

Using ESI and InCites Databases for Institutional Benchmarking Analysis of Postprints

Authors: Chen Sisi, Qiu Yuhong, Xu Dan, Xu Shuang, Han Shuang

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose/Significance] To construct a model for institutional benchmarking analysis using ESI and InCites, providing a feasible approach for comprehensive, quantitative, and publicly comparable analysis of the international publication performance of benchmark institutions.

[Method/Process] Utilizing publication counts, citation frequencies, and field-normalized citation impact indicators from ESI and InCites, and based on the varying contributions of authors to publications, we propose three concepts: institutional total publications, leading publications, and participating publications, and introduce a new metric for measuring institutional publication strength—participation impact rate. A benchmarking analysis is conducted on two institutions with comparable international publication ranking strength, examining modules of overall publications, affiliated authors, collaborating institutions, and publishing journals.

[Results/Conclusion] The metrics used in this paper to describe institutional publication quantity include publication count, while field-normalized citation impact and highly cited papers describe publication quality. The results show that Institution A excels in publication quantity and overall citation frequency, while Institution B demonstrates superiority in publication quality. Analysis of affiliated author data reveals that top authors affiliated with Institution B have greater international publication impact and higher rankings. Collaboration institution data analysis indicates that Institution B's best collaborative partnerships yield higher mutual benefits, with both parties achieving higher international rankings through cooperation. Publishing journal module analysis demonstrates that Institution B's published journals have higher impact factor levels, which may be a primary reason for its superior publication quality.

Full Text

Preamble

ChinaXiv Cooperative Journal: Academic Exploration

Institutional Benchmarking Using ESI and InCites Databases

Chen Sisi, Qiu Yuhong, Xu Dan, Xu Shuang, Han Shuang

Library of China Medical University, Shenyang 110112

Abstract

[Purpose/Significance] This study constructs a model for institutional benchmarking analysis using ESI and InCites, providing a feasible approach for comprehensive, quantitative, and publicly comparable analysis of international publication performance among benchmarking institutions. **[Method/Process]** Leveraging publication counts, citation frequencies, and Category Normalized Citation Impact (CNCI) metrics from ESI and InCites, we propose three concepts based on varying author contributions: total institutional publications, leading publications, and participating publications. We introduce a new metric—Participation Impact Rate—to measure institutional publication strength. A comparative analysis was conducted across four modules: overall publications, affiliated authors, collaborating institutions, and publishing journals for two institutions with comparable international publication rankings. **[Result/Conclusion]** The metrics used to describe publication quantity include publication counts, while publication quality is described by CNCI and highly cited papers. Results show that Institution A leads in publication quantity and total citations, while Institution B demonstrates superior publication quality. Analysis of affiliated authors reveals that Institution B's top authors have greater international publication impact and higher rankings. Collaboration data indicates that Institution B's best collaborative partnerships yield higher mutual benefits, with cooperation advancing both parties' international rankings. Journal analysis shows that Institution B's publications appear in higher-impact journals, which may be a primary reason for its quality advantage.

Keywords: ESI; InCites; Leading Publications; Participation Impact Rate; Category Normalized Citation Impact

Classification Number: G250

Citation Format: Chen Sisi, Qiu Yuhong, Xu Dan, et al. Institutional Benchmarking Analysis Using ESI and InCites Databases [J/OL]. Knowledge Management Forum, 2017, 2(4): 304-317 [citation date]. <http://www.kmf.ac.cn/p/1/150/>.

1 Introduction

Since 2012, numerous Chinese institutions have adopted the Essential Science Indicators (ESI) database [1] and InCites [2] to analyze their international pub-

lications over a ten-year period (document types limited to Articles and Reviews). ESI is a Web-based research analytics tool derived from the Web of Science database that measures research performance and tracks scientific development trends. ESI exclusively includes journals from the Web of Science Core Collection (SCI/SSCI) with document types restricted to Articles and Reviews, providing rolling data for the most recent ten-plus years. The InCites database integrates nearly 30 years of data from seven Web of Science Core Collection indexes, offering diverse metrics and rich visualization capabilities.

