WHITEPAPER USING BIBLIOMETRICS:

A GUIDE TO EVALUATING RESEARCH PERFORMANCE WITH CITATION DATA
“A LIBRARY FACED WITH COLLECTION DECISIONS, A FOUNDATION MAKING FUNDING CHOICES, OR A GOVERNMENT OFFICE WEIGHING NATIONAL RESEARCH NEEDS MUST RELY ON EXPERT ANALYSIS OF SCIENTIFIC RESEARCH PERFORMANCE.”
A guide to evaluating research performance with citation data

USING BIBLIOMETRICS:

USING BIBLIOMETRICS:

A GUIDE TO EVALUATING RESEARCH PERFORMANCE WITH CITATION DATA

WHY EVALUATE RESEARCH PERFORMANCE?

Counting, measuring, comparing quantities, analyzing measurements: quantitative analysis is perhaps the main tool of science. Scientific research itself, and recording and communicating research results through publications, has become enormous and complex. It is so complex and specialized that personal knowledge and experience are no longer sufficient tools for understanding trends or for making decisions. Yet the need to be selective, to highlight significant or promising areas of research, and to manage better investments in science is only increasing.

Those in universities, government offices and labs, and boardrooms must decide what research should be supported and what should not, or which research projects and researchers should receive more support than others. Until relatively recently, peer review was the main route by which science policymakers and research funders made policy decisions about science.

A library faced with collection decisions, a foundation making funding choices, or a government office weighing national research needs must rely on expert analysis of scientific research performance. Increasingly, universities everywhere must demonstrate their special capabilities to a variety of constituencies. For example, in some countries universities that were once run by the government are in the process of becoming private operations. For the first time, these universities must identify a mission, develop a strategy for meeting it, and market themselves to students and supporters—as do all private institutions.

In other countries, both public and private universities that seek research funding from government agencies and foundations must provide evidence of their accomplishments and capacities. And in many countries, universities, whether public or private, must account for their performance as part of national or professional accountability protocols. Indeed, every university must have a clear, evidence-based understanding of the institution’s performance toward its goals and mission. This understanding is achieved and maintained through ongoing evaluation of all of the institution’s functions.

Because research is a central function, the university must evaluate its performance. Data on research performance helps to inform strategic decisions about what areas of research to support or build. It also helps the university leaders understand the institution’s position relative to global and domestic standards of research production. How much research is conducted? What is its impact? How many of the faculty members’ articles are published in first-class journals? Is that number of publications increasing or decreasing?

With solid, objective information about production and impact, the university has a strong basis for setting goals, charting progress, making budgetary and hiring decisions, investing in facilities, and working with external agencies.

HOW CAN RESEARCH BE EVALUATED?

BIBLIOMETRICS

Bibliometrics (sometimes called Scientometrics) turns the main tool of science, quantitative analysis, on itself. There are various definitions used for “bibliometrics.” Essentially, bibliometrics is the application of quantitative analysis and statistics to publications such as journal articles and their accompanying citation counts. Quantitative evaluation of publication and citation data is now used in almost all nations around the globe with a sizeable science enterprise. Bibliometrics is used in research performance evaluation, especially in university and government labs, and also by policymakers, research directors and administrators, information specialists and librarians, and researchers themselves.

Bibliometrics and Peer Judgment: A Two-Pronged Approach

The two together—peer review and quantitative analysis of research—better inform evaluation. Quantitative analysis offers certain advantages in gathering the objective information necessary for decision-making:

- Quantitative research is global in perspective, offering a “top-down” review that puts the work in context, complementing the local perspective of peer review. Quantitative research analysis provides data on all activity in an area, summaries of these data, and a comprehensive perspective on activity and achievements.
- Weighted quantitative measures, such as papers per researcher or citations per paper, remove characteristics, such as the place of production or past reputation, that color human perceptions of quality. For example, when we think of “the best,” it is hard not to think automatically of the biggest producers, such as individuals, labs, and universities; but those locations might not be the source of the most important work. Likewise, quantitative analysis indicates current top performers, thus balancing human perceptions of reputation.

