# An Empirical Study of Scientific Evaluation Based on the Correlation Between Field-Normalized Citation Impact and Peer Review: Postprint

**Authors:** Song Liping, Wang Jianfang

**Date:** 2023-08-27T00:00:00+00:00

## Abstract

[**Purpose/Significance**] To analyze the feasibility of field-normalized citation impact in scientific evaluation and its correlation with peer review, providing insights for responsible metrics and peer review supported thereby.

[**Method/Process**] The F1000 and InCites platforms were selected to conduct correlation analysis between CNCI (Category Normalized Citation Impact) and citation frequency for 29,850 cell biology articles and 30,326 biotechnology articles, and Spearman correlation coefficient test was performed between CNCI and F1000 scores for 956 cell biology papers.

[**Results/Conclusion**] The results demonstrate that from a statistical perspective, CNCI exhibits a highly positive correlation with citation frequency and a significant positive correlation with F1000 scores, though instances of divergence between the two also exist. Therefore, CNCI can to a certain extent reflect peer review outcomes, serve as a proxy for attributing academic impact, and facilitate cross-disciplinary comparison; however, relying exclusively on either peer review or CNCI as scientific evaluation criteria would be biased, and peer review underpinned by new-generation responsible metrics indicators exemplified by CNCI will become the mainstream approach in future scientific evaluation.

## Full Text

## Preamble

**Vol. 62 No. 18 September 2018**
*ChinaXiv Cooperative Journal*

ChinaRxiv 中X

# An Empirical Study of Scientific Evaluation Based on the Correlation Between Category Normalized Citation Impact and Peer Review

**SONG Liping**[1] **WANG Jianfang**[2]
[1] School of Management, Tianjin Normal University, Tianjin 300387
[2] Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190

---

## Abstract

[**Purpose/Significance**] This study analyzes the feasibility of using Category Normalized Citation Impact (CNCI) in scientific evaluation and its correlation with peer review, aiming to provide insights for responsible metrics and peer review informed by such metrics. [**Method/Process**] Using the F1000 and InCites platforms, we conducted correlation analyses between CNCI and citation counts for 29,850 cell biology papers and 30,326 biotechnology papers, and performed Spearman rank correlation tests between CNCI and F1000 scores for 956 cell biology papers. [**Result/Conclusion**] The results demonstrate that CNCI is highly positively correlated with citation counts and significantly positively correlated with F1000 scores from a statistical perspective, though contradictory cases between the two indicators also exist. Therefore, CNCI can partially reflect peer review outcomes, substitute for the function of attributing academic impact, and enable cross-disciplinary comparisons. However, relying solely on either peer review or CNCI as the sole criterion for scientific evaluation would be biased. Peer review supported by a new generation of responsible metrics represented by CNCI will become the mainstream of future scientific evaluation.

**Classification Number:** G250
**Keywords:** peer review; responsible metrics; F1000; InCites

---

On December 18, 2014, the results of the UK's Research Excellence Framework (REF) were announced. In July 2015, "The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management" was published, authored by British scientometricians, university administrators, and public policy experts. The report proposed principles of responsible metrics and informed peer review, specifically noting that scientific evaluation is severely troubled by disciplinary factors and recommended that raw citation frequencies should not be directly compared across disciplines. The Leiden Manifesto similarly argues that citation patterns differ significantly across disciplines—for instance, mathematics journals have a maximum impact factor of 3, while cell biology journals reach as high as 30. Given these differences in citation density, normalizing citation indicators for different disciplines, publication years, and document types has become a crucial issue in scientometrics.

In 1986, T. Braun and A. Schubert published a paper in *Scientometrics* titled "Evaluating Research Performance and Citation Impact Across Different Scientific Fields," which proposed establishing relative indicators for cross-disciplinary comparison, marking the beginning of exploration and practice in normalized indicators. In 2010, H.F. Moed proposed the Source Normalized Impact per Paper (SNIP), which corrected for disciplinary differences in journal citation behavior from the perspective of average citation counts per paper, drawing on the concepts of citation potential and standardization to enable direct comparison between journals in different fields. In recent years, as UK higher education has shifted from peer review-based RAE to metrics-based REF, a "metrics tide" has swept the globe, transforming scientometrics from a subfield of information science into an important tool for science policy and research management. Consequently, Thomson Reuters (now Clarivate Analytics) introduced CNCI in its InCites research evaluation platform to address the need for cross-disciplinary comparison of academic impact.

