

The effects of performance-based research funding systems

Renewed discussions on the relationship between productivity and citation impact in research

The effects of performance-based research funding systems are among the most contested issues in both science policy studies and broader academic discussions. A key question in these debates, the relationship between publication quantity and citation impact, has recently gained renewed attention in the scholarly debates. This Policy Brief highlights major points in these debates and discusses potential policy implications.

Gunnar Sivertsen (NIFU) and Kaare Aagaard (CFA/Aarhus University)

1. Performance-based research funding systems under scrutiny

Performance-based research funding systems have gained importance in recent decades (Hicks 2012). The background has partly been an introduction of market mechanisms to the higher education sector in order to enhance the accountability, effectiveness and legitimacy of public research, and partly a growing notion of the importance of public research in the development of competitive knowledge societies (Whitley and Gläser 2007). While the first of these systems already were introduced in the late 1980s and the early 1990s, they have become prominent across a wider set of countries since the turn of the millennium (Aagaard 2015). Broadly speaking, the current systems can be placed in two general categories (Sivertsen 2017). Some countries (Italy, Lithuania, Portugal, and United Kingdom) combine the purpose of funding allocation with the purpose of research evaluation. The evaluation is organized at intervals of several years and based on expert panels applying peer review. In several other countries (Croatia, the Czech Republic, Denmark, Estonia, Flanders (Belgium), Finland, Norway, Poland, and Sweden), the funding allocation is based on a set of indicators that represent research activities. Bibliometrics is often part of the set of indicators and is also the focus of this article. The indicators are used annually and directly in the funding formula. None of these models are identical even if they may be inspired by each other. They all have their own strengths and weaknesses in terms of cost and organization, valuation methods, transparency and legitimacy, extent of allocation of resources, and presumably also in terms of impacts on research and academic publication (Aagaard, Bloch and Schneider 2015).

Although a number of studies have been undertaken over the years, our understanding of how these systems impact research is still limited (Whitley and Gläser 2007). As Butler (2010) notes:

‘Assessing the impact of PRFS is a fraught exercise, which perhaps explains the paucity of broad authoritative texts on the subject. The literature [. . .] contains relatively few concrete examples where the impact of PRFS has been examined in detail’. This quote by Butler has become even more relevant during the latest year, where a re-examination of her landmark studies from the early 2000’s has led to renewed discussions. We will highlight two of these discussions here: The first concerns the general challenges related to the assessment of causal effects of performance-based systems, while the second more specifically revolves around the relationship between “quality” and quantity. Notice, that the term “quality” is put in brackets here, as the studies under examination in fact only look at measures of citation impact. At best, such measures can be seen as no more than an imprecise and partial proxy of research quality.

2. The effects of the Australian system of the 1990’s

Linda Butler’s seminal studies demonstrated how Australian researchers presumably changed their collective publication behaviour in the 1990s in response to a new national funding model, partly based on productivity measures undifferentiated by any measure of “quality” (Butler, 2002, 2003a,b, 2004). Overall, her studies showed that publication activity increased, but mostly in lower-impact journals, leading to a general drop in general citation impact for Australia. The results have since then been highlighted as a warning of what would most likely happen if funding is linked to publication activity.

This warning has however been challenged recently in a new series of studies from Ulf Sandström, Peter van den Besselaar and Ulf Heyman (Sandström and Besselaar 2016; Besselaar, Heyman & Sandström 2017). Based partly on a critique of Butler’s studies and partly on new studies they claim that an

emphasis on quantity should be encouraged as it eventually also leads to higher “quality”. The claims from the Swedish/Dutch duo have subsequently stirred a lot of debate and resistance as the policy implications run counter to a growing awareness of the perceived negative effects of publication pressure and in fact may increase the overwhelming number of publications from public research institutions. Yet to be debated, is Besselaar & Sandström’s (2017) most recent claim that the Norwegian model (also used in Denmark and Finland) tries to stimulate quality unsuccessfully, while the former Australian model stimulated quality without trying to do so. The parallel Swedish model, which was designed by Sandström (2008) ten years ago, also remains to be analysed and debated.

