What Proportion of Excellent Papers Makes an Institution One of the Best Worldwide? Specifying Thresholds for the Interpretation of the Results of the SCImago Institutions Ranking and the Leiden Ranking

Lutz Bornmann
Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany. E-mail: bornmann@gv.mpg.de

Felix de Moya Anegón
SCImago Group, Instituto de Políticas y Bienes Públicos – Consejo Superior de Investigaciones Científicas, C/Albasanz, 26–28, Madrid 28037, Spain. E-mail: felix.moya@scimago.es

University rankings generally present users with the problem of placing the results given for an institution in context. Only a comparison with the performance of all other institutions makes it possible to say exactly where an institution stands. In order to interpret the results of the SCImago Institutions Ranking (based on Scopus data) and the Leiden Ranking (based on Web of Science data), in this study we offer thresholds with which it is possible to assess whether an institution belongs to the top 1%, top 5%, top 10%, top 25%, or top 50% of institutions in the world. The thresholds are based on the excellence rate or $PP_{top 10\%}$. Both indicators measure the proportion of an institution's publications which belong to the 10% most frequently cited publications and are the most important indicators for measuring institutional impact. For example, while an institution must achieve a value of 24.63% in the Leiden Ranking 2013 to be considered one of the top 1% of institutions worldwide, the SCImago Institutions Ranking requires 30.2%.

Introduction

Until today it has been customary in evaluating bibliometrics to use the arithmetic mean value to normalize citation data (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011). According to the results from Albarrán, Crespo, Ortuño, and Ruiz-Castillo (2011), the distribution of citations in every subject category is very skewed, however: "The mean is 20 points above the median, while 9-10% of all articles in the upper tail account for about 44% of all citations" (p. 385). The skewed distribution poses the risk that the citation statistics are dominated by a few highly cited papers (Boyack, 2004; Waltman et al., 2012a). This is not possible with statistics based on percentiles. Using percentiles to normalize citations can give better comparisons of the impact of publications from different subject areas and years of publication than normalization using the arithmetic mean. The percentile provides information about the citation impact the publication (e.g., of an institution) has had compared to other publications in a reference set (Bornmann, 2013, in press; Bornmann, Leydesdorff, & Mutz, 2013; Bornmann & Marx, 2013). A percentile is a value below which a certain proportion of observations fall: The higher the percentile for a publication, the more citations it has received compared to publications in the reference set—the publications in the same field and publication year.

This study is based on the percentile indicators $PP_{top10\%}$ (Waltman et al., 2012b) and the excellence rate (Bornmann, de Moya Anegón, & Leydesdorff, 2012). Both indicators measure the same value (in this study we use the expression excellence rate only): it is the proportion of an institution's publications which belong to the top 10% most frequently cited publications; a publication belongs to the top 10% most frequently cited if it is cited more frequently than 90% of the publications published in the corresponding reference set, that is, in the same field and in the same year. The excellence rate is seen as the most important indicator in the Leiden Ranking produced by the Centre for Science and Technology Studies (CWTS, Leiden University, The Netherlands) (Waltman et al., 2012b).

The advantage of using the excellence rate to measure the citation impact of institutions is that the results for the

institutions can be compared to an expected value of 10% (Bornmann, Mutz, Marx, Schier, & Daniel, 2011). If the percentiles for the impact of publications in a database were to be selected at random, it might be expected that 10% of the publications would be among the 10% most cited publications. Thus, it is possible to test for every institution whether it is above or below the expected value. However, only a comparison with the performance of all other institutions in a ranking makes it possible to determine how well an institution is really performing. To do this it is necessary to know, for example, which excellence rate the best institution has achieved. A rate of 100% is theoretically possible, but not achievable in practical terms. For the comparison of institutions, we report in this study on the excellence rate thresholds which will allow an institution to be assessed in two different rankings: It is possible to assess whether an institution belongs to the top 1%, top 5%, top 10%, top 25%, or top 50% of institutions in the world.

