Citation score normalized by cited references (CSNCR): The introduction of a new citation impact indicator

Lutz Bornmann [a,*], Robin Haunschild [b]

[a] *Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany*
[b] *Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany*

A R T I C L E   I N F O

A B S T R A C T

In this paper, a new field-normalized indicator is introduced, which is rooted in early insights in bibliometrics, and is compared with several established field-normalized indicators (e.g. the mean normalized citation score, MNCS, and indicators based on percentile approaches). Garfield (1979) emphasizes that bare citation counts from different fields cannot be compared for evaluative purposes, because the "citation potential" can vary significantly between the fields. Garfield (1979) suggests that "the most accurate measure of citation potential is the average number of references per paper published in a given field." Based on this suggestion, the new indicator is basically defined as follows: the citation count of a focal paper is divided by the mean number of cited references in a field to normalize citations. The new indicator is called citation score normalized by cited references (CSNCR). The theoretical analysis of the CSNCR shows that it has the properties of consistency and homogeneous normalization. The close relation of the new indicator to the MNCS is discussed. The empirical comparison of the CSNCR with other field-normalized indicators shows that it is slightly poorer able to field-normalize citation counts than other cited-side normalized indicators (e.g. the MNCS), but its results are favorable compared to two citing-side indicator variants (SNCS indicators). Taken as a whole, the results of this study confirm the ability of established indicators to field-normalize citations.

1. Introduction

In order to improve the quality of scientific research, research evaluation is an important part of scientific activity. Whereas research evaluation was dominated by the application of the peer-review process over a long time (Bornmann, 2011), the quantitative analysis of research outputs and their citation impact has become more and more important (Moed & Halevi, 2015). One main reason for the increasing importance of metrics is that "pressures on universities to be more accountable to government and public funders of research have intensified" (Wilsdon et al., 2015; p. 68). Bibliometric numbers are available in specific databases for nearly all scientific fields and can be used – following certain standards (Bornmann et al., 2014) – to investigate the performance of institutions in comparison with each other or with a world-average. Although bibliometrics are accepted by the general scientific community (Panel for Review of Best Practices in Assessment of Research et al., 2012),

---

the use of bibliometrics is also critically seen, as Wilsdon et al. (2015) outlined: "Metrics are widely seen as absolving research managers of the responsibility for making assessments based on more accurate and complete information, and as contributing to mistrust of research management more generally" (p. 80).

According to Wilsdon et al. (2015) "the most widely exploited bibliometric relies on counts of citations. Citation counts are sometimes used as an indicator of academic impact in the sense that citations from other documents suggest that the cited work has influenced the citing work in some way" (p. 5). Citations are used as an indicator of scientific impact, although scientific impact can occur which does not manifest in citations. Thus, citations are used as a measurable quantity which substitutes for impact. It is an advantage of this quantity that it is directly rooted in the publication process of scientific results: "It is one of the basic rules of scientific research that a piece of written research, in order to warrant publication, needs to be adequately situated within the existing research literature. Awareness of the existing literature, and of the decisive developments and discussions in a field, is signaled through the inclusion of a range of markers, most commonly, a combination of in-text citations and bibliographic entries" (Woelert, 2013, p. 350).

For citation analyses in the evaluative practice, it is often necessary to make comparisons between papers which have been published in different fields (Waltman, 2016). Bare citation counts cannot be used for these comparisons, because each field has developed own practices of publication, citation, and authorship (Waltman & van Eck, 2013b). These practices lead to differences in the referencing patterns (which will be explored in Section 3.1) and – as a direct consequence – to differences in average citation counts. Beginning in the mid-1980s, field-normalized indicators have been developed whereby the citation count of a focal paper is compared with the average citation count of a field, in which the focal paper was published (Aksnes, 2006). Thus, the field average is used as the expected number of citations for a focal paper (Waltman, 2016). The comparison with the expected value is intended to correct as much as possible for the effect of field-specific practices of publication, citation, and authorship. Today, it is standard in bibliometrics to use field-normalized indicators; its use is also recommended in the guiding principles for research evaluation in the Leiden manifesto (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015). In recent years, many different methods have been developed to field-normalize citations (see an overview, for example, in Mingers & Leydesdorff, 2015; Waltman, 2016).

In this paper, a new field-normalized indicator is introduced, which is rooted in early insights in bibliometrics. Garfield (1979) emphasizes that bare citation counts from different fields cannot be compared for evaluative purposes, because the "citation potential" can vary significantly between the fields (see also Moed, 2010). Garfield (1979) suggests that "the most accurate measure of citation potential is the average number of references per paper published in a given field". Based on this suggestion, our new indicator is basically defined as follows: the citation count of a focal paper is divided by the mean number of cited references in a field to normalize citations. The new indicator is called citation score normalized by cited references (CSNCR) and will be explained, justified, and compared with existing field-normalized indicators in the following. It is an advantage of the new indicator that the mean number of cited references in a field, which are used for normalizing citation counts, does no longer change after the paper has been published. In other words, having produced these numbers for various time periods, they can be used for normalization at any time. For other field-normalized indicators, these scores have to be reproduced regularly. For example, Thomson Reuters could publish tables with the mean numbers of cited references in the fields in the Essential Science Indicators, which could then be used for normalizing citation counts.

The paper is organized as follows: In the Methods Sections 2.1 and 2.2, the underlying dataset of the study is described and the indicators are explained which are used for the comparison with the CSNCR. In the Results Section 3, field-specific referencing patterns are revealed and the CSNCR is explained as well as theoretically and empirically analysed. The paper closes with a discussion of the CSNCR in the context of field normalization.

2. Methods

2.1. Dataset used

The complete publication records of the Web of Science (WoS) with the document type "article" including their cited references (which are not restricted to "articles") of papers published between 1980 and 2014 are used in this study. The bibliometric data used in this paper are from an in-house database developed and maintained by the Max Planck Digital Library (MPDL, Munich) and derived from the Science Citation Index Expanded (SCI-E), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI) prepared by Thomson Reuters (Philadelphia, Pennsylvania, USA). The in-house database was updated on the 5th of February 2016. As we can't expect the publication year 2015 to be indexed completely, citations are accounted for until the end of 2014.

2.2. Field-normalized indicators used for comparison with the citation score normalized by cited references

In recent years, several overviews on field-normalized indicators have been published (Bornmann & Marx, 2015; Vinkler, 2010; Waltman, 2016). For comparison with the CSNCR, we have selected those six indicators which have gained a degree of importance in bibliometrics and are available in the MPDL in-house database. In this study, WoS subject categories are used to define fields. The WoS subject categories are sets of journals from similar research areas.