Review of domestic literature on ESI and InCites reveals four primary research categories: (1) statistical analysis of institutional publication patterns [3-4], typically at national or provincial levels, aiming to analyze ESI inclusion status within specific regions; (2) predictive analysis [5], usually focusing on forecasting potential disciplines for research institutions; (3) comparative analysis [6], such as comparisons between domestic and foreign institutions or across different domestic regions; and (4) statistical analysis of publication journals or authors using InCites modules [7-8].

Using ESI and InCites for publication output analysis aims to produce quantified, objective, and comparable results. However, existing research suffers from several limitations: (1) limited analytical metrics are employed; (2) most researchers accept the default assumption in ESI and InCites that all authors contribute equally regardless of authorship order, without further distinguishing author contributions; and (3) there is a lack of targeted benchmarking analysis between two specific institutions, representing a gap in focused institutional competitiveness analysis.

ESI and InCites provide numerous comparative metrics, including discipline baselines, percentiles, Category Normalized Citation Impact, Journal Normalized Citation Impact, and impact relative to global averages. These comparative metrics enable quantification of institutional publication outcomes. Publicly available, easily accessible, and quantified institutional performance analysis makes benchmarking highly effective, providing objective guidance for resource allocation. We contend that institutional benchmarking will become a primary function for users of ESI and InCites databases.

Based on varying author contributions to literature, we propose three concepts: total institutional publications, leading publications (where institutional researchers serve as first or corresponding authors), and participating publications (where institutional researchers are neither first nor corresponding authors). We introduce Participation Impact Rate as a new metric to measure institutional publication strength. Combining ESI and InCites metrics for publication counts, citations, and CNCI, we conduct a benchmarking analysis of two medical universities with comparable strength across four modules: overall publications, affiliated authors, collaborating institutions, and publishing journals, aiming to identify institutional advantage disciplines and explore key issues for enhancing international publication competitiveness.

2 Tools and Methods

We utilized ESI database data updated in 2016, covering publications from 2006 to 2016. After accessing the ESI homepage, we selected the “Indicators” menu, limited search results to research fields in the left panel, selected “Institutions” as the filter, entered the English names of Universities A and B, and downloaded the results in XLS format. Sections 3.1 and 3.2 are based on data obtained through this method.

For InCites analysis, we accessed the homepage, selected the “Analytics” menu, and conducted specific analyses across four modules: People, Organizations, Research Areas, and Journals, Books, and Conference Proceedings. In the left panel, we set the dataset to InCites Dataset, filtered by institution name (Universities A and B), and restricted the timeframe to 2006-2016 with document types limited to Articles and Reviews to maintain consistency with ESI parameters. Results were downloaded in CSV format for comparative analysis. Section 3.1 primarily uses Organization module data; Section 3.3 uses Research Areas module data; Section 3.4 uses People module data; Section 3.5 uses Organization module data; and Section 3.6 uses Journals, Books, and Conference Proceedings module data.

To distinguish leading publications, we analyzed downloaded publications from both institutions to identify those with affiliated researchers as corresponding or first authors. We extracted the Web of Science (WOS) numbers of these publications, performed advanced searches in Web of Science, saved the results to InCites, changed the dataset setting to “Search Results Dataset,” and repeated the InCites search process to obtain specific data on corresponding and first authors for both institutions.

3.1 International Publication Rankings of Domestic Medical Institutions

China has 606 institutions ranked in the global top 1%, including 27 independent medical colleges. Since medical schools within comprehensive universities cannot be counted as independent entities in ESI, we analyzed the output of these 27 medical institutions across ESI’s 22 disciplines. For each institution, we sorted disciplines by citation count in descending order and selected those whose cumulative citations reached 90% of total citations, yielding nine disciplines: Biology & Biochemistry (27 instances), Clinical Medicine (26), Pharmacology & Toxicology (25), Molecular Biology & Genetics (23), Neuroscience & Behavior (19), Chemistry (12), Engineering (1), Materials Science (1), and Microbiology (1). We selected the five disciplines appearing more than 18 times and used InCites to query academic institutions in mainland China within these disciplines. The search yielded 428 contributing academic institutions, which we sorted by total citation frequency. We divided them into three tiers: the top tier included institutions whose cumulative citations accounted for 34% (approximately one-third) of total citations, the second tier comprised the next 33%,

and the remaining institutions formed the third tier.