The global and cross-sectoral indicators of citation analysis can, for example, pinpoint a paper as being in the top one percent in the world or reveal that the organization is within the top 50 in a specific area. Citation data serve as an objective and quantitative indicator for evaluating research performance, and provide additional information for assessments. While bibliometrics is a valid form of assessment, it is important that data and method of analysis selected is
appropriate to the purpose of the evaluation (Please see the Thomson Reuters document “Using Bibliometrics in Evaluating Research” for additional guidance on proper use of bibliometric indicators, at http://www.isiwebofknowledge.com/media/pdf/10Reasons.pdf).

THE GROWING USE OF BIBLIOMETRICS
For these reasons and others, nations with significant science enterprises have embraced bibliometrics. Today, bibliometrics programs with large teams of analysts are firmly established in many nations, and these groups issue bibliometric reports, often called science indicators studies, at regular intervals. A few such groups are the National Science Foundation (United States); the European Commission; France’s L’Observatoire des Sciences et des Techniques (OST); and Japan’s National Institute for Informatics (NII), National Institute for Science and Technology Policy (NISTEP), and Ministry of Economy, Trade and Industry (METI). Other nations with active bibliometrics groups include Argentina, Australia, Belgium, Brazil, Chile, China, Finland, France, Germany, Israel, Italy, The Netherlands, New Zealand, Norway, Portugal, South Africa, South Korea, Spain, Sweden, Switzerland, and Taiwan. In almost all cases, the publication and citation data of Thomson Reuters form the basis of their bibliometric analyses.

WHO AT AN INSTITUTION ENGAGES IN RESEARCH EVALUATION?

Citation data can be used in many ways, for a variety of purposes. In a competitive environment, it is important for an institution to show how performance supports its mission. Indeed, the results can be a powerful communication tool.

EVALUATION AND STRATEGIC PLANNING
An institution’s research and planning related offices often undergo periodic evaluations of research performance and produce an external or internal report. These institution-level assessments may be accompanied by departmental or researcher-level assessments for purposes of accreditation, tenure/ faculty review, etc.

Key performance indicators for research activity may be incorporated into strategic planning, or used in annual reporting to government bodies, boards of directors/trustees, alumni, etc. Performance indicators on an institution’s funding, which measures resources flowing into its research activities, can be augmented with bibliometrics, which measures the output of the research activities.

RESEARCH CENTERS
Research centers must continuously find new staff, develop lines of investigation, and compete for funds. External assessment of performance helps in making decisions, developing strategies, and demonstrating capacity. Understanding the publication activity of a research group can help to maximize funding and staffing decisions.

PUBLIC RELATIONS AND DEVELOPMENT
To publicize the university and attract students and researchers, as well as donors and other supporters, the public relations and development units of the university can benefit from objective figures about the university’s research accomplishments.

LIBRARIES
Bibliometrics can help libraries to understand their local community’s journal usage patterns, and to identify the most influential journals to subscribe to. Collection development can be informed by bibliometrics data on both the relative influence of journals in the worldwide community, and on the publication and citation patterns of an institution’s researchers themselves. Librarians too play a role in observing, tracking, and validating the institution’s research performance trends by tracking publishing output and impact.
WHAT TYPES OF DATA ARE USED IN BIBLIOMETRICS?

The results that citation analysis yields are statistically significant because of massive and systematic data collection. This is the reason that Thomson Reuters’ citation index, accessible via Web of Science®, is used as the worldwide standard for bibliometrics.

For over 50 years, Thomson Reuters has comprehensively collected data across academic fields, from natural sciences to social sciences and humanities. Highly reliable statistical analysis of the data is possible because of Thomson Reuters’ consistent policy on indexing and storing data. A citation index for science was first described in 1955 by Eugene Garfield, the founder and chairman emeritus of what was then ISI, in the journal Science. He realized his vision a few years later with the production of the 1961 Science Citation Index®. The operating principle of a citation index is this: If a researcher knows of a publication important to his or her work, a citation index would allow the researcher to identify journal articles published subsequent to that work which have cited it.

