Bibliometrics, a useful tool within the field of research

Bibliometría, una herramienta útil dentro del campo de la investigación

Karla Salinas-Ríos, & Angélica Janneire García-López

Abstract:

The activity in scientific research has been studied, compared, measured and analyzed through Scientometrics, a discipline that applies to all scientific literature, mathematical and statistical methods, thus achieving that social aspects of science can be quantified. Scientific publications (tangible products of research) are derived from the scientific literature, which are specifically studied by bibliometrics. This last one, is a branch of the Scientometrics, that is guided under the assumption that the scientific discoveries and the research results are published in scientific journals, so its unit of analysis is the scientific article. The word bibliometrics was defined for first time by Alan Pritchard in 1969, and since then, multiple concepts of this term have been developed. However, it has reached to the consensus that this methodological tool allows to know the scientific production (in quantity, quality and impact) on various topics, journals, authors and countries, among others. Its main research lines are the methodology for bibliometrics, the scientific disciplines and the health management and policies. Likewise, it has descriptive, evaluation and supervision/monitoring functions of the research activity, on which its classification into levels (micro, meso and maso) will depend directly. Because it has components from various sciences, among them the mathematics, its methodology and theory are based on mathematical models, from which the bibliometric indicators are derived. Although there are other types of research such as systematic reviews and meta-analyses, these, require a better management in the field of research and of the statistical measurement, as well as more resources. On the other hand, a bibliometric study owns the nobility of being within the reach of students and researchers due to its methodology, practicality, relevance, resource saving, potential to extend to most of the scientific areas, multiple applications and favoring the fact of not committing ethical misconduct related to research. Finally, although bibliometrics is often underestimated, its power and importance as a tool to manage evidence-based knowledge and to serve as a basis for other types of studies such as systematic reviews must be emphasized.

Keywords:
Bibliometrics, science, research, scientific production, scientific publications

Resumen:

La actividad en investigación científica se ha estudiado, comparado, medido y analizado a través de la Cienciometría, disciplina que aplica a toda la literatura de carácter científico, métodos matemáticos y estadísticos, logrando así, que aspectos sociales de la ciencia puedan ser cuantificados. De la literatura científica, se derivan las publicaciones científicas (productos tangibles de la investigación), que son estudiadas específicamente por la bibliometría. Ésta última, es una rama de la Cienciometría, que se guía bajo el supuesto de que los descubrimientos científicos y los resultados de las investigaciones son publicados en revistas científicas, por lo que su unidad de análisis es el artículo científico. La palabra bibliometría fue definida por primera vez por Alan Pritchard en 1969, y desde entonces, se han desarrollado múltiples conceptos de este término. No obstante, se ha llegado al consenso de que esta herramienta metodológica permite conocer la producción científica (en cantidad, calidad e impacto) en diversos temas, revistas, autores y países, entre otros. Sus principales líneas de investigación son la metodología para la bibliometría, las disciplinas científicas y la gestión y políticas de salud. Asimismo, tiene funciones descriptivas, de evaluación y de supervisión/monitoreo de la actividad investigadora, de las que dependerá directamente su clasificación en niveles (micro, meso y maso). Debido a que tiene componentes de diversas ciencias, entre ellas las matemáticas, su metodología y teoría se basan en modelos matemáticos, de los cuales se derivan los indicadores bibliométricos. Si bien existen otros tipos de investigación como las revisiones sistemáticas y los metaanálisis, estos, requieren un mejor manejo en el campo de la investigación y de la medición estadística, así como más recursos. Por otro lado, un estudio bibliométrico posee la nobleza de estar al alcance de estudiantes e investigadores debido a su metodología, practicidad, relevancia, ahorro de recursos, potencial para extenderse a la mayoría de las áreas científicas, múltiples aplicaciones y favorecimiento al hecho de no cometer faltas éticas relacionadas con la investigación. Finalmente, aunque la bibliometría suele subestimarse, se debe recalcar su poder e importancia como herramienta para gestionar el conocimiento basado en evidencia y para servir como base para otro tipo de estudios como las revisiones sistemáticas.