The first tier included eight institutions: Shanghai Jiao Tong University, Peking University, Fudan University, Sun Yat-sen University, Zhejiang University, Sichuan University, Huazhong University of Science and Technology, and Capital Medical University. The second tier contained 23 institutions, including Shandong University, Nanjing Medical University, Second Military Medical University, Fourth Military Medical University, Central South University, Tsinghua University, Nanjing University, Wuhan University, China Medical University, Third Military Medical University, Harbin Medical University, Soochow University, Xi'an Jiao Tong University, Tongji University, Southern Medical University, Tianjin Medical University, Jilin University, Chongqing Medical University, China Agricultural University, University of Science and Technology of China, Southeast University, Beijing Normal University, and Nankai University—positioned in the middle of the second tier. The third tier comprised 397 institutions, with the top ten being Wenzhou Medical University, Anhui Medical University, Xiamen University, Zhengzhou University, China Medical University (Taiwan), Shenyang Pharmaceutical University, Jinan University, Huazhong Agricultural University, Guangzhou Medical College, and Nanjing Agricultural University.

We selected two institutions with comparable strength for benchmarking analysis: Institution A (ranked sixth) and Institution B (ranked eighth).

3.2 Overview of Both Institutions in InCites

Using the ESI database, we first examined both institutions' presence in the global top 1%. Institution A had four disciplines in the global top 1%, while Institution B had three. Institution A led in publication count, total citations, top papers, world ranking, and China ranking, but had lower average citations per paper. Details are shown in Table 1 .

Category Normalized Citation Impact (CNCI) [9] is a valuable and unbiased impact metric that eliminates the effects of publication year, discipline, and document type, enabling comparison across different scales and mixed-discipline paper sets. A CNCI value of 1 indicates citation performance equivalent to the global average; CNCI greater than 1 indicates above-average performance; CNCI less than 1 indicates below-average performance; and CNCI equal to 2 indicates average citation performance twice the global average.

Considering that first and corresponding authors typically contribute most significantly, we categorized institutional publications as: (1) total publications—any publication with institutional participation; (2) leading publications—those with institutional researchers as first or corresponding authors; and (3) participating publications—those where institutional researchers are neither first nor corresponding authors. Total publications = leading publications + participating publications.

Based on this classification, we propose the Participation Impact Rate (P) to quantify the influence of participating publications on total publications for specific metrics. If an institution's total citation count is N and its participating publications' citation count is M, then the citation Participation Impact Rate $P = (N/M) \times 100\%$. A higher Participation Impact Rate indicates greater influence of participating publications on the institution's overall performance for that metric. We calculated Participation Impact Rates for several metrics, including publication count, CNCI, and highly cited papers, with results shown in Table 2.

Institution A leads in both total and leading publications for publication count and citation frequency, while Institution B leads in CNCI for both categories. For highly cited papers, Institution A leads in total publications but lags in leading publications. Institution A's Participation Impact Rates exceed Institution B's for all metrics except publication count.

3.3 Advantage Disciplines of Both Institutions

ESI Thresholds [10] represent the minimum citation count required to rank among top journals in a specific ESI discipline over the past decade. Table 3 presents threshold values, institutional citation frequencies, and the ratio of institutional citations to threshold values (hereafter "threshold ratio") for both institutions.

