



How can citation impact in bibliometrics be normalized? A new approach combining citing-side normalization and citation percentiles

an open access  journalLutz Bornmann Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society,
Hofgartenstr. 8, 80539 Munich, Germany**Keywords:** bibliometrics, citation analysis, citation percentiles, citing-side normalization

Citation: Bornmann, L. (2020). How can citation impact in bibliometrics be normalized? A new approach combining citing-side normalization and citation percentiles. *Quantitative Science Studies*, 1(4), 1553–1569. https://doi.org/10.1162/qss_a_00089

DOI:
https://doi.org/10.1162/qss_a_00089

Received: 8 May 2020
Accepted: 30 July 2020

Corresponding Author:
Lutz Bornmann
bornmann@gv.mpg.de

Handling Editor:
Ludo Waltman

ABSTRACT

Since the 1980s, many different methods have been proposed to field-normalize citations. In this study, an approach is introduced that combines two previously introduced methods: citing-side normalization and citation percentiles. The advantage of combining two methods is that their advantages can be integrated in one solution. Based on citing-side normalization, each citation is field weighted and, therefore, contextualized in its field. The most important advantage of citing-side normalization is that it is not necessary to work with a specific field categorization scheme for the normalization procedure. The disadvantages of citing-side normalization—the calculation is complex and the numbers are elusive—can be compensated for by calculating percentiles based on weighted citations that result from citing-side normalization. On the one hand, percentiles are easy to understand: They are the percentage of papers published in the same year with a lower citation impact. On the other hand, weighted citation distributions are skewed distributions with outliers. Percentiles are well suited to assigning the position of a focal paper in such distributions of comparable papers. The new approach of calculating percentiles based on weighted citations is demonstrated in this study on the basis of a citation impact comparison between several countries.

1. INTRODUCTION

Research systematically investigates what is (still) not known. In order to demonstrate the lag in current knowledge and the shoulders on which the exploration of the lag by new studies stand, authors of papers (ideally) cite all relevant previous publications (Kostoff, Murday, et al., 2006). On the basis of this norm in science to cite the relevant past literature, citations have been established as a proxy for scientific quality—measuring science “impact” as an important component of quality (Aksnes, Langfeldt, & Wouters, 2019). Narin (1976) proposed the term *evaluative bibliometrics* for methods using citation-based metrics for measuring cognitive influence (Moed, 2017; van Raan, 2019). Bornmann and Marewski (2019) introduced the bibliometrics-based heuristics (BBHs) concept concretizing the evaluative use of bibliometrics: “BBHs characterize decision strategies in research evaluations based on bibliometrics data (publications and citations). Other data (indicators) besides bibliometrics are not considered” (Bornmann, 2020).

According to Moed and Halevi (2015), research assessment (based on bibliometrics) is an integral part of any scientific activity these days: “it is an ongoing process aimed at improving

Copyright: © 2020 Lutz Bornmann.
Published under a Creative Commons
Attribution 4.0 International (CC BY 4.0)
license.



the quality of scientific/scholarly research. It includes evaluation of research quality and measurements of research inputs, outputs, and impacts, and embraces both qualitative and quantitative methodologies, including the application of bibliometric indicators and peer review" (p. 1988). Current research evaluation processes concern single researchers (Bornmann & Marx, 2014), research groups, institutions, organizations (Bornmann, Bowman, et al., 2014), and countries (Leydesdorff, Wagner, & Bornmann, 2014). Since the beginning of the 20th century, annually produced international university rankings have become more and more popular (Vernon, Balas, & Momani, 2018).

The analysis of citations is at the core of bibliometrics: "citation impact is an important indicator of scientific contribution because it is valid, relatively objective, and, with existing databases and search tools, straightforward to compute" (Nosek, Graham, et al., 2010, p. 1292). The problem of citation analysis is, however, that fields differ in their publication, citation, and authorship practices (Waltman & van Eck, 2013b). Crespo, Li, and Ruiz-Castillo (2013) estimated that 14% of overall citation inequality can be attributed to field-specific differences in citation practices. These and similar findings from bibliometrics research make clear that the results of citation analyses from different fields cannot be compared. Whereas single publications and researchers can be compared within one field, this is not possible with universities and many research-focused institutions. For citation analyses in which cross-field comparisons are necessary, field-normalized citation impact indicators have been developed. Field normalization aims to remove the noise that traces back to the fields while maintaining the signal that reflects (true) performance differences (Waltman & van Eck, 2019). It is an indication of advanced bibliometrics to use "reliable statistics, e.g., corrections for differences in publication and citation practices between scientific disciplines" (van Raan, 2019, p. 244).

Since the 1980s, many approaches have been developed in the scientometrics field to field-normalize citations. Although some approaches (e.g., the number of publications published by an institution that belongs to the 10% most frequently cited publications in the corresponding fields) could reach the status of quasistandards, each approach has its specific disadvantages. In this paper, an approach is introduced combining the advantages of two published approaches and smoothing their specific disadvantages. The first approach is citing-side normalization, whereby each single citation of a paper is field-normalized. The second approach is the citation percentile, which is the percentage of papers in a given set of papers with lower citation impact than the focal paper.

2. LITERATURE OVERVIEW: A SHORT HISTORY OF FIELD NORMALIZATION

Field normalization has a long tradition in bibliometrics. Literature overviews on the developments in the field can be found in Mingers and Leydesdorff (2015), Bornmann and Marx (2015), and Waltman (2016). Field normalizations start from the basic premise that "not all citations are equal. Therefore, normalization can be seen as a process of benchmarking that is needed to enhance comparability across diverse scientists, fields, papers, time periods, and so forth" (Ioannidis, Boyack, & Wouters, 2016). Many studies on field normalization either deal with technical issues (e.g., the development of improved indicator variants) or with the way fields should be defined for use in normalization (e.g., by using journal sets or human-based assignments; see Wilsdon, Allen, et al., 2015). One of the earliest attempts in bibliometrics to field-normalize citations was made by Schubert and Braun (1986) and Vinkler (1986). They proposed to calculate the average citation rate for a journal or field and to use this reference score to field-normalize (single) papers published in the journal or field (by dividing the citation counts of every single papers by the reference score). The resulting metric was named the relative citation rate (RCR) by Schubert and Braun (1986).

Since its introduction, the RCR has been criticized for its use of the arithmetic average in the normalization. The arithmetic average should not be used as a measure of central tendency for skewed citation data. According to Glänzel and Moed (2013), “the mean should certainly not be used if the underlying distribution is very skewed, and has a long tail” (p. 383). The fact that arithmetic averages of citation data and, thus, field normalized citation scores are sensitive to outliers has been named by van Raan (2019) as the *Göttingen effect*: “In 2008, a paper published by a researcher of the University of Göttingen became extremely highly cited, many thousands of times a year, within a very short time ... As a result, for several years after this publication, Göttingen achieved a very high position in ... [university] rankings” (p. 260).

To deal with the problem of skewed distributions in field normalization, McAllister, Narin, and Corrigan (1983, p. 207) already proposed in the 1980s that percentiles should be used for citation data:

the p th percentile of a distribution is defined as the number of citations X_p the percent of papers receiving X_p such that or fewer citations is equal to p . Since citation distributions are discrete, the p th percentile is defined only for certain p that occur in the particular distribution of interest. Thus we would say that a 1974 physiology paper receiving one citation falls in the 18th percentile of the distribution. This means that 82 percent (100 – 18) of all 1974 U.S. physiology papers received more than one citation. For any paper in the 18th percentile of any subject area citation distribution, 18 percent of the papers performed at a level less than or equal to the particular paper, and 82 percent of the papers in the subject area outperformed the particular paper.

