MASARYKOVA UNIVERZITA

Recommendations for working with and understanding bibliometric indicators

These recommendations are issued for the benefit of MU employees working with bibliometric indicators (impact factor, h-index, etc.). Situations where bibliometrics may affect the evaluation of an individual or group call for an understanding of bibliometric limitations and fundamental knowledge of bibliometric indicator usage. These recommendations include an overview of the general principles applicable to work with bibliometric indicators. **All recommendations are based on internationally accepted methodological concepts and best practices employed by experts in scientometrics.**[1]

Bibliometrics is a statistical analysis of written publications based on data from **bibliographic databases**. The most important data producers across a range of fields currently include primarily the following:

1. Clarivate Analytics (Web of Science platform, WoS),
2. Elsevier (Scopus platform).

In order to simplify the compilation of bibliometric indicators from the Web of Science and analyse publication activity, Masaryk University employs the **InCites** tool developed originally by Thomson Reuters (now Clarivate Analytics). **InCites** analyses are either based on a specialized dataset of Masaryk University publications and authors, prepared by the Research & Development Office of the Rector's Office on the basis of research publications available via the IS MU, or on a dataset generated according to WoS addresses. InCites facilitates a wide range of analyses at university, unit and individual levels. InCites also constitutes a basic benchmarking tool for comparing output on a worldwide scale. **Analyses involving the InCites tool will be processed in accordance with specific requirements by the Research & Development Office of the Rector's Office.**

In addition to large multidisciplinary databases, a range of field-specific databases may also be employed. These include e.g. MathSciNet (mathematics), SciFinder – CAS (chemistry, biochemistry and related fields), PubMed (medical fields) and ADS – Astrophysics Data System (astrophysics).

[1] HICKS, D, et al. Bibliometrics: The Leiden Manifesto for research metrics. Nature. 2015, vol. 520, 7548, 429–431. doi: http://dx.doi.org/10.1038/520429a. Available at: http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351; San Francisco Declaration on Research Assessment (DORA), Available at: http://www.ascb.org/dora/.

However, these databases generally present greater limitations for the compilation of bibliometric indicators than traditional producers. Specialized patent databases include Derwent (part of Web of Science), EPO- PATSTAT, USPTO, DEPATISnet and WIPO.

The utilization of **Google Scholar** for bibliometric purposes is a frequently discussed topic. This specialized bibliographic search service employs a proprietary algorithm to locate and display references to selected documents. While utilizing Google Scholar for bibliometric analysis purposes is possible only with considerable complications (see Data sources), it does allow the user to monitor response to scientific output in fields which are underrepresented or absent from traditional databases (e.g. selected social sciences and humanities).

Bibliometric outputs are also utilized – albeit with varying importance ascribed to them – by **university rankings** (e.g. Leiden Ranking, THE World University Rankings, Shanghai Ranking). In this context, however, bibliometric indicators are frequently prone to a considerable degree of misinterpretation.

The following section introduces individual databases as well as methods relevant to the utilization of certain indicators.

General principles and recommendations

1. Bibliometrics is most reliable for the purpose of collating and analysing data on a macroscopic level (i.e. country or institution). When applied to smaller units of assessment (workplaces, individuals), its suitability for comparing and analysing research performance decreases and a supplementary qualitative assessment becomes essential[2] (e.g. if bibliometric methods establish a difference between two employees in the same workplace, it does not necessarily mean that there is a real qualitative difference between the two). The accuracy of analyses conducted at the level of individuals may be increased either by means of authorisation provided by the researchers themselves or by employing ResearcherID profiles with regularly updated input data.

2. Bibliometrics is designed to serve as a basis for subsequent qualitative assessment. Bibliometrics is not suitable as the sole basis for an evaluation of individuals or groups, particularly in cases where such an assessment may affect careers and matters of employment.

3. Any bibliometric analysis should always simultaneously employ multiple sources. In addition to using a multidisciplinary database, employing a field-specific database is always recommended.

4. Bibliometric analysis outcomes should be transparent and universally understandable. The evaluated unit should always have the opportunity to verify data. Should bibliometric

[2] For example, when comparing the h-index, it is essential to take into account the field of study, age, duration of publication activities, output types and the development of publication activity over time.

analysis constitute a basis for evaluation, all criteria should be transparent and known in advance.