It did not escape Garfield’s notice that such a database could serve other purposes as well, such as monitoring and analyzing the structure and growth of science. The combination of an ever-growing corpus of publication and citation data compiled by ISI over the 1960s and 1970s and the simultaneous increase in computing power and software applications, especially those developed in the 1980s and 1990s, has made bibliometrics a practical and even cost-effective pursuit.

Web of Science now includes not only the Science Citation Index, but also the Social Science Citation Index®, the Arts and Humanities Citation Index®, Index Chemicus®, and Current Chemical Reactions®, resulting in a truly multidisciplinary citation resource. Web of Science, which now covers nearly 9,300 high-quality, core journals from every field, is used by over 3,400 organizations and universities in more than 90 countries around the world. In addition, statistical values based on the collected information serve as significant indicators for science and technology in many government organizations.

COVERING CORE JOURNALS

A relatively small core of research literature is frequently cited, while the remainder is not cited at all: 10 percent of research literature accounts for as much as 90 percent of citations. In addition, important research that earns a large number of citations in one field is also frequently cited in other fields. Because a collection of core literature in any one field may provide researchers in other fields with necessary literature, Thomson Reuters’ multidisciplinary citation index was created.

COVERING ALL TYPES OF ITEMS

In order to maintain statistical integrity, Thomson Reuters’ citation index includes all items published in the selected journals. This includes the primary vehicles of disseminating research results, such as articles and reviews, but also includes editorials, letters, book reviews, etc. One can limit a bibliometric analysis to only articles and reviews, or choose to include the more marginal document types.

COVERING ALL AUTHORS AND THEIR ORGANIZATIONS

Web of Science® consistently includes the authors’ names and the names of the organizations the authors are affiliated with, and does so according to uniform protocols.

WHAT DOES BIBLIOMETRICS MEASURE?

THE DEVELOPMENT OF PUBLICATION AND CITATION ANALYSIS

There are many activities and outcomes of research that can be counted. Perhaps the most basic and common is the number of scholarly journal publications, which may be used as a measure of output. Citations are the references researchers append to their papers to explicitly show earlier work on which they have depended to conduct their own investigations. Tracking citations and understanding their trends in context is a key to evaluating the impact and influence of research.

As described in Eugene Garfield’s reasons for citing a paper, the citations in academic papers create a record of influence. Motivations for citing a prior work can include:

- Paying homage to pioneers.
- Giving credit for related work (homage to peers).
- Identifying methodology, equipment, and the like.
- Providing background reading.
- Correcting one’s own work.
- Correcting the work of others.
- Criticizing previous work.
- Substantiating claims.
- Alerting researchers to forthcoming work.
- Providing leads to poorly disseminated, poorly indexed, or uncited work.
- Authenticating data and classes of fact (such as physical constants).
- Identifying original publications in which an idea or concept was discussed.
- Identifying the original publications describing an eponymic concept or terms.
- Arguing against the work or ideas of others.
- Disputing the claims of others to have been first with their work.

Not only can the influence of an individual research paper be traced through its citations in other papers, but the influence of a body of research in a specific domain can be determined. As Garfield explains, “The papers marked with frequent citations are regarded as useful by a comparatively large number of researchers and experiments.” Regarding the quality that citation counts measure, Garfield explains, “People talk about citation counts being a measure of the ‘importance,’ or ‘impact’ of scientific work, but those who are knowledgeable about the subject use these words in a very pragmatic sense: what they really are talking about is utility.”

Using citation analysis—bibliometrics—in this way, a university can assess the performance of its research units, gauge its contribution to the creation of knowledge and technology, and make decisions based on objective, quantitative data.
CITATION METRICS: WHAT TYPES OF DATA ARE BEST FOR WHICH PURPOSES?

There are no all-purpose indicators. It is important to start by identifying the question the results will help to answer and then collecting data accordingly. The results are only as good as the questions that prompt the assessment. The university decision-maker must first clearly define the purpose of the evaluation, the kinds of information they need, and how they will use the analyses. This extra effort is necessary and worthwhile. This clarity of purpose determines what, of the vast amount of data available from the citation index, should be used, and the kind of analysis that will be performed.