When a paper is classified into multiple disciplinary fields, the average of the ratios of actual to expected citations across all assigned fields is used. The calculation formula is:

$$CNCI = \frac{c}{e}$$

where $e$ represents the expected citation rate or baseline value, $c$ represents total citation count, $f$ represents disciplinary field, $t$ represents publication year, $d$ represents document type, and $n$ represents the number of disciplinary fields to which the paper is assigned.

If CNCI equals 1, the paper's citation performance is equivalent to the global average; if greater than 1, it exceeds the global average; if less than 1, it falls below the global average. In this sense, CNCI eliminates the influence of publication year, disciplinary field, and document type, making it a practical impact indicator.

F1000 (Faculty of 1000), the peer review platform used for comparison in this study, is a post-publication peer review platform in biomedicine founded in 2002 by V. Tracz based on the concept of a virtual team of thousands of scientists. In 2015, while lamenting the "five sins" of traditional scientific publishing, Tracz integrated F1000 into three unique services: F1000Prime, F1000Research, and F1000Posters, combining open-access journals, open data, open peer review, and open science to ensure timely dissemination of research findings. Among these, F1000Prime is currently the world's only platform that provides online recommendations for the most important research papers based on peer review by over 8,000 top scientists in biology and medicine, with the F1000 Score (FS) representing expert evaluation scores. Therefore, F1000 serves not only as a retrieval tool but also as an important scientific evaluation instrument. This study uses FS as the benchmark to explore the applicability of CNCI and similar new-generation metrics in scientific evaluation through correlation analysis between FS and corresponding paper CNCI values.

## 1. Background

Citation impact is central to scientometrics and scientific evaluation, yet different disciplines, publication years, and document types severely constrain its measurement. E. Garfield noted that raw citation counts across different disciplines cannot be directly compared. The Leiden Manifesto points out that disciplinary differences in citation are significant. In 1986, T. Braun and A. Schubert proposed establishing relative indicators for cross-disciplinary comparison in their paper "Evaluating Research Performance and Citation Impact Across Different Scientific Fields" published in *Scientometrics*, initiating exploration in normalized indicators. In 2010, H.F. Moed proposed SNIP, which corrected for disciplinary differences in journal citation behavior by considering average citation counts, citation potential, and standardization, enabling direct comparison between journals across disciplines. In recent years, as UK higher education transitioned from peer review-based RAE to metrics-based REF, the "metrics tide" has swept the globe, transforming scientometrics from an information science subfield into a crucial tool for science policy and research management. Consequently, Thomson Reuters (now Clarivate Analytics) launched CNCI in its InCites platform to meet the need for cross-disciplinary academic impact comparison.

## 2. Data Collection and Statistics

As a new-generation research analysis platform, the InCites database is built upon the seven core citation indexes of Thomson Reuters' Web of Science. CNCI for a paper is defined as the actual citation count divided by the expected citation count for papers of the same document type, publication year, and disciplinary field. The formula is:

$$CNCI = \frac{c}{e}$$

where $e$ = expected citation rate or baseline, $c$ = total citations, $f$ = disciplinary field, $t$ = publication year, $d$ = document type, and $n$ = number of disciplinary fields assigned.

Platform searches in InCites reveal significant differences in citation frequencies across 252 Web of Science disciplines, ranging from 754,184 citations in materials science to 51 citations in poetry. Figures 1 [Figure 1: see original paper] and 3 [Figure 3: see original paper] compare citation frequencies and CNCI across disciplines for papers and reviews published in 2014, while Figures 2 [Figure 2: see original paper] and 4 [Figure 4: see original paper] show average citations per paper and CNCI differences across disciplines, with the x-axis representing disciplines and the y-axis representing citation counts or CNCI values. These results demonstrate the necessity of introducing CNCI in scientific evaluation practice, as scholars such as L. Bornmann and L. Leydesdorff have extensively analyzed and validated.