With these thresholds we introduce a second level of normalization of citation impact for evaluating institutions: Percentiles are used to normalize each individual publication against the reference set of similar publications (see above) and with the aid of the thresholds presented here it is possible to evaluate every institution against the reference set of all other institutions in the ranking.

Materials and Methods

This study uses two data sets on which the SCImago Institutions Ranking (SIR; www.scimagoir.com/) and the Leiden Ranking (LR) 2011/2012 as well as 2013 (http://www.leidenranking.com/) are based. While the LR only refers to universities, the SIR also includes nonuniversity research institutions.

SCImago publishes an annual *World Report* in which a set of bibliometric indicators (e.g., number of publications, excellence rate, proportion of international collaboration) for institutions worldwide is presented after a very conscientious process of disambiguation is applied to the names of the institutions. The publication and citation data for calculating the indicators is derived from the Scopus (Elsevier) database. We base this study on the excellence rate as the most important indicator for the comparison of institutions. It is derived from all the publications of an institution between 2003 and 2011. The citation window extends from publication to the end of 2011. In order to include only those institutions in the study which we can assume to have been active in scientific research over the period we are looking at, we have taken only the institutions from the SIR with at least 1,000 publications into account.

The LR 2011/2012 was published early in 2012 (http://www.leidenranking.com/). It measures the scientific performance of 500 major universities worldwide and is based on publications in the Thomson Reuters Web of Science database for the period 2005 to 2009. The LR 2013 was published in April 2013, includes 500 major universities, and is based on publications from 2008 to 2011. Similar to the SIR,

TABLE 1. Mean, median, minimum, and maximum values on the number of publications.

Ranking	N	Mean	Median	Minimum	Maximum
SIR	2,958	7,317.33	3,131.5	1,001	333,772
LR 2011/2012	500	8,996.7	6,948.22	3,229.5	61,622.53
LR 2013	500	8,616.56	6,715	3,283	58,543

Note. N is the number of institutions.

the LR uses a set of bibliometric indicators (number of publications, mean citation score, mean normalized citation score, and excellence rate) to rank the universities (Waltman et al., 2012b). The LR is in fact a set of many rankings which vary according to the selection criteria decided by the end user of the CWTS website. The indicator values for the individual universities can be downloaded as an Excel file. The LR figures on which we based this study (the excellence rates) relate to the data set in which each institution is assigned a paper if the institution is cited in the paper. Alternatively, the LR offers fractional counting. Furthermore, publications in special types of journals were included in the calculation of the excellence rates (the LR also offers the option of excluding these publications). A journal is considered special if it meets at least one of the following two conditions:

- The journal does not publish in English or it does publish in English but authors are concentrated in one or a few countries, indicating that the journal does not have a strong international scope.
- The journal has only a small number of references to other journals in the Web of Science database, indicating that in terms of citation traffic the journal is only weakly connected to these other journals. This is the case for many journals in the humanities, but also for trade journals and popular magazines (http://www.leidenranking.com/methodology/indicators).

Results

Table 1 contains several characteristic values for the number of publications for the institutions in the ranking data sets. As the table shows, at $n = 2,958$ there are significantly more institutions for the SIR than for the LR ($n = 500$). The characteristic values for the publications also vary commensurately, such as the median number of publications. In order to test the correlation between the number of publications and the excellence rate for the institutions, we calculated a Spearman rank correlation for each data set (Bornmann & de Moya Anegón, 2011). A number of studies have already established a positive correlation between output and impact (Abramo, D'Angelo, & Costa, 2010; Hemlin, 1996). As the results show, in the SIR $r_s = .22$, in the LR 2011/2012 $r_s = .45$, and in the LR 2013 $r_s = .42$. Therefore, in line with the results of other studies, there is a positive correlation in both rankings between the indicators (a higher number of publications is therefore

TABLE 2. Thresholds for the excellence rates in SIR and LR to belong to certain top classes.