The chart below lists a number of different purposes a university might have for evaluating its research performance. Each purpose calls for particular kinds of information. The discussion following the chart describes the relationship between purpose and data in greater detail.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Metric Type</th>
<th>Metric Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity</td>
<td>Counts of papers</td>
<td>Author, group, inst.</td>
</tr>
<tr>
<td>Influence</td>
<td>h-index</td>
<td>Author, group, inst.</td>
</tr>
<tr>
<td>Indirect recognition</td>
<td>Second-generation citation</td>
<td>Author, group, inst.</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Average citations per paper</td>
<td>Author, group, inst.</td>
</tr>
<tr>
<td>Relative impact</td>
<td>Percent cited/united papers</td>
<td>Author, group, inst.</td>
</tr>
<tr>
<td>Benchmarking</td>
<td>Journal impact factor</td>
<td>Journal</td>
</tr>
<tr>
<td>Specialization</td>
<td>Percentile indicators</td>
<td>Author, group, inst.</td>
</tr>
<tr>
<td>Trend analysis</td>
<td>Time series</td>
<td>Author, group, inst.</td>
</tr>
</tbody>
</table>

These citation metrics can help to answer important questions, including:

- What is the university’s research performance?
- How competitive is our research compared to our peers?
- How can the university forecast growth?
- What are the university’s centers of excellence?
- What is our citation ranking?
- How can we demonstrate the influence of our research?

PRODUCTIVITY: Counts of papers

Paper counts, which measure productivity, are the most basic bibliometric measure and provide the raw data for all citation analysis. Ranking institutions in terms of paper counts helps to compare the productivity and volume of research output among various institutions. The number of researchers at an institution should be taken into account when comparing publication counts across institutions. Characteristics of the papers, such as document type, publication year, and categorization method, should also be considered.

Comparison of paper counts. Source: University Science Indicators
TOTAL RECOGNITION/INFLUENCE:

Citation counts

Citations measure impact and influence. Citations to papers are summed over some time period to create an aggregate citation count. Aggregate citation counts of institutions or researchers over the same time period can be useful in comparing and ranking their research impact. As with paper counts, the number of researchers at an institution will affect the total number of citations accrued.

H-index

The Hirsch index, or H-index, is a distribution-based indicator that corresponds to the number of papers at or above a given citation level equal to the value of the citation threshold. This statistic reflects the number of papers (N) in a given dataset having N or more citations. For example, an H Index of 77, indicates that 77 papers in the given set were cited at least 77 times each. The H-index of a subset of papers is always less than the H-index of the entire set and hence cannot be normalized in a ratio manner.

This measure attempts to reflect both productivity (number of papers) and impact (number of citations) in one number. This metric is useful because it discounts the disproportionate weight of highly cited papers, or papers that have not yet been cited. For example, the citations per paper for a group of 5 papers could be 12.20, but the H-index is 2 because only 2 papers out of the 5 had received at least 2 citations.

Like citation counts and citations per paper, the H-index depends on time period and discipline. A proper use of the H-index will ensure that the time period of analysis and field of research is similar among the researcher/institutions/research groups being compared.

INDIRECT RECOGNITION/INFLUENCE

Second-generation citation counts

Second-generation citation counts are the sum of the citation counts of all the papers citing a target paper. This is a measure of the long-term impact of a paper which is similar in effect to the Google PageRank. Like first-generation citation counts, second-generation citation counts need to be normalized by field.

EFFICIENCY

Average citations per paper

Citations per paper (sometimes called “impact”) is computed by dividing the sum of citations to some set of papers for a defined time period by the number of papers (paper count).

The citations per paper score is an attempt to weight impact in respect to output, since a greater number of publications tends to produce a greater number of citations. Citations per paper is a useful statistic when comparing large with small producers; however, some minimum number of publications, a threshold, ensures that one or a few highly cited papers do not skew the results. Different fields of research exhibit different average rates of citation. Sometimes the difference between fields in terms of average citation counts can be as much as 10:1, so comparing like with like is a “golden rule” of citation analysis.