Palabras Clave:
Bibliometría, ciencia, investigación, producción científica, publicaciones científicas

*Corresponding author, Maestría en Psicología de la Salud, UAEH, https://orcid.org/0000-0003-2440-8126, Email: salinaskarla93@gmail.com

Maestría en Psicología de la Salud, UAEH, https://orcid.org/0000-0002-4179-4244, Email: janneireg@gmail.com

Received: 12/12/2020, Accepted: 09/11/2021, Published: 05/01/2022
INTRODUCTION

Bibliometrics definition
In order to define what a bibliometrics is, it is necessary to conceptualize to the Scientometrics. This, studies, measures, compares, analyzes and objectifies the activity in scientific and technical research (Gauthier, 1998; Louzada-Junior & Carioca, 2011), through mathematical and statistical methods that are applied to all the literature of scientific nature, thus achieving that some social aspects of science can be quantified (López, 1996). However, evaluating scientific development is not easy, since there are different factors that can interfere with it, such as technology, culture and economy (Dávila et al., 2009).

On the other hand, bibliometrics is a branch of Scientometrics that deals specifically with the study of the scientific publications (Gauthier, 1998; Rehn & Kronman, 2008), which are tangible products of the research. Etymologically, bibliometrics comes from the greek terms biblos (that means book) and metron (which means to measure) (Dávila et al., 2009). Its analysis unit is the scientific article (Glänzel, 2003), since it is guided under the assumption that scientific discoveries and research results are published in scientific journals, at least mostly (Rehn & Kronman, 2008).

Its main background are the Statistical Bibliography (Spinak, 1996), and the first mention of the metric science Librametry, made by Ranganathan in 1948. Being Alan Pritchard, the first to define the word bibliometrics, in the year 1969 (Pérez, 2002). Until now, different definitions of this term have been developed, with the characteristic in common of applying the statistics and mathematics to be able to objectify and quantify scientific production (Tague, 1988; Dávila et al., 2009).

Methodologically, a bibliometric study is a tool that (López, 1996), through groups of knowledge, helps to build the theory empirically. In other words, from a set of documents a series of indicators are studied, whose results will evidence aspects of science that need to be explained, achieving that their conclusions acquired more transcendence at a scientific level (Schinka & Velicer, 2003). Moreover, this tool allows to know both the scientific production, in quality and quantity, as well as its impact at a social level (Rueda-Clausen et al., 2005; Dávila et al., 2009).

Bibliometrics research lines
The bibliometric studies have the potential to extend to most of the scientific areas, since they belong to a research field considered multi and interdisciplinary with components such as: mathematics, social sciences, natural sciences, engineering, computing and statistics, among other (Glänzel, 2003). For this reason, as a research type, they offer various possibilities to students of different levels (undergraduate, postgraduate) and also to health professionals, having three main approaches, approximations or research lines (Romani et al., 2011):

- Methodology for bibliometrics: it is basic research, through which they have been developed: 1) bibliometric indicators 2) mathematical models and 3) a methodology for research in bibliometrics (Gauthier, 1998; Glänzel, 2003; Aleixandre et al., 1996; Retzer & Juransinski, 2009).
- Scientific disciplines: this line uses bibliometric indicators for the study of the scientific publications distribution that focus specifically on a discipline or thematic in particular (Glänzel, 2003; López-Piñero & Terrada, 1992).
- Health management and policies: it is the bibliometric research line that can have scope at national, regional or institutional level, therefore, it is the most important of the three. It is developed by the evaluation of the research that is expressed in publications (Glänzel, 2003; Dávila et al., 2009; Lundberg, 2006).

It should be mentioned, that the previous research lines are not mutually exclusive (Romani et al., 2011), and that in general, this type of analysis is of great relevance for the psychologist within the research field (Rueda-Clausen et al., 2005).

Bibliometrics functions
The description, evaluation and supervision/monitoring of the research activity represent the three main functions of the bibliometrics. Likewise, depending on the scope of each function or study, they are classified into three levels: micro or individual (researcher or specific journal), meso (institution or investigation group) and macro (country, region, city or province) (Gauthier, 1998; Franks et al., 2006).

The bibliometric study of descriptive type provides a comparative analysis of the scientific productivity based on the quantitative information about the published articles at country, province, city and institutions level, and even at individual level; allowing to identify the most productive researchers and institutions, to know the scientific production on those topics, to describe the publications profile made (study type, journal where it was published, language, publication year, funding institutions, study population), and to know the state of the art in national research (Gauthier, 1998; Devos, 2011; Romani et al., 2011).

