Journal of Informetrics 11 (2017) 927–932

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Correspondence

A fight on epistemological quicksand: Comment on the dispute between van den Besselaar et al. and Butler

1. Introduction

Two studies using different data produce different results and draw different conclusions. So what? Since we all know all indicators to be only partial indicators of what we attempt to measure, this is the expected outcome rather than a surprise. Nevertheless, I cannot possibly forego the opportunity to annoy all parties involved and part of the audience by pointing out that the debate between van den Besselaar, Heyman & Sandström (BHS) and Butler is based on the shared erroneous assumption that causal claims can be made with the measurement strategy employed.

In a nutshell, my argument is that the causal chain between performance-based funding and changed aggregate publication behavior is rather long, and that a large number of confounding variables operates at different links of the causal chain. A causal claim could be made only if all these variables were considered in conjunction, which is impossible with bibliometric methods alone.

2. Causal arguments

In a paper on evaluation-based funding in 2002, Grit Laudel, Sybille Hinze, Linda Butler and I made a statement about a possible causal relationship that used the figure later to be published by Butler and reproduced in Fig. 1 by BHS. Referring to the picture, we stated: "The timing of this productivity increase in relation to the introduction of funding formulas suggests that there is a causal relationship." (Gläser, Laudel, Hinze, & Butler, 2002: 12). We later referred to the case studies on two Australian universities as "[providing] further support for the *assumption* that the coupling of increasing quantity and decreasing quality is due to the introduction on quantity-based funding formulas" (ibid: 14, my emphasis). Thus, we derived from Butler's analysis the hypothesis that the introduction of the publication indicator might have caused changes in publication behavior (which was correctly quoted as a hypothesis by BHS at page 6).

Butler subsequently turned this into a causal argument. Although the texts of her publications phrase her claim as "hypothesis" for which she finds "support" (Butler, 2003b: 41, 43; 2003a: 151), titles and an abstract (Butler, 2003b: 39) contain a causal statement rather than an unproven but plausible hypothesis. The titles read "Modifying publication practices in response to funding formulas", "Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts", and "What Happens when Funding is Linked to Publication Counts?" (Butler, 2004).

The BHS paper (van den Besselaar, Heyman, & Sandström, 2017) includes both hypothetical statements ("if output based research funding has an effect on research quality, it is positive and not negative", ibid: abstract) and factual statements ("our aim is . . . to reanalyze the effect of the changes in the funding system in Australia. . .", ibid: 3). They hedge their causal claim in the "conclusions and discussion" section by suggesting that "[w]e should refrain from drawing *too stark* conclusions regarding the causality of the process described . . ." It does not become clear what the 'weak' conclusions about causality are supposed to be. Ultimately, their claim is a causal counter-claim to Butler's causal claim, as becomes clear in the paragraph following the 'hedge paragraph':

> Nevertheless, the data do suggest that the new policy during the 1990s gave the Australian science system a new impulse – as funding partly became output-dependent after 1995. This initiative *did not only contribute to higher productivity*, but – as can be expected from creativity theory – *also to higher quality*. (ibid: 18, my emphasis)

But how can such a causal argument be made? With just one case, the only way to establish causality is to exclude all other possible causes by identifying the mechanism that produces changed aggregate publication and citation patterns from the introduction of the publication indicator.

http://dx.doi.org/10.1016/j.joi.2017.05.019

3. The appropriateness of measurements

Some of the indicators used by Butler and BHS pose interesting problems for establishing causality. First, the use of Australian shares in the SCI implies an enormous complication of the causal argument because shares in international collections of publications and citations are influenced by the publication and citation activity of all other countries included in the SCI. How can one establish the extent to which a changing share of publications or citations is caused directly by a national policy under these conditions? This problem has been discussed by Hicks (2005), who argued that the US were losing their leading position in the SCI due to "the determination of Asian policy makers to strengthen their R&D systems and consequently their knowledge economies" (ibid: 8, see also Hicks, 2007: 232-238). Is it possible that increased competition for publication space in the higher-impact journals crowded out some Australian publications, which moved to lower-impact journals? In the light of changing shares, one can still argue that Australian researchers were unable or were able to maintain their relative position due to the policies introduced at home. But this is a different causal argument, namely the argument that the Australian funding formula failed to make Australian university researchers outcompete researchers in other countries who were affected by different national policies.