The year 2014 was selected as the statistical year to ensure sufficient citation accumulation. Data collection proceeded as follows: (1) **Determine the sample**: Among nine disciplinary classification systems provided by InCites, we selected the commonly used Web of Science classification and identified cell biology and biotechnology as our target fields by mapping Web of Science's 252 categories to F1000Prime's 40 research areas. (2) **Search in InCites**: Using cell biology as an example, we formulated the search query as `field=cell biology and article publication date=2014`, retrieving 26,137 papers and 3,713 reviews in cell biology, and 27,660 papers and 2,666 reviews in biotechnology. We then obtained each paper's DOI, title, authors, citation count, CNCI, disciplinary percentile, journal normalized citation impact, and journal impact factor. (3) **Search in F1000Prime**: Using the same cell biology example, we searched with `article publication date=2014 and faculties=cell biology`, retrieving 3,334 cell biology papers and 2,225 biotechnology papers, along with each paper's title, DOI, and FS. The search was conducted on June 5, 2017. (4) **Match datasets**: Using Excel functions `if(countif(), "S", "F")` and VLOOKUP, we identified papers with matching DOIs and extracted their CNCI and FS values, yielding 956 cell biology papers, 20 cell biology reviews, and 91 biotechnology papers for further analysis (biotechnology reviews were too few for statistical significance and were omitted).

---

## 3. Data Processing

Using SPSS 22.0, we first performed K-S tests for normality to determine data distribution and select appropriate statistical methods. The significance level was less than 0.05, indicating non-normal distribution, thus non-parametric Spearman tests were appropriate.

### 3.1 Correlation Analysis Between Citation Count and CNCI

Spearman correlation results for 26,137 cell biology papers (biotechnology results omitted) are shown in Table 1 , where DCF represents citation frequency.

**Table 1. Correlation Between CNCI and Citation Frequency for Cell Biology Papers**

| Spearman's rho | DCF | CNCI |
|---|---|---|
| DCF Correlation Coefficient | 1.000 | .989** |
| Significance (2-tailed) | | .000 |
| CNCI Correlation Coefficient | .989** | 1.000 |

| Spearman's rho | DCF | CNCI |
|---|---|---|
| Significance (2-tailed) | .000 | |

**Note:** ** Correlation is significant at the 0.01 level (2-tailed).

For 3,713 cell biology reviews (2,666 biotechnology reviews omitted), Spearman correlation results are shown in Table 2 .

**Table 2. Correlation Between CNCI and Citation Frequency for Cell Biology Reviews**

| Spearman's rho | DCF | CNCI |
|---|---|---|
| DCF Correlation Coefficient | 1.000 | .990** |
| Significance (2-tailed) | | |
| CNCI Correlation Coefficient | .990** | 1.000 |
| Significance (2-tailed) | | |

**Note:** ** Correlation is significant at the 0.01 level (2-tailed).

### 3.2 Correlation Analysis Between CNCI and FS for Individual Papers

Using cell biology papers as an example, Spearman correlation results for 956 papers are shown in Tables 3 and 4 .

**Table 3. Correlation Between CNCI and FS for Cell Biology Papers**

| Spearman's rho | DCF | CNCI | FS |
|---|---|---|---|
| DCF Correlation Coefficient | 1.000 | .990** | .364** |
| Significance (2-tailed) | | .000 | .000 |
| CNCI Correlation Coefficient | .990** | 1.000 | .362** |
| Significance (2-tailed) | .000 | | .000 |
| FS Correlation Coefficient | .364** | .362** | 1.000 |
| Significance (2-tailed) | .000 | .000 | |

**Note:** ** Correlation is significant at the 0.01 level (2-tailed).

**Table 4. Correlation Between CNCI and FS for 97 High-Scoring Cell Biology Papers**

| Spearman's rho | CNCI | FS |
|---|---|---|
| CNCI Correlation Coefficient | 1.000 | .163** |
| Significance (2-tailed) | | .000 |

| Spearman's rho | CNCI | FS |
|---|---|---|
| FS Correlation Coefficient | .163** | 1.000 |
| Significance (2-tailed) | .000 | |