5. To obtain reliable information on an analysed unit, a range of relevant indicators is required rather than a single indicator. In order to interpret data in a responsible way, indicators must be monitored in relation to one another.

6. Bibliometrics necessitates the comparison of like with like (e.g. within one field). For example, it is impossible to compare the h-index of a historian with that of a biologist due to the vastly different citation conventions and output structures found in each field.

7. Excellence must always be located in a relevant field-specific context.

8. Metrics associated with journals (e.g. impact factor) as well as any derived metrics must be excluded from consideration when assessing funding, personnel issues (employment, promotions) and individual evaluation matters.

9. Performance must be measured and interpreted within the context of an evaluated unit's research strategy.

10. Any evaluation of research must take into account the quality and impact of all research output, i.e. not just publications, while considering a wide range of measurements including the influence of qualitative indicators (e.g. utilized methods and practices).

11. Research must be evaluated in view its contribution or impact, i.e. not purely on the basis of the journal it has been published in or on the merit of publication metrics alone.

12. The improper selection of indicators must never be allowed to directly influence the behaviour or career of individual researchers. While not ideal, informed qualitative assessment (e.g. board/peer review) remains the best science evaluation and management tool available.

13. Data must always be refined and verified. Data refinement usually focuses on detecting identical names, different variations of institutional names and addresses, disciplines, result types, etc. Unless data has been refined, it does not constitute a reliable basis for further analysis.

Data sources: Web of Science, Scopus, Google Scholar

Policies employed by data producers tend to determine generally assumed criteria for research quality and excellence. When working with databases it is therefore essential to factor in the following characteristics and limitations:

Web of Science	**Scopus**
Strict indexing criteria (fewer resources). Books indexed on a limited scale. Limited use in the case of the humanities and certain social sciences. An analysis of cited sources may make	Less stringent indexing criteria (more resources). Books indexed on a limited scale. Limited use in the case of the humanities and certain social sciences. However, Scopus includes a more

use of Cited Reference Search which includes journals and results not indexed by the WoS, including e.g. book publications. For example, Cited Reference Search may be used to accurately establish the h-index, especially in the social sciences and humanities.	extensive selection of humanities and social sciences sources than WoS.
Majority of journals from the Anglo-American environment, relatively small number of local sources.	More European-oriented, more extensive number of local sources.
WoS is not a suitable source of data for all fields of study. With certain exceptions, social sciences and humanities sources are inadequately represented and thus provide insufficient numbers of source records.	Scopus is generally broader in scope and thus more appropriate for the social sciences and humanities as well as for conference papers.
Data must always be refined and verified. Data refinement usually focuses on detecting identical names, different variations of institutional names and addresses, disciplines, etc.	Data must always be refined and verified. Data refinement usually focuses on detecting identical names, different variations of institutional names and addresses, disciplines, etc.
Impact factor used to measure source popularity.	SJR factor (SCImago Journal Rank) used to measure source popularity.

Using Google Scholar for bibliometric purposes is still quite problematic; all data must be subject to a strict critical procedure and further refinement. However, Google Scholar may well be the right choice in the case of fields not covered by traditional databases (see above). When working with Google Scholar it is therefore essential to factor in the following characteristics and limitations:

Google Scholar disadvantages	**Google Scholar advantages**
Lack of transparency: Google Scholar uses an algorithm which evaluates what is “scholarly”. This algorithm is not public and it is thus up to the researcher to determine which search results are relevant (scholarly vs popular science).	Wide range of document types: in addition to traditional articles and books, Google Scholar also includes reports, presentations, final theses and teaching materials among search results. Wide-ranging geographical and field of study coverage. A potentially useful source for the social sciences and humanities.
Google Scholar search results may include different versions of the same output depending on where it has been published.	Supplementary services (user profile, basic citation metrics).
Google Scholar does not specify which sites are included in the search. It is also impossible to control online content changes, i.e. the same search may return different results at different times.	Repository scanning. Full text access (if applicable).
Danger of data manipulation (fraudulent	Google Scholar is a modern challenge with

authorship claims, fake articles).	significant room for improvement.
No data purity control and thus the resulting need to exercise extreme caution when working with data, particularly at the level of individuals. Data must always be refined and verified.	Google Scholar search results are analysed using Harzing's Publish or Perish software as well as various Android and iOS applications and associated functionalities (e.g. UGRinvestiga); browser plug-ins.
	Academic output linked to social networks (in case authors have Google Scholar profiles) and other functionalities. Authors with Google Scholar profiles contribute to making the tool more reliable, as this essentially means that they have verified their publications.
	A significant tool for increasing visibility, especially in the case of young researchers.

