

RESEARCH ARTICLE

Measuring the Outcome of Biomedical Research: A Systematic Literature Review

Frédérique Thonon^{1,2,3*}, Rym Boulkedid^{2,4}, Tristan Delory⁵, Sophie Rousseau^{6,7}, Mahasti Saghatchian¹, Wim van Harten⁸, Claire O'Neill², Corinne Alberti^{2,3,4}

1 European and International Affairs Unit, Gustave Roussy, Villejuif, France, **2** AP-HP, Hôpital Robert Debré, Unité d'épidémiologie clinique, Paris, France, **3** Université Paris Diderot, Sorbonne Paris Cité, UMR-S 1123 and CIC-EC 1426, ECEVE, Paris, France, **4** INSERM, U 1123 and CIC-EC 1426, ECEVE, Paris, France, **5** AP-HP, Hôpital Bichat, Département d'Epidémiologie et de recherche clinique, Paris, France, **6** Direction de la Recherche Clinique, Gustave Roussy, Villejuif, France, **7** Centre Hygée, Department of Public Health, Lucien Neuwirth Cancer Institute, CIC-EC 3 Inserm, IFR 143, Saint-Etienne, France, **8** The Netherlands Cancer Institute, Amsterdam, the Netherlands

* frederique.thonon@gustaveroussy.fr



Abstract

OPEN ACCESS

Citation: Thonon F, Boulkedid R, Delory T, Rousseau S, Saghatchian M, van Harten W, et al. (2015) Measuring the Outcome of Biomedical Research: A Systematic Literature Review. PLoS ONE 10(4): e0122239. doi:10.1371/journal.pone.0122239

Academic Editor: Daniele Fanelli, Stanford University, UNITED STATES

Received: August 11, 2014

Accepted: February 10, 2015

Published: April 2, 2015

Copyright: © 2015 Thonon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was funded by the European Commission under the 7th Framework Programme (FP7) under grant agreement n° 260791 for the project entitled 'EurocanPlatform'. The URL of the funders is http://ec.europa.eu/index_en.htm and, URL of the FP7 is http://ec.europa.eu/research/fp7/index_en.cfm and the URL of the EurocanPlatform project is <http://eurocanplatform.eu>. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manual.

Background

There is an increasing need to evaluate the production and impact of medical research produced by institutions. Many indicators exist, yet we do not have enough information about their relevance. The objective of this systematic review was (1) to identify all the indicators that could be used to measure the output and outcome of medical research carried out in institutions and (2) enlist their methodology, use, positive and negative points.

Methodology

We have searched 3 databases (Pubmed, Scopus, Web of Science) using the following keywords: [Research outcome* OR research output* OR bibliometric* OR scientometric* OR scientific production] AND [indicator* OR index* OR evaluation OR metrics]. We included articles presenting, discussing or evaluating indicators measuring the scientific production of an institution. The search was conducted by two independent authors using a standardised data extraction form. For each indicator we extracted its definition, calculation, its rationale and its positive and negative points. In order to reduce bias, data extraction and analysis was performed by two independent authors.

Findings

We included 76 articles. A total of 57 indicators were identified. We have classified those indicators into 6 categories: 9 indicators of research activity, 24 indicators of scientific production and impact, 5 indicators of collaboration, 7 indicators of industrial production, 4 indicators of dissemination, 8 indicators of health service impact. The most widely discussed and described is the h-index with 31 articles discussing it.

Competing Interests: The authors have declared that no competing interests exist.

Discussion

The majority of indicators found are bibliometric indicators of scientific production and impact. Several indicators have been developed to improve the h-index. This indicator has also inspired the creation of two indicators to measure industrial production and collaboration. Several articles propose indicators measuring research impact without detailing a methodology for calculating them. Many bibliometric indicators identified have been created but have not been used or further discussed.

Introduction

There is an increasing demand for research evaluation. Research funders want to assess whether the research that they fund has an impact [1]. In addition to demonstrate accountability and good research governance, research funding organizations need to build an evidence base to inform strategic decisions on how to fund research [2]. According the Canadian Academy of Health Sciences, evaluation of research is carried out for three main purposes: accountability purposes, advocacy purposes, and learning purposes. Evaluation for accountability is usually performed by funders to assess whether the outcome of their funding has fulfilled its anticipated aim and has strong links to value-for-money issues. Evaluation for advocacy aims to increase awareness of the achievements of a research organisation in order to encourage future support. Evaluation for learning is an inward looking process that aims to identify where opportunities, challenges and successes arise for the research performed in an institution [3].

Some of the existing evaluation systems include assessments by national agencies, the Organisation for Economic Cooperation and Development (OECD) Frascati Manual, the UK Research Assessment Exercise, the Shanghai ranking, etc. . . Those systems use a set of different indicators, sometimes complemented by peer-review. An indicator is defined as ‘a proxy measure that indicates the condition or performance of a system’ [4]. Indicators are said to be more objective than peer-review assessment [5].

Nevertheless, there is an increasing call for evaluation of medical research in terms of its benefits to patients [6–12]. From a policy making perspective, this vision particularly applies to health care institutions, such as university hospitals or comprehensive cancer centers where research and patient care are integrated and sometimes carried out by the same professionals. In the view of designing an evaluation system to measure the outputs, outcomes and impact of medical research, it is necessary to have first an overview of all possible indicators, as well as their positive and negative points.

