BRIEF COMMUNICATION

Relative Citation Ratio (RCR): An Empirical Attempt to Study a New Field-Normalized Bibliometric Indicator

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

Robin Haunschild
Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany. E-mail: R.Haunschild@fkf.mpg.de

Hutchins, Yuan, Anderson, and Santangelo (2015) proposed the Relative Citation Ratio (RCR) as a new field-normalized impact indicator. This study investigates the RCR by correlating it on the level of single publications with established field-normalized indicators and assessments of the publications by peers. We find that the RCR correlates highly with established field-normalized indicators, but the correlation between RCR and peer assessments is only low to medium.

Introduction

It is standard in bibliometrics to field-normalize citation counts (Vinkler, 2010). For cited-side normalization, the citation counts of a focal article are compared with the citation counts of a reference set: all articles published in the same field and year. The fields are defined on the basis of journal sets, subject categories, which are algorithmically constructed on the basis of citation relations between publications (Ruiz-Castillo & Waltman, 2015), or expert opinions. For citing-side normalization, each citation of a focal article is weighted by the citation density of the citing article’s field (Bornmann & Marx, 2015). Hutchins, Yuan, Anderson, and Santangelo (2015)—a team of authors affiliated to the US National Institutes of Health (NIH)—proposed the Relative Citation Ratio (RCR) as a new field-normalized impact indicator. This indicator could signify an interesting new approach to normalization on the cited-side, because it relies on cocitations to generate the reference set. The articles cocited with the focal article are considered to represent the field of the article. In bibliometrics, cocitation is a similarity measure for articles which is based on citation relationships (another measure is bibliographic coupling).

This study investigates the correlation of RCR on the level of single publications with established field-normalized indicators and assessments of publications by peers (from the F1000 postpublication peer review system).

Literature Overview and Alternative Field-Normalized Indicators

The RCR has received positive comments because it could mean a move away from the use of the journal impact factor (JIF) in the biomedical area: Stefano Bertuzzi, executive director of the American Society for Cell Biology in Bethesda, “applauds the NIH for moving away from the journal impact factor (JIF). He wrote that the metric ‘evaluates science by putting discoveries into a meaningful context. I believe that the RCR is a road out of the JIF swamp’” (Bloudoff-Indelicato, 2015). The JIF measures the average citation impact of journals without normalizing citations and is frequently misused (in biomedicine) to study the impact of single publications. Many other moves away from the JIF were made, see for example the San Francisco Declaration on Research Assessment (DORA, http://www.ascb.org/dora/). Waltman (2015) criticizes the RCR using a fictitious example publication receiving citation impact from different fields. He shows that “publications may be penalized rather than rewarded for receiving interdisciplinary citations.”

Received November 25, 2015; revised December 23, 2015; accepted January 30, 2016

© 2016 ASIS&T • Published online 11 June 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23729

JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 68(4):1064–1067, 2017
citation density) could mean that the RCR of a publication is decreasing instead of increasing. Waltman (2015) regards this property of the RCR as an important disadvantage that disqualifies it from being an equitable alternative to the field-normalized indicators already in use in bibliometrics. The following three established indicators are currently in use in bibliometrics and reflect different field-normalizing methods (Bornmann & Marx, 2015): (i) The Mean Normalized Citation Score (MNCS) is a quotient composed of the citation counts of a focal article (numerator) and the average citation counts in the reference set (denominator, see earlier) (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011a, 2011b). (ii) For the calculation of Citation Percentiles (CP), the articles in the reference set are sorted by citation counts in descending order and the proportion of articles with equal or lower citation counts is calculated for every article (and also the focal article) (Bornmann, Leydesdorff, & Mutz, 2013). (iii) MNCS and CP are cited-side normalization approaches. For the SNCS(2), which is an indicator on the basis of citing-side normalization, each citation to an article is weighted with the number of cited references in the citing article (Waltman & van Eck, 2013). The idea behind this indicator is that the number of references reflects the citation density of the field in which the citing article was published. All indicators used in this study are described in detail by Bornmann and Marx (2015).