3. To what extent can we isolate the effects of performance-based research funding systems?

The first of the discussions sparked by these studies concerns the difficulties in isolating the effects of performance-based funding systems. Several contributions from international experts to volume 3/2017 of *Journal of Informetrics* discuss this question from different angles. A key issue is to what extent changes in research behaviour can be attributed to a specific policy mechanism when this mechanism functions in extremely complex systems, where disciplinary cultures interact with both local, national and international incentive structures. Different methodological designs can be utilized to address such challenges, but none are able to solve all problems. Among the available studies, Butler’s original examination of the Australian model still stands out as one of the more robust, although the data still leave room for other interpretations of the effects. But while the recent studies have led to interesting debates, there is, in general, limited support to the claims of Sandström and van den Besselaar in the discussions among international experts.

4. Should productivity be stimulated or will such measures lead to lower impact papers?

The discussions are however not only addressing this overall issue. There is also a parallel, more specific, debate concerning another challenge: Even in cases where relatively clear effects are found, this may hide a heterogeneous underlying picture and/or be the product of data limitations. A good example of this problem can be found in the discussions concerning the relationship between

R-QUEST Policy Briefs

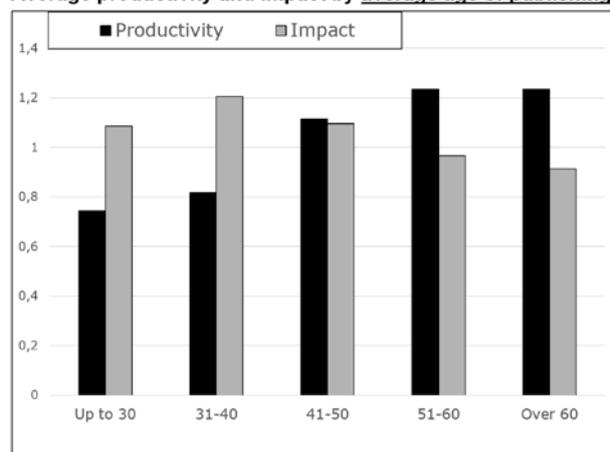
Centre for Research Quality and Policy Impact Studies ([R-QUEST](#)) is devoted to studies of the notion of research quality, conditions for high quality research and its effects on the society. Future R-QUEST policy briefs will communicate results from this research – published at www.r-quest.no.

R-QUEST is sponsored by the RCN’s [FORINNPOL](#) programme for the period 2016-2024. It is managed by [NIFU](#) and involves a number of [research partners](#) as well as [policy/user partners](#).

“quality” and quantity in general. In two independent studies first presented at ISSI 2015 and recently published in PLOS ONE, Larivière & Costas (2016) and Sandström & van den Besselaar (2016) observe similarly that productivity among individual researchers is correlated with citation impact in large datasets from Web of Science (WoS). However, they draw very different policy implications from this observation.

Larivière & Costas (2016) frame their study within the increasing literature that warns against the possible adverse effects of quantitative research assessment and funding methods. They study whether the incentive to publish as much as possible leads to lower citation impact, but find that higher productivity does not influence citation impact negatively. They explain this as a confirmation of the Mertonian theory of cumulative advantages in research and maintain that, in line with the Leiden Manifesto (Hicks et al., 2015), individuals are best assessed by qualitative peer review. Sandström & van den Besselaar (2016) also relate their discussion to the Leiden Manifesto and the policy discussion about a possible overproduction of low impact research. But they take an opposing position and conclude that increased productivity should not be seen as a perverse effect of output oriented evaluation systems, but rather as a positive development. Interpreting quantity as the number of WoS-articles and citation impact as quality, they find that there is no evidence that quantity and quality are opposed to each other. Instead, their view is that evaluation methods based on peer review of only a few publications “disadvantages the most productive and best researchers”. They argue that there is higher chance of breakthroughs with repeated tries and more experience, mentioning that e.g. Nobel laureates often have many more publications than normal researchers. A common feature of the two studies is however that they identify individual researchers by using author

