



# Novel utilization of a paper-level classification system for the evaluation of journal impact: An update of the CAS Journal Ranking

Sichao Tong<sup>1,2</sup> , Fuyou Chen<sup>1</sup> , Liying Yang<sup>1</sup> , and Zhesi Shen<sup>1</sup> 

<sup>1</sup>National Science Library, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China

an open access  journal



Citation: Tong, S., Chen, F., Yang, L., & Shen, Z. (2023). Novel utilization of a paper-level classification system for the evaluation of journal impact: An update of the CAS Journal Ranking. *Quantitative Science Studies*, 4(4), 960–975. [https://doi.org/10.1162/qss\\_a\\_00270](https://doi.org/10.1162/qss_a_00270)

DOI:  
[https://doi.org/10.1162/qss\\_a\\_00270](https://doi.org/10.1162/qss_a_00270)

Peer Review:  
[https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss\\_a\\_00270](https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00270)

Supporting Information:  
[https://doi.org/10.1162/qss\\_a\\_00270](https://doi.org/10.1162/qss_a_00270)

Received: 15 February 2021  
Accepted: 3 August 2023

Corresponding Author:  
Zhesi Shen  
[shenzhs@mail.las.ac.cn](mailto:shenzhs@mail.las.ac.cn)

Handling Editor:  
Vincent Larivière

Copyright: © 2023 Sichao Tong, Fuyou Chen, Liying Yang, and Zhesi Shen. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



**Keywords:** CAS Journal Ranking, Citation Success Index, field normalization, journal impact factor, paper-level classification

## ABSTRACT

Since its first release in 2004, the CAS Journal Ranking, a ranking system of journals based on a citation impact indicator, has been widely used both in selecting journals when submitting manuscripts and in conducting research evaluation in China. This paper introduces an upgraded version of the CAS Journal Ranking released in 2020 and the corresponding improvements. We will discuss the following improvements: a) the CWTS paper-level classification system, a fine-grained classification system utilized for field normalization; b) the Field Normalized Citation Success Index (FNCSI), an indicator that is robust against not only extremely highly cited publications but also wrongly assigned document types; and c) document type difference. In addition, this paper will present part of the ranking results and an interpretation of the features of the FNCSI indicator.

## 1. INTRODUCTION

The CAS Journal Ranking<sup>1</sup>, an annually released journal ranking by the Center of Scientometrics (CoS), National Science Library of Chinese Academy of Sciences (CAS), is a journal ranking system widely recognized in China. It ranks journals included in Clarivate's Journal Citation Reports (JCR), based on citation indicators. In this paper, we will sketch its history and introduce the upgraded version of the CAS Journal Ranking.

In Section 1.1 we will briefly describe the history of the CAS Journal Ranking; then we will introduce the old version of the CAS Journal Ranking in Section 1.2. In Section 1.3 we point out its limitations and corresponding improvements. A literature review of normalized journal indicators is presented in Section 2. Data and the method used in the upgraded version of the CAS Journal Ranking are presented in Section 3. Ranking results and characteristics of the upgraded version are described in Section 4. In Section 5, we discuss the usage of the CAS Journal Ranking and some future directions.

<sup>1</sup> The website (<https://www.fenqubiao.com/>) is for registered institutional users. For individual researchers, scan the QR code on the website and follow the WeChat official account.

### 1.1. History of the CAS Journal Ranking

The origin of the CAS Journal Ranking is related to the misuse of the journal impact factor (JIF) in China. The idea of a JIF to measure journal impact was first proposed by Garfield and Sher (1963). Later, the JIF was gradually accepted in the sciences and some social sciences. In the late 1990s, the JIF was first introduced to China, and provided a reference for the Chinese research community, especially for junior researchers, to estimate the quality of journals indexed in the Web of Science (WoS). At that time, one feature of the JIF, namely that the JIF varies significantly across disciplines, was ignored by the Chinese research community and policymakers, resulting in JIFs from different disciplines being compared. Beginning in 2000, to promote the proper use of JIF in terms of heterogeneity across different disciplines, the CoS started to share the CAS Journal Ranking with some Chinese universities. In 2004, the first edition of the CAS Journal Ranking (the old CAS Journal Ranking) was officially released and continually updated until now. After years of developing the ranking method and indicators, the upgraded CAS Journal Ranking was released in January 2020.

The original intention of the CAS Journal Ranking was to provide a reference for researchers when selecting journals for publishing. Twenty years ago, the degree of internationalization in China was much lower than it is now. A considerable number of researchers, especially junior researchers, had only rare opportunities to participate in international communication, so it was hard for them to interpret journal impact metrics (e.g., JIF). After the first release, the CAS Journal Ranking was approved and became widespread in China. Later, in 2007, the JCR released its quartile ranking. A survey<sup>2</sup> among Chinese researchers conducted by Elsevier's STM Journals China Program team in 2021 shows that the CAS Journal Ranking is the most recommended journal list in China. Moreover, several CAS institutions and Chinese universities utilize the CAS Journal Ranking as one of the references to evaluate citation impact at the institutional level.

How to utilize quantitative data properly in research performance evaluation is a challenge for many countries, not just for China. Although the original intention of the CAS Journal Ranking is not relevant to this task, the CAS Journal Ranking has, inevitably, been involved in this controversy. This is a problem that the CAS Journal Ranking has to deal with. However, in this paper, we will not discuss research evaluation at large but focus on the improvement of the journal evaluation method from a methodological aspect.

### 1.2. The Old CAS Journal Ranking

The main idea of the CAS Journal Ranking is to group journals in each discipline category into tiers, following the idea of Jin and Wang (1999). The primary method of the old CAS Journal Ranking is as follows. First, WoS journals were grouped by discipline category. Second, the journals of each category are divided into four tiers based on the descending order of their indicators. Here we use the average of three recent JIFs to rank journals, i.e.,  $IF_Y = (JIF_Y + JIF_{Y-1} + JIF_{Y-2})/3$ .

The top 5% of journals are classified as Tier 1, and then the remaining journals are classified into Tiers 2 to 4 to make sure that the total impact in each tier is equal. The distribution of impact indicators for each category varies somewhat, so the fraction of journals in each tier is slightly different, yet they still have a rough pyramid-like structure (5%–20%–45%–100%). Journals within the same tier can be compared across disciplines. The idea of putting journals

<sup>2</sup> [https://www.elsevier.com/connect/behind-the-rising-influence-of-chinese-research?utm\\_source=twitter&utm\\_medium=social&utm\\_campaign=dt](https://www.elsevier.com/connect/behind-the-rising-influence-of-chinese-research?utm_source=twitter&utm_medium=social&utm_campaign=dt).