Beyond its four top 1% disciplines (Clinical Medicine, Neuroscience & Behavior, Biology & Biochemistry, Pharmacology & Toxicology), Institution A's threshold ratios rank highest in Molecular Biology & Genetics and Immunology. Beyond its three top 1% disciplines (Clinical Medicine, Biology & Biochemistry, Pharmacology & Toxicology), Institution B's threshold ratios rank highest in Molecular Biology & Genetics and Neuroscience & Behavior. We identified Clinical Medicine, Neuroscience & Behavior, Molecular Biology & Genetics, Biology & Biochemistry, Pharmacology & Toxicology, and Immunology as established and potential advantage disciplines for both institutions.

We plotted ten-year trends (2006-2015) for three metrics—publication count, citation frequency, and CNCI—for both institutions' total publications and leading publications across all disciplines and the six established/potential advantage disciplines, as shown in Figures 1 [Figure 1: see original paper] through 21 [Figure 21: see original paper].

Institution A leads in publication quantity and citation frequency, while Institution B leads in CNCI. Comparative analysis reveals that Institution A's quantity advantages lie in Clinical Medicine, Neuroscience & Behavior, Pharmacology & Toxicology, and Immunology. Institution B's quantity advantage is in Molecular Biology & Genetics. Institution A's impact advantage is in Immunology, while Institution B's impact advantages are in Clinical Medicine and Biology & Biochemistry, with both institutions showing consistent performance in Pharmacology & Toxicology. Additionally, both institutions show annual increases

in total and leading publication counts across all six disciplines, with rising impact trends in Clinical Medicine, Neuroscience & Behavior, and Pharmacology & Toxicology. Molecular Biology & Genetics impact peaked in 2006-2007, declined in 2008, then rose annually thereafter.

3.4 Analysis of Affiliated Authors' Publications

Table 4 shows the distribution of affiliated authors by publication count for both total and leading publications at both institutions. Table 5 presents the distribution by citation frequency.

For author distribution by publication count, Institution A leads in the 100+ and 10-99 ranges for both total and leading publications. For author distribution by citation frequency, Institution B leads across all ranges.

We extracted the top 10 authors with total publication citation frequencies of 500 or higher, designating them as top authors for total publications, as shown in Table 6 . Similarly, we extracted the top 10 authors with leading publication citation frequencies of 500 or higher, designating them as top authors for leading publications, as shown in Table 7 .

For both total and leading top authors, Institution B leads in average citation frequency, impact, and highly cited papers, while Institution A leads in average publication count. This indicates that Institution B's top authors have greater international publication impact and higher rankings.

3.5 Analysis of Collaborative Institution Publications

Institution A produced 8,517 total publications and 6,484 leading publications, while Institution B produced 7,207 total and 5,119 leading publications. Collaboration rate—the percentage of collaborative publications relative to total publications—measures institutional cooperation intensity. Institution B leads in both domestic and international collaboration rates for total and leading publications, as shown in Table 8 .

Tables 9 and 10 show collaborative institution distributions by publication count and citation frequency for both total and leading publications. For collaborative institution distribution, Institution B leads in the number of institutions with 50+ publications. For citation distribution, Institution B leads in institutions with 500+ citations, with a gap of approximately 550 institutions between the two universities, while maintaining higher average citations.

We extracted the top 10 collaborative institutions with total publication citation frequencies of 500 or higher, designating them as optimal collaborative institutions for total publications, as shown in Table 11 . Similarly, we identified optimal collaborative institutions for leading publications, as shown in Table 12 .

For total publications, Institution A leads in quantity while Institution B leads in quality and citation frequency. For leading publications, Institution B leads in quantity, highly cited papers, and average impact. This suggests that Institution B's best collaborative partnerships generate higher mutual benefits, advancing both parties' international rankings through cooperation.

3.6 Journal Impact Factor Analysis

Impact factors (IF) were obtained from the 2015 Journal Citation Reports (JCR), using the most recent available JCR value when 2015 data was unavailable. We categorized journals into three tiers: high ($IF > 6$), medium ($3 \leq IF \leq 6$), and low ($IF < 3$). Table 13 shows the distribution of total and leading publications across these impact factor tiers.