Percent cited/uncited papers- Absolute

This metric refers to the percentage of cited or uncited papers in a sample. A “cited” paper has received at least one citation. This metric can reveal the amount of publications with no or very little influence.

Journal Impact Factor

Journal impact factor is a measurement applied to journals. The impact factor of a journal represents the average citation count of the articles published in the journal during a two-year period. Please see the section “Is impact factor used in bibliometrics?” which discusses the journal impact factor in more detail.
RELATIVE IMPACT/BENCHMARKING
Percent cited/uncited papers - Relative
Percent cited/uncited papers can be considered relative to the field of research, a country, institution, etc. This method provides further context to percent cited/uncited. For example, the rates of citedness vary across disciplines. Relativizing the measure enables you to judge the influence of the papers in light of the norm in their field, or the norm in their country or institution.

Field baselines and relative impact
Field baselines are average citations per paper for papers in a field (usually a journal set) defined for a specific time period. Since different fields exhibit different average rates of citation, the mean for the field should be used to gauge the relative impact of one or a group of papers. By dividing the actual number of citations by the average, a ratio is obtained. A measure of 1.0 would therefore designate average performance. This relative measure exhibits all the strengths and weaknesses (in regard to skewed distributions) of a mean.

Expected citation rate (ECR) and crown indicator (C-index)
Expected citation rate (ECR) indicates how often a paper is expected to be cited based on its year of publication, journal, and article type. ECR represents the average citations per paper for the cohort of papers having these same attributes. Just like fields, different article types exhibit different average rates of citation. For example, letters to the editor, meeting abstracts, and correction notices are generally much less cited than original research reports and reviews.

When ECRs are summed for a group of papers, this sum can be used as an expected rate for the group. The sum of the actual citations divided by the sum of expected citations equals the crown index (C-index). The C-index takes into account field differences in citation rates, since each paper in a set is normalized for the average in its journal, even if those journals are from many different fields. Because it is a ratio, a C-index of greater than 1 indicates performance that is better than expected.

Percentile indicators: paper percentiles, average percentiles, relative percentile thresholds indicators
A paper percentile is determined by taking the year and journal category of the paper, creating a citation frequency distribution for all the papers in that year and category (finding the number of papers cited N times, and arranging these in descending order), and determining the percentage of papers at each level of citation. The percentile then indicates how a paper has performed relative to others in its field.

For a set of papers, each with an associated percentile, an average percentile can be calculated. This can apply to any set of papers, such as an author’s body of work. In this case, the average percentile would indicate the relative performance of the author’s works, having been normalized for whichever fields they fall into. Therefore, the average percentile of researchers from different fields can be compared, despite field differences in citation rates.
A percentile threshold indicator is created by determining the citation threshold value at some predefined percentile, e.g., 1%, for a sample group of papers in comparison to some baseline or reference group of papers. The ratio of thresholds gives the percentile threshold indicator.

These are relative measures that more exactly reveal position, with regard to influence and rank, within a field.

**SPECIALIZATION**

**Collaboration indicators**

Metrics for collaboration include rates of co-authorship for pairs of authors, institutions, countries, etc. They can include standard series such as the percentage of papers with 1, 2, 3, etc. authors over time, as well as the computation of impact and relative impact indicators for specific country or institutional collaboration pairs. These metrics can help identify where collaboration has and has not taken place. Such analysis can also reveal dynamic and influential points of collaboration, as well as those that have been less influential.

**Disciplinarity index**

This metric indicates the concentration or dispersion of a group of papers over a set of field categories. This can be expressed as the sum of squared fractions of papers over some set of categories (disciplinarity index). A value of 1 indicates total concentration in a single field category.

This metric helps to view multi- or interdisciplinary research output. It represents a response to the problem of field definition (a set of journals defining a field or category), which sometimes poses difficulties due to the constantly changing nature of science.