Second, some of the indicators used by BHS fail to support a convincing counter claim. BHS use the top 10% cited papers because "this indicator yields a direct measure of impact, while the RCI (and the Incites® variant of it, the NCI) is dependent of variations in low or non-cited papers, papers that do not contribute to impact" (van den Besselaar et al, 2017: 9). They thus exclude the phenomenon that is at the core of Butler's argument – a disproportionate growth of low-impact papers. This makes it impossible to refute her claim. Acknowledging this, BHS look at the development of non-cited papers "in order to cover the various arguments by Butler" (ibid: 14, note 9). Unfortunately, this still doesn't cover Butler's argument because BHS only look at the poles of the spectrum of Australian publications, while Butler's argument refers to all of them.

4. Causality and complexity

Fig. 1 illustrates the methodological problem of deriving causal claims from bibliometric analyses. Both sides assume the operation of a causal chain which produces the observed output. Butler observes that at least some universities respond to funding formulas by internally mirroring them (an observation I can confirm, see below), and suggests that these internal policies made academics produce more publications. BHS suggest that the performance-based funding of Australian universities stimulated researchers to become more productive and also increased the quality of their production. They do not further specify the causal chain.

Neither Butler nor BHS empirically investigate the assumed translation of the funding formula in university policies and of university policies into changed publication behavior. While it is not impossible to investigate these translations, it is impossible to investigate them *bibliometrically*.

Fig. 1 demonstrates that neither party is able to produce sufficient evidence for the claim that the assumed causal chain operates. My main objection goes one step further by arguing that the assumed causal chain is very unlikely to operate at all because it disregards a large number of factors that may affect publication behavior.

Both BHS and Butler are aware of the problem of confounding variables but only Butler does something about them. She discusses and excludes two alternative explanations for the changed publication pattern, namely the higher education reform at the end of the 1980s, which turned colleges of advanced education and institutes of technology into 'new' universities or amalgamated them with existing universities, and the rise of numbers of staff, which was considerable in the period under consideration (Butler, 2003a: 149). In order to further strengthen her claim, Butler (2003b) compares the publication output of the university sector to the outputs of the hospital and the government sector and finds that "the 1989-93 period does not mark a turning point in trends for either of these sectors" (ibid: 43).

BHS did not use any additional information of this kind. They are aware of possible confounding variables, mention a study that refuted a similar causal claim (Osuna, Cruz-Castro, & Sanz-Menendez, 2011), and list some possible confounding variables on page 18. However, they do not provide evidence for or against a causal role of any of these factors.

What other factors would need to be taken into consideration? I limit my discussion to those factors for which I can suggest plausible mechanisms (Fig. 2). For the operation of some of these mechanisms I have micro-level empirical evidence from a study on the impact of the Australian performance-based funding system conducted by Grit Laudel and myself from 2002 to 2006 (Gläser & Laudel, 2007; Gläser, Lange, Laudel, & Schimank, 2010a, Gläser, Lange, Laudel, & Schimank, 2010b).

First, it is not clear why so much steering power is ascribed to the publication indicator although it weighs only ten percent in the funding formulas. Would one not expect universities to respond most strongly to the indicator 'external funding', which controlled 68% of the research block grants for universities compared to the 10.3% of the publication indicator in 1996?[1] The distinct but limited role of the publication indicator is represented by a white strip in the otherwise grey box of the funding formulas.

Second, the formula-based funding occurred in the context of other higher education policies. Two of the most consequential policies were absolute reductions in government funding of higher education, which began in 1997 (Meek, 2005), and the higher education reforms discussed by Butler. I am not putting forward these policy changes as alternative explanations

[1] This can be calculated from the weights of the indicators and sums allocated provided by Butler (2003b: 40).

