Postprint: Effectiveness Analysis of Citation Evaluation Methods for Research Field Identification Using Co-citation Networks

Authors: Song Liping, Wang Jianfang, Fu Jie, Yuan Shanshan

Date: 2023-04-01T16:02:59+00:00

Abstract

[Purpose/Significance] To investigate paper-level academic standardized citation impact indicators, this study analyzes the effectiveness of Relative Citation Ratio (RCR), a citation evaluation method that identifies research fields based on co-citation networks. [Method/Process] Using iCite, F1000, and InCites databases, with 739,604 documents as samples, R language was employed to conduct correlation analysis between RCR and Category Normalized Citation Impact (CNCI) based on a priori classification, Journal Impact Factor (JIF), etc., to examine the validity of RCR in academic impact evaluation. [Results/Conclusion] RCR shows high correlation with CNCI and weak correlation with JIF, thereby demonstrating both similarities and differences with CNCI and revealing its operability in implementing the San Francisco Declaration and moving away from "evaluating papers based on journals". The inverse relationship between the correlation of RCR and JIF and the number of interdisciplinary fields in cross-disciplinary science, along with the finding that the correlation between RCR and JIF is lower than that in single disciplines, indicates that RCR possesses certain applicability and effectiveness in evaluating the academic impact of interdisciplinary papers. Therefore, RCR's algorithm for identifying research fields through co-citation networks provides a novel approach for paper impact evaluation, particularly for assessing interdisciplinary scientific research achievements.

Full Text

Preamble

Volume 65, Issue 23, December 2021
ChinaXiv Collaborative Journal

Analysis of the Effectiveness of Citation Evaluation Methods for Identifying Research Fields Through Co-Citation Networks

Song Liping[1], **Wang Jianfang**[2], **Fu Jie**[1], **Yuan Shanshan**[1]
[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 effectiveness of Relative Citation Ratio (RCR), a citation evaluation method that identifies research fields through co-citation networks, to explore field-normalized reference influence indicators at the paper level. [**Method/Process**] Using iCite, F1000, and InCites databases, we examined 739,604 articles and employed R language to conduct correlation analyses between RCR and Category Normalized Citation Impact (CNCI) based on a priori classification, as well as journal impact factor (JIF), to test the validity of RCR in academic impact evaluation. [**Result/Conclusion**] RCR shows high correlation with CNCI and weak correlation with JIF, indicating both similarities and differences between RCR and CNCI, and revealing its practical feasibility for implementing the San Francisco Declaration and moving beyond "evaluating papers by their journals." The inverse relationship between the correlation of RCR/JIF and the number of interdisciplinary subjects in cross-disciplinary fields, along with lower RCR-JIF correlations compared to single disciplines, demonstrates RCR's applicability and effectiveness in evaluating the academic impact of interdisciplinary papers. Therefore, RCR's algorithm for identifying research fields through co-citation networks provides a new approach for evaluating paper influence, particularly for interdisciplinary research outputs.

Keywords: CNCI; RCR; F1000; Journal Impact Factor
Classification Number: G250.252
DOI: 10.13266/j.issn.0252-3116.2021.23.011

Based on R.K. Merton's conception of "citations as the currency in which scientists pay their scholarly debts," citation has become a weighty measure of academic influence in the scientific evaluation system [**?**]. However, citation distributions are inherently skewed: due to differences in citation density, Times Cited (TC) exhibits highly skewed distributions across disciplines, publication years, and document types. As early as 1979, Dr. E. Garfield noted that raw citation counts from different disciplines cannot be directly compared [**?**]. Since 2012, under the impetus of Europe's responsible metrics movement, normalized citation indicators have become a research focus to enable cross-disciplinary comparisons. This paper analyzes the Relative Citation Ratio (RCR) proposed by B.I. Hutchins et al. in 2016, which identifies research fields through co-citation networks [**?**], to provide reference for evaluating academic paper impact, particularly for interdisciplinary research outcomes.

1 Research Background

In 1986, Hungarian bibliometricians A. Schubert and T. Braun proposed establishing normalized indicators for cross-disciplinary paper comparison based on the concept of bibliographic coupling [**?**]. In 2016, B.I. Hutchins et al. from the U.S. National Institutes of Health (NIH) built upon this work to propose RCR, which identifies paper disciplines through temporal and spatial measures provided by co-citation networks. To quantify publication X's impact, RCR identifies all publications co-cited with X as its field. The metric is calculated as the ratio of Article Citation Rate (ACR) to Expected Citation Rate (ECR), where ACR equals citation count divided by years since publication, and ECR represents the expected citation rate derived through regression [**?**].