Some reviews of indicators measuring research production have been conducted [13–16]. Those review focus mainly or exclusively on bibliometric indicators or research input indicators (such as research funding). The scope of our systematic review is different as it focuses exclusively on indicators measuring the production, output and outcome of medical research, and that it intends to go beyond bibliometric indicators in order to include indicators measuring long-term impact of medical research. In this article we define outputs as “the immediate tangible results of an activity” and outcomes as “longer-term effects such as impact on health” [17]. We use the definition of impact proposed by the Canadian Institute of Health Research: “In the context of evaluating health research, the overall results of all the effects of a body of research have on society. Impact includes outputs and outcomes, and may also include additional contributions to the health sector or to society. Impact includes effects that may not have been part of the research objectives, such as contributions to a knowledge based society or to

economic growth” [18]. In this article we make a distinction between scientific impact and health service impact.

We conducted a systematic review with the following objectives: (1) to identify all existing indicators that can be used to measure the output or outcome of medical research performed by institutions, and (2) to list, for all indicators, their positive and negative points, as well as the comments on their validity, feasibility and possible use.

Methodology

We wrote a protocol prior to the start of the study. We chose to undertake a review of all indicators, including those used to measure research areas outside the biomedical field.

1. Search strategy

We searched in Pubmed, Scopus and Web of science, using the following terms: [“research outcome*” OR “research output*” OR “research impact*” OR bibliometric* OR scientometric* OR “scientific production*”] AND [indicator* OR index* OR evaluation* OR metric* OR “outcome* assessment”], as terms in the abstract, title or keywords, with no time limit. On those three databases we applied filters on language (including only articles written in French or English) and on type of document (including only articles and reviews).

Through snowballing, we added articles that seemed relevant in the bibliography of selected articles. Two of us (FT and RB) undertook the search independently, assessed articles on the basis of title and abstract and compared our search results. Differences in results were discussed and resolved.

2. Inclusion criteria

Considering the scope of our project to develop indicators designed to measure the production and outcome of research undertaken by institutions (such as hospitals, research centres or research units. . .) we set up inclusion and exclusion criteria accordingly. We included articles written in French or English, that presented, discussed or evaluated indicators measuring the scientific production of an institution. We excluded:

- articles that presented or assessed only indicators measuring research inputs (such as funding or human resources),
- articles that presented, discussed or evaluated indicators measuring the scientific production of an individual researcher or a country,
- articles that presented a bibliometric or scientometric study,
- articles that presented, discussed or evaluated indicators measuring the quality of a scientific journal and
- articles in languages other than French or English.

We assessed the relevance or quality of articles using a list of criteria. Each article should at least contain one criterion to be selected:

- The article presents an indicator and clearly describes how it is calculated
- The articles evaluates the validity or reliability of the indicator
- The articles evaluates the feasibility of the indicator
- The article discusses the validity of the indicator to measure scientific production

- The article contains proof on the implementation: It describes the possible perverse consequences of measuring the indicator
- The article relates how the indicator was developed and implemented

We noted the reason for exclusion of articles and presented the results according to the PRISMA reporting method [19].

3. Information retrieved

We developed and tested two data extraction forms to collect and organise data about (1) the type of articles selected and the information they produced and (2) details of indicators presented in the articles. The templates of the data extraction forms are available in annex 1 and annex 2 ([S1 Appendix](#); [S2 Appendix](#)).

Using the first form, we retrieved information about each article: the article presents the results of surveys to select one/several indicator(s) (Yes/No), the article relates the development of one/several indicator(s) (Yes/No), the article evaluates the feasibility of one/several indicator(s) (Yes/No), the article evaluates the validity of one/several indicator(s) (Yes/No), the article evaluates the impact of measuring one/several indicator(s) (Yes/No), the article presents any other form of evaluation of the indicator (Yes/No), number and names of indicators. We also collected for each article the name of the journal in which it was published, the impact factor and type of journal, and the domain of research.

Using the second data extraction form, we retrieved the following information for each article: name of the indicator, references in articles, definition of the indicator and details how it is calculated, rationale of the indicator (why it was created), in what context the indicator is used, positive and negative points of the indicator, impact or consequences of measuring the indicator and any additional comments.

When the article presented a mix of relevant indicators and irrelevant indicators (example: inputs indicators), we only retrieved information about the relevant indicators.

The full text of every article was read and data were extracted a first time by FT and proof-read by different authors. The allocation of articles to proofreaders was executed randomly, however the number of articles read differed by reviewers: TD reviewed 16% articles (N = 12), SR reviewed 48% of articles (N = 36), RB reviewed 18% articles (N = 14), CO reviewed 18% articles (N = 14). All differences in opinion were discussed and resolved through discussion.

Results

1. Number and characteristics of articles

After applying filters we retrieved 8321 articles and selected 114 on the basis of title and abstract. Then, 45 articles were excluded after reading the full text, either because of the quality or type of the article (such as commentaries), or because the indicators presented were irrelevant (such as indicators of research inputs, indicators of a journal. . .), because the articles presented a scientometric or bibliometric study and did not contain indicator description, or because the subjects of the articles were irrelevant (articles presenting a policy analysis on research evaluation or exclusively relating the development of indicators). In addition, 5 articles were added by reference, 2 articles were added by the second reviewer. A total of 76 articles were selected. [Fig 1](#) describes the selection of articles ([Fig 1](#)).