Institution A's mean Participation Impact Rate for high-tier publications is 0.7, compared to 0.63 for Institution B. Medium- and low-tier Participation Impact Rates are below 0.4 for both institutions. Figures 22 [Figure 22: see original paper] through 25 [Figure 25: see original paper] illustrate ten-year trends (2006-2015) for journal impact factor distributions.

As shown in Figures 3 and Table 13, both institutions show annual increases across all three impact factor tiers. Institution B has led in high-tier publications since 2010, with higher total volume over the decade. In the medium tier, the institutions alternate annual leadership, though Institution B leads in total volume. Institution A leads annually and in total volume for low-tier publications.

For high-tier publication proportions, Institution B has led since 2010 with higher overall proportion. Both institutions show increasing medium-tier proportions, with Institution B leading since 2011 and showing a more pronounced upward trend. Low-tier proportions are declining for both, with Institution A leading since 2011 and maintaining higher overall proportion. Low-tier publications remain substantial for both institutions.

4 Discussion

The metrics employed include publication count, citation frequency, CNCI, and highly cited papers. Publication count describes quantity, while CNCI and highly cited papers describe quality. Publication quality essentially refers to impact—higher quality indicates greater influence—and represents an average measure of publication impact. Citation frequency is a composite metric involving both quantity and quality, influencing international rankings.

We propose Participation Impact Rate as a derived metric to measure the extent to which institutional leading publications reflect true research strength—representing publication capacity without external assistance. Participation Impact Rate quantifies the influence of participating publications on overall institutional metrics, indicating the degree to which ESI/InCites performance relies

on external collaboration.

Findings indicate: (1) Institution A has one more discipline in the global top 1% and higher ESI rankings across all disciplines and ESI-indexed fields; (2) InCites metrics show Institution A leads in total and leading publication quantity and citation frequency, while Institution B leads in quality. Participating publications significantly impact Institution A's quality metrics and Institution B's quantity metrics; (3) Both institutions show optimistic decade-long growth in quantity and quality across Clinical Medicine, Neuroscience & Behavior, Molecular Biology & Genetics, Biology & Biochemistry, Pharmacology & Toxicology, and Immunology.

Affiliated author analysis forms the core of institutional competitiveness assessment. Talent represents the most flexible competitive component, making institutional analysis essentially an analysis of human capital. Both institutions show consistent patterns: Institution A leads in quantity, Institution B in quality, with Institution B leading in the composite citation metric. For top authors in total and leading publications, Institution A leads in quantity while Institution B leads in quality and citation frequency, indicating that Institution B's top authors have greater international impact and higher rankings.

Collaborative institution analysis addresses the collaborative spirit emphasized in contemporary academia. Institution B demonstrates higher collaboration rates for both total and leading publications. For collaborative institution distribution, Institution B leads in institutions with 50+ publications and 500+ citations, with a gap of approximately 550 institutions in the 500+ citation range, while maintaining higher average citations. For optimal collaborative institutions in leading publications, Institution B leads in quantity, quality, and citation frequency. For total publications, Institution A leads in quantity while Institution B leads in quality and citations. This indicates that Institution B's best collaborations generate higher mutual benefits, advancing international rankings for both partners.

Impact factor represents a major concern in Chinese academia, with most institutions assuming higher-impact journals publish more valuable papers. In this benchmarking analysis, Institution B leads in both total and leading publications in high- and medium-tier journals. Both institutions show increasing proportions in the medium tier and decreasing proportions in the low tier (though low-tier remains substantial). Institution B leads in high- and medium-tier proportions, with participating publications influencing 70% of Institution B's high-tier (IF > 6) journal distribution. Institution B's superior journal placement likely explains its quality advantage. Therefore, we recommend targeting higher-impact journals as a strategy for improving institutional publication quality.

Future work will further analyze affiliated authors, incorporating additional appropriate ESI and InCites metrics to address talent recruitment and institutional performance evaluation.