RCR has attracted significant academic attention. In practice, NIH and the scientific data platform Dimensions have adopted this indicator; the Wellcome Trust and Italy's Fondazione Telethon have used RCR to analyze their funding portfolios [**?**]. Simultaneously, scholars have validated its applications. In 2019, Elsevier's A. Purkayastha et al. extended RCR beyond biology by examining its correlation with FWCI [**?**]. In 2020, N. Steck et al. from the University of Bern explored the feasibility of replacing JIF with RCR in paper impact evaluation [**?**]. In 2021, researchers from the University of Pittsburgh Medical Center studied RCR's validity in spine neurosurgery literature through median comparisons [**?**]. Theoretically, Hutchins et al. tested RCR against expert peer review, MNCS, and SNCS, concluding that RCR offers advantages in technical implementation and usability over existing citation-based metrics [**?**]. L. Waltman noted transparency issues in RCR's theoretical model and methodology [**?**], while L. Bornmann et al. found high correlations between RCR and field-normalized percentiles but low correlations with expert peer review scores, expressing concerns about the complex normalization process and field definition [**?**]. In 2020, J.T. Seppänen from the University of Jyväskylä proposed modifications to RCR [**?**], and UberResearch recognized RCR as an important step forward in cross-disciplinary evaluation while calling for further validation [**?**]. Domestic scholar Chen Sisi et al. acknowledged RCR's advancement while urging further verification of its reliability and validity [**?**]. Thus, despite adoption in some medical systems, RCR's effectiveness requires continued exploration [**?**]. This paper addresses this need and elaborates on its application in interdisciplinary contexts.

2 Statistical Analysis of Differences Between RCR and Typical Evaluation Indicators

2.1 Data Collection and Processing

This study selected iCite, F1000, and InCites databases for analysis, focusing on 2017 publications to avoid RCR instability within 18 months of publication while capturing citation windows of 2-3 years post-publication. The data collection process involved: (1) downloading and splitting 1,068,541 papers from the

iCite open-source dataset to obtain DOI, RCR, and journal ISSN; (2) collecting 12,873 journals from the 2020 JCR, deduplicating to 12,186 unique journals; (3) matching datasets (1) and (2) via ISSN using VLOOKUP to obtain 5,721 journals; (4) retrieving paper DOI, CNCI, JIF, and other indicators from InCites using the 5,721 ISSNs; (5) merging results from (1) and (4) by DOI, removing invalid data to obtain 739,604 papers with DOI, RCR, CNCI, TC, and JIF, classified by WoS journal categories into 337,476 interdisciplinary papers (with multiple classification tags) and 402,128 single-discipline papers; (6) crawling 10,199 papers from F1000's online peer review system using WebScraper to obtain weighted expert scores (FFa); (7) matching (5) and (6) by DOI to obtain 7,963 papers with FFa, RCR, CNCI, TC, and JIF (2,739 interdisciplinary, 5,224 single-discipline) for further analysis.

Based on this data, we address: (1) similarities and differences between RCR and CNCI to highlight RCR's theoretical advancement; (2) correlation between RCR and JIF to analyze RCR's practical application in paper impact evaluation; (3) RCR's performance in F1000 to assess its effectiveness in representing interdisciplinary paper impact.

2.2 Correlation Analysis of RCR, CNCI, TC, and JIF

Kolmogorov-Smirnov tests (process omitted) indicated non-normal distribution ($p < 0.05$), warranting Spearman correlation coefficients suitable for skewed data. [Figure 1: see original paper] presents Spearman correlation results for 739,604 papers. RCR correlates with CNCI at 0.77, with TC at 0.82, and with JIF at 0.20. shows that for both interdisciplinary (337,476) and single-discipline (402,128) papers, JIF correlates substantially lower with RCR and CNCI than TC does; moreover, CNCI shows higher correlations with TC and JIF than RCR does.

2.3 RCR Performance Analysis in F1000

In F1000, correlation results for 5,224 single-discipline and 2,739 interdisciplinary papers appear in ; correlations across interdisciplinary subsets by number of crossed disciplines appear in . Key findings include: (1) RCR-JIF correlations in both single-discipline (0.519) and interdisciplinary (0.443) subsets exceed the overall sample (0.20); (2) interdisciplinary RCR-CNCI and RCR-JIF correlations are significantly lower than single-discipline correlations; (3) inverse relationships exist between RCR/CNCI-JIF correlations and the number of crossed disciplines: RCR-JIF correlations are 0.481, 0.358, and 0.242 for papers crossing 2, 3, and 4 disciplines respectively, while CNCI-JIF correlations are 0.558, 0.442, and 0.331.