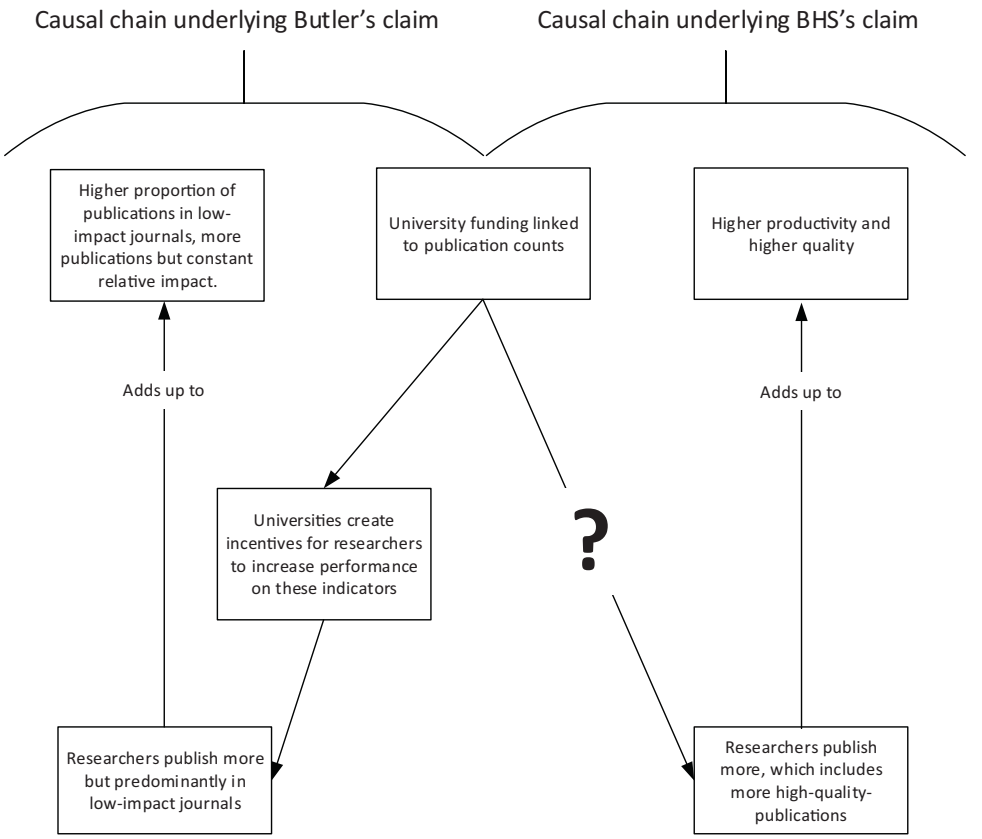

Fig. 1. Hypothetical causal chain and measurements.

for changed publication behavior. However, the question how these policies contributed to behavioral changes cannot be easily dismissed. In particular, higher education reforms confronted many staff with the new expectation to conduct research and to publish (Meek, 1991), and it is not unlikely at all that this led to a disproportionate share of 'new' publications in low-impact journals.

Third, the success rates of competitive grant funding in Australia were notoriously low at about 22% (Bazeley, 1998), which created enormous competitive pressure for both universities and individual researchers.

Fourth, all these factors combined created the environment universities responded to by developing internal performance management strategies. We found in 2002–2005 that university strategies prioritized external funding, which was consistent with the weight of the indicator and the overall scarcity of funding. We also found that universities internally introduced funding formulas similar to those applied to them but that little or no money was distributed to individual researchers on that basis. There simply was not enough money to distribute. Thus, there were no financial incentives for researchers to perform well in the indicators applied to universities. However, the indicators were translated in promotion criteria, which mattered to academics. This is why there is some white in the university policy box. We found scattered empirical evidence that Australian academics adapted their behavior to promotion criteria, which included a version of the publication indicator. Where publications were addressed, university policies focused on 'DEST points', i.e. publications that were counted by the funding formulas.[2]