Percentile rank class	Ranking		
	SIR	LR 2011/2012	LR 2013
Maximum	49.53	26.44	27.23
Top 1%	30.20	22.87	24.63
Top 5%	23.86	19.73	20.50
Top 10%	21.21	17.92	18.21
Top 25%	16.77	14.85	15.59
Top 50%	12.07	12.42	13.11
Minimum	0.62	3.18	2.47

accompanied by a higher excellence rate). However, if we apply Cohen's (1988) instructions on interpretation, the correlation in the SIR should be considered small/medium and large in the LR.

Table 2 shows thresholds that will put an excellence rating of institutions in a specific impact class. These classes are geared towards an evaluation scheme in the Science & Engineering Indicators of the National Science Foundation of the USA (National Science Board, 2012) prepared biannually by the American corporation ipIQ. In this scheme the focus is on frequently cited papers (Leydesdorff, Bornmann, Mutz, & Opthof, 2011): The table shows the relevant thresholds for the excellence rate to qualify for the top 1%, top 5%, top 10%, top 25%, and top 50% of institutions throughout the world. While for example an institution in the LR must achieve an excellence rating of 22.87% (2011/2012) or 24.63% (2013), respectively, to belong to the top 1% of institutions worldwide, the SIR requires 30.2%. The table also shows the minimum and maximum for the excellence rates in each ranking.

As the results in the table show, institutions in both rankings require an excellence rate of around 12% to perform better than the average for all the institutions. This value is slightly higher than the expected value of 10%. This deviation is undoubtedly due to a subset of institutions with a considerable number of publications being selected for both rankings. As there is a positive correlation between the excellence rate and the number of publications (see above), a higher excellence rate is to be expected from the subset selected here than from all (other) institutions.

All the other thresholds in Table 2 (in addition to the top 50%) are more or less different in SIR and LR (both editions) and have the following pattern: The higher the excellence rate, the more difference there is between the thresholds. While the difference in the top 25% is only around 1 percentage point, it is around 6 percentage points for the top 1%. The differences can be explained, on the one hand, by the use of different databases with Scopus and the Web of Science. Furthermore, the number of institutions included is very different. When more institutions are taken into account in a ranking, the variation between the institutions can generally be expected to be higher. On the other

hand, the results reflect the differences in the calculation of the excellence rate. These differences are related mainly to the problem of the unambiguous assignment of publications to the class of excellent publications (Bornmann et al., 2013). If it is not possible to determine the 10% most cited publications unambiguously in a reference set, because the set is characterized in this range by many publications with the same citation counts (which therefore cannot be clearly ranked), there are various options for solving the problem.

SCImago introduces a further publication attribute (in addition to the subject area and the publication year) into the percentile calculation: Where citation counts are equal, the SJR2 (Guerrero-Bote & de Moya-Anegon, 2012) of the journal which has published the publications is used as a second sort key (from highest to lowest). This journal metric takes into account not only the prestige of the citing scientific publication but also its closeness to the cited journal. In the LR, a fractional counting approach is used to assign publications unambiguously to the category of excellent publications (Waltman & Schreiber, 2013). For example, if it were not possible to assign four publications in one reference set with the same citation counts unambiguously to the excellent publications, they would be assigned to the 10% most cited publications with a weighting of 0.525 and to the remaining 90% publications with a weighting of 0.525. The publications are therefore counted with weightings.

Both proposals ensure that one always obtains a proportion of 10% for the excellence rate and a proportion of 90% for the rest with one reference set.

As examples from both rankings, Tables 3 and 4 show those institutions that belong to the top 1% in SIR and LR 2013.

