



RESEARCH ARTICLE

Powerful numbers: Exemplary quantitative studies of science that had policy impact

Diana Hicks¹ and Kimberley R. Isett²

¹School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30332

²Joseph R. Biden Jr. School of Public Policy and Administration, Newark, DE 19711

an open access  journal



Citation: Hicks, D., & Isett, K. R. (2020). Powerful numbers: Exemplary quantitative studies of science that had policy impact. *Quantitative Science Studies*, 1(3), 969–982. https://doi.org/10.1162/qss_a_00060

DOI: https://doi.org/10.1162/qss_a_00060

Corresponding Author:
Diana Hicks
dhicks@gatech.edu

Handling Editors:
Loet Leydesdorff, Ismael Rafols,
and Staša Milojević

Copyright: © 2020 Diana Hicks and Kimberley R. Isett. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: Butler, evidence, Mansfield, Martin and Irvine, Narin

ABSTRACT

Much scientometric research aims to be relevant to policy, but such research only rarely has a notable policy impact. In this paper, we examine four exemplary cases of policy impact from quantitative studies of science. The cases are analyzed in light of lessons learned about the use of evidence in policy making in health services, which provides very thorough explorations of the problems inherent in policy use of academic research. The analysis highlights key dimensions of the examples, which offer lessons for those aspiring to influence policy with quantitative studies of science.

1. INTRODUCTION

Over the past few decades, bibliometrics in the policy realm has become welded to the broader neoliberal, new public management agenda of quantifying university and researcher performance in order to foster competition. Many, though not all, national performance-based university funding schemes (Hicks, 2012) use metrics of university performance to determine part of university research funding, and quantitative university rankings have become influential. Metrics have also become a standard component of research program evaluation. More recently, Google Scholar's deployment of the h-index, ResearchGate's posting of oeuvre metrics, and PLOS ONE and others' posting of article metrics have enabled researchers to self-quantify. These developments have not gone unchallenged. A large number of critical analyses and attempts to improve practice have been published (e.g., DORA, 2012; Leiden Manifesto, 2015). Thus has discussion of policy use of quantification of science become overly focused on evaluative bibliometrics.

Here we explore something else, namely the potential of quantification to serve advocates arguing against damaging policies or for increased research funding. We explore four cases of quantitative analyses of science that influenced policy for research. The cases are Edwin Mansfield's calculation of the rate of return to public funding of academic research, Ben Martin and John Irvine's argument that U.K. science was in decline, Francis Narin's analysis of the sources of journal articles cited in patents, and Linda Butler's analysis of the effects of Australia's university evaluation scheme. In each case, the analysts were arguing for science against government policy or were supporting the case for public research funding.

Our discussion of the cases is framed using the three characteristics of knowledge that influence its use by policymakers: relevance, legitimacy, and accessibility (Contandriopoulos, Lemire, et al., 2010), which are discussed in light of a list of best practices drawn from a review

of literature on the use of evidence in health policy. The literature on health-related evidence translation in public decision-making is one of the best developed—more so than that in many other policy-related areas. Therefore, it is a rich source of insights into how research can influence policymaking. We began with two influential articles in the field: Greenhalgh, Robert, et al. (2004) and Jewell and Bero (2008). Then a snowball approach was used to collect review and framework articles that were cited by or cited these two articles using both PubMed and Google Scholar through December 2017. Articles were screened for a focus on evidence translation in public decision-making, and 32 articles were found. Articles were read to identify common elements and create a synthesis of best practices. We will first discuss how the best practices illustrate the framework and then look for the elements identified in the four case studies. We finish with a discussion of the tensions between the three characteristics and draw out lessons for bibliometricians aspiring to policy impact.