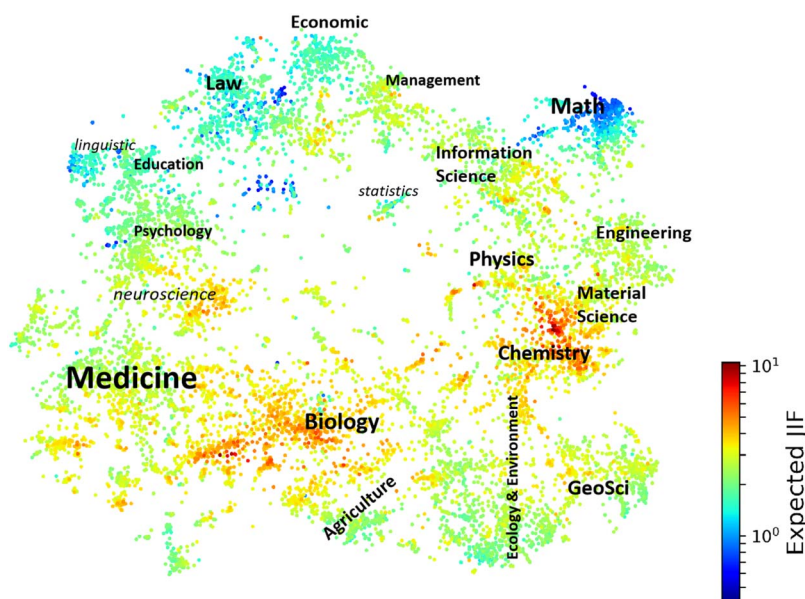
into tiers and comparing journals using tiers is adopted in the upgraded version of the CAS Journal Ranking.

### 1.3. Improvements in the Upgraded CAS Journal Ranking

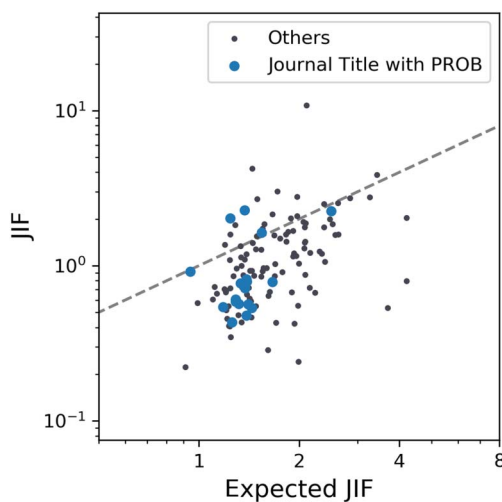
In the old CAS Journal Ranking, three significant limitations exist. One is related to the JIF. For a journal, because the citation distributions are skewed (Larivière, Kiermer et al., 2016; Milojević, Radicchi, & Bar-Ilan, 2017; Seglen, 1992, 1997), the JIF can be vastly affected by a tail of highly cited papers and may thus vary a lot across years. We previously utilized a 3-year average JIF to alleviate such fluctuations. However, this solution is still not robust enough against occasional highly cited papers and cannot accurately reveal the average impact of journals.

The second limitation is that the citation potential varies greatly among different document types (de Solla Price, 1965). Research articles generally have a lower citation potential than review papers. Therefore, journals with a higher proportion of reviews can attract more citations. Hence, it is unfair to compare journals with different proportions of reviews. In the old version, we tried to solve this problem by ranking research journals and review journals separately. However, many research journals also publish a relatively large percentage of review articles.

The third limitation is that the discipline categorization is not fine grained enough in the old CAS Journal Ranking. Initially, the CAS Journal Ranking adopted a 13-field categorization, including Medicine, Physics, Biology, etc. Indeed, citation differs significantly within fields (Figure 1). The JCR subject categories, a more fine-grained categorization, was adopted by the CAS Journal Ranking in 2008 to eliminate citation differences within these 13 fields. However, the problem still existed in the JCR subject categorization (van Eck, Waltman et al., 2013) (Figure 2). This limitation will reduce comparability for journals belonging to the same category.



**Figure 1.** Map of scientific journals with expected JIF. The color of each dot is related to the value of the corresponding journal's expected JIF: the darker the color of red/blue, the larger/smaller the value of the expected JIF.



**Figure 2.** Correlation of JIF and expected JIF for journals in the Statistics & Probability category. The dashed line is  $y = x$ . A dot positioned above the dashed line means that its actual JIF is larger than its expected JIF, and vice versa.

To illustrate the effect of the third limitation, we plotted a scatter map of journals from all fields (Figure 1), with each dot representing a journal and the color representing its potential citation.

For this map, we continue to use the same layout based on the journal citation network in earlier work (Shen, Chen et al., 2019). Moreover, we use the journal's expected JIF to indicate its potential citation. Regarding expected JIF, we also want to briefly introduce two basic definitions in Waltman's (2016) study, as below.

- Expected number of citations for a publication: the average number of citations for total publications from the related topic, the same year, and the same document type; here, we use article or review type.
- Moreover, for a journal, the expected JIF is the average of expected number of citations for the total publications of the journal.

See Section 3 for a more detailed formula for expected JIF. Figure 1 indicates a clear distinction in potential citations among research fields. We can see that citations differ between different areas within the medical fields and many other fields; for example, the upper part and the lower part of the Math category obviously perform differently.

We then take journals from the JCR category Statistics & Probability as an example. In Figure 2, each dot represents a journal, and we color journals titled with *probability* in blue. We can see that most blue dots have smaller expected JIF, indicating a difference in citation potential between journals from different topics within the Statistics & Probability category (e.g., probability-related journals have much smaller citation potential).

To overcome the limitations mentioned, we released the upgraded version of the CAS Journal Ranking. This version includes the following refinements:

- Instead of JIF, a new indicator, the Citation Success Index (CSI), was used. Compared with other citation indicators (e.g., the 3-year average JIF shown in earlier editions), CSI excels not only in robustness to ultrasmall numbers of extremely highly cited publications but also in robustness to wrongly assigned document types.

- In addition, we consider the document type when performing normalization (i.e., calculating the indicator of articles and reviews, respectively).
- The CWTS paper-level classification system, a more fine-grained system, has been utilized to classify each paper into the corresponding cluster (topic) (see Section 3) to ensure the CSI indicator's calculation from a paper level.