Data Set

F1000Prime is a postpublication peer review system of articles from the medical and biological areas (see http://f1000.com/prime). The articles included in the system are rated by faculty members (leading scientists and clinicians) as “Good,” “Very good,” or “Exceptional,” which are equivalent to scores of 1, 2, or 3, respectively. In many cases an article is not evaluated by one faculty member alone but by several. F1000Prime provided one of the authors of this study with data on all ratings made and the bibliographic information for the corresponding articles in their system (Bornmann, 2014, 2015a, 2015b; Bornmann & Haunschild, 2015). Because the data set does not contain any citation impact scores, it was matched with a bibliometric in-house database at the Max Planck Society (MPG), which is administered by the Max Planck Digital Library (MPDL) and is based on the Web of Science (WoS, Thomson Reuters). To enable a citation window of at least 3 years for every publication (Glänzel, 2008), those published later than 2012 are discarded. Thus, publications published between 1996 and 2012 are included. For this study, the F1000Prime data set is reduced to only those publications with at least two faculty members’ scores each. In order to increase the reliability of the scores, publications with only one score are excluded. Assessments by experts might be personally biased and the consideration of more than one score should increase the reliability of the assessments (Bornmann & Marx, 2014). In order to have only one total score from the faculty members, an average is calculated over the members’ scores for one and the same article. The consideration of only those articles published before 2013 with (i) at least two scores of faculty members and (ii) bibliometric data available in the in-house database reduce the F1000Prime data set to $n = 16,557$ articles for this study. For $n = 16,521$ publications an RCR score could be retrieved from https://icite.od.nih.gov/ on November 10, 2015.
approach pursued by Hutchins et al. (2015) is to use the sum over all F1000 scores for each article. Using the sum over all F1000 scores for each article results in slightly increased correlation coefficients: They range between $r_s = 0.27$ for CP and $r_s = 0.38$ for SNCS$_{(23)}$, but do not differ significantly from the correlation coefficients in Table 1 where the average value over all F1000 scores per article is used.

### Discussion

Our results show that RCR correlates highest with MNCS, closely followed by CP and SNCS$_{(23)}$. This is surprising considering the valid criticism by Waltman (2015). Probably, the fictitious example in Waltman (2015) does not occur often in the publication set studied here. Articles recommended by F1000 are usually from the biomedical fields and usually receive their citations from there, so that the fictitious example in Waltman (2015) does not occur in many cases within the F1000Prime publication set.

Hutchins et al. (2015) also studied the correlation of RCR with F1000 scores and reported a correlation coefficient of $r = 0.44$ (in the supplemental information). This coefficient is somewhat higher than the one calculated here ($r = 0.29$). However, their validation was done on a smaller ($n = 2,193$ articles) and more selective publication set (R01-funded articles published in 2009). Our study includes a larger publication set ($n = 16,521$ articles) with a range of publication years (from 1996 to 2012) and imposes no funding restrictions at all.

Even larger scale studies are desirable when RCR scores are also available for publications outside the biomedical area. The database PubMed focuses on this area and one needs PubMed IDs to retrieve RCR scores. If the RCR scores were available for publications covering more areas (e.g., publications without a PubMed ID), it could be investigated whether the RCR can field-normalize citation counts better than established field-normalized indicators in bibliometrics. Furthermore, it would be necessary for a larger study that RCR scores be obtained for larger publication sets (without the restriction of processing 200 articles at a time, see https://icite.od.nih.gov/). It would be very helpful if an application programming interface (API) were provided by the NIH for the purpose of comparing RCR values of a large amount of articles with the field-normalized indicators currently in use in bibliometrics. There are various methods available which can be used to study the ability of the RCR to normalize citation counts (Bormann, de Moya Anegón, & Mutz, 2013; Waltman & van Eck, 2013). These methods should be used to compare the RCR with established field-normalized indicators.