For Schreiber (2013) “percentiles ... have become a standard instrument in bibliometrics” (p. 822) in current bibliometrics. Percentiles are recommended in the Leiden manifesto which includes 10 principles to guide research evaluation (Hicks, Wouters, et al., 2015). The most recent field-normalizing percentile approach has been published by Bornmann and Williams (2020).

One of the biggest challenges in field normalizing citations is the selection of the system categorizing papers to fields. The overview by Sugimoto and Weingart (2015) shows that existing systems emphasize cognitive, social, or institutional orientations of fields to a different extent. Various field categorization schemes are in use to normalize citations and there exists no standard use in bibliometrics. The most frequently used schemes are multidisciplinary schemes that span all fields (Sugimoto & Weingart, 2015; Wang & Waltman, 2016). These schemes are typically based on journal sets: the Web of Science (WoS) subject categories of Clarivate Analytics and the Scopus subject areas of Elsevier. The use of journal sets can be justified quite well: according to Milojević (2020, p. 184) “journals often serve as anchors for individual research communities, and new journals may signify the formations of disciplines.” Each journal is a well-crafted folder sustained by editors, reviewers, and authors who usually know and use that outlet. Authors typically direct their manuscripts in an informed way to reach the appropriate audience for the content and argument.

There are two problems with these schemes, however, which is why Waltman and van Eck (2012) proposed a new method for algorithmically constructing classification systems (see also Boyack & Klavans, 2010): (a) Because journals publish many different papers, journals are usually assigned to more than one category; and (b) journal sets represent broad fields which is why papers from specific fields might be misclassified (see Strotmann & Zhao, 2010). The results by Shu, Julien, et al. (2019) reveal that about half of the papers published in a journal are not from the field to which the journal has been assigned.

The system proposed by Waltman and van Eck (2012) is based on citation relations between single publications. The advantages of the system are that (a) it assigns single publications (and not

journals) to fields and (b) it provides a fine-grained categorization scheme of publications. Ruiz-Castillo and Waltman (2015) demonstrate the use of the system for field normalization. The system, however, has not remained without criticism: because

“fields” are algorithmic artifacts, they cannot easily be named (as against numbered), and therefore cannot be validated. Furthermore, a paper has to be cited or contain references in order to be classified, since the approach is based on direct citation relations ... However, algorithmically generated classifications of journals have characteristics very different from content-based (that is, semantically meaningful) classifications ... The new Leiden system is not only difficult to validate, it also cannot be accessed or replicated from outside its context of production in Leiden (Leydesdorff & Milojević, 2015, p. 201).

As the recent results by Sjögarde, Ahlgren, and Waltman (2020) show, at least the labeling problem of the fields can be solved.

Another critical point is that the field assignments based on citation relations change with new citations. The approach does not lead to stable results, and it is elusive why the field assignment of a paper should change. Further critical remarks can be found in Haunschild, Schier, et al. (2018). Based on the critique of the system proposed by Waltman and van Eck (2012), Colliander and Ahlgren (2019) introduced an item-oriented approach that avoids clustering, but uses publication-level features to estimate subject similarities. The empirical comparison of this approach with standard approaches in bibliometrics by the authors revealed promising results. Future independent studies will demonstrate whether these first positive results can be confirmed.

As an alternative to multidisciplinary schemes, monodisciplinary schemes have been proposed for field normalization. The advantages of these schemes are that papers are usually assigned to a single research field and human indexers (field experts or authors of papers) assign the relevant field to a paper intellectually (Bornmann, Marx, & Barth, 2013). In recent years, studies have used different monodisciplinary schemes to field-normalize citations in certain fields: Bornmann and Wohlrabe (2019) used *Journal of Economic Literature* classification (JEL) codes in economics, Bornmann, Schier, et al. (2011) and Bornmann and Daniel (2008) used *Chemical Abstracts* (CA) sections in chemistry and related areas, Radicchi and Castellano (2011) used Physics and Astronomy Classification Scheme (PACS) codes in physics and related areas, and Smolinsky and Lercher (2012) used the MathSciNet’s Mathematics Subject Classification (MSC) system in mathematics. The disadvantages of monodisciplinary schemes are that they are restricted to single fields and the assignments by the indexers may be affected by subjective biases.

One problem that affects many field classification systems (mono- and multidisciplinary) is that they exhibit different aggregation levels, and it is not clear which level should be used to field-normalize citations (Waltman & van Eck, 2019; Wouters, Thelwall, et al., 2015). In bibliometrics, different results and opinions have been published as to whether an aggregation level change has any (significant) influence on the field-normalized scores: Zitt, Ramanana-Rahary, and Bassecoulard (2005) report a lack of stability of these scores; Colliander and Ahlgren (2011) arrive at another conclusion. Wang (2013) holds the opinion that “normalization at finer level is still unable to achieve its goal of improving homogeneity for a fairer comparison” (p. 867).

3. CITING-SIDE NORMALIZATION

The literature overview in section 2 has shown that there are many problems with field normalization in bibliometrics and it has not yet been possible to establish a standard. One can expect that some problems will remain unsolved without finding a perfect solution. For example, it will

remain a normative decision as to which field categorization scheme is used (and on what level). Independently of the system that is used, fields are not isolated and research based on between-field collaborations is common (Ioannidis et al., 2016). “With the population of researchers, scientific literature and knowledge ever growing, the scientific endeavour increasingly integrates across boundaries” (Gates, Ke, et al., 2019, p. 34). According to Waltman and van Eck (2013a), “the idea of science being subdivided into a number of clearly delineated fields is artificial. In reality, boundaries between fields tend to be rather fuzzy” (p. 700).

A possible solution to these problems might be to avoid the use of field categorization schemes (Bornmann, Marx, et al., 2013), clustering (Waltman & van Eck, 2012), and similarity approaches (Colliander & Ahlgren, 2019), and for each focal paper (that is assessed) to manually search some papers for comparison that are thematically similar (Kostoff, 2002; Waltman, 2016). This solution corresponds to the judgement by Hou, Pan, et al. (2019) that field normalization cannot be solved by statistical techniques. The manual collection of papers for the comparison with a focal paper might be possible in the evaluation of small sets of papers; however, it is not practicable for large sets (e.g., all papers published by a university over several years). Furthermore, one needs experts from the fields to find the papers for comparison.

Another solution that can be applied to large sets of papers is not to normalize citation impact based on expected citations from the reference sets, but to normalize single citations directly. So-called citing-side field normalizing approaches have been proposed in recent years that normalize each single citation of a focal paper. van Raan (2014) sees these “field-independent normalization procedures” (p. 22) as an important and topical issue in bibliometrics. The simplest procedure is to divide each citation by the number of cited references of the citing paper. The use of the number of cited references is intended to reflect the disciplinary context of the citing paper and to standardize the citation field specifically. It is a decisive advantage of citing-side normalization that it “does not require a field classification system” (Waltman & van Eck, 2013a, p. 700). Citing-side normalization, thus, solves the problem with the selection of a field-categorization scheme by refraining from it.

Citing-side normalization might be a reasonable approach for citation analysis, as the goal of field normalization is the normalization of citation impact (see Waltman, van Eck, et al., 2013). Given the different directions of the two basic field normalization approaches, citing-side approaches are more focused on the aim of field normalization than approaches that are based on reference sets on the cited side: Citing-side approaches normalize each single citation of a focal paper. Bornmann and Marx (2015) demonstrated the problem of field normalization based on cited-side normalization by using the well-known paper by Hirsch (2005) on the *h*-index as an example. This paper is a typical bibliometrics paper (it introduces a new indicator based on publication and citation data), but receives citations from many fields (not only from the bibliometrics field). If a focal paper is attractive for authors publishing in other fields with high citation density, it has an advantage over another focal paper that is not as attractive for these fields. Although both focal papers might belong to the same field (viewed from the cited-side perspective), they have different chances of being cited.