Discussion

University rankings generally present users with the problem of placing the results given for an institution in a comparison with others. Only a comparison of the excellence rates for a specific institution with those of all other institutions makes it possible to say exactly where an institution stands. To interpret the excellence rates, we offer in this study thresholds with which to classify an institution in a certain rank class. It is possible to assess whether an institution belongs to the top 1%, top 5%, top 10%, top 25%, or top 50% of institutions in the world. Furthermore, the user learns the maximum excellence rate that can be achieved in the ranking: The best institution does not achieve the theoretically possible value of 100% but values of around 50% (SIR) and 26% (LR 2011/2012) or 27% (LR 2013), respectively. The availability of figures for comparison is very helpful in allowing a meaningful estimate of institutional excellence rates. These figures are primarily required for the SIR, as the *World Report* (SCImago Research Group, 2012) in which the results are published is not (nor can it be) sorted by the excellence rate.

The thresholds we introduce in this study can be used for more than interpreting university rankings. They can also be

TABLE 3.   Institutions belonging to the top 1% in SIR.

Institution	Excellence rate
Broad Institute of MIT and Harvard	49.53
Whitehead Institute for Biomedical Research	47.92
Cold Spring Harbor Laboratory	41.64
Howard Hughes Medical Institute	41.62
Institute for Systems Biology	40.10
J. Craig Venter Institute	37.49
Wellcome Trust Sanger Institute	37.16
Harvard-MIT Division of Health Sciences and Technology	36.73
Salk Institute for Biological Studies	34.68
Institute of Electrical and Electronics Engineers, USA	34.14
Novartis Institutes for Biomedical Research	33.98
National Bureau of Economic Research	33.67
The Rockefeller University	33.40
European Molecular Biology Laboratory Heidelberg	33.28
Flanders Interuniversity Institute for Biotechnology	33.23
Novartis Pharma SA, East Hanover	33.20
Dana Farber Cancer Institute	32.95
F. Hoffmann-La Roche, Ltd	32.47
Group Health Cooperative	32.37
Microsoft Research Cambridge	32.08
International Agency for Research on Cancer	31.21
Nathan S. Kline Institute for Psychiatric Research	31.10
American Cancer Society	31.01
FOM Institute for Atomic and Molecular Physics	31.00
Medical Research Council	30.73
Scripps Research Institute	30.57
London Business School	30.51
World Health Organization Switzerland	30.38
Perimeter Institute for Theoretical Physics	30.32
Cancer Research UK	30.20

TABLE 4.   Universities belonging to the top 1% in LR 2013.

University	Excellence rate
Massachusetts Institute of Technology	27.23
Stanford University	24.91
Princeton University	24.88
University of California, Santa Barbara	24.81
University of California, Santa Cruz	24.77

used when a bibliometrician has calculated excellence rates in an evaluation of certain institutions and would like to assess which class of institutions worldwide they belong to. However, an application of the thresholds such as this should take account of the fact that excellence rates can be calculated in different ways (as previously discussed). The user should therefore have employed a method comparable to the threshold in question. Even though the thresholds presented here relate to certain publication periods, we are assuming that they will be valid for others. It should therefore be possible to apply them to other periods without any problems. The excellence rates for an institution are as a rule relatively stable over the years (Bornmann & Marx, 2013), since the excellence rates calculated as part of the rankings

are based mostly on large sets of publications which substantially overlap. As the example of the LR in this study shows, the thresholds between the two editions only differ slightly, although they are based on different publication years.

There have already been numerous calls for thresholds for comparative assessment in order to evaluate individual scientists (Garfield, 1979; Kreiman & Maunsell, 2011). Something that is very difficult to implement for individuals (Coleman, Bolumole, & Frankel, 2012; El Emam, Arbuckle, Jonker, & Anderson, 2012) can be easily achieved at the institutional level with the data from the rankings, as we have shown in this study.