Our observations of university policies also raise the question of timing. Butler is right when stating in her response that actors anticipate events and may respond to them before they actually occur. However, we also found the institutionalization of performance-related internal university strategies to be a slow process, which was still ongoing in the early 2000s. This is consistent with Aagaard's (2015) observation that about ten years after the introduction of the Norwegian performance-based funding scheme, there is still a wide variation in the use of the publication indicators in departmental performance management strategies. Thus, very prompt responses by all Australian universities may have occurred with regard to publication reporting but not with regard to setting incentives for increased publication numbers.

[2] The standard phrase "DEST Points" refers to Australia's Department of Education, Science, and Training, which was in charge of the formula-based funding at the time of our investigation.

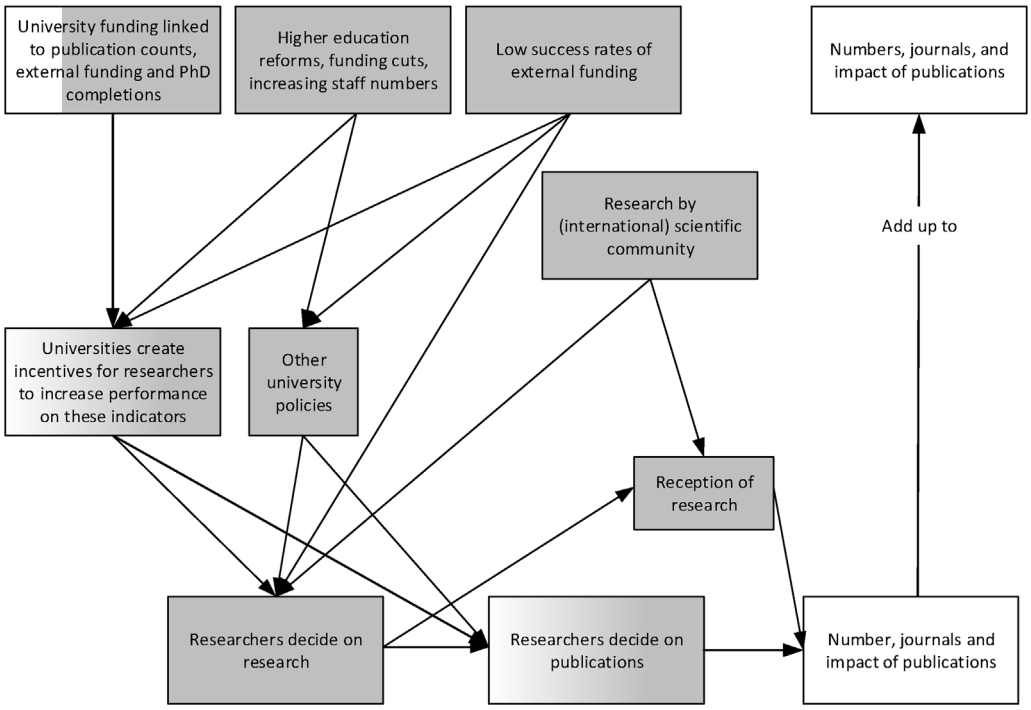

Fig. 2. A more complex picture of causal relationships (factors considered by Butler or BHS remain white, factors not taken into account are shaded grey).

Fifth and most importantly, there are the responses by researchers to consider. Researchers first and foremost decide on their research, which means that decisions on publications are almost always secondary. Decisions on research are made in response to the researcher's scientific community's work (state of the art, methodological standards, research priorities and so on) and in response to perceived opportunities to continue research (Gläser, 2012).

One of the main purposes of these decisions is to secure the continuous conduct of research. In order to maintain their identity as researchers, academics must continuously work on contributions to their community's knowledge. Continuity is also necessary because it keeps researchers' knowledge and capabilities at the level that is necessary to produce relevant contributions. In some fields, being excluded from the research of their community even for a short time can lead to a permanent lock-out from research.