The paper by Hirsch (2005) is concerned with another “problem” (for field normalization): It was published in the *Proceedings of the National Academy of Sciences of the United States of America*. This is a multidisciplinary journal and is assigned to another journal set than most of the papers published in bibliometrics (which are assigned to library and information science). Thus, by using journal sets as a field categorization scheme, the paper would not be compared with its “true” reference papers, but with various papers from many different fields, which are usually published in multidisciplinary journals. An appropriate reference set for this paper would be

all papers published in journals in the library and information science set. If one decides to manually collect the reference papers for comparison (see above), the ideal reference set for the paper by Hirsch (2005) would consist of all papers publishing a variant of the h -index or all papers having introduced an indicator combining the number of publications and the number of citations in a single number.

The idea of citing-side normalization has been introduced by Zitt and Small (2008). They proposed a modification of the Journal Impact Factor (JIF) by fractional citation weighting. The JIF is a popular journal metric that is published in the Journal Citation Reports by Clarivate Analytics. The indicator measures the average citation rate of papers published in a journal within 1 year. Citing-side normalization is also named as source normalization, fractional counting of citations, or a priori normalization (Waltman, 2016; Waltman & van Eck, 2013a). The method focuses on the citation environment of single citations and weights each citation depending on its citation environment: A citation from a field with high citation density (on average, authors in these fields include many cited references in their papers) receives a lower weight than a citation from a field with low citation density (on average, authors in these fields include only a few cited references in their papers). The basic idea of the method is as follows: Each citation is adjusted for the number of references in the citing publication or in the citing journal (as a representative for the entire field). In the recent decade, some variants of citing-side indicators have been published (Waltman, 2016; Waltman & van Eck, 2013a). These variants are presented in the following based on the explanations by Bornmann and Marx (2015).

$$\text{SNCS1} = \sum_{i=1}^c \frac{1}{a_i} \quad (1)$$

The first variant has been named *SNCS1* (Source Normalized Citation Score 1). In the formula, a_i is the average number of linked references in those publications that appeared in the same journal and in the same publication year as the citing publication i . Linked references are the part of cited references that refers to papers from journals covered by the citation index (e.g., WoS or Scopus). The limitation to linked references (instead of all references) is intended to prevent a situation in which fields that frequently cite publications are not indexed in WoS are disadvantaged (see Marx & Bornmann, 2015). The calculation of the average number of linked references in *SNCS1* is restricted to certain referenced publication years. Imagine a focal paper published in 2008 with a citation window covering a period of 4 years (2008 to 2011). In this case, every citation of the focal paper is divided by the average number of linked references to the four previous years. In other words, a citation from 2010 is divided by the linked cited references from the period 2007 to 2010. This restriction to recent publication years is designed to prevent fields that cite rather older literature from being disadvantaged in the normalization (Waltman & van Eck, 2013b).

$$\text{SNCS2} = \sum_{i=1}^c \frac{1}{r_i} \quad (2)$$

SNCS2 is the second variant of citing-side indicators. Here, each citation is divided by the number of linked cited references in the citing publication. Therefore, the journal perspective is not considered in this variant. The selection of the reference publication years is analogous to *SNCS1*.

$$\text{SNCS3} = \sum_{i=1}^c \frac{1}{p_i r_i} \quad (3)$$

SNCS3 is a combination of SNCS1 and SNCS2. r_i is equally defined as in SNCS2. p_i is the share of papers that contain at least one linked cited reference among the following papers: from the same journal and publication year as the citing paper i . The selection of the referenced publication years is analogous to SNCS1 and SNCS2.

Whereas Leydesdorff, Radicchi, et al. (2013) concluded that cited-side normalization outperforms citing-side normalization, the empirical results of Waltman and van Eck (2013a) and Bornmann and Marx (2015) demonstrated that citing-side normalization is more successful in field-normalizing citation impact than cited-side normalization. Therefore, it seems reasonable for reaching the goal of field normalization to weight each citation “based on the referencing behavior of the citing publication or the citing journal” (Waltman & van Eck, 2013a, p. 703). The comparison of the three citing-side approaches by Waltman and van Eck (2013b, p. 842) revealed that

SNCS(2) should not be used. Furthermore, the SNCS(3) approach appears to be preferable over the SNCS(1) approach. The excellent performance of the SNCS(3) approach in the case of classification system C ... suggests that this approach may be especially well suited for fine-grained analyses aimed for instance at comparing researchers or research groups active in closely related areas of research.

The results by Bornmann and Marx (2015), however, did not reveal these large differences between the three indicator variants.

Cited-side normalization is frequently confronted with the problem that the used field categorization scheme assigns papers to more than one field. Thus, it is necessary to consider these multiple assignments in the calculation of field-normalized indicators (see Waltman, van Eck, et al., 2011). As multiple assignments are not possible with citing-side normalization, this problem is no longer existent—a further decisive advantage of the approach.

4. PURPOSE OF THE STUDY—THE COMBINATION OF CITING-SIDE NORMALIZATION AND CITATION PERCENTILES

In section 3, the advantages of field normalization using citing-side approaches have been demonstrated based on the previous literature. Although these advantages have been reported in several papers over many years, these approaches have not been established as standard indicators in (applied) bibliometrics. For example, the Leiden Ranking (see <https://www.leidenranking.com>) does not consider citing-side indicators, but percentile-based cited-side indicators. One important reason for the avoidance of citing-side indicators might be that these indicators are more complicated to understand (and explain) than many cited-side indicators and indicators that are not based on field normalization. The results by Hammarfelt and Rushforth (2017) show that “simple and well-established indicators, like the JIF and the h-index, are preferred” (pp. 177–178) when indicators are used in practice. Jappe, Pithan, and Heinze (2018) similarly wrote that “the journal impact factor (JIF) ... and the Hirsch Index (h-index or HI) ... have spread widely among research administrators and funding agencies over the last decade.” According to the University of Waterloo Working Group on Bibliometrics (2016), “there is often a demand for simple measures because they are easier to use and can facilitate comparisons” (p. 2).

This study is intended to propose a field normalization approach that combines citing-side normalization and citation percentiles. The advantage of the combination lies in the abandonment of a field classification system (by using citing-side normalization) and the realization of field normalized scores (percentiles) that are relatively simple to understand and being applied in

research evaluation. In the first step of the approach, weighted citation counts are calculated based on the formula (see above) presented by Waltman and van Eck (2013a). In this study, the SNCS3 is used, as Waltman and van Eck (2013b) recommended its use (based on their empirical results). However, the approach is not bound to this SNCS variant. In the second step, the percentile approach proposed by Bornmann and Williams (2020) is used to calculate citation percentiles based on SNCS3. In this step, too, it is possible to use another percentile approach such as those proposed by Bornmann, Leydesdorff, and Mutz (2013) or Bornmann and Mutz (2014). This study prefers the approach by Bornmann and Williams (2020), because the authors point out the advantages of their approach over previous approaches.