3 Effectiveness Analysis of RCR in Scientific Evaluation

3.1 The Advanced Design Philosophy of RCR

The 0.77 correlation between RCR and CNCI indicates moderate positive correlation (same directional variation), suggesting similarity in paper impact evaluation. Fundamentally, both RCR and CNCI, as TC derivatives, share citation-based origins and employ classification-based normalization to correct disciplinary skewness, giving them homologous properties and similar evaluation effects. The 0.82 RCR-TC correlation indirectly validates RCR's value and validity in scientific evaluation, given academic consensus on citations as a quantifiable impact measure.

However, as University of Toronto's G. Eysenbach notes, correlation does not imply proportionality: while $r = 1$ indicates 100% predictable functional relationship, $r = 0.5$ yields only 25% predictability ($r^2 = 0.25$) [**?**]. Thus, RCR and CNCI share 59.3% predictability. Similarity represents a unity of commonality and individuality. The data also reveal differences stemming from distinct disciplinary field definition standards.

CNCI is a bias-free impact indicator excluding publication year, discipline, and document type effects, but relies on a priori classification systems, making it dependent on and sensitive to disciplinary taxonomy [**?**]. This requires clear disciplinary boundaries that academic communities have yet to consensus on. C.R. Sugimoto and S. Weingart interpret discipline across cognitive, social, communicative, and institutional dimensions [**?**], while French metrician M. Zitt asserts no optimal method exists for defining disciplines [**?**]. InCites alone includes nine classification systems (ESI, WoS, Australian ERA, UK RAE, Brazilian FAPESP, UNESCO Frascati, etc.), ranging from ESI's broad 22 categories to WoS's 255 subcategories. This fluid disciplinary definition constrains CNCI's algorithm, particularly as interdisciplinary research highlights the limitations of rigid classification systems. Papers are typically classified by their journal's category, yet interdisciplinary journal classification remains problematic. While CNCI uses disciplinary means for interdisciplinary papers, it obscures focus by treating each disciplinary component equally. Thus, the paradox between classification uniqueness and literature's multidimensional themes amplifies CNCI's limitations.

RCR, grounded in the revealing relationship between co-citation networks and scientific structure, describes both static structures at specific time points and dynamic disciplinary evolution across time, interpreting scientific research patterns and trajectories [**?**]. This dynamic field definition transcends Schubert and Braun's 1986 bibliographic coupling-based relative citation ratio [**?**], breaking traditional exclusive disciplinary boundaries and aligning with interdisciplinary research paradigms, thereby conferring theoretical advancement.

3.2 RCR Demonstrates High Discriminative Power for High-Quality Papers

The 0.20 RCR-JIF correlation indicates weak association. shows substantially lower RCR-JIF than RCR-TC correlations in both disciplinary contexts, demonstrating JIF's divergence from paper-level impact. Since the late 20th century, JIF has been reified as a proxy for paper quality and universally applied to all articles in a journal, obscuring within-journal variation.

Statistical analysis supports JIF's non-equivalence to paper impact. Among 739,604 papers, JIF ranges from 0.073 to 292.278, with 12,384 papers (1.67%) from high-impact journals (JIF $\geq$ 28). These show RCR values from 0 to 1,078.68 (median 0.46, IQR 0–3.68) and JIF median 42.78 (IQR 33.65–45.54). Using the established threshold of RCR $\geq$ 3 (approximately the mean for prestigious journals) yields 54,376 papers (7.35% of total) with JIF ranging 0.592–292.278 (RCR median 4.27, IQR 3.48–5.97; JIF median 3.64, IQR 2.47–5.68). The median comparisons further highlight RCR-JIF divergence.

If JIF were accepted as measuring paper impact and quality, 93.6% of high-impact papers in lower-JIF journals would be ignored, while 71.86% of low-impact papers in higher-JIF journals would be overstated. This prompted the San Francisco Declaration's call to decouple paper impact from JIF [**?**]. Despite the Declaration's eight-year campaign—endorsed by institutions like Imperial College London, Elsevier, Austrian Science Fund, French National Research Agency, and EU Commission—and the publication of implementation roadmaps, entrenched evaluation systems persist because JIF-based reward mechanisms are embedded in academic ecosystems [**?**]. A February 2020 *Nature* article reported China's ban on cash rewards for publications, noting that SCI-centric incentives drove China's SCI output from ~120,000 (2009) to 450,000 (2019) papers annually, causing quality decline and misconduct [**?**]. While new policies from China's Ministry of Science and Technology and Ministry of Education aim to eliminate "publish or perish" drivers, alternative evaluation frameworks remain exploratory [**?**].