The articles were found in 42 different journals and the median impact factor of all articles was 2.1. Almost half of the articles (N = 35) emanated from journals specialized in information science or scientometrics. Overall, 36% of other articles (N = 28) belonged to a medical or public health journal. [Table 1](#) shows the characteristics of the journals in which the articles were

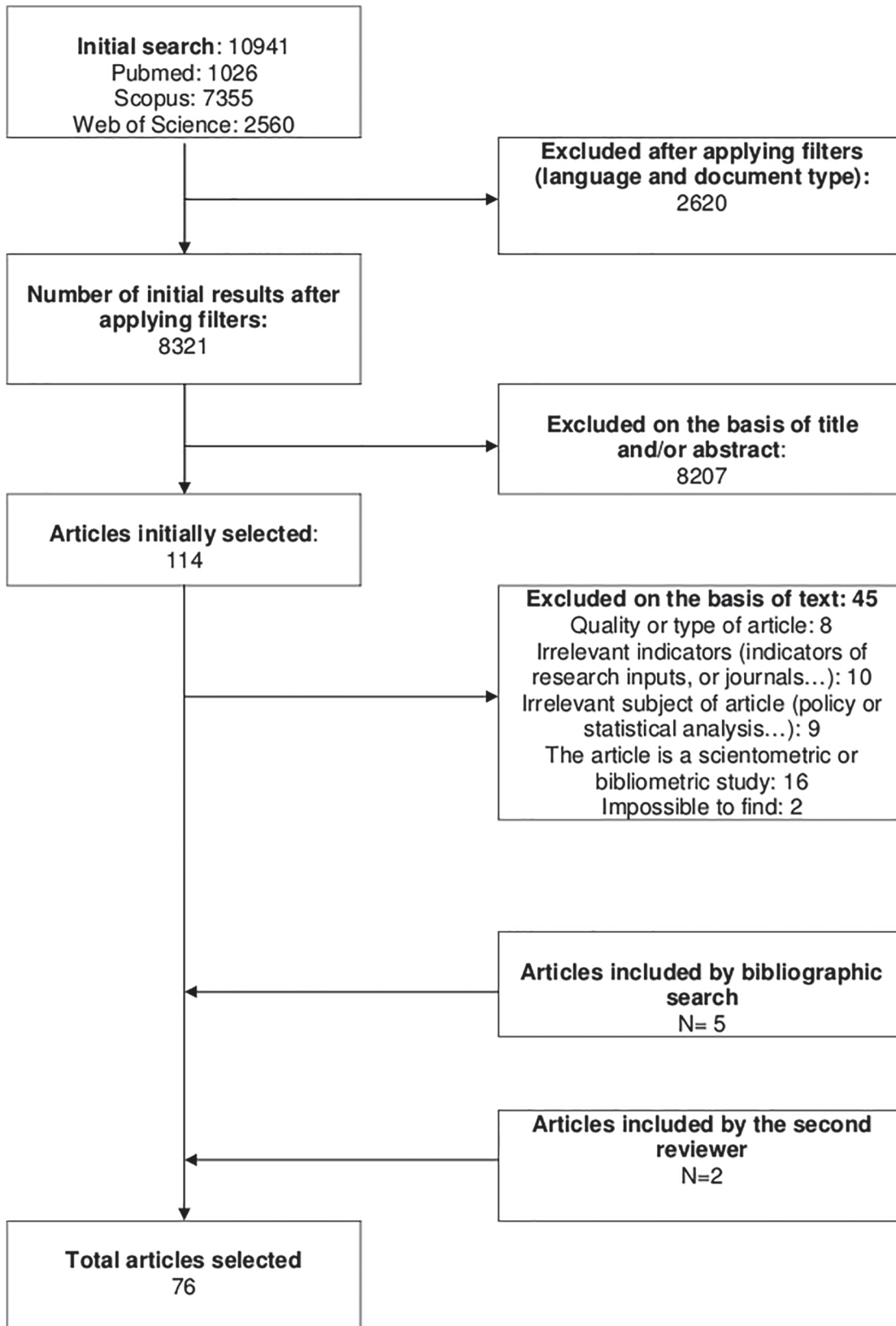


Fig 1. PRISMA flowchart.

doi:10.1371/journal.pone.0122239.g001

Table 1. Type of journals and area of research measured by indicators.

Characteristics (N = 76)	N (%)
Type of journal	
Information science/scientometrics	35 (46%)
Medical speciality	16 (21%)
Public health	6 (8%)
General science	5 (7%)
General medicine	6 (8%)
Biology	4 (5%)
Others (social science, research methodology, pharmacological)	4 (5%)
Research area measured by indicator(s)	
General	47 (62%)
Biomedical research	16 (21%)
Public health	3 (4%)
Psychiatry	2 (3%)
Surgery	2 (3%)
Cancer research	1 (1%)
Chemistry	1 (1%)
Biobanks	1 (1%)
Bioinformatics	1 (1%)
Technological transfer	1 (1%)
Translational research	1 (1%)

doi:10.1371/journal.pone.0122239.t001

published, as well as the research area covered by the article (Table 1: Type of journals and area of research measured by indicators).

2. Content of selected articles

Among all the articles found, 1 article presented the results of a survey to select indicators, 5 articles related the development of one or more indicator(s), 12 evaluated the feasibility of one or more indicator(s), 24 evaluated the reliability or validity of one or more indicator(s), no studies evaluated the impact of measuring one or more indicator(s). Among all articles, 34 studies undertook any other form of evaluation of one or more indicator(s).

3. Indicators

We found 57 indicators presented or discussed in all the articles. We classified those indicators into 6 categories: indicators of research activity, indicators of scientific production and impact, indicators of collaboration, indicators of dissemination, indicators of industrial production and indicators of health service impact, (Table 2: Number of indicators by category). Table 3 summarises the indicators identified, the number of articles in which those indicators are discussed, whether a definition and a methodology for indicator measurement are provided, and whether the positive and negative points of this indicator are mentioned (Table 3: Summary of indicators identified). Complete synthesis of the indicators is enlisted in annex 3. This synthesis is based on reported data and includes the definition of each indicator, the rationale for its creation or its use, and its positive and negative points (S3 Appendix).