Bornmann and Williams (2020) calculated cumulative frequencies in percentages (CPs) as demonstrated in Table 1 based on the size-frequency distribution (Egghe, 2005) to receive citation percentiles. The table shows the citation counts and SNCS3 for 24 fictitious papers. For example, there are five papers in the set with 12 citations and a weighted citation impact of 0.45 each. Note that not all papers with five citations have an SNCS3 score of 0.45 and vice versa. For the indicator CP-EX_{WC} (the subscript WC stands for weighted citations), the first percentage (for papers with 1 citation) is set at 0. The calculation of the cumulative percentage starts in the second row with the percentage of the lowest citation count (16.67%). By setting the first row to zero, CP-EX_{WC} measures exactly the percentage of papers with lower citation impact in the set of papers. For example, CP-EX_{WC} = 95.83 means that exactly 95.83% of the papers in the set of 24 papers received a citation impact—measured by SNCS3—that is below the weighted citation impact of 4.51. 16.67% of the papers received less impact than the weighted citations of 0.20.

CP-EX_{WC} can be calculated for all papers in a database (e.g., all WoS papers) with SNCS3 scores included (or the scores based on another variant). Because (weighted) citation impact depends on the length of the citation window, CP-EX_{WC} should be calculated based on all papers in 1 year (i.e., separated by publication years). With CP-EX_{WC} calculated using SNCS3, one receives a field-normalized indicator that is simple to understand—because the scores are

Table 1. Cumulative percentages (CP-EX_{WC}) based on 24 fictitious papers

Citations	SNCS3 (rounded)	Number of papers	Percentage	Cumulative percentage (CP-EX _{WC})
1	0.00	4	16.67	0
4	0.20	3	12.50	16.67
15	0.37	1	4.17	29.17
12	0.45	5	20.83	33.33
7	0.48	2	8.33	54.17
17	0.67	2	8.33	62.50
25	1.16	3	12.50	70.83
30	1.63	1	4.17	83.33
22	2.17	2	8.33	87.50
6	4.51	1	4.17	95.83
Total		24	100.00	

cumulative percentages—and it is based on an advantageous method of field normalization (see above). The definition of CP-EX_{WC} for a focal paper is that *x*% of papers published in the same year received a lower weighted citation impact than the focal paper. Weighted citation impact means that each citation of the focal paper is weighted by the citation behavior in its field. This definition is simple to understand, not only by bibliometric experts but also by laypersons.

As citation impact is dependent not only on the publication year but also on the document type of the cited publication (see, e.g., Lundberg, 2007), the CP-EX_{WC} calculation should not only be separated by publication year, but also by document type. In this study, it was not necessary to consider the document type in the calculation, because only articles were included.

5. METHODS

The bibliometric data used in this paper are from an in-house version of the WoS used at the Max Planck Society (Munich, Germany). In this study, all papers are included from this database with the document type “article” and published between 2011 and 2015. The data set contains *n* = 7,908,584 papers; for *n* = 914,472 papers no SNCS3 values are available in the in-house database. Thus, the study is based on *n* = 6,994,112 papers. The SNCS3 scores and CP-EX_{WC} values have been calculated as explained in the sections above. In the calculation of the SNCS3 indicator, we followed the procedure as explained by Waltman and van Eck (2013b). Whereas Waltman and van Eck (2013b), however, only included selected core journals from the WoS database in the SNCS3 calculation, the SNCS3 scores for the present study were calculated based on all journals in the WoS database.

6. RESULTS

Figure 1 shows the distribution of SNCS3 scores for 6 years using frequency distributions. It is clearly visible that the SNCS3 distributions are very skewed and characterized by outliers (articles with very high weighted citation impact).

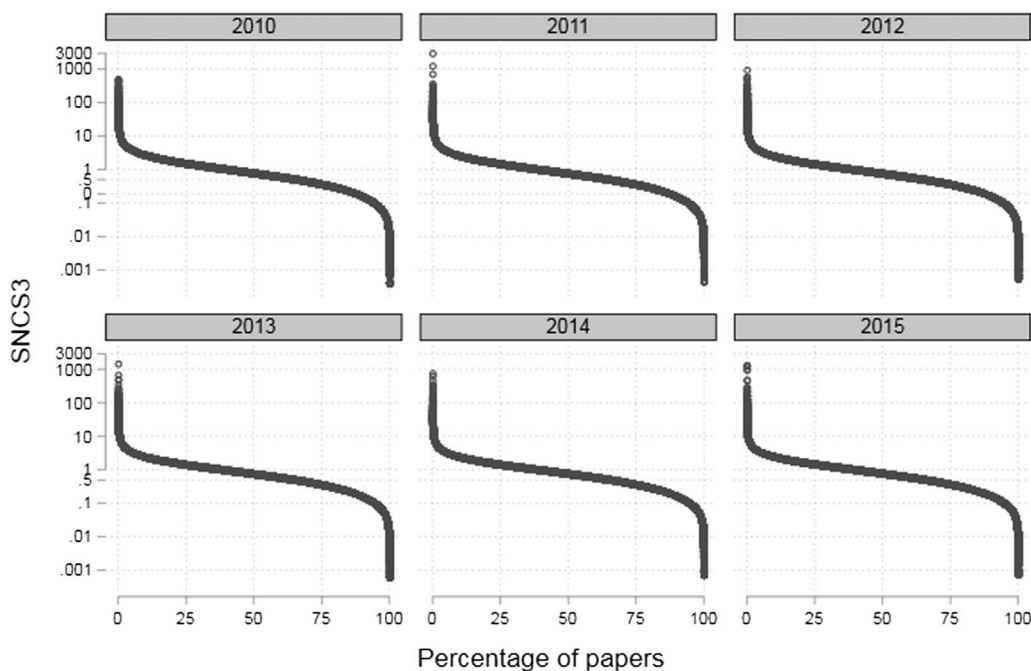


Figure 1. Quantile plots (Cox, 2005) of SNCS3 scores for papers published between 2011 and 2015.

Against the backdrop of these skewed distributions (despite citation weighting by citing-side normalization), it sounds reasonable (more than ever) to calculate percentiles based on SNCS3 scores. According to Seglen (1992), skewed citation distributions “should probably be regarded as the basic probability distribution of citations, reflecting both the wide range of citedness values potentially attainable and the low probability of achieving a high citation rate” (p. 632). This basic probability distribution does not appear to be valid only for citation distributions, but also weighted citation distributions (based on SNCS3). Similar to citations, the SNCS3 indicator appears to follow the so-called “bibliometric laws” (de Bellis, 2009, p. xxiv). This is a set of regularities working behind citation processes according to which a certain number of citations is related to the authors generating them (in their papers). The common feature of these processes (and similar processes based on the number of publications or text words) is an “amazingly steady tendency to the concentration of items on a relatively small stratum of sources” (de Bellis, 2009, p. xxiv).

One of these regularities leading to skewed citation distributions might be (larger) quality differences between the research published in the papers (Aksnes et al., 2019). A second regularity might be the type of contribution made by the paper: For example, one can expect many more citations for methods papers than for papers contributing empirical results (Bornmann, 2015; van Noorden, Maher, & Nuzzo, 2014). A third regularity might be a cumulative advantage effect by which “already frequently cited [publications] have a higher probability of receiving even more citations” (van Raan, 2019, p. 239). According to Ruocco, Daraio, et al. (2017), “Price’s [Derek J. de Solla Price] assumption was that the papers to be cited are chosen at random with a probability that is proportional to the number of citations those same papers already have. Thus, highly cited papers are likely to gain additional citations, giving rise to the rich get richer cumulative effect.”

Figure 2 shows the distribution of CP-EX_{WC} values for papers published between 2010 and 2015. Comparison of Figure 2 with Figure 1 reveals that the scores are no longer skewed.