In this context, RCR's weak JIF correlation and statistical evidence demonstrate its superior discriminative power for high-quality papers. As NIH's G. Santangelo stated, "RCR can identify influential papers regardless of where they are published," offering a new operational tool for implementing the San Francisco Declaration and escaping "evaluating papers by their journals" [**?**].

3.3 RCR's Applicability in Interdisciplinary Paper Evaluation

RCR values range from 0 to 1,078.68, JIF from 0.073 to 292.278. Using JIF $\geq$ 28 as a threshold [**?**], the high-impact journal subset (12,384 papers, 1.67% of total) shows RCR-JIF correlation of 0.087 (single-discipline: 0.117; interdisciplinary: -0.092). RCR median is 0.46, while CNCI-JIF correlation is -0.022 with CNCI median 1.799. The negative correlations further reveal RCR/CNCI-JIF divergence in interdisciplinary contexts, with lower RCR medians and weaker

RCR-JIF correlations indicating RCR's strong discriminative power for interdisciplinary papers.

F1000 data analysis yields several insights: (1) F1000 was designed to develop JIF alternatives, and its recent Faculty Opinions platform displays RCR alongside field percentiles and peer review scores. The overall 7,963-paper sample shows FFa-CNCI correlation of 0.194 and FFa-RCR correlation of 0.147, suggesting RCR has not fully met peer-review standards as advocated by the Leiden Manifesto [**?**, **?**]. (2) Lower RCR-CNCI correlations in interdisciplinary versus single-discipline contexts () highlight RCR-CNCI differences in interdisciplinary evaluation. (3) The inverse relationship between RCR/CNCI-JIF correlations and number of crossed disciplines (), plus lower interdisciplinary RCR-JIF correlations, aligns with interdisciplinary knowledge integration characteristics and demonstrates greater RCR-JIF decoupling. (4) Lower interdisciplinary RCR-JIF than CNCI-JIF correlations indicate RCR's superior discriminative power and thus greater applicability and effectiveness for interdisciplinary papers. (5) Deep analysis of the JIF $\geq$ 28 subset confirms RCR's strong identification capability for interdisciplinary work.

4 Conclusions

4.1 RCR's Unique Design Philosophy

Overall, RCR's high correlation with CNCI and lower correlations with expert review and TC suggest no significant advantage. However, while CNCI's a priori approach is simple and understandable, it is constrained by disciplinary classification. RCR's dynamic field identification using citation contexts and co-citation networks offers new solutions for correcting disciplinary differences [**?**].

4.2 RCR's High Discriminative Power for High-Quality Papers Compared to JIF

Correlation analysis and median comparisons confirm JIF's divergence from paper-level impact, demonstrating RCR's strong identification capability for high-quality papers. RCR provides a new operational tool for implementing the San Francisco Declaration, helping to escape the "swamp of evaluating papers by their journals" [**?**].

4.3 RCR's Applicability and Effectiveness in Interdisciplinary Paper Impact Evaluation

F1000 data reveal inverse relationships between RCR-JIF correlations and interdisciplinary breadth, plus lower interdisciplinary RCR-JIF correlations, highlighting RCR's advantages for interdisciplinary evaluation. However, limitations remain: RCR is currently applied mainly in biology and medicine, and its "black box" algorithmic opacity hinders full acceptance. Nonetheless, RCR

offers a new perspective and operational tool for field-normalized impact evaluation. Future research should refine and optimize RCR [?] and extend it to other disciplines, particularly for interdisciplinary impact evaluation. As T. Braun cautioned, "Different indicators often yield different results, underscoring the need for multi-dimensional, multi-indicator approaches" [?]. Thus, RCR can inform China's efforts to promote interdisciplinary integration and optimize disciplinary layouts [?].