The crucial role of the imperative to continuously conduct research is often overlooked. It brings us back to external funding, which in the early 2000s was not only the overriding concern of Australian universities but also of many researchers. Academics conducting empirical research in the sciences and social sciences depended on external funding, which is why all our interviewees from these fields were worried about it. There is a clear link between this major concern of researchers and publications because an applicant's 'track record' is evaluated in decisions on grants. In Australia, this assessment of the track record is formalized. The assessment of the track record is given a weight of 10% in the overall assessment of a grant application. For the 1990s, it was observed that the number of publications contributed to the application's success (Bazeley, 1998). According to our investigation, the indirect pressure to publish emerging from the necessities of grant applications was much stronger than the direct pressure exercised by university policies. However, university policies contributed to shaping some publication decisions, which is why there is some white in this box.

I introduced these additional processes because they very likely overlaid and modified the causal chain assumed by Butler and BHS. The purpose of this exercise is to show that without assessing all these factors, no causal statement is possible. This includes our own work. We were able to show that

- the performance-based funding of universities was not translated into effective financial incentives for academics due to scarcity,
- the translation into promotion policies was effective in individual cases but occurred unevenly and in longer time frames across universities,
- the continuation of research and the acquisition of external funding mattered most to academics, and had the strongest impact on their behavior, and that
- the decision practices of funding councils and their anticipation by researchers made them prefer mainstream research and applied research, and led to a narrowing of individual research portfolios.

We were unable, however, to draw any conclusions about macro-level changes because this is impossible with qualitative methods alone. We developed the hypothesis that an aggregate effect of individual adaptations to university policies concerning external funding and to funding council priorities might be a decrease in the diversity of research at the national level. However, we could not measure this effect at the macro level or trace the aggregation mechanisms. Both tasks are interesting challenges for bibliometrics. If macro-level effects are aggregate effects of individual behavioral changes, qualitative methods alone cannot establish macro-level causality, either.

5. Conclusions

The impact of performance-based funding on universities is overlaid and mediated by other policy changes, which is why it is difficult to assess its effect on university strategies. University strategies, in turn, are only one among several factors that affect academics' decisions on research and publishing. Decisions on research are much more strongly affected by the work of scientific communities and opportunities for acquiring external funding. The reception of research and subsequent citations are dependent on the competition of researchers from all over the world for attention to their work. All these conditions and mechanisms affect a country's publication output. This is why associating changes in that output with a preceding political event is not sufficient to establish causality. The causal claims by Butler, BHS, and others are not justified (see Gläser & Laudel, 2016: 129-134).

In order to establish one phenomenon as a cause of another phenomenon, we need to empirically identify the mechanism that produces the effect from the cause and either include or exclude all other possible causes, which is best done by identifying the generating mechanism. Since causal mechanisms that change research content or output operate through researchers, we need to follow the assumed macro-level trigger down to the researcher, identify behavioral change, and trace the aggregation of these changes. A macro–macro causal link can only be established through macro-micro-macro process tracing.

This cannot be done exclusively with bibliometric methods. Bibliometric methods are indispensable in such research for two reasons. First, they are the only methods we have for identifying macro-level patterns in publication behavior or properties of knowledge. In other words, we need bibliometric methods in order to raise interesting questions. Second, bibliometric methods are our best bet for addressing an unresolved problem, namely linking back changes in individual behavior to the macro level. Sociologists are able to identify changes in problem choices and publication decisions at the individual level as well as conditions that shape these decisions. They cannot, however, causally link micro-level behavioral changes to macro level changes in knowledge or publication patterns. This micro-macro link, which is both theoretically and politically important, cannot yet be established because we lack the methods for generalizing conditions under which individual researchers work to the meso-level of larger communities. Bibliometrics appears to be the methodology of choice because some of its methods work on both the individual and the community level, and should thus be able to link the two.