## 2. LITERATURE REVIEW

In this section, we will briefly review the development of normalized journal indicators and embed the CAS Journal Ranking in this developing timeline. The literature review is mainly based on the review by Waltman (2016). The normalization of journal indicators is mainly towards disciplinary differences, document types, and the skewness of citation distribution. In Figure 3, we present several representative works on normalized journal indicators related to the CAS Journal Ranking.

Considering the field difference of JIFs, Sen (1992) and Marshakova-Shaikovich (1996) proposed normalizing JIF using maximum JIF or a few of the highest JIFs in each subject. Van Leeuwen and Moed (2002) proposed the Journal to Field Impact Score considering the normalization of the document type, field, and citation window. Jin and Wang (1999) proposed putting journals in each category into three tiers of equal size and comparing journals across disciplines using tiers. Pudovkin and Garfield (2004) proposed the rank-normalized impact factor (mIF), which uses journals' relative positions ordered by JIF within each JCR category to facilitate comparison among fields. In the same year, the first version of the CAS Journal Ranking was released and proposed to group journals within categories based on JIF into four

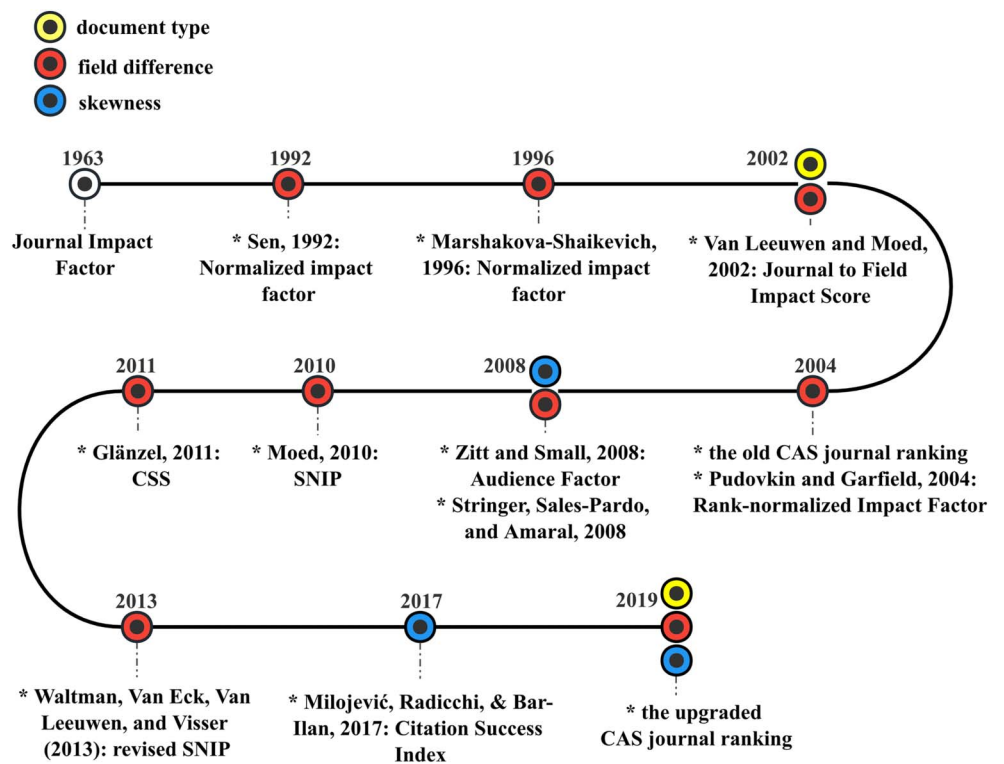


Figure 3. Timeline of normalized journal indicators with the CAS Journal Ranking embedded.

tiers and compare journals using tiers across fields. Glänzel (2011) proposed normalizing JIF using the parameters extracted from Characteristic Scores and Scales (CSS). Zitt and Small (2008) proposed the audience factor, which uses the journal-level citing-side (or source-side) normalization. Later, Moed (2010) presented the Source Normalized Impact per Paper (SNIP), and Waltman, van Eck et al. (2013) introduced a revised SNIP indicator. Toward the problem of skewness, researchers turn to the ranking-based indicators, such as percentile rank (Pudovkin & Garfield, 2009), percentile rank classes (Bornmann, Leydesdorff, & Mutz, 2013; Leydesdorff, Bornmann et al., 2011), the proportion of highly cited papers, indicators added in the Leiden Ranking (Waltman, Calero-Medina et al., 2012) and Clarivate's InCites<sup>3</sup>, in which relative ranking matters instead of the absolute citation value. Stringer, Sales-Pardo, and Amaral (2008) proposed the probability that a randomly selected paper published in one journal has received more citations than a randomly selected paper published in another journal to test the effectiveness of comparing journals using JIF. Later, Milojević et al. (2017) defined this probability as the Citation Success Index (CSI) and found an S-type relation between JIF and CSI. Shen, Yang, and Wu (2018) found that the relationship between JIF and CSI mainly results from the lognormal distribution of citations. Another way to alleviate the skewness problem is logarithmic conversion (Lundberg, 2007).

In the past, most colleagues used a journal-level classification system (e.g., WoS Subject Category) for normalization. However, the drawbacks of these classification systems for normalization have been revealed on many occasions. With the improvement of the accessibility of large-scale bibliometric data sets and an increase in computing power, data-based paper-level classification is constructed and used for normalization gradually (Waltman & van Eck, 2012, 2013a, 2013b). To address the problems of journal indicators and combining the recent advances in scientometric methods, the CAS Journal Ranking released its upgraded version in 2020.

### 3. METHOD AND DATA

We will take the 2019 version of the CAS Journal Ranking as an example to show the method and data, as well as the results.

#### 3.1. Journals and Citation Data

The CAS Journal Ranking (2019 version) includes the journals contained in Clarivate's Journal Citation Reports (JCR) (2019 version according to Clarivate<sup>4</sup>). For citation data, we use the Journal Impact Factor contributing items released by the JCR. It contains the citations in year  $Y$  for each article and review, published in years  $Y - 1$  and  $Y - 2$ , that counted towards the JIF.

#### 3.2. Paper-Level Classification Data

We utilized the results of CWTS paper-level classification, which means that each paper belongs to a certain cluster (topic). The article and review papers in the Web of Science Core Collection—Science Citation Index Expanded and Social Sciences Citation Index between 2000 and 2018—were collected. For the details of constructing the CWTS paper-level classification system, we refer to (Waltman & van Eck, 2012, 2013a, 2013b) for an exhaustive introduction to the classification methods for exploring the relatedness of publications and clustering them into groups. Depending on the granularity, this classification system consists

---

<sup>3</sup> <https://incites.clarivate.com/>.

