

BIBLIOMETRIC RESEARCH OUTPUT INDICATORS AND UNIVERSITY FUNDING IN THE NORDIC COUNTRIES

Hampus Rabow

Funding of research and the role of governments

Traditionally, decisions about research funding have mainly been informed by three basic criteria. One has been the potential practical utility of the research, another, the quality of the research as perceived by peers, and finally the conformity with the preconceptions or imagination of the funders. The last criterion is generally perceived as illegitimate and will not be further considered here. The first criterion has with the advancement of science and technology comes to be more and more dependent on the second. In other words, it has become increasingly more difficult for a non-peer to determine the probability that a certain research project will provide some beneficial applications.

There are of course usually easily discernible differences between research primarily devoted to some intra-scientific problem, and that with some extra-scientific application in mind. Between these two extremes, there are however a whole spectrum of possibilities. It is also frequently the case that some more theoretical research is discovered to have practical applications, whereas many applied research projects turn out to be practically useless. Another important aspect is that practical applications of research frequently appear in other contexts than those originally intended. This is not the place to discuss the complex relationship between theoretical research and possible applications, but it is important to note that the need to protect basic research from the demand for immediate results has been of fundamental importance for the development of the modern research system.

When, towards the end of WWII, Vannevar Bush wrote his famous report, which stressed the need for government funding of science and education, and outlined, what would eventually become, the *National Science Foundation*, he also formulated five essential principles which may be summarized (somewhat simplified) as follows: (1) The funding must remain (more or less) stable over several years, (2) decision makers should be non-partisan professionals with a good understanding of science and education, (3) the funding agency may only distribute funds to outside institutions such as colleges, universities, and research

institutes, (4) as long as the provisions of the funding application are followed, the funding agency can not exercise any influence over the research, and (5) the funding agency should be assured “complete independence and freedom for the nature, scope, and methodology of research” while at the same time being responsible to the executive and legislative branches of government.¹ These five principles were derived from the necessity to sustain “basic research”. From a global perspective, basic research was required to make progress in applied research possible, and from a national perspective, basic research was required to maintain a competitive edge. As Bush put it: “A nation which depends upon others for its new basic scientific knowledge will be slow in its industrial progress and weak in its competitive position in world trade, regardless of its mechanical skill.”²

The road towards bibliometrics

The system proposed by Vannevar Bush turned out to be quite stable and successful, and was, with some minor modifications, duplicated in many countries around the world. Yet, the increased demand for government funding of science also created a demand for increased accountability, which would seem to require a more efficient and transparent (to the taxpayers) process for allocating resources. Apart from the lack of transparency to outsiders, the peer review process for grant applications tended to be either too unreliable or too costly. A reliable review required many days of work both by the applicants and the reviewers. In the early 1960s, political criticism of the *NIH* led to a large evaluation carried out by experienced research administrators. The final controversial report suggested that NIH reviewers lacked competence in certain areas and proposed that Bush's third principles should be implemented. It also recommended an increase use of “administrative devices” in the decision process.³

At the same time, the complexity and size of the research divisions of some large corporations had grown so much, that the research managers needed

¹Bush (1945) esp. pp 32-33

²Ibid p 19

³For a summary see Greenberg (1965). Criticism in Cooper (1965)

some systematic and science-based approach to monitoring and quality control. In 1958, the main US organization for research directors began publishing an academic journal called *Research Management* adopting approaches from operations research and management science to the problems of research administration. During this period, new successful methods for the quality control of production had been developed in the Japanese manufacturing industry, and many research managers believed that similar methods could benefit research & development.⁴ This was also the period of the Cold War, and there was a widespread fear that the Soviet ability to steer resources to prioritized areas of military research would give them an advantage in the arms race. During the early 1960s both the US navy and NASA sponsored large projects to increase the efficiency of research management.⁵

In 1963 the physicist and historian of science Derek de Solla Price proposed a new research programme in order to solve many of the above-mentioned problems. This research programme was first called the “science of science” and later “scientometrics”, and it laid the foundation for modern applications of *bibliometrics* in the areas of science and technological innovation.⁶ Price wanted a unified approach where knowledge from the history, sociology and psychology of research would inform the statistical models used in research administration and thus provide support for decision makers in the realms of research management and science policy.⁷ Although the overall programme had very limited success, some of Price’s core ideas gradually became more and more influential. He proposed that the new bibliographic database called *Science Citation Index* could be used as the main data source for a statistical analysis of scientific research. It would be possible to examine and measure the spread of scientific information, model the structure of science (in terms of research fields and their relatedness) and study patterns of cooperation between individuals and institutions.