### Acknowledgements

We would like to thank Ros Dignon and Iain Hrynaszkiewicz from F1000 for providing one of the authors with the F1000Prime data set. The bibliometric data used in this article are from an in-house database developed and maintained by the Max Planck Digital Library (MPDL, Munich) and

### Table 1. Spearman rank-order correlation coefficients for the correlations between MNCS, CP, SNCS$_{(23)}$, Citations, F1000 score, and RCR.

| RCR | MNCS | CP | SNCS$_{(23)}$ | Citations | F1000 score |
|-----|------|----|--------------|-----------|-------------|
| 1.00| 0.88 | 0.85| 0.85         | 0.70      | 0.29        |
| 0.88| 0.99 | 0.78| 0.78         | 0.64      | 0.25        |
| 0.85| 0.99 | 1.00| 0.75         | 0.61      | 0.23        |
| SNCS$_{(23)}$ | 0.85 | 0.78 | 0.75 | 1.00 |
| Citations | 0.70 | 0.64 | 0.61 | 0.88 | 1.00 |
| F1000 score | 0.29 | 0.25 | 0.23 | 0.31 | 0.25 | 1.00 |

Note: CP is calculated on the basis of the formula proposed by Hazen (1914).
derived from the Science Citation Index Expanded (SCI-E), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI) provided by Thomson Reuters. We would like to thank Ludo Waltman and Ian Hutchins for their valuable feedback on an earlier version of our article.

References

Bloudoff-Indelicato, M. (2015). NIH metric that assesses article impact stirs debate. The US biomedical funding agency is using the metric to evaluate funded research. Retrieved from http://www.nature.com/news/nih-metric-that-assesses-article-impact-stirs-debate-1.18734

Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. Journal of Informetrics, 8(4), 935–950.

Bornmann, L. (2015a). Inter-rater reliability and convergent validity of F1000Prime peer review. Journal of the Association for Information Science and Technology, 66(12), 2415–2426.

Bornmann, L. (2015b). Usefulness of altmetrics for measuring the broader impact of research: A case study using data from PLOS and F1000Prime. Aslib Journal of Information Management, 67(3), 305–319. doi: 10.1108/AJIM-09-2014-0115.

Bornmann, L., de Moya Anegón, F., & Mutz, R. (2013). Do universities or research institutions with a specific subject profile have an advantage or a disadvantage in institutional rankings? A latent class analysis with data from the SCImago ranking. Journal of the American Society for Information Science and Technology, 64(11), 2310–2316.

Bornmann, L., & Haunschild, R. (2015). Which people use which scientific papers? An evaluation of data from F1000 and Mendeley. Journal of Informetrics, 9(3), 477–487.

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. (2014). The wisdom of citing scientists. Journal of the American Society for Information Science and Technology, 65(6), 1288–1292.

Bornmann, L., & Marx, W. (2015). Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts? Journal of Informetrics, 9(2), 408–418.

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

Glänzel, W. (2008). Seven myths in bibliometrics. About facts and fiction in quantitative science studies. Paper presented at the Proceedings of WIS 2008, Berlin. Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, Berlin, Germany.

Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. Transactions of American Society of Civil Engineers, 77, 1539–1640.

Hutchins, B.I., Yuan, X., Anderson, J.M., & Santangelo, G.M. (2015). Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level. Retrieved from http://biorxiv.org/content/early/2015/10/22/029629

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. Journal of Informetrics, 9(1), 102–117. doi: http://dx.doi.org/10.1016/j.joi.2014.11.010.

Vinkler, P. (2010). The evaluation of research by scientometric indicators. Oxford, UK: Chandos Publishing.

Waltman, L. (2015). NIH’s new citation metric: A step forward in quantifying scientific impact? Retrieved from http://www.cwts.nl/blog/article=nq4294&title=nihs-new-citation-metric-a-step-forward-in-quantifying-scientific-impact#hash.w1KC3A1O.dpuf

Waltman, L., van Eck, N., van Leeuwen, T., Visser, M., & van Raan, A. (2011a). Towards a new crown indicator: An empirical analysis. Scientometrics, 87(3), 467–481. doi: 10.1007/s11192-011-0354-5.

Waltman, L., & van Eck, N.J. (2013). A systematic empirical comparison of different approaches for normalizing citation impact indicators. Journal of Informetrics, 7(4), 833–849.

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S., & van Raan, A.F.J. (2011b). Towards a new crown indicator: Some theoretical considerations. Journal of Informetrics, 5(1), 37–47. doi: 10.1016/j.joi.2010.08.001.