- **Research activity indicators.** We found 8 indicators measuring research activity. Indicators of research activity describe the size and diversity of the research pipeline and assess progress towards established milestones. With the exception of one indicator, most of those

Table 2. Number of indicators by category.

Category of indicator	Number of indicators presented (n = 57)
Indicators of activity	8 (14%)
Indicators of scientific production and impact	24 (42%)
Indicators of collaboration	5 (9%)
Indicators of dissemination	4 (7%)
Indicators of industrial production	7 (12%)
Indicators of health service impact	9 (16%)

doi:10.1371/journal.pone.0122239.t002

indicators are presented in one article [20] that does not discuss their positive and negative points. A general comment warns against using solely those kinds of indicators as they would reward an organisation that keeps projects in the pipeline even if they don't appear to be destined for a successful outcome.

• **Indicators of scientific production or impact.** Most of the indicators we found are indicators of scientific production and impact (N = 23). Definition and methodology were provided for all those indicators. For most of them, rationale (N = 20), positive points (N = 20), and negative points (N = 14) were mentioned. The most discussed indicator of that category is the h-index (31 articles). This indicator was created by Hirsch in 2005 and combines measures of quantity (number of publications) and impact (number of citations). It was created to overcome the flaws of other classical bibliometric indicators such as the number of publications (which does not measure the importance of papers), the number of citations (which may be influenced by a small number of very highly cited papers), or the number of citations per paper (which rewards low productivity) [21]. Some of the reported advantages of the h-index include its easy calculation [22–24], its insensitivity to a few highly or infrequently cited papers [22], and the fact that it favours scientists that publish a continuous stream of papers with good impact [25]. Studies have tested this indicator and found evidence that it can predict the future achievement of a scientist [26]; it correlates with peer review judgement [27] and shows a better validity than publication count or citation count alone [28]. Some of its reported flaws include its failure to take into account the individual contribution of each researcher [29], or its low resolution (meaning that several researchers can have the same h-index) [30]. As a result, several indicators have been created to overcome those flaws: the j-index [23], the central index [31], the w-index [32], the e-index [30], the r-index [33], the m-index [34], the m-quotient [24], the citer h-index [35], the q² index [34], the g-index [24] and the hg-index [36].

One criticism of several indicators based on citations (such as the h-index, number of citations, journal impact factor. . .) is that citation practices vary between disciplines [5]. In the field of medicine, for example, basic research is cited more than clinical research [37]. Hence the creation of indicators adjusting the citation rate by disciplines such as the mean normalised citation score [38], b index [24] and the crown indicator [13].

Another indicator subject to much controversy is the journal impact factor. Although this indicator was created to measure the visibility of a journal, it is commonly used to measure scientists and institutions. Some of its criticisms point out that it is influenced by discipline, language, open access policy [5], and can be manipulated with the number of articles [39]. It has been suggested this indicator should be used only to measure journals and not scientists.

• **Indicators of collaboration.** Five indicators were found to measure research collaboration: the dependence degree, the partnership ability index, the number of co-authored publications, the number of articles with international collaboration and the proportion of long-distance collaboration. The rationale for the use of those indicators is that research benefits

Table 3. Summary of indicators identified.

Category	Indicators	Number of articles	Rationale provided	Definition provided	Methodology or calculation provided	Positive and negative points discussed
Indicators of research activity	Number of clinical trials	2	Yes	Yes	No	No
	Number of patients in clinical trial	1	No	No	No	No
	Number of biological samples transmitted	1	No	No	No	No
	Number of research projects ongoing	1	No	No	No	No
	Number of biomarker identified	1	No	No	No	No
	Number of assays developed	1	No	No	No	No
	Number of databases generated	1	No	No	No	No
	Number of visits to the EXPASY server	1	Yes	Yes	Yes	Yes (+/-)
Indicators of scientific production and impact	h-index	31	Yes	Yes	Yes	Yes (+/-)
	Number of publications	16	Yes	Yes	Yes	Yes (+/-)
	Number of citations	14	Yes	Yes	Yes	Yes (+/-)
	Journal impact factor	10	Yes	Yes	Yes	Yes (+/-)
	g-index	6	Yes	Yes	Yes	Yes (+/-)
	Crown indicator	4	Yes	Yes	Yes	Yes (+/-)
	m-quotient	4	Yes	Yes	Yes	Yes (+/-)
	hg-index	3	Yes	Yes	Yes	Yes (+)
	Citer h-index (Ch-index)	2	No	Yes	Yes	Yes (+)
	Mean citations per papers	2	No	Yes	Yes	Yes (+/-)
	b-index	2	No	Yes	Yes	No
	AWCR (age-weighted citation ratio)	1	Yes	Yes	Yes	Yes (+/-)
	Mean normalised citation score	1	No	Yes	Yes	No
	z-factor	1	Yes	Yes	Yes	Yes (+)
	j-index	1	Yes	Yes	Yes	Yes (+/-)
	SP-index	1	Yes	Yes	Yes	Yes (+)
	Number of publications in the top-ranked journals	1	Yes	Yes	Yes	Yes (+/-)
	x-index	1	Yes	Yes	Yes	Yes (+/-)
	Central index	1	Yes	Yes	Yes	Yes (+/-)
	w-index	1	Yes	Yes	Yes	Yes (+/-)
e-index	1	Yes	Yes	Yes	No	
r-index	1	Yes	Yes	Yes	Yes (+)	
m-index	1	Yes	Yes	Yes	Yes (+)	
Q ² index	1	Yes	Yes	Yes	No	
Indicators of collaboration	Number of co-authored publication	4	No	Yes	Yes	No
	Number of articles with international collaboration	2	No	Yes	Yes	No
	Proportion of long-distance collaborative publication	1	No	Yes	Yes	No
	Partnership Ability Index (PHI-index)	1	No	Yes	Yes	Yes (+)
	d-index (dependence degree)	1	Yes	Yes	Yes	Yes (-)
Indicators of dissemination	Reporting of research in the news/media	4	Yes	Yes	No	Yes (+/-)
	Citation in medical education books	2	Yes	Yes	No	Yes (+/-)
	Number of presentations at key selected conference	2	Yes	Yes	No	No
	Number of conference held	1	No	No	No	No
Indicators of industrial production	Number of patents	6	Yes	Yes	Yes	Yes (+/-)
	Number of public-private partnerships	2	Yes	Yes	No	No
	Patent citation count	2	Yes	Yes	Yes	Yes (+/-)
	Number of spin-off companies created	2	Yes	Yes	No	Yes (-)
	Citation of research in patents	2	Yes	Yes	No	No
	Number of papers co-authored with the industry	1	Yes	Yes	Yes	Yes (+/-)
	Patent h-index	1	Yes	Yes	Yes	Yes (+/-)