One of the major areas of application for the new bibliometric methods was, just as Price had envisioned, for research evaluation and funding decisions. At the same time, it was initially difficult to find the right place for this kind of studies as an aid in funding decisions. The bibliometric indicators were often seen as crude and unreliable in comparison with well established methods of peer review. The main use of bibliometrics in this context has been to calculate performance indicators that help ascertain the

quantity, quality and impact of previous research. A number of studies have shown that simple bibliometric indicators correlate well with the results of large complex evaluations based on peer review. Problems only occurred when too few of the most important publications were included in the citation database.⁸

What you measure is what you get

Administrative decision makers in many types of organizations often use performance indicators to create negative and positive feedback loops that counteract or promote certain types of behaviour. The indicators may also be used simply to see how the input in terms of resources and funding corresponds to the output in terms of some useful products. In the scientific research system, bibliometric research indicators may thus be used to steer the research activity in a certain direction or simply to see how much funding is required to produce a certain quantity of scientific output or impact.

In their classic paper on “the balanced scorecard” Kaplan and Norton stressed the importance of the positive feedback loop with the use of performance indicators. As they succinctly put it: “What you measure is what you get”.⁹ Outcomes not reflected in the indicators may be ignored in funding decisions, and individuals or units tend to give priority to the activities that are counted. Rather than trying to find extremely sophisticated indicators that are sensitive to all important activities, Kaplan and Norton’s solution was to use a complex scorecard, where different perspectives were balanced against each other. Thus most forms of “Larsen effects”, where a particular activity is given too much priority, may be avoided. On the other hand, a more complex system may be seen as less transparent, and the direct steering effects may be smaller. Ronald L. Straight has stressed the need to clearly define the goals, metrics, and weights, in any performance metric system.¹⁰ Applying his principles to bibliometric performance indicators, the goals implied in the metric must be generally accepted among the scientists, the metric must match these goals as closely as possible, and the weights should be explicitly adjusted to match research priorities. Looking at the bibliometric indicators most frequently used in research evaluation from the 1960s and onwards, they fulfil Straight’s first criterion reasonably well in most cases. Goals like “scholarly impact”, “submission to qualified peer review”, and “productivity” are widely accepted in the scientific and scholarly communities. For the second and third criteria, however, the situation has been more ambiguous. This may be one of the reasons why the

⁴See Montgomery (2007) for an introduction to modern quality control. It includes a brief historical overview.

⁵For more examples see e.g. Dedijer (1966)

⁶Price (1963). See also Garfield (1987) for an appraisal.

⁷Price (1965)

⁸E.g. Norris & Oppenheim (2010), Oppenheim (2007)

⁹Kaplan & Norton (1992)

¹⁰Straight (2000)

general use of bibliometric indicators in funding decisions has been slow to gain acceptance in the Nordic countries

Bibliometrics and government funding in Nordic countries

In the early 90s several international bodies promoted a greater international conformity of the evaluation of higher education and research. The most important initiative for the Nordic countries was probably a pilot project launched by the European Commission in 1994. The aim was to develop a unified approach to academic evaluation in all European countries. All Nordic countries took part in the the project despite some not being EU members.¹¹ These evaluations did not generally make much use of bibliometrics, but they were an important step towards standardized performance indicators (even if only in the form of qualitative reviewer scores). At this time there were also many *ad hoc* research evaluations performed for various purposes and such evaluations, especially in the fields of technology, medicine and natural science frequently made use of bibliometrics as a complement to more qualitative peer review. This increased focus on academic evaluations also led to attempts to follow up on such evaluations with strategic funding of institutions with a high performance or potential. Finland created special *Centres of Excellence*, and Sweden transferred the money from the employee funds to a new foundation for “strategic research”, in order to support research “of the highest international standard”.¹² On the other hand, the core research funding for colleges and universities were largely left intact, except in so far as that funding depended upon external grants.

