts externally and show that the correlation of percentiles with F1000 scores is higher than with any other indicator [6].

Rodríguez-Navarro proposes creating quotients from a research indicator and the corresponding research expenditures. The question with these indicators is, of course, can we assume that their validity has already been tested on a broad database? We doubt that. Bibliometricians face great problems in considering research efforts in their studies—for example, by the number of researchers and scope of research resources—particularly when institutions are being compared. These data are usually difficult to obtain, frequently incomplete and always inconsistent. As an external criterion for reviewing the validity of bibliometric indicators, Rodríguez-Navarro proposes using the number of Nobel Prizes. This indicator is also problematic, as the Noble Prize is too infrequent an event; it is often not clear to which institution a Nobel Prize should be awarded; and the prizes are awarded many years after the researcher has delivered the performance in question.

Second, Rodríguez-Navarro criticizes our choice of the proportion and number of top 10% most cited papers. Generally speaking, we have found that when comparing institutions, we get similar results with the top 5% or the top 20%. We used the top 10% most-cited papers because these are equated with highly cited publications [7–10]. Furthermore, this indicator is

already used for the Leiden Ranking [11] and the SCImago Institutions Ranking [12].

Third, Rodríguez-Navarro states that a bibliometric indicator should be able to distinguish between normal and revolutionary science. But what is normal and revolutionary science? Kuhn [13] used examples from history to explain the difference between the two, but fails to shed any light on how it could be determined in modern science. We have already addressed Kuhn’s concept in bibliometric terms. Our results, using the example of the big bang theory [14] and plate tectonics [15], show that regarding the development of research as two phases does not reflect its complexity appropriately. We have therefore proposed the enhancement of Kuhn’s paradigm concept with the Anna Karenina principle [16]: a scientific revolution can only be expected when several key conditions have been fulfilled. For example, solid evidence to answer basic questions must be presented and taken up by colleagues, and must be amenable to verification by means of independent data and methods.

Fourth, Rodríguez-Navarro incorrectly mixes societal and scientific impact measurements of research. The indicators proposed by us measure scientific impact—no more and no less. Indicators of research “that is relevant to society”, have not yet been developed [17,18]. At present, the case study approach, which is very complex, is seen as the best method to measure societal impact. We developed a more simple approach where scientists write assessment reports (such as the IPCC) that explain how their results could be used and/or applied in society [19].

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Rodríguez-Navarro A (2013) *EMBO Rep* **14**: [Epub ahead of print] doi:10.1038/embor.2013.60
2. Rodríguez-Navarro A (2011) *PLoS ONE* **6**: e20510
3. Harnad S (2009) *Scientometrics* **79**: 147–156
4. Bornmann L, Daniel H-D (2005) *Scientometrics* **65**: 391–392
5. Bornmann L, Wallon G, Ledin A (2008) *Res Eval* **17**: 149–156
6. Bornmann L, Leydesdorff L (2013) *J Informetr* **7**: 286–291
7. Bornmann L, de Moya Anegón F, Leydesdorff L (2012) *J Informetr* **6**: 333–335
8. Sahel JA (2011) *Sci Transl Med* **3**: 84cm13
9. Tijssen R, van Leeuwen T (2006) Centres of research ‘excellence’ and science indicators. In *Ninth International Conference on Science and Technology Indicators* (ed. Glänzel W), pp 146–147. Katholieke Universiteit Leuven
10. Tijssen R, Visser M, van Leeuwen T (2002) *Scientometrics* **54**: 381–397
11. Waltman L et al (2012) *J Am Soc Inf Sci Technol* **63**: 2419–2432
12. SCImago Research Group (2012) *SIR World Report 2012*. University of Granada
13. Kuhn TS (1962) *The Structure of Scientific Revolutions*. University of Chicago Press
14. Marx W, Bornmann L (2010) *Scientometrics* **84**: 441–464
15. Marx W, Bornmann L (2013) *Scientometrics* **94**: 595–614
16. Bornmann L, Marx W (2012) *J Am Soc Inf Sci Technol* **63**: 2037–2051
17. Bornmann L (2012) *EMBO Rep* **13**: 673–676
18. Bornmann L (2013) *J Am Soc Inf Sci Technol* **64**: 217–233
19. Bornmann L, Marx W (2013) *Scientometrics* [Epub ahead of print] doi:10.1007/s11192-013-1020-x

Lutz Bornmann is at the Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society in Munich, Germany.

Werner Marx is at the Max Planck Institute for Solid State Research, Stuttgart, Germany. E-mail: w.marx@fkf.mpg.de

EMBO reports (2013) **14**, 493–494; published online 14 May 2013; doi:10.1038/embor.2013.63

Response: ‘Antibody crossreactivity between the tumour suppressor PHLPP1 and the proto-oncogene β -catenin’

Lobert and co-workers [1] reported a consequential flaw in the immunoreactivity of the antibodies A300-660A and IHC-00382 to PH domain leucine-rich repeat protein phosphatase 1 (PHLPP1) manufactured by Bethyl Laboratories, Inc.

(Montgomery, Texas, USA). The flaw is the crossreactivity of these preparations with β -catenin. This response is to acknowledge the validity of the report of Lobert and co-workers, to report that testing at Bethyl Laboratories has confirmed their findings