Average productivity and impact by average age of publishing



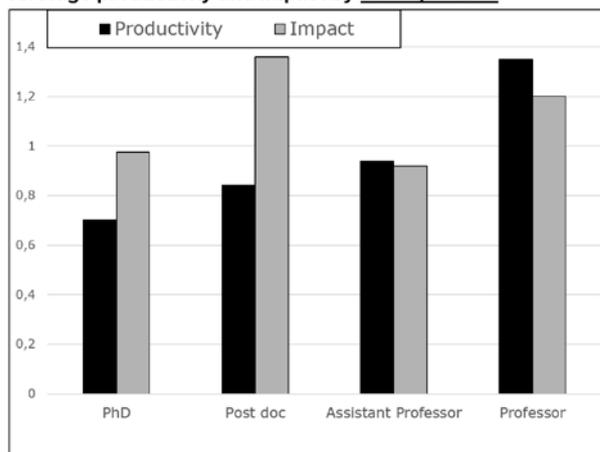
name disambiguation within data from WoS, and that they do not go beyond this database for further information about the researchers. This has clear implications for their results.

5. A new study based on more complete data

The claims of these two studies have very recently been re-examined by Lin Zhang and Gunnar Sivertsen (2017) in a study presented at the 2017 ISSI conference in Wuhan, China. Using combined datasets comprising 17,750 researchers in Norway, Zhang and Sivertsen include data on the age, gender and position of the researchers, as well as their former career and educational background in the higher education sector. In addition, they include all peer reviewed scholarly and scientific publications of these researchers – not only those covered by WoS. The study shows that a more complete measurement of productivity is in fact very weakly correlated with citation impact in Web of Science, even within fields where the database coverage is very high. In addition, they observe that while average productivity increases with seniority, average citation impact does not. It is the highest for post-docs and in general for researchers in their 30's. Men are on the average more productive than women, but the difference in citation impact is small, as found in earlier studies (Aksnes et al. 2011).

Examples of the results of the new study are shown in the two figures above. Here, citation impact is measured by a field-normalized relative citation indicator by which the world average in the same year and field is equal to 1. Productivity is measured (on the same scale in the figure) as the average annual number of publications. Multi-authored publications are divided between authors using the square root of the authors' fractions, a method that

Average productivity and impact by latest position



has been demonstrated to balance well between different publishing practices in datasets that are extended to non-WoS peer reviewed scholarly publications (Sivertsen 2016b).

The policy implications of the new study would be that if funding should primarily follow productivity, as the earlier studies might advise, there would be a risk of prioritizing older male professors at the cost of young female researchers without stimulating citation impact in general.

Policy implications

- The discussions highlighted above raise a key research policy question: **To what extent can meaningful policy recommendations be deduced from estimated aggregated effects of specific policy mechanisms** – given both the challenges related to isolating these effects in complex systems and the fact that aggregated effects often may hide a heterogeneous underlying picture?
- First, this calls for more caution than is normally seen in the way such results are interpreted. Both researchers, administrators and policy makers **must explicitly acknowledge the unavoidable uncertainties** associated with such exercises. Obviously, thorough and well-designed studies can reduce these challenges, but only to a certain extent.
- Secondly, more attention should also be paid to the fact that even in cases where relatively certain aggregated effects can be attributed to a specific policy mechanism, the general pattern may hide a very different underlying picture. When incentive systems have varying effects across institutions, fields, career stage, sex, age, etc. **there is a high risk of drawing misleading conclusions about the actual effects at the individual level.**

Such ecological fallacies may arise when inferences are made about individuals based on aggregate data for a group. This seems exactly to be the case in relation to the relationship between productivity and citation impact, where the aggregated trends hide a surprising underlying picture.

- The current discussions thus remind us that aggregated effects seldom represent a certain and nuanced picture of the dynamics within complex systems. The measures and indicators created at aggregated levels will accordingly almost by definition be crude and imprecise. If such measures nevertheless are used in incentive systems the **redistributive effects must remain marginal in order to limit potential perverse effects and in order to allow for local translation and adjustment of the general incentives**. In this local adaption process it should also be kept in mind that even within relatively confined groups we may find important variations related to factors such as age, sex and career stage.
- Finally, the current discussions also underline, that while we will never see a perfect study of effects of performance-based systems, it is still highly relevant to study them at the best of our ability – and equally important to discuss the results as it is done currently (Aagaard and Schneider 2017).