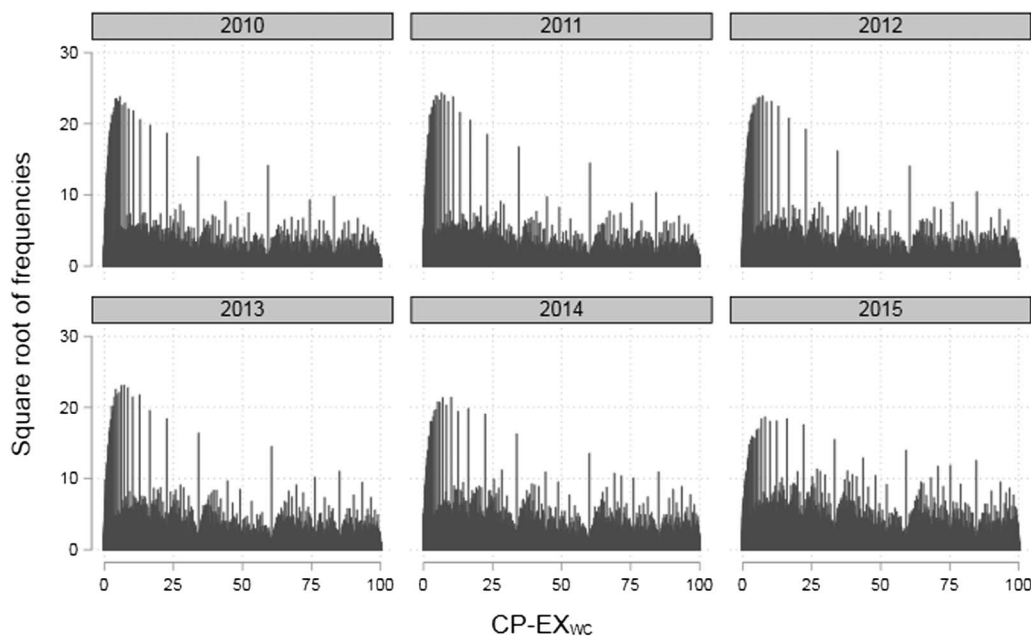


Figure 2. Distribution of CP-EX_{WC} values for papers published between 2010 and 2015.

Papers with low citation impact (i.e., low $CP-EX_{WC}$ scores) are prevalent, but the distributions approximate a uniform distribution.

In this study, the proposed indicator $CP-EX_{WC}$ has been exemplarily applied to publication and citation data of some countries: Switzerland, United Kingdom, United States, Germany, China, and Japan. The results are shown in Figure 3. The upper graph in the figure is based on full counting of the countries' papers. Thus, each paper contributes to the citation impact of a country with a weight of 1—independent of the additional number of countries involved. The score for a country shown in Figure 3 is its $CP-EX_{WC}$ median value. The dotted line in the graph marks the worldwide average. The score for Switzerland in the upper graph is above that line and means, for example, that on average, 60.85% of the papers worldwide have a weighted citation impact that is below the weighted citation impact of papers with a Swiss address.

The results in the upper graph correspond to results based on other (field-normalized) citation-based indicators (e.g., Bornmann & Leydesdorff, 2013; Bornmann, Wagner, & Leydesdorff, 2018; Stephen, Stahlschmidt, & Hinze, 2020). When citation impact is measured size independently, certain small countries such as Switzerland show an excellent performance (the Netherlands is another example, although it is not considered here). It follows the United Kingdom in the upper graph of Figure 3, which has exceeded the United States in citation impact in recent years. China and Japan are at the bottom of the country list. Although these results come as no real surprise, differences from previous results are also observable. One difference refers to the performance differences between the countries that do not appear to be very large. For example, the differences between Switzerland, the United Kingdom, and the United States exceed no more than four percentage points. Another difference from previous studies concerns the performance level. In previous studies, countries such as Switzerland show an excellent performance far away from midlevel performance. If we assume that the dotted line in Figure 3 represents a midlevel

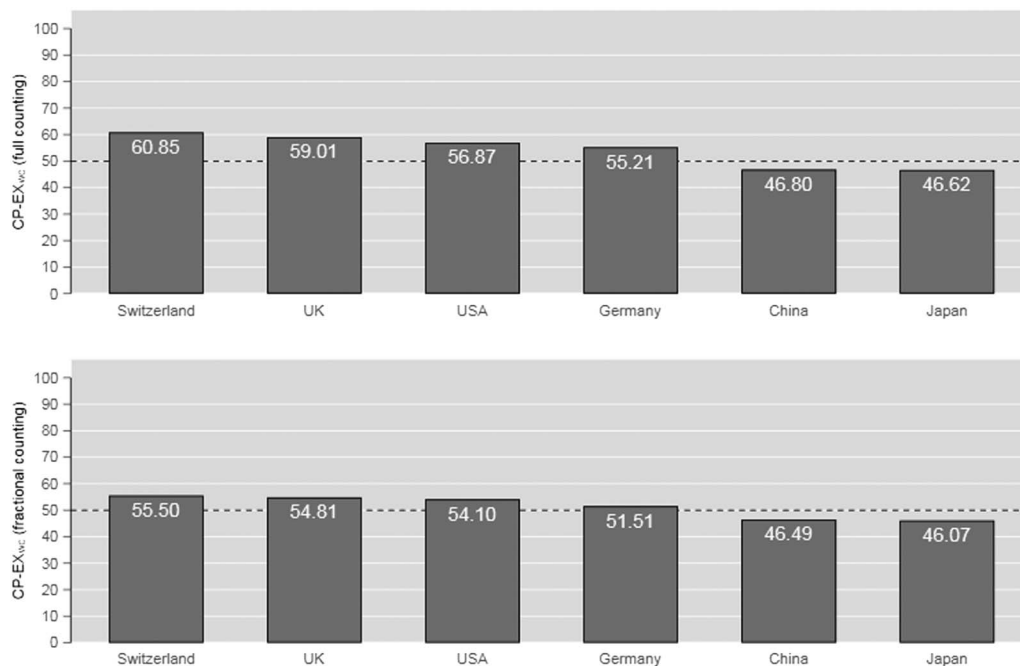


Figure 3. $CP-EX_{WC}$ for papers published between 2010 and 2015 by six countries: Switzerland ($n = 138,947$), United Kingdom ($n = 540,287$), United States ($n = 1,949,391$), Germany ($n = 510,207$), China ($n = 1,096,608$), and Japan ($n = 394,328$). The national numbers of papers are based on full counting.

performance (50% of the papers worldwide exhibit a lower performance), the best countries (and also the worst) are not far away from 50%. On average, for example, papers from Switzerland are (only) around 10 percentage points above the midlevel performance.

The lower graph in Figure 3 is based on fractional counting. Thus, it has been considered that many papers were published by more than one country. In this study (which is based on the SNCS3 impact indicator), the CP-EX_{WC} score for a paper has been weighted by the number of countries given on a paper (Bornmann & Williams, 2020).

The following formula leads to a fractionally counted mean CP-EX_{WC} score for a country:

$$mCPEX(F) = \left(\frac{(CPEX_1 * FR_1) + (CPEX_2 * FR_2) + \dots + (CPEX_y * FR_y)}{\sum_{i=1}^y FR_i} \right) \quad (4)$$

where CPEX₁ to CPEX_y are weighted by the number of countries given on a paper. For example, if a paper was published by authors from four countries, the paper is weighted by 0.25. The fractional assignment (weighting) is included by the notation FR_i for paper *i* = 1 to paper *y*. The sums of the CP-EX_{WC} scores for paper 1 to paper *y* published by the unit are divided by the sums of the weightings for paper 1 to paper *y*.