An evaluation research has diagnostic and monitoring purposes. Despite of not being designed to evaluate the results and methodological quality of the studies it collects, it allows to know the quantity of the published works and their “impact” or dissemination among the research community measured through the number of citations they receive. In this way, it is possible to evaluate the research programs, the efficiency and efficacy of their implementation, and to determine if the objectives are being achieved, in addition to recommending the necessary adjustments. Another important application of the bibliometric studies, related to the evaluation, is its use for a) support the justification for the allocation of research funds b) review the methods and aspects both the positive as negative of the journals c) stimulate the promotion of literary decisions (López-Piñero & Terrada, 1992; Dávila et al., 2009; Klein & Bloom, 2005).
The bibliometric analysis oriented to the supervision or monitoring of the science and technology, are in charge of identifying the research areas that are being developed or leaving aside, in order to show what research topics are being little addressed or require more investigation by the scientific community. Furthermore, at the same time they offer a report of the most studied research phenomena (Gauthier, 1998; Devos, 2011).

Worth mentioning that the finality of a bibliometric study many times define its complexity, being the descriptive type those of less difficulty. Despite this, none of them ceases to be important for the science (Romani et al., 2011).

**Bibliometrics application fields**

Directly related to the functions, are the application fields of this branch. As a first point, it can be mentioned that it is a tool for aspects such as the decision making, the direction of an investigation and the designation of economic resources. As for the researchers, this type of analysis contributes to the prestige of the scientist (Koskinen et al., 2008; Hood & Wilson, 2001).

Regarding to the scientific activity, it can be obtained a panorama of its performance and repercussion in a certain region. Also, they can be measured the differences between the productivity in different areas of science and their contribution to development. Secondary to this, decisions are made to create policies or allocate resources, the journals can decide which authors publish, and low-productivity specialties can receive support (Sylvan, 1999).

The bibliometrics also serve to select books and periodicals, identify thematic characteristics of the literature, evaluate bibliographies and collections, know the history and sociology of the science, determine core journals in a specific area and identify the countries, institutions and authors that are more productive in a given period (Koskinen et al., 2008; Sylvan, 1999).

**Basic bibliometric theory**

Under the assumption that the scientific activity is reported through the scientific literature, and particularly through the scientific publications (articles), it is a requirement in the bibliometrics field that researchers publish their investigations (Rehn & Kronman, 2008). Linked to this assumption, within the bibliometrics there are four main axioms that serve to define and use the bibliometric indicators, same that will be explained later (Glänzel, 2003):

- In a time t, an article receives at least one quotation;
- In a time t, an author publishes only one article;
- An article doesn’t receive prior quotations to its publication;
- The quotation link between two articles is unique.

Although the scientific article is the analysis unit of this science, the evaluation of its “impact” is measured through some indirect indicators such as the number of quotations. At this point, it is important to clarify that (Romani et al., 2011):

- Within the field of the bibliometric research, it is assumed that the majority of citations of an author or work are given due to the prestige and quality of the articles.
- There is a difference between reference and quotation. A reference serves to point out and locate the scientific literature that supports the information declared by the author in his research, this is placed at the end of each article. Instead, a citation, is that reference realized in another article published later to the initial article (that is, when an article is used as a reference in certain research, it is being quoting).

**Bibliometrics mathematical principles**

As it mentioned in previous paragraphs, the bibliometrics methodology comprises components of different sciences, among them, the math (Romani et al., 2011) that, through models, have helped to the comprehension of various aspects of this discipline and also to generate laws to relate two variables (Hubert, 1981).

Some models serve to explain: the scientific productivity, the distribution of articles in journals, the growth of the scientific literature, the co-citation analysis, the scientific collaboration analysis, among others (Glänzel, 2003). Some of them are mentioned and explained below.

**Lotka’s Law** (Romani et al., 2011).

- Postulate: only some authors publish the majority of articles on a topic;
- Verbal formulation: the authors number (An) that publish n articles about a topic is inversely proportional to n articles squared.
- Mathematical formulation: \( An = A1/n^2 \), where An is the authors number with n articles, A1 is the authors number with only one publication on the topic and n is the articles number.