(Continued)

Table 3. (Continued)

Category	Indicators	Number of articles	Rationale provided	Definition provided	Methodology or calculation provided	Positive and negative points discussed
Indicators of health service impact	Citation of research in clinical guidelines	7	Yes	Yes	suggested	Yes (+/-)
	Contribution to reports informing policy makers	4	Yes	No	No	No
	Citation of research in policy guidelines	3	Yes	Yes	No	Yes (+/-)
	Patients outcomes	3	Yes	Yes	No	Yes (-)
	Public knowledge about a health issue	2	No	Yes	No	Yes (-)
	Changes in legislation/regulations	2	Yes	Yes	No	Yes (+/-)
	Generation of clinical guidelines	1	Yes	Yes	No	No
	Changes in clinical practice	1	Yes	Yes	suggested	Yes (+/-)
	Measures of improved health services	1	Yes	Yes	No	No

In the 'positive and negative points discussed' column: Yes(+) = Yes, only positive points are mentioned, Yes(-) = Yes, only negative points are mentioned, Yes (+/-) = Yes, positive and negative points are mentioned

doi:10.1371/journal.pone.0122239.t003

from collaboration between institutions because it brings new ideas and methods and multifaceted expertise can be reached [40], and therefore evaluation metrics should focus on the interactions between researchers rather than on the outputs of individual researchers [41]. A definition and methodology was provided for all those indicators, but we found little critical discussion about the advantages and disadvantages of using them.

• **Indicators of dissemination.** There were 4 indicators measuring the dissemination of research: citation in medical education books, number of presentations at key selected conferences, number of conferences organised and reporting of research in news or media. Although the rationale and definition was provided for 3 of them, none proposed a methodology. The most discussed indicator was the reporting of research in the news/media. The rationale for the development or use of that indicator is that media are often influential in terms of public opinion and public formation and media reporting of research allows patients to be better informed [7]. It is argued that scientists interacting most with mass media tend to be scientifically productive and have leadership roles and media have a significant impact on public debate [37]. However, criticisms of this indicator include its bias (for example, clinical research is over-represented), its lack of accuracy [37] and lack of evidence that it leads to an actual policy debate [9].

• **Indicators of industrial production.** There were 7 indicators measuring industrial production: the number of public-private partnerships, the number of patents, the number of co-authored publications, the number of papers co-authored with the industry, the patent citation count, the patent h-index, the citation of research into patents, and the number of spin-off companies created. The most widely discussed indicator was the number of patents with 6 articles discussing it and 3 indicators derived from it (patent citation count, patent h-index and citation of research in patents). Although it is mentioned that patent protection enhances the value of an organisation's output and attracts future commercial investment, more criticisms of this indicator are acknowledged, such as its lack of reliability in measuring patent quality and subsequent impact, and its adverse effects on the quality of patents produced by a university [9]. To overcome this flaw, the patent citation count is sometimes used; however most patents are never cited or cited only once or twice. As a corrective, the patent h-index has been

created, that combines measures of quantity and impact of a patent. Other measures of industrial production or collaboration are scarcely discussed and rarely used.

• **Indicators of health service impact.** We found 9 indicators measuring research impact on health or health service. Several articles proposed to measure the impact of medical research in terms of various measures of patients' outcomes (such as mortality, morbidity, quality of life. . .). However those indicators are very challenging to measure [12], and pose the problem of attribution (how to link health improvements to one particular research finding) [9]. Other intermediate outcome indicators for health research have been suggested, such as changes in clinical practice, improvement of health services, public knowledge on a health issue, changes in legislation and clinicians' awareness of research. However no article gave a clear methodology for calculating those indicators, or a way to tackle the attribution problem. The authorship of clinical guidelines or researchers' contribution to reports informing policy makers are other possible indicators of medical research outcome, although there is little discussion on their advantages and disadvantages. More has been written on two indicators: citation of research in clinical guidelines and citation of research in policy or public health guidelines. The indicator 'citation of research in clinical guidelines' has been most widely discussed in that category. It is reported as being easy to calculate and correlated with other quality indicators such as impact factor [37]. But this indicator can only measure research several years after it has been published. It also favours clinical research compared with basic research.