By applying fractional counting, citation impact benefits arising from collaborations are adjusted. As the results in the lower graph in Figure 3 show, fractional impact counting changes the national results differently: Whereas larger differences are visible for Switzerland, the United Kingdom, and Germany, the differences are smaller for Japan and China. Compared with the upper graph in Figure 3, China and Japan do not really profit from controlling international collaborations in the lower graph: The CP-EX_{WC} scores only change from 46.80% to 46.49% (China) and 46.62% to 46.07% (Japan). In contrast to China, Switzerland appears to profit significantly in terms of citation impact from international collaboration: Its CP-EX_{WC} decreases from 60.85% (upper graph) to 55.5% (lower graph). The other two countries that also appear to profit from international collaboration are the United Kingdom and Germany (around four percentage points).

7. DISCUSSION

Because only experts from the same field can properly assess the research of their colleagues, the peer review process is the dominant research evaluation method. Since around the 1980s, the use of indicators in research evaluation has become increasingly popular. One reason might be that “direct assessment of research activity needs expert judgment, which is costly and onerous, so proxy indicators based on metadata around research inputs and outputs are widely used” (Adams, Loach, & Szomszor, 2016, p. 2). For Lamont (2012), another reason is that “governments have turned to new public management tools to ensure greater efficacy, with the result that quantitative measures of performance and benchmarking are diffusing rapidly” (p. 202). However, peer review and the use of indicators do not have to be incompatible approaches; it is seen as the “ideal way” in research evaluation to combine both methods in the so-called informed peer review process. According to Waltman and van Eck (2016, p. 542), “scientometric indicators can ... be used by the peer review committee to complement the results of in-depth peer review with quantitative information, especially for scientific outputs that have not been evaluated in detail by the committee.” In the confrontation of peer review and bibliometrics, one should consider that both methods are related: “citations provide a built-in form of peer review” (McAllister et al., 1983, p. 205).

Citation analysis is one of the most important methods in bibliometrics, as the method appears to measure quality issues: “at high frequency, citations are good indicators of utility, significance,

even the notion of impact. The late sociologist of science, Robert Merton likened citations to repayments of intellectual debts. The normative process in science requires authors to acknowledge relevant previous contributions" (Panchal & Pendlebury, 2012, p. 1144). One of the major challenges of citation analyses is the field dependency of citations. If larger units in science are evaluated that are working in many fields, it is necessary to consider these differences in the statistical analyses (Bornmann, 2020). According to Kostoff (2002), "citation counts depend strongly on the specific technical discipline, or sub-discipline, being examined ... The documentation and citation culture can vary strongly by sub-discipline. Since citation counts can vary sharply across sub-disciplines, absolute counts have little meaning, especially in the absence of absolute citation count performance standards" (p. 53; see also Fok & Franses, 2007).

One solution to the problem of field-specific differences in citation counts is to contextualize the results of citation analyses "case by case, considering all the relevant information" (D'Agostino, Dardanoni, & Ricci, 2017, p. 826). According to Waltman and van Eck (2019), one can "use straightforward non-normalized indicators and to contextualize these indicators with additional information that enables evaluators to take into account the effect of field differences" (p. 295). This might be the best solution if smaller research groups or institutions working in clearly definable fields are evaluated. For this solution, however, it is necessary to involve not only a bibliometric expert in the evaluation but also an expert from the evaluated field to contextualize these indicators. For example, for the identification of research groups working in the same field as the focal group, it is necessary for an expert to identify these groups that can be used for comparison of the focal group. This solution of contextualizing the number of times when research is cited is stretched to its limits when large units such as organizations or countries are addressed in evaluations. These units are multidisciplinary by nature.

Since the 1980s, many different methods have been proposed to field-normalize citations. It has not been possible to establish a standard method until now. In this study, an approach is proposed that combines two previously introduced methods: citing-side normalization and percentiles. The advantage of combining two methods is that their advantages can be integrated into a single solution. Based on citing-side normalization, each citation is field weighted and, therefore, contextualized in its field. The most important advantage of citing-side normalization is that it is not necessary to work with a specific field categorization scheme. The disadvantages of citing-side normalization—the calculation is complex and the values elusive—can be compensated by calculating percentiles based on the field-weighted citations. On the one hand, percentiles are well understandable: It is the percentage of papers published in the same year with lower citation impact. On the other hand, weighted citation distributions are skewed distributions including outliers. Percentiles are well suited to assigning the position of a focal paper in such skewed distributions including a field-specific set of papers.

Many different approaches of percentile calculation exist (Bornmann, Leydesdorff, et al., 2013). According to Schreiber (2013, p. 829) "all the discussed methods have advantages and disadvantages. Further investigations are needed to clarify what the optimal solution to the problem of calculating percentiles and assigning papers to PRCs [percentile rank classes] might be, especially for large numbers of tied papers." Bornmann and Williams (2020) appear to have found a percentile solution with comparably good properties. In this study, their percentile approach based on weighted citations (CP-EX_{WC}) has been applied to the analysis of several countries. The country results are similar to many other published results. This correspondence in the results can be interpreted as a good sign for the new approach: It appears to measure field-normalized citation impact in a similar way to other indicators. However, the approach also reveals the importance of measuring citation impact based on fractional counting. Several countries are strongly internationally oriented, which has a larger influence on the results.

Further studies are necessary to investigate the new approach introduced here. These studies could also focus on other units than those considered in this study (e.g., institutions and research groups). Furthermore, it would be interesting to know how the new approach can be understood by people who are not bibliometric experts: Is it as easy to understand as expected, or are there difficulties in understanding it?

ACKNOWLEDGMENTS

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), and Arts and Humanities Citation Index (AHCI) prepared by Clarivate Analytics, formerly the IP & Science business of Thomson Reuters (Philadelphia, Pennsylvania, USA).

COMPETING INTERESTS

The author has no competing interests.

FUNDING INFORMATION

No funding has been received for this research.

DATA AVAILABILITY

The data cannot be made available in a data repository because the provider of the data (Clarivate Analytics) does not allow this.

REFERENCES

- Adams, J., Loach, T., & Szomszor, M. (2016). *Interdisciplinary research: Methodologies for identification and assessment. Do we know what we are measuring?* London, UK: Digital Science.
- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *Sage Open*, 9(1). DOI: <https://doi.org/10.1177/2158244019829575>
- Bornmann, L. (2015). Nature's top 100 revisited. *Journal of the Association For Information Science and Technology*, 66(10), 2166. DOI: <https://doi.org/10.1002/asi.23554>
- Bornmann, L. (2020). Bibliometrics-based decision trees (BBDTs) based on bibliometrics-based heuristics (BBHs): Visualized guidelines for the use of bibliometrics in research evaluation. *Quantitative Science Studies*, 1(1), 171–182. DOI: https://doi.org/10.1162/qss_a_00012
- Bornmann, L., Bowman, B. F., Bauer, J., Marx, W., Schier, H., & Palzenberger, M. (2014). Bibliometric standards for evaluating research institutes in the natural sciences. In B. Cronin & C. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 201–223). Cambridge, MA: MIT Press.
- Bornmann, L., & Daniel, H.-D. (2008). Selecting manuscripts for a high impact journal through peer review: A citation analysis of communications that were accepted by *Angewandte Chemie International Edition*, or rejected but published elsewhere. *Journal of the American Society for Information Science and Technology*, 59(11), 1841–1852. DOI: <https://doi.org/10.1002/asi.20901>
- Bornmann, L., & Leydesdorff, L. (2013). Macro-indicators of citation impacts of six prolific countries: InCites data and the statistical significance of trends. *PLOS ONE*, 8(2), e56768. DOI: <https://doi.org/10.1371/journal.pone.0056768>, PMID: 23418600, PMCID: PMC3572076
- 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. DOI: <https://doi.org/10.1016/j.joi.2012.10.001>
- Bornmann, L., & Marewski, J. N. (2019). Heuristics as conceptual lens for understanding and studying the usage of bibliometrics in research evaluation. *Scientometrics*, 120(2), 419–459. DOI: <https://doi.org/10.1007/s11192-019-03018-x>
- Bornmann, L., & Marx, W. (2014). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. *Scientometrics*, 98(1), 487–509. DOI: <https://doi.org/10.1007/s11192-013-1161-y>
- 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. DOI: <https://doi.org/10.1016/j.joi.2015.01.006>
- Bornmann, L., Marx, W., & Barth, A. (2013). The normalization of citation counts based on classification systems. *Publications*, 1(2), 78–86. DOI: <https://doi.org/10.3390/publications1020078>
- Bornmann, L., & Mutz, R. (2014). From P100 to P100': A new citation-rank approach. *Journal of the Association For Information Science*