The Norwegian model

Not until around the year 2000 was the next step taken towards a more direct feedback from university research performance to government funding. In the Norwegian government report *Frihet med ansvar* (“Freedom with responsibility”), the committee proposed that the core funding to the institutions should be made partly dependent on “the results of the institutions’ activities”.¹³ This proposal was later followed by a decision to create a national standardized publication database and a bibliometric output indicator for Norwegian research. In 2003 the university of Oslo introduced a very simple output indicator called the “publication score”. This publication score was calculated by giving different weights to publications depending on whether the *publication channel* (journal, book or series) had an international scope or not. The publication channel

¹¹Thune & Staropoli (1997)

¹²SOU 1996:29 p 186-187

¹³NOU 2000:14 p 46

also had to have a generally acknowledged scholarly or scientific importance for the publication to be counted at all. The amount A_f given to each faculty f was calculated from the publication score P and a “cost factor” C according to the following formula: $A_f = A * (P_f / \sum P_t)$, where $P_t = P * C_f$. A was the total amount of funding to distribute, and the cost factor C (which was optional) was supposed to reflect different research costs at different faculties.

The Oslo model was clearly too simplistic to be useful, but it were to provide the basis for the new national output indicator.¹⁴ Only scholarly and scientific publication channels with at least a national reach were accepted. Out of these, around 20% with an especially high international importance were placed in “level 2” and assigned a higher weight. The selection of publication channels was done by specially selected committees of researchers for around 40 different subject areas. Three different publication types were used: (1) articles in a journal or series, (2) monographs, and (3) book chapters in anthologies. Thus no distinction was made between e.g. different types of articles or between different kinds of anthologies or between conference proceedings and yearly reviews. Once the type of a publication and the level of its channel had been determined, a publication score was assigned as follows:

Publication type	Level 1	Level 2
Monograph	5	8
Article in journal or book series	1	3
Article in anthology	0.7	1

The Swedish model

In Sweden, a new model for resource allocation was proposed first in general in 2005, and then in details in 2007. Rather than using the simple Norwegian output indicator, based upon the quantity of output and the general importance of the publication channels, a citation-based model was proposed. The citation model had been developed by Ulf and Erik Sandström and was based upon standard bibliometric indicators such as the Leiden “Crown Indicator”, but with some modifications¹⁵. The Sandström model was approximately as follows (the actual model has been a little simplified here): A citation score was calculated based on all publications from each college or university in the *Web of Science* during a 4-year period. The number of citations to these articles was divided by the world average for the same research area

¹⁴See Sivertsen (2003) and Sivertsen (2007)

¹⁵Sandström & Sandström (2008) and SOU 2007:81

and the same type of article. When many universities had contributed to the same publication, the citation score was divided equally among them. The total citation score would thus indicate the total impact of a university or college in the corresponding research area. In order to adjust for the difference in productivity between different research areas, the total citation impact was adjusted by a “Waring factor” corresponding to the estimated average productivity for each of 23 different research areas.

There were several reasons why Sweden preferred the Sandström model to the Norwegian model. The division into publication types and publication channels was seen as too crude and overly simplistic. Another problem was the disregard for differences in productivity between different research areas. A study by the Swedish Research Council had shown how an average doctoral dissertation in Medicine only gave 28% of the score of a Humanities dissertation. A third problem was that the Norwegian model ignored the actual scientific impact of the research. Seglen and others have shown that there is very little correlation between the impact of individual papers and the overall impact of their respective research channels.¹⁶ Finally, it was felt that the Norwegian model was too conservative and failed to send any clear signals to researchers.

The Sandström model was not completely uncontroversial in Sweden. The Research Council argued that it had several flaws, and especially that the adjustments made for differences in productivity between different subject areas were based on an erroneous model. Another difficulty was that some research areas, especially in the Humanities, only had few of their publications in the form of journal articles, and only some of these journals were indexed in the *Web of Science*. Thus, it was decided that actual citation impact would be ignored for the humanities.