Working with selected bibliometric indicators

1. Impact factor

The impact factor (Thomson Reuters) was originally created as a tool to help librarians locate influential magazines suitable for purchase. It measures the popularity of a journal rather than the scientific impact of an article or individual. This is apparent from its calculation:

$$\textit{Impact factor in year } X = \frac{\textit{all citations of all articles published in a given journal between X-1 and X-2}}{\textit{number of “citable” outputs (articles, notes and reviews) published in a given journal between X-1 and X-2.}}$$

At present, the impact factor (IF) constitutes a highly influential figure, as it is frequently used thanks to its simplicity and clarity, though sometimes in an incorrect context. Arguments in favour of IF include the fact that (in most areas) there is a correlation between IF and output quality. **Journals with high IF are subject to demanding and constantly developing peer review.** When evaluating the work of an individual, it is essential to assess which journals he or she has published in (i.e. with what IF). However, this cannot be used as a primary indicator of quality and it is always necessary to take into consideration additional bibliometric data and perform a qualitative assessment (see below). Arguments against IF include the fact that IF does not constitute the sole motivation for many authors. Some authors publish in lower IF journals, e.g. industry-specific journals, in order to reach a specific target audience. In addition, many fields do not focus on IF at all and even top scientists may not consider IF to be a relevant criterion.

In the case of research areas with rapidly ageing data, scholars may also be motivated to target less prestigious journals with simpler and faster review processes. Some fields (e.g. informatics) even tend to publish results primarily in the form of conference papers and conference proceedings, likewise due to the generally employed more rapid approach. **Therefore, IF should be used exclusively for the purpose it is suitable for, i.e. to evaluate journals. The use of IF as the sole**

indicator for the evaluation of articles, institutions or persons is avoided by the professional scientometric community as well as by the data producers themselves.[3]

The impact factor has a number of limitations:

a) While IF draws attention to a journal, it does not guarantee that an article published in the journal will be cited more often.
b) The number of citable articles (i.e. the denominator) may be influenced by the size of the magazine.
c) Older and more frequently issued journals are at an advantage compared to newer and less frequently published ones.
d) The impact factor may be manipulated as a result of editorial policy.[4] Review journals have an advantageous position as well (critical review articles are generally cited more frequently).
e) Imbalance (IF strongly influenced by a minority of frequently cited articles).[5]
f) IF properties are specific to different fields. It is impossible to compare absolute IF values between journals from separate fields.
g) The selection of criteria for calculating IF is not a transparent process.

IF is calculated to three decimal places, which by itself evokes an impression of precision. However, the above characteristics clearly indicate that the calculation may be affected by a number of "soft" factors and that – especially in the case of journals with only a slight difference in absolute IF – it is impossible to determine a journal's quality purely on the basis of IF.

Accumulating journal impact factors in order to evaluate a certain unit (i.e. **cumulative impact factor)** is a good example of IF being misused for inappropriate purposes. While this approach will return the sum of impact factors pertaining to individual journals an author has published in, it says nothing about the scientific qualities of the author or the impact of his or her articles. The value of the cumulative IF increases with an author's age. However, such data cannot be used or compared in any meaningful way, not even within a single field.

Sample evaluation of two authors on the basis of quantitative data:

Author A	**Author B**
- one article from 2012 in a journal with IF 12, interdisciplinary journal. - Number of citations: 9. - Consortium of 55 authors.	- one article from 2012 in a journal with IF 4, specialized professional journal. - Number of citations: 30. - Bilateral cooperation with a foreign university. - Article shared among the scientific community through Twitter.
H-index 1	H-index 1

[3] e.g. State of Journal Evaluation 2015. Available at: http://stateofinnovation.thomsonreuters.com/state-of-journal-evaluation.
[4] The PLoS Medicine Editors. The impact factor game. PLoS Medicine. 2006, 3, 6, e291, doi: http://dx.doi.org/10.1371/journal.pmed.0030291.
[5] Editorial. Not so deep impact. Nature 2005, 435, 1003–1004. doi: http://dx.doi.org/10.1038/4351003b.