<sup>4</sup> <https://jcr.clarivate.com/>.



of three levels—macro-, meso-, and microlevels. Here we use the microlevel with about 4,000 clusters.

It should be noted that in the released results for the CWTS paper-level classification system, trade journals and some local journals are excluded because their citation links are too weak. Because we are trying to include as many journals as possible, we retrieved the related records from the WoS for these excluded journals and put them back into the corresponding clusters based on the clusters of the retrieved related records using the majority rule. In total, the upgraded version includes 99% of articles and reviews indexed by JCR. From the perspective of journals, 98% of journals with more than 90% of their total publications are included.

### 3.3. Journal Ranking Indicators/Journal Indicators Utilized in This Article

In the upgraded version, we follow the idea of the Citation Success Index (CSI) and extend CSI to a field-normalized indicator. The original CSI, presented to compare the citation capacity between two journals (Milojević et al., 2017; Shen et al., 2018, 2019, 2023; Stringer et al., 2008), is defined as the probability of a randomly selected paper from one journal having more citations than a randomly selected paper from the other journal. Following the same idea, we propose the Field Normalized Citation Success Index (FNCSI). The FNCSI can be defined as the probability that the citation of a paper from a journal is larger than a random paper on the same topic and with the same document type from other journals. More details will be introduced below.

Before the upgraded version was completed, to investigate the differentiation between the new indicator (FNCSI) and the old indicator (JIF), we proposed the Field Normalized Impact Factor (FNIF). It should be noted that we did not use FNIF in the upgraded version; it is only used here for comparison.

#### 3.3.1. Field Normalized Citation Success Index (FNCSI)

For journal  $A$ , the probability  $P(c_a > c_o | a \in A, o \in O)$  that the citation of a paper from journal  $A$  is larger than a random paper on the same topic and with the same document type from other journals is defined as below:

$$S_A = P(c_a > c_o | a \in A, o \in O) = \sum_{t,d} P(A^{t,d}) P(c_a > c_o | a \in A^{t,d}, o \in O^{t,d})$$

where  $c_a$  is the citation count of one paper from journal  $A$ ,  $t \in \{\text{topic}_1, \text{topic}_2, \dots\}$ ,  $d \in \{\text{article, review}\}$ ,  $A^{t,d}$  represents the publications belonging to journal  $A$  in topic  $t$  with document type  $d$ , and  $O$  represents the publications from the other journals. For a specific research topic  $t$ , its FNCSI is calculated as below:

$$S_A^t = \frac{1}{N_{A^t}} \sum_d N_{A^t,d} \left[ \frac{\sum_{a \in A^{t,d}, o \in O^{t,d}} 1(c_a > c_o) + \sum_{a \in A^{t,d}, o \in O^{t,d}} 0.5(c_a = c_o)}{N_{A^t,d} N_{O^t,d}} \right]$$

where weight 1 is for  $c_a > c_o$  and weight 0.5 for ties (i.e.,  $c_a = c_o$ ). Journal  $A$  usually involves several research topics at the microlevel of the CWTS paper-level classification system, so that the total FNCSI of Journal  $A$  can be summed from its affiliated topics as below:

$$S_A = \frac{1}{N_A} \sum_t N_{A^t} S_A^t$$

where  $N_{A^t}$  is the number of publications of journal  $A$  in topic  $t$ . The FNCSI is based on a two-year citation window (but this window can be adapted).

For a better understanding of FNCSI, examples of FNCSI formulation represented in an analogy way are presented in Supplementary material A.

### 3.3.2. Field Normalized Impact Factor (FNIF)

The Field Normalized Impact Factor (FNIF) uses the same classification system as FNCSI but uses the average citations based on a normalization approach (i.e., each citation is normalized by the average citation of papers in the same topic cluster and with the same document type). For instance, the FNIF of journal *A* is defined as

$$F_A = \frac{\sum_{t,d} \sum_{a \in A^{t,d}} c_a / \mu_{t,d}}{N_A}$$

where  $\mu_{t,d}$  is the average citation of papers in topic *t* with document type *d*. By comparing the results of FNCSI and FNIF, we can see the advantages of CSI.

### 3.3.3. Expected JIF

As we mentioned earlier, for each journal, we use the expected JIF as an indicator of citation potential:

$$E_A = \frac{\sum_{t,d} \mu_{t,d} N_{A^{t,d}}}{N_A}$$

where  $\mu_{t,d}$  is the average citation of papers in topic *t* with document type *d*.

## 4. RESULTS

### 4.1. Ranking Results

This section presents the results of the CAS Journal Ranking based on FNCSI and the comparisons with other indicators<sup>5</sup>. Table 1 shows the top 20 journals ranked according to FNCSI. Here we only list journals mainly publishing research articles. This list is dominated by *Nature*-titled journals, *Lancet*-titled journals, and *Cell*-titled journals. The top five journals are well acknowledged in the natural and life sciences. The other journals belong to different fields and do not concentrate on a single field or narrow group of fields.

The corresponding rankings based on FNIF values of these top 20 journals are also presented in Table 1. Among these journals, the rankings of *Cancer Cell*, *Nature Neuroscience*, *Cell Metabolism*, and *Nature Immunology* are boosted most by the FNCSI indicator; they all climb more than 20 positions. Only *Lancet Oncology* shows a slight drop in position from the FNCSI indicator. Overall, journals from medical-related categories have a relatively big gap between these indicators. In Supplementary material Table C1, we present the top 20 journals according to FNCSI and FNIF, respectively.

The correlation among relative rankings by FNCSI and FNIF is shown in Figure 4, with values closer to 0 representing better-ranked journals. We can see that FNCSI and FNIF are highly correlated (Spearman correlation: 0.98, *p* value: 0.00). In the lower part of Figure 4, we highlight several journals that have worse rankings in FNCSI compared with FNIF. These journals share a common property in that they each have one or several highly cited papers, and a majority of poorly cited papers. For example, *Chinese Phys C* has one paper cited more than

<sup>5</sup> All journals' FNCSI results for 2019 version are available at <https://doi.org/10.57760/sciencedb.08419>.