References

- [1] Essential Science Indicators [EB/OL]. [2016-11-05]. <http://ipscience-help.thomsonreuters.com/InCitesLiveESI/ESIGroup/overviewESI.html>.
- [2] InCites help [EB/OL]. [2016-11-05]. <http://ipscience-help.thomsonreuters.com/InCites2Live/overviewGroup>
- [3] Liu Hong, Xu Jiaying. Evaluation of International Influence of Shanghai Universities' Disciplines: A Bibliometric Analysis Based on InCites Database Discipline Mapping [J]. Fudan Education Forum, 2014, 12(4): 29-34.
- [4] Liang Musheng, Wang Bingzhong. Analysis of Chinese Universities Entering ESI Top 1% Disciplines [J]. Library Work and Research, 2016(5): 71-77.
- [5] Gu Donglei, Wu Ying, Qiu Jiaxue, et al. Research on Discipline Development Prediction Based on WOS [J]. Modern Information, 2014, 34(6): 32-40.
- [6] He Pei, Zheng Zhong, He Dezhong, et al. Comparison of Discipline Development Between C9 Universities and World-Class University Groups: A Bibliometric Analysis Based on ESI Database [J]. Academic Degrees & Graduate Education, 2012(12): 64-69.
- [7] Zhang Li, Zhu Qinghua, Xu Xiaojuan. Analysis of Growth Characteristics and Evolution Patterns of International Scientific Talents: A Bibliometric Analysis [J]. Journal of Intelligence, 2014, 33(9): 64-71.
- [8] Gu Donglei, Qiu Jiaxue. Comparative Study on the Competitiveness of Scientific Journals in China, Japan, Russia, and South Korea Based on ESI [J]. Chinese Journal of Scientific and Technical Periodicals, 2014, 25(12): 1504-1509.
- [9] Category Normalized Citation Impact [EB/OL]. [2016-11-05]. <http://ipscience-help.thomsonreuters.com/InCites2Live/indicatorsGroup/aboutHandbook/usingCitationIndicatorsWisely/norm>
- [10] Citation Thresholds-ESI Threshold [EB/OL]. [2016-11-05]. <http://ipscience-help.thomsonreuters.com/InCitesLiveESI/ESIGroup/aboutESI/citationThresholdsAction.action>.

Author Contributions

Chen Sisi: Study design and manuscript writing, data analysis, figure preparation; Qiu Yuhong: Data coordination and analysis; Xu Dan: Data download and preprocessing; Xu Shuang: Data download and preprocessing; Han Shuang: Study design and manuscript revision.

Institutional Benchmarking by Using ESI and InCites Databases

Chen Sisi, Qiu Yuhong, Xu Dan, Xu Shuang, Han Shuang
Library of China Medical University, Shenyang 110112

Abstract: [Purpose/significance] This paper aims to build a model for benchmarking analysis of two institutions by using ESI and InCites databases. It

provides a feasible approach for comprehensive, quantitative and public comparison and analysis of the international delivery performance of benchmarking institutions. [Method/process] We used the index provided by ESI and InCites, including Web of Science documents, citation times, Category Normalized Citation Impact. According to the different literature contribution, we proposed three new concepts: all documents, leading documents, and not-leading documents. We put forward a new index—participation rate to measure the paper competitiveness of the agencies. We compared the general paper competitiveness, the affiliated author, the cooperative organization, and the publication of journals of two universities. [Result/conclusion] In this paper, Web of Science documents are used to describe the quantity of the document issued by the organization. The Category Normalized Citation Impact and citation times are used to describe the quality of the document issued by the organization. For A university, the number of documents is dominant, and for B university quality is dominant. According to the author's data analysis, B's top authors of international publications have greater influence, ranking in the front row. According to the data analysis of the cooperation institution, the cooperation between the best cooperation institutions in B is more efficient, and the cooperation between the two sides makes the international ranking more advanced. It is the main reason for the quality of published papers of B university that the impact factors of the periodicals where papers of B university published are much higher.

Keywords: ESI; InCites; Leading Documents; Participation Rate; Category Normalized Citation Impact

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

Source: ChinaXiv — Machine translation. Verify with original.