**Note:** ** Correlation is significant at the 0.01 level (2-tailed).

---

# 4. Results and Discussion

## 4.1 Statistical Test Results

Based on the criterion that $|r| \geq 0.8$ indicates high correlation, $0.5 \leq |r| < 0.8$ moderate correlation, $0.3 \leq |r| < 0.5$ low correlation, and $|r| < 0.3$ no correlation, the correlation analysis yields the following findings:

1. **High correlation between CNCI and citation counts**: Whether for cell biology or biotechnology, papers or reviews, CNCI shows high positive correlation with citation counts, with correlation coefficients of 0.989, 0.990, 0.993, and 0.992 respectively.
2. **Significant but low correlation between FS and CNCI**: Using FS as the benchmark, both 956 cell biology papers and 91 biotechnology papers show significant positive correlation between FS and CNCI, with coefficients of 0.362 and 0.443 respectively, indicating general consistency in their trends. However, the low correlation coefficients suggest that discrepancies exist between the two variables—some papers receive high expert ratings but low citation counts/CNCI, and vice versa.
3. **Divergence in high-impact papers**: Given the low correlation between FS and CNCI, we further analyzed 97 cell biology papers with FS $\geq$ 5. The correlation decreased from 0.362 to 0.163, shifting from low to no correlation, indicating that contradictions become more pronounced among high-impact papers. For instance, among 956 papers, the paper ranked second by FS was 611th by CNCI, while the top CNCI paper was 420th by FS, and the second-ranked CNCI paper was 666th by FS.

## 4.2 Analysis of Results

1. **CNCI's role and limitations**: CNCI demonstrates certain advantages in cross-disciplinary comparison and can substitute for citation analysis in information filtering and academic impact attribution. However, given its sensitivity to disciplinary classification, optimizing its normalization algorithm (Mean Normalised Citation Score) to better accommodate the skewed distribution of citations across disciplines without altering data properties remains an open research question. In this study, the matching rate between F1000 and InCites for cell biology papers was only 3.65%,

partly due to different selection principles: Web of Science is known for highly selective journal curation, while F1000 operates on the principle that "good papers can come from anywhere." This also indicates that disciplinary classification significantly affects CNCI. To address this, InCites provides nine classification systems including Web of Science, ESI, Frascati, UK RAE, and Australian ERA. The implementation of REF 2014 pushed CNCI from theoretical exploration to practical application, making disciplinary evaluation a priority. In China, scholars such as Wu Yishan have long proposed similar comparative principles for research evaluation, namely classified management and classified evaluation. The 2015 "Implementation Plan for Deepening Reform of the Science and Technology System" and the 2016 "Guidelines for Evaluating Philosophy and Social Science Research in Higher Education Institutions" both advocate for establishing classified evaluation systems, demonstrating that classified evaluation has become a consensus in the scientific community worldwide. However, classification systems can interfere with CNCI calculations, prompting metrologists to both caution about selecting appropriate classification systems and actively explore CNCI variants. Algorithms developed by scholars such as L. Waltman, L. Bornmann, L. Leydesdorff, H.F. Moed, and M. Thelwall, and the Relative Citation Ratio (RCR) proposed by S. Haustein, represent attempts in this direction.

2. **Combining peer review and metrics**: Given the subjectivity and "Matthew effect" in peer review, citation-based indicators can assist peer review to achieve deeper public participation, democratize research evaluation, and improve peer review quality while reducing bias and risk. However, how to effectively combine them remains an exploratory question. CNCI represents one such attempt, meeting the need for responsible metrics and disciplinary evaluation, partially reflecting peer review results, and thus providing strong support for peer review.

3. **Integration as the future**: The organic combination of peer review and CNCI concretizes the principles of the Leiden Manifesto and responsible metrics. The Leiden Manifesto's overarching principle is to maintain qualitative evaluation (represented by FS) as primary, while disciplinary normalization has become one of its ten core principles for scientific evaluation. Fundamentally, the main purpose of scientific evaluation is to identify, promote, and support excellence, with scientometrics providing a series of indicators and methods centered on this goal. CNCI represents a scientific evaluation indicator built on the principle of responsible metrics, ensuring the implementation of metrics-informed peer review. In this sense, peer review supported by a new generation of responsible metrics represented by CNCI will become the mainstream approach in future scientific evaluation, guaranteeing the practical implementation of metrics-based scientific assessment.