Discussion

Interpretation of results

The aim of our study was to obtain a comprehensive view of existing indicators used to measure the output or outcome of medical research. We found 57 indicators developed, with a majority of bibliometric indicators present in most articles. This finding is consistent with previous review of research indicators [16]. We also found a diversity of indicators, measuring different elements of medical research. We decided to classify them into 6 categories (indicators of research activity, indicators of scientific production, indicators of collaboration, indicators of dissemination, indicators of industrial production and indicators of health research impact). Few articles discussed research activity indicators. The positive and negative points of those indicators were not discussed and no methodology for calculating the indicator was provided, except for the indicator 'number of visits to EXPASY server'. Therefore the second objective of our study (noting the positive and negative points and remarks about feasibility, validity or use of indicators) could not be completed for this category of indicators. More research is needed on this aspect.

Not surprisingly, bibliometric indicators of scientific production were the most represented and discussed category of indicators. The most discussed indicator of that category was the h-index. Several indicators have been developed to improve it or complement it. The h-index has inspired the creation of different indicators belonging to other categories, such as the patent h-index or the partnership ability index. However the focus of bibliometric to measure research production has been criticized. Indeed some of the critics point out that those indicators do not reflect truly the impact of a research work on the scientific community and on public health. For example, Edward Lewis, a scientist famous for his work on the role of radiation on cancer and who won a Nobel Prize had a small publication count and a very low h-index [42]. Another one is the discovery of the pathogenic role of the *Helicobacter Pylori* bacterium, which was first published in the *Medical Journal of Australia*, a journal with an impact factor below 2 [43].

Five indicators measure inter-institutional research collaboration. Some of those indicators are used alongside other bibliometric indicators in evaluation systems such as the Leiden

Ranking [38]. According to Abramo [44], there has been a trend towards an increase in collaboration between institutions. This trend could be attributed to many factors, such as specific policies favouring research collaboration, specifically the EU Research Framework Programmes at the international level [44], or the increased division of labour among scientists [45]. Several studies have measured the impact of co-authored papers and found that they are more highly cited than papers authored by a single institution, and papers with international co-authorships have an even higher citation rate [44–48]. However, it has also been argued that co-authorship is a poor indicator of actual collaboration [49; 50].

In the category ‘indicators of dissemination’, one indicator has been widely reported upon is the citation of research in the mass media. The criticism on its bias and lack of accuracy is consistent with other researches. A study on the reporting of cancer research on BBC website [51] has shown that this media does not broadcast themes of cancer research consistently with their epidemiological burden. For example many cancers such as lung cancer and upper gastro-intestinal tract fare poorly in their media exposure despite an important incidence or mortality. Another research on the reporting of mental health research found similar results [52]. And a study found evidence of poor media reporting of health interventions despite recent improvements [53]. We found some indicators of health service impact. But they seem difficult to measure and present the challenge of attributing particular research findings to health improvements.

Strength and limitations of the study

This study has limitations. We have been able to identify indicators belonging to a broad spectrum and that can measure the outcome of medical research from various perspectives. However this is also one limitation of the study. We chose to design our analysis in order to obtain a broad view of indicators. However, we might not have been able to give an in-depth analysis of the bibliometric indicators. They were not the focus of our study.

We have decided to classify the indicators found into 6 categories but several indicators could belong to more than one category.

Policy implications

Several lessons can be drawn from this study. Given the fact that all indicators have flaws or are incomplete, several studies [54; 55; 27] stressed the importance of using a mix of several indicators rather than just one to measure research outputs. An evaluation system should follow this recommendation. Another important step in the development of indicators is to assess their validity and feasibility. According to the OECD, an indicator is valid when it accurately measures what it is intended to measure and it is reliable when it provides stable results across various populations and circumstances [56]. There are three conditions to assess the feasibility of an indicator: existence of prototypes (whether the measure is in use), availability of internationally-comparable data across countries and cost or burden of measurement [56].

Conclusion

We have drawn a comprehensive list of indicators measuring the output and outcomes of medical research. Not all indicators are suitable to evaluate the outcome of translational research carried out in health facilities. In order to operate a selection of indicators, we plan on investigating the view of concerned researchers about the key indicators to select. We also need to test the feasibility and validity of the selected indicators.

Supporting Information

S1 PRISMA Checklist. PRISMA checklist.

(DOC)

S1 Appendix. Data Extraction Form 1.

(DOC)

S2 Appendix. Data Extraction Form 2.

(DOC)

S3 Appendix. Details on each indicator.

(DOC)

S1 Protocol. Study protocol.

(DOC)

Author Contributions

Conceived and designed the experiments: FT RB MS WvH CA. Performed the experiments: FT RB TD SR CO. Analyzed the data: FT RB TD SR CO. Contributed reagents/materials/analysis tools: FT RB CA. Wrote the paper: FT RB TD SR MS WvH CO CA.