Establishing causality thus turns into an interdisciplinary enterprise in which bibliometricians have to collaborate with political scientists and sociologists. We somehow lost this kind of collaboration during the last three decades. This is unfortunate because bibliometrics is more effective as part of an integrated effort.

References

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

Bazeley, P. (1998). Peer review and panel decisions in the assessment of Australian Research Council project grant applicants: what counts in a highly competitive context? *Higher Education*, *35*(4), 435–452.

Butler, L. (2003a). Explaining Australia's increased share of ISI publications–the effects of a funding formula based on publication counts. *Research Policy*, *32*, 143–155.

Butler, L. (2003b). Modifying publication practices in response to funding formulas. *Research Evaluation*, *12*(1), 39–46.

Butler, Linda. (2004). What happens when funding is linked to publication counts? In Henk F. Moed, Wolfgang Glänzel, & Ulrich Schmoch (Eds.), *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T systems* (pp. 389–405). Dordrecht: Kluwer.

Gläser, J., & Laudel, G. (2007). Evaluation without evaluators: the impact of funding formulae on australian university research. In R. Whitley, & J. Gläser (Eds.), *The changing governance of the sciences: the advent of research evaluation systems* (pp. 127–151). Dordrecht: Springer.

Gläser, J., & Laudel, G. (2016). Governing science: how science policy shapes research content. *European Journal of Sociology/Archives Européennes De Sociologie*, *57*(01), 117–168.

Gläser, J., Laudel, G., Hinze, S., & Butler, L. (2002). *Impact of evaluation-based funding on the production of scientific knowledge: what to worry about, and how to find out (Expertise für das BMBF)*. http://www.sciencepolicystudies.de/dok/expertise-glae-lau-hin-but.pdf

Gläser, J., Lange, S., Laudel, G., & Schimank, U. (2010a). Informed authority? The limited use of research evaluation systems for managerial control in universities. In R. Whitley, J. Glaser, & L. Engwall (Eds.), *Reconfiguring knowledge production: changing authority relationships in the sciences and their consequences for intellectual innovation* (pp. 149–183). Oxford: Oxford University Press.

Gläser, J., Lange, S., Laudel, G., & Schimank, U. (2010b). The limits of universality: how field-specific epistemic conditions affect authority relations and their consequences. In R. Whitley, J. Glaser, & L. Engwall (Eds.), *Reconfiguring Knowledge Production: changing authority relationships in the sciences and their consequences for intellectual innovation* (pp. 291–324). Oxford: Oxford University Press.

Gläser Jochen. (2012). How does Governance change research content? On the possibility of a sociological middle-range theory linking science policy studies to the sociology of scientific knowledge. In *The technical university technology studies working papers, TUTS-WP-1-2012*.

Hicks, D. (2005). America's Innovative Edge At Risk? *Research-Technology Management*, *48*(6), 8–12.

Hicks, D. M. (2007). Global research competition affects measured U.S. academic output. In Paula E. Stephan, & Ronald G. Ehrenberg (Eds.), (pp. 234–253). Science and the Unviersity, Madison: University of Wisconsin Press.

Meek, V. L. (1991). The transformation of Australian higher education from binary to unitary system. *Higher Education*, *21*(4), 461–494.

Meek, L. (2005). Country paper on Australia: cross-border higher education in Australia. In UNESCO (Ed.), *UNESCO forum occasional paper series paper no. 9: implications of WTO/GATS on higher education in asia & the pacific Paris* (pp. 45–85). United Nations Educational, Scientific and Cultural Organisation (UNESCO).
Osuna, C., Cruz-Castro, L., & Sanz-Menendez, L. (2011). Overturning some assumptions about the effects of evaluation systems on publication performance. *Scientometrics*, *86*, 575–592.
van den Besselaar, P., Heyman, U., & Sandström, U. (2017). Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics*.

Jochen Gläser
Center for Technology and Society, TU Berlin, Germany
E-mail address: Jochen.Glaser@ztg.tu-berlin.de
Available online 29 June 2017