- and Technology, 65(9), 1939–1943. DOI: <https://doi.org/10.1002/asi.23152>
- Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2011). Is interactive open access publishing able to identify high-impact submissions? A study on the predictive validity of *Atmospheric Chemistry and Physics* by using percentile rank classes. *Journal of the American Society for Information Science and Technology*, 62(1), 61–71. DOI: <https://doi.org/10.1002/asi.21418>
- Bornmann, L., Wagner, C., & Leydesdorff, L. (2018). The geography of references in elite articles: What countries contribute to the archives of knowledge. *PLOS ONE*, 13(3), e0194805. DOI: <https://doi.org/10.1371/journal.pone.0194805>, PMID: 29579088, PMCID: PMC5868817
- Bornmann, L., & Williams, R. (2020). An evaluation of percentile measures of citation impact, and a proposal for making them better. *Scientometrics*, 124, 1457–1478. DOI: <https://doi.org/10.1007/s11192-020-03512-7>
- Bornmann, L., & Wohlrabe, K. (2019). Normalisation of citation impact in economics. *Scientometrics*, 120(2), 841–884. DOI: <https://doi.org/10.1007/s11192-019-03140-w>
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404. DOI: <https://doi.org/10.1002/asi.21419>
- Colliander, C., & Ahlgren, P. (2011). The effects and their stability of field normalization baseline on relative performance with respect to citation impact: A case study of 20 natural science departments. *Journal of Informetrics*, 5(1), 101–113. DOI: <https://doi.org/10.1016/j.joi.2010.09.003>
- Colliander, C., & Ahlgren, P. (2019). Comparison of publication-level approaches to ex-post citation normalization. *Scientometrics*, 120(1), 283–300. DOI: <https://doi.org/10.1007/s11192-019-03121-z>
- Cox, N. J. (2005). Speaking Stata: The protean quantile plot. *Stata Journal*, 5(3), 442–460. DOI: <https://doi.org/10.1177/1536867X0500500312>
- Crespo, J. A., Li, Y. R., & Ruiz-Castillo, J. (2013). The measurement of the effect on citation inequality of differences in citation practices across scientific fields. *PLOS ONE*, 8(3). DOI: <https://doi.org/10.1371/journal.pone.0058727>, PMID: 23516542, PMCID: PMC3597729
- D’Agostino, M., Dardanoni, V., & Ricci, R. G. (2017). How to standardize (if you must). *Scientometrics*, 113(2), 825–843. DOI: <https://doi.org/10.1007/s11192-017-2495-7>
- de Bellis, N. (2009). *Bibliometrics and citation analysis: From the Science Citation Index to cybermetrics*. Lanham, MD: Scarecrow Press.
- Egghe, L. (2005). *Power laws in the information production process: Lotkian informetrics*. Kidlington, UK: Emerald Group Publishing Limited. DOI: [https://doi.org/10.1108/S1876-0562\(2005\)05](https://doi.org/10.1108/S1876-0562(2005)05)
- Fok, D., & Franses, P. H. (2007). Modeling the diffusion of scientific publications. *Journal of Econometrics*, 139(2), 376–390. DOI: <https://doi.org/10.1016/j.jeconom.2006.10.021>
- Gates, A. J., Ke, Q., Varol, O., & Barabasi, A. L. (2019). Nature’s reach: Narrow work has broad impact. *Nature*, 575(7781), 32–34. DOI: <https://doi.org/10.1038/d41586-019-03308-7>, PMID: 31695218
- Glänzel, W., & Moed, H. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, 96(1), 381–394. DOI: <https://doi.org/10.1007/s11192-012-0898-z>
- Hammarfelt, B., & Rushforth, A. D. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26(3), 169–180. DOI: <https://doi.org/10.1093/reseval/rvx018>
- Haunschild, R., Schier, H., Marx, W., & Bornmann, L. (2018). Algorithmically generated subject categories based on citation relations: An empirical micro study using papers on overall water splitting. *Journal of Informetrics*, 12(2), 436–447. DOI: <https://doi.org/10.1016/j.joi.2018.03.004>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. DOI: <https://doi.org/10.1038/520429a>, PMID: 25903611
- Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. DOI: <https://doi.org/10.1073/pnas.0507655102>, PMID: 16275915, PMCID: PMC1283832
- Hou, J., Pan, H. X., Guo, T., Lee, I., Kong, X. J., & Xia, F. (2019). Prediction methods and applications in the science of science: A survey. *Computer Science Review*, 34. DOI: <https://doi.org/10.1016/j.cosrev.2019.100197>
- Ioannidis, J. P. A., Boyack, K., & Wouters, P. F. (2016). Citation metrics: A primer on how (not) to normalize. *PLOS Biology*, 14(9), e1002542. DOI: <https://doi.org/10.1371/journal.pbio.1002542>, PMID: 27599158, PMCID: PMC5012555
- Jappe, A., Pithan, D., & Heinze, T. (2018). Does bibliometric research confer legitimacy to research assessment practice? A sociological study of reputational control, 1972–2016. *PLOS ONE*, 13(6), e0199031. DOI: <https://doi.org/10.1371/journal.pone.0199031>, PMID: 29902239, PMCID: PMC6002049
- Kostoff, R. N. (2002). Citation analysis of research performer quality. *Scientometrics*, 53(1), 49–71. DOI: <https://doi.org/10.1023/A:1014831920172>
- Kostoff, R. N., Murday, J. S., Lau, C. G. Y., & Tolles, W. M. (2006). The seminal literature of nanotechnology research. *Journal of Nanoparticle Research*, 8(2), 193–213. DOI: <https://doi.org/10.1007/s11051-005-9034-9>
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38(1), 201–221. DOI: <https://doi.org/10.1146/annurev-soc-070308-120022>
- Leydesdorff, L., & Milojević, S. (2015). The citation impact of German sociology journals: Some problems with the use of scientometric indicators in journal and research evaluations. *Soziale Welt*, 66(2), 193–204. DOI: <https://doi.org/10.5771/0038-6073-2015-2-193>
- Leydesdorff, L., Radicchi, F., Bornmann, L., Castellano, C., & de Nooy, W. (2013). Field-normalized impact factors (IFs): A comparison of rescaling and fractionally counted IFs. *Journal of the American Society for Information Science and Technology*, 64(11), 2299–2309. DOI: <https://doi.org/10.1002/asi.22911>
- Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2014). The European Union, China, and the United States in the top-1% and top-10% layers of most-frequently cited publications: Competition and collaborations. *Journal of Informetrics*, 8(3), 606–617. DOI: <https://doi.org/10.1016/j.joi.2014.05.002>
- Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics*, 1(2), 145–154. DOI: <https://doi.org/10.1016/j.joi.2006.09.007>
- Marx, W., & Bornmann, L. (2015). On the causes of subject-specific citation rates in Web of Science. *Scientometrics*, 102(2), 1823–1827. DOI: <https://doi.org/10.1007/s11192-014-1499-9>
- McAllister, P. R., Narin, F., & Corrigan, J. G. (1983). Programmatic evaluation and comparison based on standardized citation scores. *IEEE Transactions on Engineering Management*, 30(4), 205–211. DOI: <https://doi.org/10.1109/TEM.1983.6448622>
- Milojević, S. (2020). Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines.