**Research fronts**

A research front is a group of highly cited papers, referred to as core papers, in a specialized topic defined by a cluster analysis. A measure of association between highly cited papers is used to form the clusters. That measure is the number of times pairs of papers have been co-cited, that is, the number of later papers that have cited both of them. Clusters are formed by selecting all papers that can be linked together by a specified co-citation threshold. Research fronts can reveal emerging areas of science through citation patterns.

**TREND ANALYSIS**

**Time Series**

Time series are powerful depictions of citation data. Whereas single period statistics provide a snapshot of research performance, time-series provide insight into the change in output and impact over time. Nearly all in-depth bibliometric studies provide some type of time-series analysis. For citation metrics, both citing and cited periods must be defined.

Changes in staffing or funding can be correlated with subsequent changes in publication and citation metrics.
IS “IMPACT FACTOR” USED IN BIBLIOMETRICS?

“Journal Impact factor” (JIF) is a measurement applied to journals, based on the citation index database, Web of Science. The impact factor of a journal represents the average citation count of the articles published in the journal. The JIF is updated annually and published in Journal Citation Reports®.

The JIF is a journal-level metric designed for one purpose—to compare the citation impact of one journal with other journals.

Regarding the use of the JIF in research evaluation, it can take one of two main forms:
- An indicator of success achieved in having an article accepted by a prestigious journal.
- A surrogate for a more carefully derived direct measure of citation impact.

While the first use may have some utility, the second appears difficult to justify. The first use has some justification. There is a hierarchy of journals within subject areas. This hierarchy broadly corresponds to impact rankings, and, in a formal communication system based on peer review, acceptance of an article for publication in one of these journals is an important scholarly achievement. Rewards based on this achievement can be seen as encouraging scholars to aim high in their publishing goals. The data is thus appropriate to the question being asked.

It is not recommended to use the JIF as a surrogate for more direct measures in such concepts as a “total impact factor” or other calculations in which the JIF stands in for article performance, and further calculations are performed on lists of JIFs. It is very hard to see how such data, so manipulated, are appropriate for any question related to evaluation or comparison. We must remember that publication in journals with a high impact factor does not necessarily indicate a high citation count for one’s articles. A small group of very highly cited papers in a journal can cause a high impact factor.

In response to issues raised in the perceived misuse of the JIF in evaluation, calls for the use of “article-level metrics” have been made. One of the popular metrics at the moment is the H-index, as well as the other article-level metrics described in the previous section. Thomson Reuters recommends using a combination of multiple indicators for assessing individual papers or researchers, including the expected citation rates and percentiles described herein.

THOMSON REUTERS PRODUCTS

The table below serves as a guide to which Thomson Reuters products can help achieve specific goals:

<table>
<thead>
<tr>
<th>User Role</th>
<th>Goals</th>
<th>Data Need</th>
<th>Thomson Reuters Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrators (e.g. provosts, chancellors, research directors)</td>
<td>Evaluate, benchmark, promote</td>
<td>Summary, comparative, high-level information</td>
<td>University Science Indicators&lt;br&gt;Essential Science Indicators</td>
</tr>
<tr>
<td>Department Chairs/Deans</td>
<td>Departmental and individual review</td>
<td>Detailed information, yet with options to compare higher scales</td>
<td>Institutional Citation Report&lt;br&gt;University Science Indicators&lt;br&gt;Essential Science Indicators</td>
</tr>
<tr>
<td>Librarians</td>
<td>Information retrieval from the detailed to the high-level, and assessments on how the library is utilized</td>
<td>Paper-level information with access to summary level information for context</td>
<td>Web of Science&lt;br&gt;Journal Use Reports&lt;br&gt;Journal Citation Reports&lt;br&gt;Institutional Citation Report</td>
</tr>
<tr>
<td>Researchers</td>
<td>Information retrieval from detailed to summary</td>
<td>Detailed data targeted to field of work</td>
<td>Web of Science&lt;br&gt;Topical/National Citation Reports&lt;br&gt;University Science Indicators&lt;br&gt;National Science Indicators</td>
</tr>
</tbody>
</table>