2. BEST PRACTICES

2.1. Relevance

Information relevant to policy decision-making is timely, salient, and actionable (Contandriopoulos et al., 2010, p. 460) and includes both qualitative and quantitative components (Jewell & Bero, 2008). On the qualitative side, stories provide an emotional hook and an intuitive appeal (Brownson, Chiqui, & Stamatakis, 2009; Troy & Kietzman, 2016). However, stories have a limited role in science policy; anecdotes about the invention of the laser go only so far in justifying hundreds of millions of dollars in research expenditure. The health policy literature suggests that quantitative evidence provides three things important in policy making: descriptions of a problem that highlight disparities in the population (Stamatakis, McBride, & Brownson, 2010); the cost of policy inaction and the distribution of program costs and benefits (Hanney, Gonzalez-Block, et al., 2003; Jewell & Bero, 2008; Stamatakis et al., 2010; Stone, 1989); and finally, for policy areas where outcomes are long term, Fielding and Briss (2006) suggest including intermediate measures of benefits to provide a shorter time to payoff for policymakers forced to work on short policy cycles. In addition, geography matters; information specific to policymakers' jurisdiction is more relevant to policymakers than statements about a general need or a pervasive phenomenon (Brownson, Dodson, et al., 2016; Fielding & Frieden, 2004; Hanney et al., 2003; Laugesen & Isett, 2013; Murthy, Shepperd, et al., 2012; van de Goor, Hamalainen, et al., 2017). Policymakers also want to see how they compare to their peers (Stamatakis et al., 2010; Stone, 1989)—those jurisdictions that they think are similar or competitors to themselves—and such comparisons can compel decision-makers to act, to avoid being left behind.

2.2. Legitimacy

Legitimacy refers to the credibility of information. While scholars like to think that they know good evidence when they see it, the criteria that public servants use are somewhat different and more nuanced. Public decision-makers want to know how good the evidence is that something will affect outcomes in a meaningful way (Atkins, Siegel, & Slutsky, 2005)—that is effect size, not statistical significance. Strictly speaking, credibility “involves the scientific adequacy of the technical evidence and arguments” (Cash, Clark, et al., 2003, p. 8086). However, policy audiences without the requisite specialist expertise to make a technical judgment instead tend to assess credibility through face validity of the messenger, or the research team (Brownson, Chiqui, & Stamatakis, 2009; Dodson, Geary, & Brownson, 2015; Lavis, Oxman, et al., 2008; Lavis, Robertson, et al., 2003).

2.3. Accessibility

The literature is clear on the need for high-quality communication. This is an area that many scholars shy away from, feeling the data should speak for itself (Pisano, 2016). Given that busy decision-makers have limited time and cognitive resources to identify the material they need to make decisions, they must rely on both heuristics (about source and content) and the summaries provided to them by others (Cyert & March, 1963; Dagenais, Laurendeau, & Briand-Lamarche, 2015; Ostrom, 1998). Systematic reviews may be an academic gold standard, but policymakers see these kinds of documents as long, complicated, and difficult to understand (Tricco, Cardoso, et al., 2016)—so the nature and presentation of the summaries are important (Dodson et al., 2015).

The ability to transparently and credibly distill information is key to getting the attention of decision-makers (Burris, Wagenaar, et al., 2010; Cyert & March, 1963; Hanney et al., 2003; Murthy et al., 2012). Information presented in a straightforward way (without jargon) that is quick and easy to understand and absorb is more likely to be used (Burris et al., 2010; Coffman, Hong, et al., 2009; Gamble & Stone, 2006). Information needs to be targeted, with the scope of the information explicit and relevant to the decision at hand, shorn of all secondary and tangential information, and inserted into the process when most useful (Brownson, Fielding, & Maylahn, 2009; Burris et al., 2010; Coffman et al., 2009; Hanney et al., 2003; Lavis et al., 2008). Further, the benefits of policy adoption should be visible and unambiguously presented (Atkins et al., 2005; Gamble & Stone, 2006).

Studies illustrate that quantitative data is inaccessible to most public decision-makers, who were trained to do things other than sort and interpret data. (Brownson, Chricqui, & Stamatakis, 2009). Furthermore, while there is some validity to the idea that showing decision-makers and their staffs evidence or teaching them how to access it will address capacity deficits (Redman, Turner, et al., 2015; VanLandingham & Silloway, 2016), there is limited evidence about the extent to which this works (Murthy et al., 2012). Thus, the craft of messaging empirical research is crucial to getting research used. Importantly, messages must be tweaked for multiple audiences (Troy & Kietzman, 2016; van de Goor et al., 2017). The details given to a policymaker differ from those produced for agencies, advocacy, or the public and take into consideration their different foci, authority, and scope of operations (Hanney et al., 2003; Lavis et al., 2003; Oliver et al., 2014; Sabatier & Jenkins-Smith, 1993). While multiple messaging artifacts should be internally consistent, the language, highlights, and modes of communication differ.