This simple example clearly illustrates that no meaningful evaluation can be carried out on the basis of a single value. Should a department head evaluate an employee on the basis of the IF of the journal he or she has published in, author A will appear to outperform author B and any personnel implications will likely be unfair as a result. While the publication of an article in a journal with high IF may be viewed as a success, it is necessary to assess the circumstances and impact of the article itself and evaluate the specific contributions of individual co-authors, i.e. perform a qualitative assessment or include additional indicators. All relevant circumstances taken into account, the scientific impact of author B would likely be considered more substantial. However, real-life situations tend to be significantly more complex.

2. h-index

The h-index is an indicator which measures the productivity and citation impact of a researcher or scholar. The value of *h* is an expression of the number of articles with *h* or more citations. However, though the h-index is not an all-purpose metric, it does blur distortion to a certain extent, especially between a single highly cited article and articles which have not been cited yet. An analysis of the h-index over time can help locate consistently high-quality research in a given field. It is also useful for detecting the presence of "rising stars" in various fields. H-index disadvantages include its strong dependence on industry-specific citation rates.

H-index characteristics:

- The h-index varies across individual databases as well as within a single database (several h-indices may be calculated for a single author according to a selected database and filter). Up to three different h-indices may be found in the WoS in specific instances, depending on the employed approach: 1) Web of Science Core Collection, 2) All Databases and 3) Cited Reference Search. Since no universally applicable methodology for obtaining the h-index is available, it is always necessary to specify where the h-index was obtained.
- Google Scholar also provides an h-index; it is typically higher than WoS since Google Scholar includes a much larger volume of sources of inconsistent quality.
- The h-index is industry-dependent and values thus cannot be compared between different fields.
- The h-index does not diminish over time. [6]

3. Additional bibliometric indicators

The choice of monitored indicators varies greatly with respect to the focus of the analysed unit, field of study, purpose of the analysis and observed phenomena, database size and many other variables. It is thus impossible to select a suitable set of indicators for specific purposes. The following list thus only includes examples of the most common indicators. Indicators must always be interpreted in relation to one another.

To obtain an overview of the **publication activities** of an analysed unit, simple statistics showing the number of publications in databases and a structured list of outcomes and their representation in

[6] The h-index may of course be calculated for a given period of time, e.g. the past 10 years.

databases is generally sufficient. In case **publishing power** is monitored, the above indicators must be viewed in the light of e.g. the number of FTE employees, amount of expenditures, etc. Another group of commonly monitored data includes **cooperation** indicators, i.e. the extent of international cooperation or lists of cooperating countries, institutions or individuals. An analysis of collaborative networks may also yield additional information, such as data on new research trends.

Metrics of influence based on the citation impact of individual publications must be normalized and viewed in the context of a given field. In this case, relevant indicators include different variants of citation impact, e.g. normalized citation impact (ratio of obtained and expected citations within a field, journal or country). Furthermore, the inclusion of an article in the top 1 % or 10 % of most cited articles in a given field is a good indicator of quality; the derived percentile values indicate what percentage of articles has been cited more frequently.
An analysis of visibility, expressed by indicators linked to utilized sources – including e.g. the structure of journals which a unit has published in, the popularity of these journals within a given field, impact factor and associated indicators – is also a suitable element.

Bibliometric services

The choice of monitored indicators varies greatly with respect to the focus of the analysed unit, field of study, purpose of the analysis and observed phenomena, database size and many other variables. We monitor indicators, focusing on elements such as activity, publishing performance, collaboration and impact. As it is generally impossible to establish a set of universally applicable all-purpose indicators, the Research & Development Office of the Rector's Office thus offers **bibliometric report services** which utilize the **InCites tool** to assess faculty units and individuals based on submitted requirements. Analyses are always preceded by an assessment of specific needs. Analysis outcomes do not constitute an evaluation but rather detailed information which may be used for strategic management, continuous monitoring, as a basis for improving publication strategy, etc. Consultations are provided on request by the Research & Development Office of the Rector's Office.

Contact information:

Mgr. Michal Petr
e-mail: petr@rect.muni.cz
tel: 549 49 5887

http://vyzkum.rect.muni.cz/

Compiled by Michal Petr, 3 November 2015

Research & Development Office, Rector's Office
Masaryk University Rector's Office

Žerotínovo náměstí 9
601 77 Brno