**Conclusions:** 1. CNCI offers advantages for cross-disciplinary comparison and can substitute for citation analysis in information filtering and impact attribution. However, given its sensitivity to disciplinary classification, optimizing its normalization algorithm to better accommodate skewed citation distributions across disciplines is a key research question. 2. In scientific evaluation practice, citation-based indicators like CNCI can assist peer review, democratize evaluation, and reduce bias. CNCI partially reflects peer review results and can support peer review effectively. 3. The organic integration of CNCI and peer review represents the future mainstream of scientific evaluation, ensuring the practical implementation of responsible metrics.

---

## References


[1] WILSDON J. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* [M]. London: SAGE Publishing, 2016: 137-158.

[2] GARFIELD E. *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities* [M]. New York: John Wiley & Sons, 1979.

[3] HICKS D, WOUTERS P, WALTMAN L. Bibliometrics: The Leiden Manifesto for Research Metrics [J]. *Nature*, 2015, 520(7548): 429-431.

[4] HAUNSCHILD R, BORNMANN L. Normalization of Mendeley reader counts for impact assessment [J]. *Journal of Informetrics*, 2016, 10(1): 62-73.

[5] TRACZ V, LAWRENCE R. Towards an open science publishing platform [EB/OL]. [2018-07-20]. https://f1000researchdata.s3.amazonaws.com/manuscripts/8575/0c0c20d6-8920-4d1c-b564-e81901b268f4_{7968}___{{{rebecca}}_{{lawrence}}}.pdf?doi=10.12688/f1000research.7968.

[6] InCites [EB/OL]. [2018-04-17]. https://clarivate.com.cn/products/incites/.

[7] LEYDESDORFF L, RADICCHI F, BORNMANN L, et al. Field-normalized impact factors (IFs): A comparison of rescaled and fractionally counted IFs [J]. *Journal of the American Society for Information Science & Technology*, 2013, 64(11): 2299-2309.

[8] BORNMANN L, HAUNSCHILD R. Measuring field-normalized impact of papers on specific societal groups: An altmetrics study based on Mendeley data [J]. *Research Evaluation*, 2016, 26(3): 10-35.

[9] InCites [EB/OL]. [2017-06-03]. https://incites.thomsonreuters.com/.

[10] F1000 [EB/OL]. [2017-06-05]. http://f1000.com/prime.

[11] BORNMANN L, HAUNSCHILD R. Relative Citation Ratio (RCR): An empirical attempt to study a new field-normalized bibliometric indicator [J]. *Journal of the Association for Information Science & Technology*, 2017, 68(4): 1064-1067.

[12] DERRICK GE, PAVONE V. Democratising research evaluation: Achieving greater public engagement with bibliometrics-informed peer review [J]. *Science & Public Policy*, 2013, 40(5): 563-575.

## Author Contributions

**Song Liping**: Conceptualized the research, collected data, and wrote the main manuscript.
**Wang Jianfang**: Collected relevant literature and performed statistical analysis.

## English Title and Abstract

### An Empirical Study of Scientific Evaluation Based on the Correlation Between CNCI and Peer Review

**Abstract:** [Purpose/significance] In order to make some recommendations for responsible metrics and peer review, the paper analyses the feasibility of Category Normalized Citation Impact (CNCI) and its correlation with peer review. [Method/process] By choosing F1000 and InCites, this paper conducted the correlation analysis between CNCI and citation counts of 29,850 cell biology papers and 30,326 biotechnology papers, and performed the test of Spearman rank correlation coefficient between the CNCI and F1000 scores of 956 cell biology papers. [Result/conclusion] The results show that CNCI is positively correlated with citation counts, and significantly correlated with F1000 rating, but there are contradictory situations for the two indicators at the same time. Therefore, CNCI can reflect the results of peer review, substitute for the indication of academic influence, and also can be used for comparison among disciplines. However, peer review or CNCI will be biased for scientific evaluation purpose alone. Then, tools that link responsible metrics indicators, such as CNCI and informed peer reviews would be valuable additions to the field of research evaluation.

**Keywords:** peer review; responsible metrics; F1000; InCites

*Note: Figure translations are in progress. See original paper for figures.*