Other Nordic countries

Denmark began to implement their model slightly after Sweden and thus had the option to base their model on any of the previous two. In the end, they decided to use a version of the Norwegian model, though it is not entirely clear why. When reading the Danish discussion of the Swedish model, it is difficult to avoid the conclusion that it is partly based on misunderstandings. The Danish model contained some improvements, notably some clarification of ambiguities and a more complete list of publication types.

Recently Finland has also decided to use a version of the Norwegian model with the full implementation beginning in 2015. Finland has added an additional

level for publication channels, meant to indicate a channel of the highest international importance.

The situation in Iceland is a bit special with only one large university and six smaller. Only a small amount of the total public R&D expenditures could be labelled “competitive funding”, mostly through council grants and centres of excellence.

Discussion

When discussing bibliometric output indicators in relation to funding, it is important to distinguish clearly between two different functions. The bibliometric scores may, on the one hand, be used to indicate the *overall volume* of various research activities, and on the other to indicate the fulfilment of some *normative criteria* for these research activities. The main purpose of using bibliometric in the first case is simply to relate output to input. In other words, the funders have a legitimate interest to know what they get for their money. If an institution produces comparative little in relation to the resources given, there are legitimate questions to be asked. The indicators are used mainly as a gauge of efficiency. In the second case, funding is used to steer research activity according to certain policies or epistemic norms. Here the indicators are used mainly to regulate behaviour.

As was mentioned above, bibliometric performance indicators have often been used with a limited adherence to Ronald Straight's second and third principles. In the case of the Norwegian model, there is also considerable ambiguity concerning the first principle. Is the indicator meant to have some kind of positive effect on the *kind* or *quality* of research produced, or is it primarily intended to relate the supplied resources to the achieved outcomes? With the Swedish model, the authors have been very clear that it has a dual function. It is not only meant to indicate to what extent the objectives have been fulfilled, but also to reward research institutions with a significant international impact, and to punish institutions whose research fails to impress the wider research community.¹⁷ As concerns the second principle, the choice of adequate metrics, it is of course impossible to evaluate the metrics without a clear notion of the purpose. But if the only purpose is to have an indicator of efficiency and help estimate the amount of resources required for a certain volume of research, then it should be imperative that the research indicator actually correlates well with the actual amount of research activity, with adjustments for differences in efficiency. In the development of the Swedish model, this problem has been at the forefront, even if the solution is far from perfect. In the development of the

¹⁶Seglen 1997

¹⁷Sandström & Sandström (2008) p 8. Compare also Vetenskapsrådet (2012)

Norwegian model, the problem has been largely ignored. Straight's third principle relates to priorities. Here it is essential that the funders and recipients agree what the priorities are. This obviously does not mean that all researchers should agree with the government's priorities, but they should be aware what they are and be able to see how they are reflected in the metrics.

The current Swedish model has recently come under review, and Anders Flodström's report has suggested that Sweden move towards something more akin to the Norwegian model.¹⁸ Unfortunately Flodström's discussion of bibliometrics is often quite difficult to follow. One example: "Either the quality is measured through how many times the publication has been cited in other publications or an impact factor for the journal where the publication occurs is used."¹⁹ Perhaps this is simply a very careless statement, which should not be given too much weight, but bibliometrics is based on *statistics* and *indicators*. Bibliometrics can never "measure" the "quality" of a scientific publication. Another example is when Flodström discusses the adjustments of the citation score with regard to the the different levels of productivity (as measured by publications) in different research areas. As was mentioned above, the solution provided by Ulf and Erik Sandström was far from perfect, but Flodström has apparently not given it much thought at all. He writes as follows: "The current distribution system makes use of the Waring model, which takes account of different research areas through the field adjustments of publications and field normalization." Perhaps this is simply a grammatical mistake, but taken together with other statements²⁰, and hardly any discussion of the issues involved, it is difficult to avoid the above conclusion.