References

1. Lavis J, Ross S, McLeod C, Gildiner A. Measuring the impact of health research. *J Health Serv Res Policy*. 2003 Jul 1; 8(3):165–70 PMID: [12869343](#)
2. Wooding S, Hanney S, Buxton M, Grant J. Payback arising from research funding: evaluation of the Arthritis Research Campaign. *Rheumatology*. 2005 Sep 1; 44(9):1145–56. PMID: [16049052](#)
3. Panel on return on investment in Health Research. Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research. Canadian Academy of Health Sciences, Ottawa, ON, Canada. 2009. Available: <http://www.caahs-acss.ca/wp-content/uploads/2011/09/ROI-FullReport.pdf>.
4. Battersby J. Translating policy into indicators and targets. In: Pencheon D, Guest C, Melzer D, Muir Gray JA, editors. *Oxford Handbook of Public Health Practice*- second edition. Oxford: Oxford University Press. 2006. Pp.334–339.
5. Adams J. The use of bibliometrics to measure research quality in UK higher education institutions. *Arch Immunol Ther Exp (Warsz)*. 2009 Feb; 57(1):19–32. doi: [10.1007/s00005-009-0003-3](#) PMID: [19219531](#)
6. Lascurain-Sánchez ML, García-Zorita C, Martín-Moreno C, Suárez-Balseiro C, Sanz-Casado E. Impact of health science research on the Spanish health system, based on bibliometric and healthcare indicators. *Scientometrics*. 2008; 77(1):131–46.
7. Lewison G. From biomedical research to health improvement. *Scientometrics*. 2002; 54(2):179–92.
8. Mostert SP, Ellenbroek SP, Meijer I, van Ark G, Klasen EC. Societal output and use of research performed by health research groups. *Health Res Policy Syst*. 2010; 8:30. doi: [10.1186/1478-4505-8-30](#) PMID: [20939915](#)
9. Ovseiko PV, Oancea A, Buchan AM. Assessing research impact in academic clinical medicine: a study using Research Excellence Framework pilot impact indicators. *BMC Health Serv Res*. 2012; 12:478. doi: [10.1186/1472-6963-12-478](#) PMID: [23259467](#)
10. Smith R. Measuring the social impact of research: Difficult but necessary. *British Medical Journal*. 2001; 323(7312):528. PMID: [11546684](#)
11. Weiss AP. Measuring the impact of medical research: moving from outputs to outcomes. *Am J Psychiatry*. 2007 Feb; 164(2):206–14. PMID: [17267781](#)
12. Wells R, Whitworth JA. Assessing outcomes of health and medical research: do we measure what counts or count what we can measure? *Aust New Zealand Health Policy*. 2007; 4:14. PMID: [17597545](#)
13. Durieux V, Gevenois PA. Bibliometric indicators: quality measurements of scientific publication. *Radiology*. 2010 May; 255(2):342–51. doi: [10.1148/radiol.09090626](#) PMID: [20413749](#)

14. Froghi S, Ahmed K, Finch A, Fitzpatrick JM, Khan MS, Dasgupta P. Indicators for research performance evaluation: An overview. *BJU International*. 2012; 109(3):321–4. doi: [10.1111/j.1464-410X.2011.10856.x](https://doi.org/10.1111/j.1464-410X.2011.10856.x) PMID: [22243665](https://pubmed.ncbi.nlm.nih.gov/22243665/)
15. Joshi MA. Bibliometric indicators for evaluating the quality of scientific publications. *J Contemp Dent Pract*. 2014; 15(2):258–62. PMID: [25095854](https://pubmed.ncbi.nlm.nih.gov/25095854/)
16. Patel VM, Ashrafian H, Ahmed K, Arora S, Jiwan S, Nicholson JK, et al. How has healthcare research performance been assessed?: a systematic review. *J R Soc Med*. 2011 Jun; 104(6):251–61. doi: [10.1258/jrsm.2011.110005](https://doi.org/10.1258/jrsm.2011.110005) PMID: [21659400](https://pubmed.ncbi.nlm.nih.gov/21659400/)
17. Academy of Health Sciences, Medical Research Council, Wellcome Trust. Medical research: assessing the benefits to society- A report by the UK Evaluation Forum, supported by the Academy of Medical Sciences, Medical Research Council and Wellcome Trust; 2006. Available: <http://www.acmedsci.ac.uk/policy/policy-projects/medical-research-assessing-the-benefits-to-society/>. Accessed: 2014 Jan 10.
18. Canadian Institutes of Health Research. Developing a CIHR Framework to Measure the Impact of Health Research; 2005. Available: <http://publications.gc.ca/collections/Collection/MR21-65-2005E.pdf> Accessed 01/10/2014
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009 Oct; 62(10):1006–12. doi: [10.1016/j.jclinepi.2009.06.005](https://doi.org/10.1016/j.jclinepi.2009.06.005) PMID: [19631508](https://pubmed.ncbi.nlm.nih.gov/19631508/)
20. Pozen R, Kline H. Defining Success for Translational Research Organizations. *Sci Transl Med*. 2011 Aug 3; 3(94):94cm20. doi: [10.1126/scitranslmed.3001970](https://doi.org/10.1126/scitranslmed.3001970) PMID: [21813756](https://pubmed.ncbi.nlm.nih.gov/21813756/)
21. Hirsch JE. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(46):16569–72. PMID: [16275915](https://pubmed.ncbi.nlm.nih.gov/16275915/)
22. Alonso S, Cabrerizo FJ, Herrera-Viedma E, Herrera F. hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices. *Scientometrics*. 2010; 82(2):391–400.
23. Todeschini R. The j-index: a new bibliometric index and multivariate comparisons between other common indices. *Scientometrics*. 2011 Jun; 87(3):621–39.
24. Egghe L. The hirsch index and related impact measures. *Annual Review of Information Science and Technology*. 2010; 44 (1):65–114
25. Bornmann L, Daniel H-D. What do we know about the h index? *Journal of the American Society for Information Science and Technology*. 2007; 58(9):1381–5.
26. Hirsch JE. Does the h index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(49):19193–8. PMID: [18040045](https://pubmed.ncbi.nlm.nih.gov/18040045/)
27. Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*. 2006 Jun; 67(3):491–502.
28. Sharma B, Boet S, Grantcharov T, Shin E, Barrowman NJ, Bould MD. The h-index outperforms other bibliometrics in the assessment of research performance in general surgery: a province-wide study. *Surgery*. 2013 Apr; 153(4):493–501. doi: [10.1016/j.surg.2012.09.006](https://doi.org/10.1016/j.surg.2012.09.006) PMID: [23465942](https://pubmed.ncbi.nlm.nih.gov/23465942/)
29. Franco G. Research evaluation and competition for academic positions in occupational medicine. *Archives of Environmental and Occupational Health*. 2013; 68(2):123–7. doi: [10.1080/19338244.2011.639819](https://doi.org/10.1080/19338244.2011.639819) PMID: [23428063](https://pubmed.ncbi.nlm.nih.gov/23428063/)
30. Zhang C-T. The e-index, complementing the h-index for excess citations. *PLoS ONE*. 2009; 4(5): e5429. doi: [10.1371/journal.pone.0005429](https://doi.org/10.1371/journal.pone.0005429) PMID: [19415119](https://pubmed.ncbi.nlm.nih.gov/19415119/)
31. Dorta-González P, Dorta-González M-I. Central indexes to the citation distribution: A complement to the h-index. *Scientometrics*. 2011; 88(3):729–45.
32. Wu Q. The w-Index: A Measure to Assess Scientific Impact by Focusing on Widely Cited Papers. *J Am Soc Inf Sci Technol*. 2010 Mar; 61(3):609–14.
33. Romanovsky AA. Revised h index for biomedical research. *Cell Cycle*. 2012 Nov 15; 11(22):4118–21. doi: [10.4161/cc.22179](https://doi.org/10.4161/cc.22179) PMID: [22983124](https://pubmed.ncbi.nlm.nih.gov/22983124/)
34. Derrick GE, Haynes A, Chapman S, Hall WD. The Association between Four Citation Metrics and Peer Rankings of Research Influence of Australian Researchers in Six Fields of Public Health. *PLoS ONE*. 2011; 6(4).
35. Franceschini F, Maisano D, Perotti A, Proto A. Analysis of the ch-index: An indicator to evaluate the diffusion of scientific research output by citers. *Scientometrics*. 2010; 85(1):203–17.
36. Franceschini F, Maisano D. Criticism on the hg-index. *Scientometrics*. 2011; 86(2):339–46.
37. Lewison G. Beyond outputs: New measures of biomedical research impact. *Aslib Proceedings*. 2003; 55(1–2):32–42.