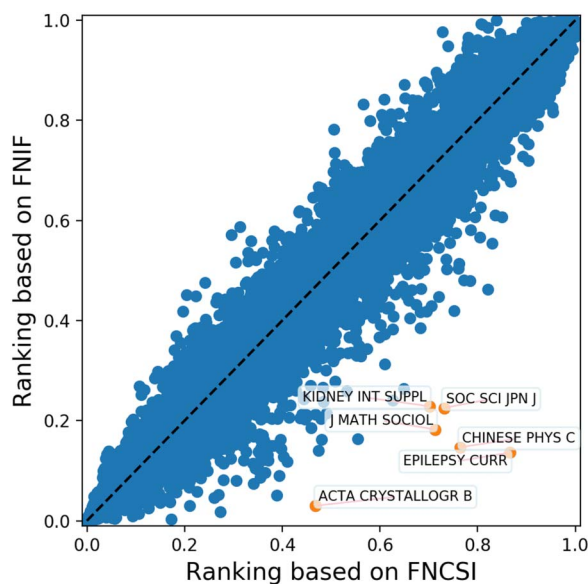
**Table 1.** Top 20 ranked journals according to FNCSI

Journal	Category-WoS	FNCSI	FNIF
<i>Lancet</i>	MEDICINE, GENERAL & INTERNAL	1	3
<i>Nature</i>	MULTIDISCIPLINARY SCIENCES	2	5
<i>JAMA</i>	MEDICINE, GENERAL & INTERNAL	3	4
<i>Science</i>	MULTIDISCIPLINARY SCIENCES	4	9
<i>Cell</i>	BIOCHEMISTRY & MOLECULAR BIOLOGY/CELL BIOLOGY	5	15
<i>World Psychiatry</i>	PSYCHIATRY	6	8
<i>Lancet Neurology</i>	CLINICAL NEUROLOGY	7	11
<i>Nature Photonics</i>	OPTICS/PHYSICS, APPLIED	8	17
<i>Nature Genetics</i>	GENETICS & HEREDITY	9	13
<i>Nature Medicine</i>	BIOCHEMISTRY & MOLECULAR BIOLOGY/CELL BIOLOGY/MEDICINE, RESEARCH & EXPERIMENTAL	10	21
<i>Nature Materials</i>	MATERIALS SCIENCE, MULTIDISCIPLINARY/CHEMISTRY, PHYSICAL/PHYSICS, APPLIED/PHYSICS, CONDENSED MATTER	11	12
<i>Lancet Oncology</i>	ONCOLOGY	12	10
<i>Cancer Cell</i>	ONCOLOGY/CELL BIOLOGY	13	38
<i>Nature Chemistry</i>	CHEMISTRY, MULTIDISCIPLINARY	14	31
<i>Nature Neuroscience</i>	NEUROSCIENCES	15	36
<i>Cell Metabolism</i>	CELL BIOLOGY/ENDOCRINOLOGY & METABOLISM	16	51
<i>Lancet Respiratory Medicine</i>	CRITICAL CARE MEDICINE/RESPIRATORY SYSTEM	17	22
<i>Nature Immunology</i>	IMMUNOLOGY	18	58
<i>Lancet Diabetes &amp; Endocrinology</i>	ENDOCRINOLOGY & METABOLISM	19	27
<i>Nature Nanotechnology</i>	NANOSCIENCE & NANOTECHNOLOGY/MATERIALS SCIENCE, MULTIDISCIPLINARY	20	23

2,000 times, but approximately 70% of papers are not cited<sup>6</sup>. This result is consistent with the difference in definition between FNCSI and FNIF.

In Section 1.3, we discussed the difference in citation potential among journals within the Statistics & Probability category, as shown in Figure 2. The paper-level classification system eliminates the effect of citation potential on journal rankings. In Table 2, we present the top 20 journals (mainly publishing research articles) according to FNCSI in this category. We can see that journals perform weakly in JIF due to low citation potential (low rank based on expected JIF) revealed reasonably by FNCSI, such as several well-acknowledged journals such as *Annals of Statistics*, *Annals of Probability*, and *Biometrika*.

<sup>6</sup> <https://jcr.clarivate.com/>.



**Figure 4.** Correlation of relative rankings based on FNCSI and FNIF; the smaller the better.

## 4.2. Robustness

### 4.2.1. Robust against extremely highly cited publications

The robustness of an indicator represents its sensitivity to changes in the set of publications based on which it is calculated. A robust indicator will not change much against the occasional ultrasmall number of highly cited publications. To measure the robustness of an indicator, we construct several sets of publications for each journal with the bootstrapping method and recalculate the indicators and rankings accordingly. For instance, for a journal with  $N$  publications, we randomly select  $N$  publications with replacement, calculate these indicators, and get a new ranking for each journal. We simulate this procedure 100 times and thus obtain 100 rankings for each journal. Figure 5A shows the distribution of the obtained rankings of *Chinese Physics: C*. We can see that the ranking range of FNCSI is much smaller than FNIF. The citation distribution of *Chinese Physics: C* is highly skewed, with one paper cited about 2,000 times and about 70% of papers not cited. Thus, FNIF depends strongly on whether this highly cited paper is included in the calculation or not.

To get an overview of the indicators' robustness, we calculate the relative ranking change for these indicators. The relative change in ranking is defined as

$$\Delta = \frac{1}{N} \sum_j \frac{\max\{R_j\} - \min\{R_j\}}{\text{avg}\{R_j\}}$$

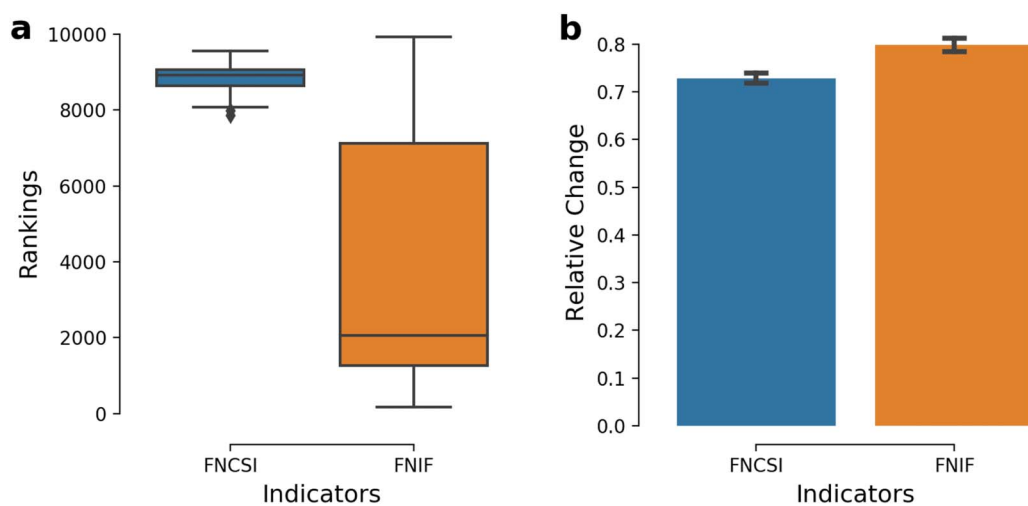
where  $\{R_j\}$  are the rankings of journal  $j$  obtained from the above simulation. As shown in Figure 5B, the relative change of FNCSI is smaller than FNIF, implying that FNCSI is more robust than FNIF. FNCSI mainly focuses on the central tendency of the citation distribution and is not easily affected by occasional highly cited papers, which implies that FNCSI can reflect the journal impact more stably.