A fundamental problem with all performance indicators is that adequacy frequently must be sacrificed for simplicity and transparency. A more advanced indicator may conform better to the desired outcomes ("what you measure is what you get"), but it may be difficult to fully understand by administrators and practitioners. Also it may make comparisons over time, or with other similar organizations more difficult. There are today many advanced bibliometric indicators that reflect intuitive concepts like "scientific impact" or "productivity" fairly well. On the other hand, a funding system based on such indicators would probably be rejected, since the considerations involved would be insufficiently understood outside the world of bibliometric research. Thus it may be preferable to use simpler indicators, such as those in the Sandström model, although with some improvements. One of the major problems with citation based measures has been the lack of coverage for certain academic fields. Recently, however, the *Web of Science*, has been complemented with a *Book Citation Index*. Thus this problem will now only be a major issue for a few research areas, and for these areas some improved version of the Norwegian model is probably to be preferred. A major weakness with the Sandström model is that it ignores the status of the publication channels altogether. This has the effect that there are several years of delay in the feedback loop, and that world class research may be ignored, if bad luck with the timing prevents it from being cited. For these reasons, an improved version of the Sandström model with the addition of a channel based score would probably be the best solution for the Nordic countries.

Hampus Rabow Malmö Högskola, Sweden, hampus.rabow@mah.se

¹⁸Flodström (2011)

¹⁹Flodström (2011) p 34

²⁰ e.g. "With the use of the Waring method the citations are field normalized" and "the productivity per researcher may be seen directly" (p 35)

References

- Bush, V., 1945. Science, the endless frontier. AYER Co. Pub.
- Cooper, J.D., 1965. Onward the Management of Science: The Wooldridge Report. Science 148, 1433–1439.
- Dedijer, S., 1966. The science of science: A programme and a plea. Minerva 4, 489–504.
- Flodström, A.. 2011. Prestationsbaserad resurstilldelning för universitet och högskolor. Stockholm.
- Garfield, E., Essays of an Information Scientist, Vol: 10, p. 72, 1987 Current Contents, # 11, p. 3, March 16, 1987.
- Greenberg, D.S., 1965. NIH Study: Wooldridge Committee Praises Past Efforts, But Urges Major Organizational Revisions. Science, New Series 147, 1556–1559.
- Karlsson, S & Jacobsson C., 2007. Vad är en svensk doktorsavhandling värd enligt det norska systemet attvärdera publikationer? Bilaga 7 SOU 2007:81 Resurser för kvalitet.
- Kaplan, R.S., Norton, D.P., 1992. The balanced scorecard—measures that drive performance. Harvard business review 70, 71–79.
- Montgomery, D.C., 2007. Introduction to statistical quality control. Wiley-India.
- Norris, M., Oppenheim, C., 2010. Peer review and the h-index: Two studies. Journal of Informetrics 4, 221–232.
- NOU 2000:14 Frihet med ansvar : om høgre utdanning og forskning i Norge : utredning fra et utvalg oppnevnt ved kongelig resolusjon 30. april 1998. Statens forvaltningstjeneste, Informasjonsforvaltning :, Oslo.
- Oppenheim, C., 1997. The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. Journal of documentation 53, 477–487.
- Price, D. de S., 1963. Little Science, Big Science. New York: Columbia University Press.
- Price, D. de S. J., 1965. The Science of Science. Bulletin of the Atomic Scientists 2–8.
- Sandström U. & Sandström E., 2008. Resurser för citeringar Rapport 2008:18 R Högskoleverket.
- Seglen, PO 1997. Why the impact factor of journals should not be used for evaluating research -- Seglen 314 (7079): 497 -- bmj.com 314.
- Sivertsen, G., 2003. Bibliografiske datakilder til dokumentasjon av vitenskapelige publikasjoner: En utredning for Utdannings-og forskningsdepartementet. Oslo: NIFU.
- Sivertsen, G., 2007. "Den norske model": Et bidrag til diskusjonen i Sverige.
- SOU 1996:29 Forskningsfinansieringsutredningen. *Forskning och pengar: slutbetänkande*. Stockholm: Fritze
- SOU 2007:81 Resursutredningen. *Resurser för kvalitet: slutbetänkande*. Stockholm.
- Thune, C., Staropoli, A., 1997. «The European Pilot Project for Evaluating Quality in Higher Education». Standards and quality in higher education 198–204.
- Vetenskapsrådet (2012) Remissvar till rapporten Prestationsbaserad resurstilldelning för universitet och högskolor.
http://www.vr.se/download/18.13384c8f135aad61b5516f6/GD-beslut+28_120313_Remissvar+resurstilldelning.pdf