38. Waltman L, Calero-Medina C, Kosten J, Noyons ECM, Tijssen RJW, van Eck NJ, et al. The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *J Am Soc Inf Sci Technol*. 2012 Dec; 63(12):2419–32
39. Wallin JA. Bibliometric methods: Pitfalls and possibilities. *Basic Clin Pharmacol Toxicol*. 2005 Nov; 97(5):261–75 PMID: [16236137](#)
40. Koskinen J, Isohanni M, Paajala H, Jääskeläinen E, Nieminen P, Koponen H, et al. How to use bibliometric methods in evaluation of scientific research? An example from Finnish schizophrenia research. *Nord J Psychiatry*. 2008; 62(2):136–43. doi: [10.1080/08039480801961667](#) PMID: [18569777](#)
41. Schubert A. A Hirsch-type index of co-author partnership ability. *Scientometrics*. 2012 Apr; 91(1):303–8.
42. Lawrence PA. Lost in publication: How measurement harms science. *Ethics in Science and Environmental Politics*. 2008; 8(1):9–11.
43. Baudoin L, Haeffner-Cavaillon N, Pinhas N, Mouchet S, Kordon C. Bibliometric indicators: Realities, myth and prospective. *Medecine/Sciences*. 2004; 20(10):909–15. PMID: [15461970](#)
44. Abramo G, D'Angelo CA, Solazzi M. The relationship between scientists' research performance and the degree of internationalization of their research. *Scientometrics*. 2011; 86(3):629–43.
45. Frenken K, Hölzl W, Vor FD. The citation impact of research collaborations: The case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management—JET-M*. 2005; 22(1–2):9–30.
46. Kato M, Ando A. The relationship between research performance and international collaboration in chemistry. *Scientometrics*. 2013; 97(3):535–53.
47. Vanecek J, Fatun M, Albrecht V. Bibliometric evaluation of the FP-5 and FP-6 results in the Czech Republic. *Scientometrics*. 2010; 83(1):103–14.
48. Kim Y, Lim HJ, Lee SJ. Applying research collaboration as a new way of measuring research performance in Korean universities. *Scientometrics*. 2014; 99(1):97–115.
49. Katz JS, Martin BR. What is research collaboration? *Research Policy*. 1997; 26(1):1–18.
50. Lundberg J, Tomson G, Lundkvist I, Skår J, Brommels M. Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics*. 2006; 69(3):575–89.
51. Lewison G, Tootell S, Roe P, Sullivan R. How do the media report cancer research? A study of the UK's BBC website. *Br J Cancer*. 2008 Aug 19; 99(4):569–76. doi: [10.1038/sj.bjc.6604531](#) PMID: [18665166](#)
52. Lewison G, Roe P, Wentworth A, Szmukler G. The reporting of mental disorders research in British media. *Psychol Med*. 2012 Feb; 42(2):435–41. doi: [10.1017/S0033291711001012](#) PMID: [21676283](#)
53. Wilson AJ, Bonevski B, Jones A, Henry D. Media reporting of health interventions: Signs of improvement, but major problems persist. *PLoS ONE*. 2009; 4(3).
54. Costas R, Bordons M. Is g-index better than h-index? An exploratory study at the individual level. *Scientometrics*. 2008; 77(2):267–88.
55. Waltman L, Van Eck NJ. The Inconsistency of the h-index. *J Am Soc Inf Sci Technol*. 2012 Feb; 63(2):406–15.
56. Kelly E, Hurst J. Health Care Indicators Project Conceptual Framework Paper. OECD Publishing; 2006. Available: <http://www.oecd-ilibrary.org/docserver/download/519t19m240hc.pdf?expires=1407506858&id=id&accname=guest&checksum=2829E0055019EAF33465ADA42DA6A46B>. Accessed: 2014 Jan.