### 4.2.2. Robust against wrongly labeled document type

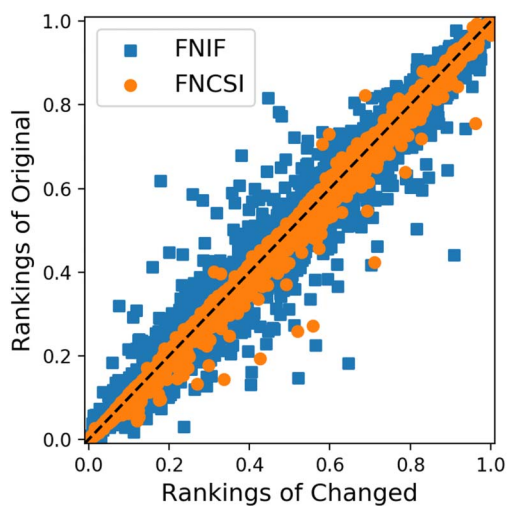
When conducting normalization for article/review, we need to assign the proper document type for each paper. Previous studies have shown that the document types assigned by WoS

**Table 2.** Top 20 ranked journals in the Statistics and Probability category according to FNCSI. Their rankings based on Expected JIF and JIF are also presented

Journal	Rank-FNCSI	Rank-expected JIF	Rank-JIF
<i>Econometrica</i>	1	69	2
<i>Journal of the Royal Statistical Society B</i>	2	45	3
<i>Annals of Statistics</i>	3	63	7
<i>Probability Theory and Related Fields</i>	4	86	10
<i>Annals of Probability</i>	5	103	15
<i>Finance and Stochastics</i>	6	99	22
<i>Journal of the American Statistical Association</i>	7	32	4
<i>International Statistical Review</i>	8	39	16
<i>Journal of Quality Technology</i>	9	104	29
<i>Journal of Statistical Software</i>	10	20	1
<i>Annals of Applied Probability</i>	11	60	28
<i>Stochastic Environmental Research and Risk Assessment</i>	12	7	8
<i>British Journal of Mathematical and Statistical Psychology</i>	13	9	20
<i>Technometrics</i>	14	56	21
<i>Biometrika</i>	15	49	33
<i>Bayesian Analysis</i>	16	44	35
<i>Bernoulli</i>	17	92	41
<i>Insurance: Mathematics and Economics</i>	18	65	43
<i>Extremes</i>	19	58	25
<i>Econometric Theory</i>	20	91	52



**Figure 5.** (A) Ranking variability of *Chinese Physics: C* for FNCSI and FNIF. (B) Average relative change of rankings based on FNCSI and FNIF.



**Figure 6.** Robustness against document type for FNCSI and FNIF.

(here for articles and reviews) are not fully correct. For example, some review papers are assigned as articles, and some articles are assigned as reviews (Colebunders & Rousseau, 2013; Donner, 2017; Harzing, 2013; Yeung, 2019; Zhu, Shen et al., 2022). To test the sensitivity of indicators against wrongly labeled document types, here we generate a virtual data set.

(Theoretically, we can randomly select  $N$  publications with replacement: Article to Review or Review to Article. In practice, extreme cases were chosen to reveal this error's impact better. For each journal, we turn the document type of its most highly cited paper to the opposite (i.e., Article to Review or Review to Article).)

Then, we recalculate the journal indicators and obtain the new rankings based on FNCSI and FNIF, respectively. The comparison of rankings based on this changed data with the original rankings is shown in Figure 6. We can see that almost all the orange dots (FNCSI-based) are located close to the diagonal line, but the blue squares (FNIF-based) spread much more broadly, which implies that rankings based on FNCSI are more robust against wrongly labeled document type than rankings based on FNIF. From the point of fault tolerance, FNCSI performs better than FNIF.

## 5. CONCLUSION AND DISCUSSION

In this paper, we first reviewed the history of the CAS Journal Ranking. We also listed the main limitations in earlier editions of the CAS Journal Ranking. For example, the old factor (the 3-year average JIF) is not robust enough against occasional highly cited papers; and the old discipline categories are not fine grained enough to achieve a more sophisticated field normalization. To improve these deficiencies, in 2019, two critical improvements were adopted in the upgraded version: The CWTS paper-level classification system is utilized to replace the JCR subject category; and FNCSI was introduced to replace JIF. Furthermore, a comparison of FNCSI and FNIF showed that FNCSI successfully addressed the issues mentioned above to a large extent.

The measures we applied in the upgraded version of the CAS Journal Ranking offer a novel idea to evaluate journal impact. In this study, the reported results demonstrate that our measures are effective overall. We also received some positive feedback (received through internal

communication via emails, telephone, social media, and interviews) from Chinese researchers. For example, some researchers stated that in the upgraded version, the ranking of journals is more similar to their subjective feelings, and some “good” journals from basic research fields with low potential citations are revealed in the upgraded version.