In general, the authors number (An) decreases when the productivity is high (large n). This law doesn’t take into consideration factors that can influence in the productivity, this is, it was formulated as a natural law.

**Models of scientific production growth**

The production characteristic of scientific articles can be explained through the following mathematical models: exponential, logistic and deterministic (Glänzel, 2003). In general terms, these models include the cumulative number of scientific articles as a function of time \( p(t) \) (Romani et al., 2011).

**Bradford’s Law**

- Postulate: the journals can be divided into a core, that are those that publish particularly on a topic. There are also other groups with the same number of articles as the main group, but with a greater number of journals (Koskinen et al., 2008).
● Formulation: the mathematical ratio of the number of journals of the core towards the subsequent zones is 1.
● Mathematical formulation: \( 1: n: n^2 \).

This law is also known as Bradford distribution, dispersion law, or core and dispersion. It focuses on the distribution of the articles in the scientific journals and/or on the order of the journals according to their number of quotations (Louzada-Junior & Carioca, 2011; Glänzel, 2003; Aleixandre et al., 1996).

**Literature half-life**

It is the time in which half of the references have been published, considered as a reflection of the aging of the literature that has been quoted. It is also known as Burton and Kebler’s semi-period or obsolescence index (Culebras-Fernández et al., 2008). Its calculation is made through the determination of the median of the publication year of the references (Romani et al., 2011).

**Price’s index**

Its calculation is made from the percentage of references with antiquity lesser to five years. In addition, this indicator complements the previous one (Romani et al., 2011). It is worth mentioning that the mentioned laws and models are fulfilled only when the analyzed literature has been carefully compiled. Furthermore, when it comes to social phenomena, their verification is made in an approximate way and not with absolute mathematical precision (Romani et al., 2011).

**Methodology for conducting a bibliometric study**

Like of other scientific productions, the bibliometrics has a structure whose process allows the achievement of goals of each study. Although the methodology for conducting a bibliometric research may not be very complex, the adequate execution of each component must be congruent with the aim and the search criteria of the topic to be analyzed, otherwise, the results may be limited, generalized or biased, diminishing its scientific function. Its methodological components are the following (Romani et al., 2011):

**Research objective setting**

The approach of a clearly established research problem represents the route on which the study aim of the research is based. In the elaboration of a bibliometric project, it will be most useful to delimit what is the current state of the research, who participates, how they interact and what relationships exist between them that possibly have allowed the increase in their production. Its analysis and management will have a positive impact on the development of the research (Romera, 1992).

The aim is not limited only at the time of determining how many articles are published in certain country or journal. It is essential to specify what information will be analyzed (number of publications, productivity of the authors, institutions, regions or countries, or even deepen the analysis of the research methodology, among other characteristics) and to establish the recovery period of the studies, as well as the set of journals or the characteristics of the articles that will be included. Moreover, it is important to note if you want to study a particular research area or phenomenon and delimit it (Romani et al., 2011).

A clear aim, settles the search strategies, a results analysis plan and the knowledge of the scopes or limitations of the research.

**Search for information**

Once the research aim is settled, the next step will be the search for information, which can ensure the effectiveness of a bibliometric study. This is why such a search must be rigorously systematized, otherwise, literature that, in fact, be of interest to the research can be omitted. Some points to consider at this methodological stage are that, the search for information (Ardanuy, 2012):

- It will be realized in certain databases, which will be chosen based on the research aims and not on the researcher’s “facilities”
- It will collect the higher quality articles through the use of various search methods or strategies, which must be informed and described (keywords, MESH methodology, use of Booleans, among others)
- Normally, it will be guided by the judgment, the interests and the professional practice experience of the researcher
- It is not necessarily realized by topics, they can also be done searches by authors, institutions, journals and countries

**Information registry**

Once the search for information and the data filtering have been carried out, a list composed of a series of bibliographic registries is obtained, which constitute the population object of the bibliometric study. From each of these records, a series of fields are extracted, the most common being the following: authors, title, publication journal, publication year, authors institutional adscription, language, type of document, thematic classification, number of citations received, financing source, ambit of the study and type of population studied, among others. The additional information fields are determined based on the object of the research study. This process requires of the content analysis of one or more of these sections, being necessary for this to realize a careful treatment of the data in order to homogenize the information collected (Romani et al., 2011). Finally, once this process has been completed, the bibliometric indicators are obtained, that is, a count or statistical treatment of data collected is carried out.