Conclusion

It is the objective of our paper to present thresholds (for the excellence rate indicator) in order to interpret the results of the SIR and the LR properly. Using our thresholds it is possible to assess whether an institution belongs to the top 1%, top 5%, top 10%, top 25%, or top 50% of institutions in the world. For example, one knows now that the ETH Zurich with an excellence rate of 19.8% belongs to the top 10% institutions worldwide in the current LR. Furthermore, the user of both university rankings learns the maximum excellence rate that can be achieved worldwide (in the rankings): In the SIR, a maximum of around 50% and in the LR, a maximum of around 27% is achievable.

We would like to encourage with this study that university rankings present as a general rule thresholds based on percentile rank classes for a better interpretation of their results on institutional performance.

References

Abramo, G., D'Angelo, C.A., & Costa, F.D. (2010). Testing the trade-off between productivity and quality in research activities. Journal of the American Society for Information Science and Technology, 61(1), 132–140.

Albarrán, P., Crespo, J., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. Scientometrics, 88(2), 385–397.

Bornmann, L. (2013). How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes and top-cited papers. Journal of the American Society for Information Science and Technology, 64(3), 587–595.

Bornmann, L. (In press). Assigning publications to multiple subject categories for bibliometric analysis: an empirical case study based on percentiles. Journal of Documentation.

Bornmann, L., & de Moya Anegón, F. (2011). Some interesting insights from aggregated data published in the World Report SIR 2010. Journal of Informetrics, 5(3), 486–488.

Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2012). The new Excellence Indicator in the World Report of the SCImago Institutions Rankings 2011. Journal of Informetrics, 6(2), 333–335.

Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. Journal of Informetrics, 7(1), 158–165.

Bornmann, L., & Marx, W. (2013). How good is research really? EMBO Reports, 14(3), 226–230.

Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H.-D. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: Do the editors of a high-profile journal select manuscripts that are highly cited after publication? Journal of the Royal Statistical Society, Series A (Statistics in Society), 174(4), 857–879.

Boyack, K.W. (2004). Mapping knowledge domains: Characterizing PNAS. Proceedings of the National Academy of Sciences of the United States of America, 101, 5192–5199.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Coleman, B.J., Bolumole, Y.A., & Frankel, R. (2012). Benchmarking individual publication productivity in logistics. Transportation Journal, 51(2), 164–196.

El Emam, K., Arbuckle, L., Jonker, E., & Anderson, K. (2012). Two h-index benchmarks for evaluating the publication performance of medical informatics researchers. Journal of Medical Internet Research, 14(5).

Garfield, E. (1979). Citation indexing—its theory and application in science, technology, and humanities. New York: John Wiley & Sons.

Guerrero-Bote, V.P., & de Moya-Anegon, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. Journal of Informetrics, 6(4), 674–688.

Hemlin, S. (1996). Research on research evaluations. Social Epistemology, 10(2), 209–250.

Kreiman, G., & Maunsell, J.H.R. (2011). Nine criteria for a measure of scientific output. Frontiers in Computational Neuroscience, 5.

Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables in citation analysis one more time: Principles for comparing sets of documents. Journal of the American Society for Information Science and Technology, 62(7), 1370–1381.

National Science Board. (2012). Science and engineering indicators 2012. Arlington, VA: National Science Foundation (NSB 12-01).

SCImago Reseach Group. (2012). SIR World Report 2012. Granada, Spain: University of Granada.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., van Eck, N.J. . . . (2012a). The Leiden Ranking 2011/2012: data collection, indicators, and interpretation. Retrieved from: http://arxiv.org/abs/1202.3941

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., van Eck, N.J. . . . (2012b). The Leiden Ranking 2011/2012: data collection, indicators, and interpretation. Journal of the American Society for Information Science and Technology, 63(12), 2419–2432.

Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. Journal of the American Society for Information Science and Technology, 64(2), 372–379.

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S., & van Raan, A.F.J. (2011). Towards a new crown indicator: Some theoretical considerations. Journal of Informetrics, 5(1), 37–47.