### 5.1. Future Research of CAS Journal Ranking

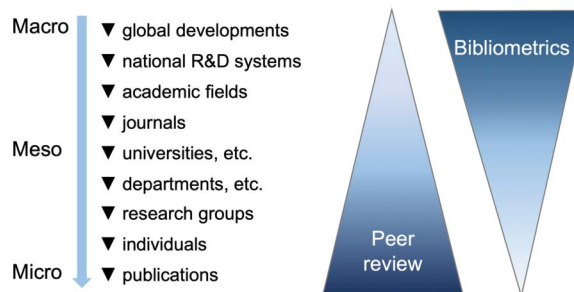
As far as future research is concerned, we propose the following:

- Eliminate the influence on the differentiated distribution of the FNCSI score among disciplines. FNCSI is a normalized indicator by the discipline of CWTS paper-level classification. Theoretically, the scores of FNCSI can be compared across different disciplines. However, based on the observation of the FNCSI score among disciplines, the distribution of the FNCSI score is differentiated among disciplines (see [Supplementary material E](#)). For example, in this situation, when the journals are divided into different disciplines, their rankings will be different. However, the basis and criteria for assigning journals to different disciplines are imprecise. Therefore, next, we will investigate how to eliminate the influence of this problem.
- Determine the optimal number of tiers. As mentioned in the introduction, for both the earlier and upgraded version, we adopted four tiers. The initial motivation for using four tiers is empirical: the more the number of tiers, the lower the distinction among different tiers and vice versa. Further, we found that having four tiers leads to a better distinction in some disciplines. After adopting FNCSI, we started to rethink whether using four tiers is the best option. Why isn't it three or five tiers? In particular, for the FNCSI score, optimal solutions for different disciplines may be different. The essential requirement for comparing disciplines is that they all have the same number of tiers. How to balance the trade-offs and find the optimal solution?
- Explore the influence of the paper-level classification system and FNCSI, respectively. There are two main measures adopted in the upgraded CAS Journal Ranking. A fine-grained paper-level classification system and the new field-normalized indicator FNCSI were utilized on the upgraded CAS Journal Ranking. Theoretically, both the classification system and the indicator will influence the ranking results. It is valuable to detect the respective consequences of the two factors and to compare which one has the most significant effect. In further studies, we will explore this point and investigate the other properties of the CAS Journal Ranking, such as the robustness towards covidization (Liu, Yang, & Shen, 2023; Liu, Zhang et al., 2023; Zhang, Liu et al., 2022).

In addition, in this paper, although we did not focus on discussing the issue of research performance evaluation and how to use journal ranking properly, we are on the way to correcting the improper use of the CAS Journal Rankings. One measure that has been taken is extending journal rankings to journal profiles based on CWTS paper-level classification and FNCSI, which will provide comprehensive information rather than only metrics.

### 5.2. Discussion of the Use of the CAS Journal Ranking

The CAS Journal Ranking, a fully bibliometric ranking system, should be used following the common principles of bibliometrics (e.g., the compatibility of objective data and peer review across different levels of granularity). [Figure 7](#) illustrates the consensus among scholars in the realm of research evaluation. When evaluating meso- and macrolevel entities such as research institutions and disciplinary domains, bibliometric indicators, such as citation impact-based



**Figure 7.** Suitability of bibliometrics and peer review at different levels (based on Hinze, 2014). Accordingly, the use of the CAS Journal Ranking can be aggregated into different levels.

journal evaluation, provide more reliable information. Conversely, when evaluating microlevel entities, such as individual researchers, peer review should play a leading role and bibliometrics should play a supporting role.

- **At the macro- or mesolevel**

- i. Comparing the research performance of countries or institutions based on the statistics of rankings or tiers of the journals published, such as the role played by the *Nature* index. A better journal ranking system will provide a more accurate macro- or mesolevel estimation of performance.
- ii. Facilitating librarians to select which journals to subscribe to. A better journal ranking system may help librarians to better allocate resources for subscriptions.
- iii. Facilitating researchers to select journals for paper reading or manuscript submission. Researchers can start their literature survey from papers published in high-ranking journals and trace the citation flow then. This usage is especially for junior researchers or graduates who are not fully familiar with all the journals in their areas.

- **At the microlevel**

- i. Evaluating researchers (e.g., for promotion, rewards, grant funding) based on the rankings of the journals they used for their publications (e.g., using the number or fraction of highly ranking journals, using the total score transformed from the ranking of journals) (Quan, Chen, & Shu, 2017).
- ii. Evaluating the quality or impact of papers based on the rankings of the journals they are published in (e.g., selecting the best paper in an area).

These microlevel direct utilizations of the CAS Journal Ranking are not recommended. In recent years, China's Ministry of Education and Ministry of Science and Technology have released policies to implement reforms to encourage more qualitative evaluation of research. Conducting high-quality peer review requires a set of supporting measures, such as establishing clear guidelines and accountability on peer review, providing education and training on high-quality peer review, and educating on the proper use of quantitative metrics to become bibliometric-wise. In addition, the path to responsible evaluation requires efforts by all stakeholders.

#### ACKNOWLEDGMENTS

We thank the generous help from CWTS (The Center for Science and Technology Studies of Leiden University, Netherlands) and Dr. Nees Jan van Eck for providing the paper-level classification data, and thank Ms. M. Zhu and Dr. Ronald Rousseau for valuable discussion. We



would also like to thank the editor and anonymous reviewers for their helpful and constructive comments on the manuscript.

#### AUTHOR CONTRIBUTIONS

Sichao Tong: Formal analysis, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing. Fuyou Chen: Data curation, Writing—review & editing. Liying Yang: Conceptualization, Methodology, Supervision, Writing—review & editing. Zhesi Shen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing.

#### COMPETING INTERESTS

The authors have no competing interests.

#### FUNDING INFORMATION

This study was partly funded by The National Social Science Foundation of China “Research on Semantic Evaluation System of Scientific Literature Driven by Big Data” (Grant No. 21&ZD329).

#### DATA AVAILABILITY

All journals’ FNCSI results for 2019 version are available at <https://doi.org/10.57760/sciencedb.08419>.