Further reading

Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*, 42(5), 725-737.

Aagaard, K., Bloch, C.W., Schneider, J.W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation* 24(2): 106-117.

Aagaard, K., & Schneider, J. W. (2017). Some considerations about causes and effects in studies of performance-based research funding systems. *Journal of Informetrics*, 11(3), 923-926.

Aksnes, D.W., Rørstad, K., Piro, F., & Sivertsen, G. (2011). Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology*, 62, 628-636.

van den Besselaar, P., Heyman, U., & Sandström, U. (2017). Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics*, 11(3), 905-918.

Bloch, C., Schneider J.W. (2016). Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication level at the individual level. *Research Evaluation* 25(4): 371-382.

Butler, L. (2002) 'A List of Published Papers is no Measure of Value—The Present System Rewards Quantity, Not Quality - But Hasty Changes Could be as Bad', *Nature*, 419/6910: 877.

Butler, L. (2003a) 'Explaining Australia's Increased Share of ISI Publications—The Effects of a Funding Formula Based on Publication Counts', *Research Policy*, 32/1: 143-55.

Butler, L. (2003b) 'Modifying Publication Practices in Response to Funding Formulas', *Research Evaluation*, 12/1: 39-46.

Butler, L. (2004) 'What happens when funding is linked to publication counts?' In: Moed H. et al. (eds) *Handbook of Quantitative Science and Technology Research*, pp. 389-405. Netherlands: Kluwer Academic Publishers.

Butler, L. (2010) 'Impacts of Performance-based Research Funding Systems: A Review of the Concerns and the Evidence', in *Performance-based Funding for Public Research in Tertiary Education Institutions*, pp. 127-65. Workshop Proceedings. Paris: OECD.

de Rijcke, S., Wouters, P.F., Rushforth, A.D., Franssen, T.P., Hammarfelt, B. (2016). Evaluation practices and effects of indicator use – a literature review. *Research Evaluation* 25(2): 161-169.

Hicks D, Wouters P, Waltman L, Rijcke SD, Rafols I. *Bibliometrics: The Leiden Manifesto for research metrics*. *Nature*. 2015;520(7548):429-31.

Journal of Informetrics: Volume 11, Issue 3, Pages 599-944 (August 2017). Special Section on Performance-based research funding systems

Larivière, V. & Costas, R. (2016). How many is too many? On the relationship between research productivity and impact. *PLoS ONE* 11(9): e0162709. doi:10.1371/journal.pone.0162709.

Sandström, U. & van den Besselaar, P. (2016). Quantity and/or quality? The importance of publishing many papers. *PLoS ONE* 11(11): e0166149. <https://doi.org/10.1371/journal.pone.0166149>.

Sandström, U. & Sandsström, E. (2008). Resurser för citeringar. Rapport 2008:18 R, Högskoleverket.

Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, 3: 17078, doi:10.1057/palcomms.2017.78

Sivertsen G (2016a). Publication-Based Funding: The Norwegian Model. In: Ochsner M, Hug SE, Daniel HD (eds). *Research Assessment in the Humanities: Towards Criteria and Procedures*. Springer Open: Zürich, 79-90.

Sivertsen, G. (2016b). A bibliometric indicator with a balanced representation of all fields. In Ràfols, I., Mollas-Gallart, J., Castro-Martínez, E., Woolley, R. (Eds.), *Proceedings of the 21ST International Conference on Science and Technology Indicators* (pp. 910-914). Valencia: Editorial Universitat Politècnica de València.

Sivertsen, G., & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential. *Scientometrics*, 91(2), 567-575.

Schneider, J. W., Aagaard, K., & Bloch, C. W. (2015). What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation*, 25(3), 244-256.

Whitley, R., & Glaeser, J. (2007). The Changing governance of the sciences. *Sociology of the Sciences Yearbook*.

Zhang, L. Sivertsen, G. (2017). Productivity versus citation impact: A study of persons, not just authors. 16th International Conference on Scientometrics & Informetrics (issi2017.org), Conference Proceedings: 970-975.