**Bibliometric indicators**

The bibliometric indicators are indices or calculations whose main function is to provide quantitative and objective (measurable) information on scientific activity (Lutman, 1992). More specifically, they can be used to measure the scientific publications quality (such as the productivity) or the impact of
the works (such as the number of citations received or the trend analysis) (Chaviano, 2004; Campanario, 2003). These, have been created through a rigorous process; they are simply straightforward and guided by logic; or they are arbitrary (Romani et al., 2011).

It is relevant to mention, that most of these indicators arise from the number of citations (citation index), this is, from the number of times that an article, published in a journal and belonging to a certain author, is quoted by other authors (Feather & Sturges, 2003). These resources (the citations), have become in the basic accounting element of visibility of the documents that receive them, but not of their scientific quality, at least not directly, so they should not be used for this purpose (Osca-Lluch et al., 2017).

The bibliometric indicators are classified as follows: of production, of visibility and impact, and of collaboration (Romani et al., 2011). Likewise, within this classification, the best known are: impact factor, immediacy index and half-life of the scientific articles (Caballero et al., 2006).

**Production indicators**

As their names indicates, they determine the magnitude of the articles production by country, institution or author; or well, the magnitude of articles production over time (per year, per GDP, per million inhabitants, among others) (Romani et al., 2011).

**Visibility and impact indicators**

These, are the ones that enjoy the greatest popularity. Some of them are: citations and references; impact factor; H index; and clinical implementation indicators, among others (Romani et al., 2011).

- **Impact factor**: it is an indicator that is calculated annually and is aimed at journals, this is, it reflects the quality of a journal, but not of the articles or authors (Garfield, 2003; Dávila et al., 2009). It can be defined as the quotient of the number of articles quoted from the publications of a journal in a certain time interval (normally the last two years), divided among the total number of articles published in the same period (Alvis-Guzmán & De la Hoz-Restrepo, 2006).

\[
\text{Impact factor (year } x) = \frac{\text{Citations in (year } x) \text{ of published documents in (period)}}{\text{Published documents in (period)}}
\]

This is, the numerator would be the total number of citations that receive during certain year all the documents that are publish in a journal in the last two years. On the other hand, the denominator would be the total number of published quotable articles in said journal, also in the last two years. 

Exemplifying (Garfield, 2003):

\[
\text{Impact factor (2008)} = \frac{\text{Citations in (2008) of published documents in (2006 y 2007)}}{\text{Published quotable articles (2006 y 2007)}}
\]

An article is considered quotable when the journal catalogues it as an original article or as a review. On the contrary, the documents classified as letters to the editor, editorials and bibliographic recensions, among others, are excluded (Caballero et al., 2006; Andersen et al., 2006; Campanario, 2003).

Currently, the Web of Science, previously called Institute for Scientific Information (ISI), generates and publishes the impact factor of some journals through the Journal Citation Report. However, this bibliometric indicator can also be found on the covers or web pages of the publications (González de Dios et al., 2007; Seglen, 1997). This index, can be useful to choose a journal to publish an article or to subscribe (Dávila et al., 2009).

Despite this factor is the most popular, successful and used of the visibility and impact indicators (Osca-Lluch et al., 2017; Osca-Lluch et al., 2019), it has also received some criticism related to the fact that the development of some sciences is slower, so that their investigations require of a longer period of time to be quoted. This results in some major journals obtaining lows impact factors (González de Dios et al., 2007; Seglen, 1997).

- **H index (HIRSCH)**: it is a distribution indicator, aimed at the researchers and that reflects the number of citations an article receives, taking into account the citations threshold (Pendlebury, 2008). This, is, what matters is the weighted one, which refers to the assigned value of the number of articles (productivity) that at the same time are highly quoted (quotations or “impact”) (Hirsch, 2005).

Exemplifying:

“A researcher has ten publications that have been quoted at least once, but only three have received more than two citations, the first received twenty citations (his article of more impact), the second four citations and the third three citations; in that case his H index is of 3, which means that he has “at least” three articles with three citations” (Romani et al., 2011, p. 60).