#### REFERENCES

- 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. <https://doi.org/10.1016/j.joi.2012.10.001>
- Colebunders, R., & Rousseau, R. (2013). On the definition of a review, and does it matter? In *14th International Society of Scientometrics and Informatics Conference* (pp. 2072–2074). Vienna, Austria.
- de Solla Price, D. J. (1965). Networks of scientific papers. *Science*, 149(3683), 510–515. <https://doi.org/10.1126/science.149.3683.510>, PubMed: 14325149
- Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. *Scientometrics*, 113(1), 219–236. <https://doi.org/10.1007/s11192-017-2483-y>
- Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, 14(3), 195–201. <https://doi.org/10.1002/asi.5090140304>
- Glänzel, W. (2011). The application of characteristic scores and scales to the evaluation and ranking of scientific journals. *Journal of Information Science*, 37(1), 40–48. <https://doi.org/10.1177/0165551510392316>
- Harzing, A.-W. (2013). Document categories in the ISI Web of Knowledge: Misunderstanding the social sciences? *Scientometrics*, 94(1), 23–34. <https://doi.org/10.1007/s11192-012-0738-1>
- Hinze, S. (2014). *Bibliometrics between ambition and responsibility*. Reykjavik, Iceland: Nordic Workshop on Bibliometrics and Research Policy.
- Jin, B., & Wang, S. (1999). The division of SCI journal grades and the distribution of Chinese papers [in Chinese]. *Scientific Research Management*, 20(2), 1–7. <https://doi.org/10.3969/j.issn.1000-2995.1999.02.001>
- Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., ... Curry, S. (2016). A simple proposal for the publication of journal citation distributions. *bioRxiv*. <https://doi.org/10.1101/062109>
- Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables on 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. <https://doi.org/10.1002/asi.21534>
- Liu, Y., Yang, L., & Shen, Z. (2023). The robustness analysis of CAS Journal Ranking against “Covidization” of research [preprint]. *27<sup>th</sup> International Conference on Science, Technology and Innovation Indicators (STI 2023)*.
- Liu, Y., Zhang, J., Yang, L., & Shen, Z. (2023). Contribution of papers on COVID-19 to journal impact factor [in Chinese]. *Chinese Journal of Scientific and Technical Periodicals*, 34(6), 775–785.
- Lundberg, J. (2007). Lifting the crown-citation z-score. *Journal of Informetrics*, 1(2), 145–154. <https://doi.org/10.1016/j.joi.2006.09.007>
- Marshakova-Shaikovich, I. (1996). The standard impact factor as an evaluation tool of science fields and scientific journals. *Scientometrics*, 35(2), 283–290. <https://doi.org/10.1007/BF02018487>
- Milojević, S., Radicchi, F., & Bar-Ilan, J. (2017). Citation success index – An intuitive pair-wise journal comparison metric. *Journal of Informetrics*, 11(1), 223–231. <https://doi.org/10.1016/j.joi.2016.12.006>
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265–277. <https://doi.org/10.1016/j.joi.2010.01.002>
- Pudovkin, A. I., & Garfield, E. (2004). Rank-normalized impact factor: A way to compare journal performance across subject categories. *Proceedings of the American Society for Information*

- Science and Technology*, 41(1), 507–515. <https://doi.org/10.1002/meet.1450410159>
- Pudovkin, A. I., & Garfield, E. (2009). Percentile rank and author superiority indexes for evaluating individual journal articles and the author's overall citation performance. *Collnet Journal of Scientometrics and Information Management*, 3(2), 3–10. <https://doi.org/10.1080/09737766.2009.10700871>
- Quan, W., Chen, B., & Shu, F. (2017). Publish or impoverish: An investigation of the monetary reward system of science in China (1999–2016). *Aslib Journal of Information Management*, 69, 486–502. <https://doi.org/10.1108/AJIM-01-2017-0014>
- Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628–638. [https://doi.org/10.1002/\(SICI\)1097-4571\(199210\)43:9<628::AID-ASI5>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0)
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079), 498–502. <https://doi.org/10.1136/bmj.314.7079.497>, PubMed: 9056804
- Sen, B. K. (1992). Normalised impact factor. *Journal of Documentation*, 48(3), 318–325. <https://doi.org/10.1108/eb026901>
- Shen, Z., Yang, L., Di, Z., & Wu, J. (2019). Large enough sample size to rank two groups of data reliably according to their means. *Scientometrics*, 118, 653–671. <https://doi.org/10.1007/s11192-018-2995-0>
- Shen, Z., Yang, L., & Wu, J. (2018). Lognormal distribution of citation counts is the reason for the relation between impact factors and Citation Success Index. *Journal of Informetrics*, 12(1), 153–157. <https://doi.org/10.1016/j.joi.2017.12.007>
- Shen, Z., Chen, F., Yang, L., & Wu, J. (2019). Node2vec representation for clustering journals and as a possible measure of diversity. *Journal of Data and Information Science*, 4(2), 79–92. <https://doi.org/10.2478/jdis-2019-0010>
- Shen, Z., Yang, L. & Wu, J. (2023). Two indicators rule them all: Mean and standard deviation used to calculate other journal indicators based on log-normal distribution of citation counts. *Data Science and Informetrics*, 3(2), 30–39. <https://doi.org/10.59494/DSI.2023.2.3>
- Stringer, M. J., Sales-Pardo, M., & Amaral, L. A. N. (2008). Effectiveness of journal ranking schemes as a tool for locating information. *PLOS ONE*, 3(2), e1683. <https://doi.org/10.1371/journal.pone.0001683>, PubMed: 18301760
- van Eck, N. J., Waltman, L., van Raan, A. F. J., Klautz, R. J. M., & Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLOS ONE*, 8(4), e62395. <https://doi.org/10.1371/journal.pone.0062395>, PubMed: 23638064
- Van Leeuwen, T. N., & Moed, H. F. (2002). Development and application of journal impact measures in the Dutch science system. *Scientometrics*, 53(2), 249–266. <https://doi.org/10.1023/A:1014808709694>
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. <https://doi.org/10.1016/j.joi.2016.02.007>
- Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., ... Wouters, P. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419–2432. <https://doi.org/10.1002/asi.22708>
- 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. <https://doi.org/10.1002/asi.22748>
- Waltman, L., & van Eck, N. J. (2013a). A smart local moving algorithm for large-scale modularity-based community detection. *European Physical Journal B*, 86(11). <https://doi.org/10.1140/epjb/e2013-40829-0>
- 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. <https://doi.org/10.1016/j.joi.2013.08.002>
- 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. <https://doi.org/10.1016/j.joi.2012.11.011>
- Yeung, A. W. K. (2019). Comparison between Scopus, Web of Science, PubMed and publishers for mislabelled review papers. *Current Science*, 116(11), 1909–1914. <https://doi.org/10.18520/cs/v116/i11/1909-1914>
- Zhang, J., Liu, Y., Yang, L., & Shen, Z. (2022). Covid-related papers contribute 50% on the JIF of high impact medicine journals. *Journal of Data and Information Science*, 7(4), 1–2. <https://doi.org/10.2478/jdis-2022-0020>
- Zhu, M., Shen, Z., Chen, F., & Yang, L. (2022). The influence of review's document type marking on the results of research evaluation [in Chinese]. *Science Focus*, 17(5), 59–67. <https://doi.org/10.15978/j.cnki.1673-5668.202205005>
- 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. <https://doi.org/10.1002/asi.20880>