References

[?] WALTMAN L. A review of the literature on citation impact indicators[J]. Journal of informetrics, 2016, 10(2): 365-391.
[?] GARFIELD E. The history and meaning of the journal impact factor[J]. Journal of the American Medical Association, 2006, 295(1): 90-93.
[?] HUTCHINS BI, YUAN X, ANDERSON JM, et al. Relative citation ratio (RCR): a new metric that uses citation rates to measure influence at the article level[J]. PLoS biology, 2016, 14(9): e1002541.
[?] SCHUBERT A, BRAUN T. Relative indicators and relational charts for comparative assessment of publication output and citation impact[J]. Scientometrics, 1986, 9(5/6): 281-291.
[?] 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.
[?] WALTMAN L, VAN EN J. Field normalization of scientometric indicators[M]// GLANZEL W, MOED HF. Springer handbook of science and technology indicators. Berlin: Springer, 2019: 281-300.
[?] Is the relative citation ratio a better metric to evaluate scientific papers?[EB/OL]. [2021-08-06]. https://www.enago.com/academy/is-the-relative-citation-ratio-a-better-metric-to-evaluate-scientific-papers/.
[?] PURKAYASTHA A, PALMAROE, FALK-KRZESINSKI HJ, et al. Comparison of two article-level, field-independent citation metrics: field-weighted citation impact (FWCI) and relative citation ratio (RCR)[J]. Journal of informetrics, 2019, 13(2): 635-642.
[?] STECK N, STALDER L, EGGER M. Journal- or article-based citation measure? a study of academic promotion at a Swiss university[EB/OL]. [2021-08-06]. https://f1000research.com/articles/9-1188.
[?] GROGAN D, REDDY V, GUPTA A. Trends in academic spine neurosurgeon productivity as measured by the relative citation ratio[J]. World neurosurg, 2021, 3(147): e40-e46.
[?] SEPPÄNEN JT, VARRI H, YIONEN I. Co-citation percentile rank and JYUcite: a new network-standardized output-level citation influence metric and its implementation using dimensions API[EB/OL]. [2021-08-06]. https://www.biorxiv.org/content/10.1101/2020.09.23.310052v1.
[?] Chen Sisi, Liu Chunli. A new indicator for evaluating publication impact based on co-citation networks: relative citation ratio[J]. Information Studies:

Theory & Application, 2020, 43(7): 75-80.
[?] JANSSENS A, GOODMAN M, POWELL KR, et al. A critical evaluation of the algorithm behind the relative citation ratio (RCR)[J]. PLoS biology, 2017, 15(10): e2002536.
[?] EYSENBACH G. Can Tweets predict citations? metrics of social impact based on twitter and correlation with traditional metrics of scientific impact[J]. Journal of medical Internet research, 2011, 13(4): e123.
[?] GLANZEL W, SCHUBERT A. A priori vs. a posteriori normalization of citation indicators. the case of journal ranking[J]. Scientometrics, 2005, 63(2): 373-401.
[?] SUGIMOTO C, WEINGART S. The kaleidoscope of disciplinarity[J]. Journal of documentation, 2015, 71(4): 775-794.
[?] ZITT M, RAMANANA-RAHARYS, BASSECOULARDE. Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field-normalisation[J]. Scientometrics, 2011, 87(2): 415-424.
[?] Qiu Junping, Liu Guohui. Research status and prospects of coupling analysis methods in China[J]. Library and Information Service, 2014, 58(7): 137-142.
[?] San Francisco declaration on research assessment[EB/OL]. [2021-04-23]. http://www.ascb.org/dora/.
[?] GUÉDON JC, KRAMER B, LAAKSO M, et al. Future of scholarly publishing and scholarly communication: report of the expert group to the European commission[EB/OL]. [2021-05-08]. https://op.europa.eu/en/publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1.
[?] MALLAPATY S. China bans cash rewards for publishing papers[J]. Nature, 2020, 579(7797): 18-19.
[?] ZANOTTO ED, CARVALHO V. Article age- and field-normalized tools to evaluate scientific impact and momentum[J]. Scientometrics, 2021, 126(4): 2865-2883.
[?] BORNMANN L, MARX W. Critical rationalism and the search for standard (field-normalized) indicators in bibliometrics[J]. Journal of informetrics, 2018, 12(3): 598-604.
[?] SCHUBERT A, BRAUN T. Cross-field normalization of scientometric indicators[J]. Scientometrics, 1996, 36(3): 311-324.
[?] ALISA S, SPORES. The relative citation ratio: what is it and why should medical librarians care?[J]. Journal of the Medical Library Association, 2018, 106(4): 508-513.

Author Contributions:
Song Liping: Conceptualization and manuscript revision;
Wang Jianfang: Manuscript revision and data analysis;
Fu Jie: Data collection and English abstract writing;
Yuan Shanshan: Data collection.

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