- Quantitative Science Studies*, 1(1), 183–206. DOI: https://doi.org/10.1162/qss_a_00014
- Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, 246(1), 1–19. DOI: <https://doi.org/10.1016/j.ejor.2015.04.002>
- Moed, H. F. (2017). *Applied evaluative informetrics*. Heidelberg, Germany: Springer. DOI: <https://doi.org/10.1007/978-3-319-60522-7>
- Moed, H. F., & Halevi, G. (2015). The multidimensional assessment of scholarly research impact. *Journal of the American Society for Information Science and Technology*, 66(10), 1988–2002. DOI: <https://doi.org/10.1002/asi.23314>
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ: Computer Horizons.
- Nosek, B. A., Graham, J., Lindner, N. M., Kesebir, S., Hawkins, C. B., Hahn, C., ... Tenney, E. R. (2010). Cumulative and career-stage citation impact of social-personality psychology programs and their members. *Personality and Social Psychology Bulletin*, 36(10), 1283–1300. DOI: <https://doi.org/10.1177/0146167210378111>, PMID: 20668215
- Panchal, H., & Pendlebury, D. A. (2012). David A. Pendlebury. *Current Science*, 103(10), 1144–1145.
- Radicchi, F., & Castellano, C. (2011). Rescaling citations of publications in physics. *Physical Review E*, 83(4). DOI: <https://doi.org/10.1103/PhysRevE.83.046116>, PMID: 21599249
- 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: <https://doi.org/10.1016/j.joi.2014.11.010>
- Ruocco, G., Daraio, C., Folli, V., & Leonetti, M. (2017). Bibliometric indicators: The origin of their log-normal distribution and why they are not a reliable proxy for an individual scholar's talent. *Palgrave Communications*, 3, 17064. DOI: <https://doi.org/10.1057/palcomms.2017.64>
- Schreiber, M. (2013). How much do different ways of calculating percentiles influence the derived performance indicators? A case study. *Scientometrics*, 97(3), 821–829. DOI: <https://doi.org/10.1007/s11192-013-0984-x>
- Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9(5–6), 281–291. DOI: <https://doi.org/10.1007/BF02017249>
- Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628–638. DOI: [https://doi.org/10.1002/\(SICI\)1097-4571\(199210\)43:9<628::AID-ASIS>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASIS>3.0.CO;2-0)
- Shu, F., Julien, C.-A., Zhang, L., Qiu, J., Zhang, J., & Larivière, V. (2019). Comparing journal and paper level classifications of science. *Journal of Informetrics*, 13(1), 202–225. DOI: <https://doi.org/10.1016/j.joi.2018.12.005>
- Sjögårde, P., Ahlgren, P., & Waltman, L. (2020). Algorithmic labeling in hierarchical classifications of publications: Evaluation of bibliographic fields and term weighting approaches. Retrieved July 29, 2020, from <https://arxiv.org/abs/2004.08090>
- Smolinsky, L., & Lercher, A. (2012). Citation rates in mathematics: A study of variation by subdiscipline. *Scientometrics*, 91(3), 911–924. DOI: <https://doi.org/10.1007/s11192-012-0647-3>
- Stephen, D., Stahlschmidt, S., & Hinze, S. (2020). *Performance and structures of the German science system 2020. Studies on the German innovation system No. 5-2020*. Berlin, Germany: German Centre for Higher Education Research and Science Studies (DZHW).
- Strotmann, A., & Zhao, D. (2010). Combining commercial citation indexes and open-access bibliographic databases to delimit highly interdisciplinary research fields for citation analysis. *Journal of Informetrics*, 4(2), 194–200. DOI: <https://doi.org/10.1016/j.joi.2009.12.001>
- Sugimoto, C. R., & Weingart, S. (2015). The kaleidoscope of disciplinary. *Journal of Documentation*, 71(4), 775–794. DOI: <https://doi.org/10.1108/Jd-06-2014-0082>
- University of Waterloo Working Group on Bibliometrics. (2016). *White paper on bibliometrics, measuring research outputs through bibliometrics*. Waterloo, Canada.
- van Noorden, R., Maher, B., & Nuzzo, R. (2014). The top 100 papers. *Nature*, 514(7524), 550–553. DOI: <https://doi.org/10.1038/514550a>, PMID: 25355343
- van Raan, A. F. J. (2014). Advances in bibliometric analysis: Research performance assessment and science mapping. In W. Blockmans, L. Engwall, & D. Weaire (Eds.), *Bibliometrics: Use and Abuse in the Review of Research Performance* (pp. 17–28). London, UK: Portland Press.
- van Raan, A. F. J. (2019). Measuring science: Basic principles and application of advanced bibliometrics. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 237–280). Cham, Switzerland: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-02511-3_10
- Vernon, M. M., Balas, E. A., & Momani, S. (2018). Are university rankings useful to improve research? A systematic review. *PLOS ONE*, 13(3). DOI: <https://doi.org/10.1371/journal.pone.0193762>, PMID: 29513762, PMCID: PMC5841788
- Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics*, 10(3–4), 157–177. DOI: <https://doi.org/10.1007/BF02026039>
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. DOI: <https://doi.org/10.1016/j.joi.2016.02.007>
- Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12), 2378–2392. DOI: <https://doi.org/10.1002/asi.22748>
- Waltman, L., & van Eck, N. J. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, 96(3), 699–716. DOI: <https://doi.org/10.1007/s11192-012-0913-4>
- Waltman, L., & van Eck, N. J. (2013b). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics*, 7(4), 833–849. DOI: <https://doi.org/10.1016/j.joi.2013.08.002>
- Waltman, L., & van Eck, N. J. (2016). The need for contextualized scientometric analysis: An opinion paper. In I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, & R. Woolley (Eds.), *Proceedings of the 21st International Conference on Science and Technology Indicator* (pp. 541–549). València, Spain: Universitat Politècnica de València.
- Waltman, L., & van Eck, N. J. (2019). Field normalization of scientometric indicators. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 281–300). Heidelberg, Germany: Springer. DOI: https://doi.org/10.1007/978-3-030-02511-3_11
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., & Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics*, 7(2), 272–285. DOI: <https://doi.org/10.1016/j.joi.2012.11.011>
- 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. DOI: <https://doi.org/10.1016/j.joi.2010.08.001>
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), 851–872. DOI: <https://doi.org/10.1007/s11192-012-0775-9>
- Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, 10(2), 347–364. DOI: <https://doi.org/10.1016/j.joi.2016.02.003>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., ... Johnson, B. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. Bristol, UK: Higher Education Funding Council for England (HEFCE). DOI: <https://doi.org/10.4135/9781473978782>
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., & Franssen, T. (2015). *The metric tide: Literature review (supplementary report I to the independent review of the role of metrics in research assessment and management)*. London, UK: Higher Education Funding Council for England (HEFCE).
- Zitt, M., Ramanana-Rahary, S., & Bassecouard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics*, 63(2), 373–401. DOI: <https://doi.org/10.1007/s11192-005-0218-y>
- Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, 59(11), 1856–1860. DOI: <https://doi.org/10.1002/asi.20880>