Within its advantages it can be mentioned that it is relatively simple to calculate and that it eliminates extreme values. On the contrary, its disadvantages include that its understanding can be complex and that over time there will always be an increase or maintenance in the index of an author (Hirsch, 2005). If it is desired to obtain easily the H index of the authors, currently, the SCOPUS database serves as tool for said purpose (Romani et al., 2011).

- **Immediacy index**: this indicator reflects the approximate number of times the document of a journal is quoted the same year of its publication (Caballero et al., 2006). This, is calculated through a division, where the numerator or quotient is the number of citations that receives the published article in the year that it is wanted to calculate; and the denominator, is the total number of quotable articles in the same year (van Raan, 2003).

\[
\text{Immediacy index (year } x) = \frac{\text{Number of citations of published articles in (year } x)}{\text{Number of published quotable articles in (year } x)}
\]
Indirectly, this indicator shows the speed with which an article is quoted since it is published, considering some aspects that can influence in said process: the publication medium and its capacity for dissemination (Caballero et al., 2006; Prakashan et al., 2006).

**Collaboration indicators**

They are networks graphics that allow to visualize the existing relations between countries, institutions and authors. In this way, they can be determined aspects such as: the collaboration degree, the proximity and who are the commissioned of the networks flow (Newman, 2004). Their main advantage is that they can serve as support for the development of strategies directed to improve the collaborations between countries, institutions and authors (González-Alcaide et al., 2008). Among these indicators they can be mentioned the collaboration networks and the bibliometric maps (Romani et al., 2011).

It is important to mention that, according to Romani et al., (2011), it is recommended that the bibliometric indicators seen above be selected from the beginning of the research, as different data are needed for the analysis of each of them. Moreover, there is a specific bibliometric indicator for each specialty. By the last, it is also relevant to know their limitations, since the reliability of the results of this type of analysis will largely depend on their correct application (Dávila et al., 2009).

**CONCLUSIONS**

Currently, the scientific development is constantly growing, including the number of scientific publications. It is for this reason that it is practically inevitable and above all necessary that the results derived from the scientific research be evaluated in an objective way. In this sense, the bibliometrics is an appropriate, relevant and practical evaluation tool for that purpose. This, can be used to: a) have a global vision of how scientific research is found at country, journal or author level b) compare the scientific development between regions, authors and journals c) objectively measure the scientific development in terms of growth or recoil d) make decisions in different fields e) grant resources to research projects or journals f) give prestige to authors or colleges g) create development policies in countries, cities and colleges, among other functions (Dávila et al. 2009).

Furthermore to the multiple functions and applications of the bibliometric studies, these, have the advantage of being studies that involve little investment and favor the fact of not committing ethical misconduct related to the research due to they work with already published databases. Nevertheless, they also have some disadvantages such as a) probability that some institutions highlight their ego b) advertise unduly c) making wrong decisions. Due to these last disadvantages, it is necessary to realize an ethical analysis of the information and to handle impartially the results. Finally, it is important to mention that the bibliometrics can be underestimated (in its applicability and relevance) and not considered as real research. However, its power and importance as a tool to manage the evidence-based knowledge and to serve as a basis for other type of studies such as the systematic reviews must be emphasized (Romani et al., 2011; Holden, Rosenberg, & Barker, 2005).

**REFERENCES**


Bibliometrics as a research field: a course on theory and application of bibliometric indicators. Course Handouts: Bélgica.

gical_journals


An index to quantify an individual’s scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569–16572. doi: 10.1073/pnas.0507655102

Bibliometrics: a potential decision making aid in hiring, reappointment, tenure and promotion decisions. Social Work in Health Care, 41(3-4), 67-92. doi: 10.1300/J010v41n03_03

The literature of bibliometrics, scientometrics, and informetrics. Scientometrics, 52(291), 291-314. doi: https://doi.org/10.1023/A:1017919924342

General bibliometric models. Library Trends, 30(1), 65-82. Recuperado de https://www.ideals.illinois.edu/bitstream/handle/2142/7188/librarytrendsv30i1h_opt.pdf?sequence=1&isAllowed=y

Bibliometrics: The best available information? Social Work in Health Care, 41(3-4), 117-121. doi: 10.1300/J010v41n03_07


Introducción a la bibliometría. Promolibro: España.


Anales de Psicología, 35(1), 166-174. doi: 10.6018/analesps.35.1.329211


_WP.pdf