Because the requirements of multiplex messaging are somewhat at odds with the requirements of scholarly incentives, and because any single study is rarely definitive enough to guide policy by itself, intermediaries play a large role in facilitating use of research in policy (Dagenais et al., 2015; Lavis et al., 2003; Meagher & Lyall, 2013; Tricco et al., 2016). Intermediaries bundle related studies, contextualizing and interpreting the information for salience to and easy processing by the decision-making body (c.f. Dodson et al., 2015; Dutton, 1997). Known intermediaries are thought of as “honest brokers” that credibly produce syntheses that are useful and unbiased. Users know the organizations in their domains that produce broad syntheses and so can quickly find relevant reviews (Lemay & Sá, 2014). The bottom line is that research findings do not stand alone on their merits. Instead, they must be interpreted for use and comprehension by decision-makers.

3. BEST PRACTICES ILLUSTRATED WITH SCIENCE POLICY CASES

The importance of effectively communicating high-quality information relevant to decision-makers will be explored using four well-known instances in which quantitative studies of

science had demonstrable policy impact. Our cases exemplify work with high academic credibility: Most were highly cited in the scholarly literature. The cases are international, with two U.S. cases, one British, and one Australian.

3.1. Case 1: Mansfield's Rate of Return (United States)

The first case of policy impact comes from the work of economist Edwin Mansfield, who was the first to empirically estimate the social rate of return to public research spending, which he calculated to be 28% (Mansfield, 1991a, 1998). This is probably the most influential number in the history of research policy. Mansfield was encouraged to produce this study by the Policy Studies Unit in the National Science Foundation (NSF), who funded the work. Mansfield's *Research Policy* (1991a) paper concluded

A very tentative estimate of the social rate of return from academic research during 1975–78 is 28 percent, a figure that is based on crude (but seemingly conservative) calculations and that is presented only for exploratory and discussion purposes. It is important that this figure be treated with proper caution and that the many assumptions and simplifications on which it is based (as well as the definition of a social rate of return used here) be borne in mind.

The paper was very highly cited, as well as influential in the policy world. Crucial to the influence of this analysis is that Mansfield did put forth a number—a bold move, and one avoided by many scholars. Nevertheless, in the article, Mansfield surrounds the number with scholarly caveats: “treat with caution,” “only exploratory,” and so forth. As the number moved into the policy world, what happened to the number and what happened to the caveats?

The following year, in an interview in *Science* magazine, President George H. W. Bush is quoted as saying

*Our support of basic research in these and other agencies is an investment in our future, but by its very nature it is impossible to predict where, when, or to whom the benefits will flow. Nevertheless, we can be sure that these benefits will be substantial. Professor Edwin Mansfield of the University of Pennsylvania has found that the social rate of return from such investments in academic research can very conservatively be estimated at 28%. (*Science policy*, 1992, pp. 384–385.)*

In using the number to argue for the value of research funding, the president dropped the caveats, which do not work well in presidential interviews.

In 1993 the Congressional Budget Office (CBO) reviewed Mansfield's work in response to a request from a House Committee. The CBO positioned Mansfield's work as a validation of the vision of Vannevar Bush, the patron saint of U.S. basic research funding, and included the caveats:

[The House Committee on Science, Space, and Technology] asked the Congressional Budget Office to comment on the policy relevance and statistical accuracy of Edwin Mansfield's estimates of the social rate of return from academic research. Since World War II, U.S. science policy has been guided by Vannevar Bush's vision that, if funded and left to set their own agenda, scientists would amply reward the nation for its

investment. Mansfield has shown that, on average, academic scientists have indeed kept their part of the bargain. The return from academic research, despite measurement problems, is sufficiently high to justify overall federal investments in this area.

Nevertheless, the very nature of the estimating methodology, as Mansfield has noted in his articles, does not lend itself to use in the annual process of setting the level of federal investment in R&D, nor to allocating that investment among its many claimants. Furthermore, given the nature of the assumptions, definitions, and other methodological questions, as Mansfield notes, his result is more properly regarded as indicating a broad range of likely orders of magnitude of the return from academic R&D than as a point estimate (28 percent) of the return from federal investment in this area. (Webre, 1993)

In 1998, Mansfield produced an update in *Research Policy* and his influence grew. In 1998 the CBO did another report:

One study that received a great deal of attention was performed by Edwin Mansfield, who tried to compensate for the inherent bias of benefit-cost studies by using conservative assumptions and offsetting known errors. Mansfield estimated that academic R&D gives society a 28 percent return on its investment; given the uncertainties involved, a more appropriate summary of the study is a range from 20 percent to 40 percent. Since most of the funding of those academic researchers came from the federal government, the returns should apply, at least roughly, to federal programs that fund academic research. (Alsalam, Beider, et al., 1998, p. 38)

It is unclear where the “range from 20 percent to 40 percent” in this document originated. Mansfield did not mention it in his 1991 or 1998 papers, nor was it in the 1993 CBO review of Mansfield’s analysis referenced in this paragraph. Nevertheless, the 20 to 40 percent range seemed to become the canonical Mansfield reference in subsequent policy documents.

In 2006 a report of the Task Force on Innovation, which is an advocacy organization, not a government department, emphasized the high end of the estimated range:

It is no wonder that economist Edwin Mansfield calculated as much as a 40% rate of return for the Federal investment in basic university based research. (Task Force on the Future of American Innovation, 2006)

In 2007 the range appeared in testimony before the House Committee on Financial Services:

Mansfield concluded that the average annual rate of return to society from academic research was anywhere from 28 to 40 percent. The Congressional Budget Office, in a 1993 review of Mansfield’s estimates, said that “the return from academic research, despite measurement problems, is sufficiently high to justify overall federal investments in this area.” (Role of Public Investment in Promoting Economic Growth, 2007, p. 39)

Although this is no doubt an incomplete record, it does establish both the enduring influence of Mansfield’s number and the evolution of that number in the hands of intermediaries.

As in most policy arenas, those arguing for the value of publicly funded research do not lack for anecdotes—the internet, the laser, and MRI exemplify public research that created tangible public value. However, in the early 1990s, quantitative evidence was scarce, and the long time lags between research and application provided an extra challenge to gathering it. In calculating the benefit that U.S. firms derive from publicly funded research, Mansfield quantified what were generally considered diffuse and intangible benefits to society, namely that “research is good for innovation,” thereby adding some clarity to an otherwise historically ambiguous public good. The protagonists in the analysis, U.S. firms, are an important Congressional constituency benefiting from public research and so benefits to firms are a “good.” The intermediate measure of benefit to firms provides a view to benefits without asking policymakers to wait for the more diffuse and longer-term societal benefits.

Mansfield put forward a number useful to those seeking to establish the value of research funding. A clear and precise message has a higher probability of being used than a diffuse message. In many cases, that “clear message” can be embodied in a number. Alone of the four cases, Mansfield’s paper offered a single, clear number in the conclusions, though surrounded with caveats. Users of Mansfield’s results repeatedly referred to the 28% figure in advocating for the value of research to the nation in the yearly competition for Congressional attention and funding.

3.2. Case 2: Narin’s Patents Citing Papers (United States)

The second case is Francis Narin’s discovery that patents were increasingly referencing scientific papers and that 73% of the papers cited by U.S. industry patents are public sector science (Narin, Hamilton, & Olivastro, 1997). Because this can be interpreted as industry using the research that government funds, it can be used to establish the value of publicly funded research. Like Mansfield’s work, and for similar reasons, this study was noticed and used by the media, advocates, and policymakers. A 1997 *New York Times* article focusing solely on this paper was headlined: “Study finds public science is pillar of industry.” There was again a CBO commentary in a report on the economic effects of federal spending:

CHI Research, a patent-citation consultancy, has collected indirect evidence on that point.⁽⁶⁵⁾ Patent applications include two types of citations: to other patents and to scientific literature. Of the scientific papers cited in patents, 73 percent were articles written by academic scientists or scientists at governmental or other institutions developing what the authors call “public science.” The authors argue that industry has increased its reliance on public science over the last decade and that public science is, to a large extent, the product of federal funds. (Alsalam et al., 1998)

Following the pattern set by the Mansfield number, Narin’s number was also misquoted, this time in a report from the House of Representatives:

The above examples of basic research pursuits which led to economically important developments, while among the most well known, are hardly exceptions. Other instances of federally funded research that began as a search for understanding but gave rise to important applications abound. In fact, a recent study determined that 73 percent of the applicants for U.S. patents listed publicly-funded research as part or all of the foundation upon which their new, potentially patentable findings were based. (Committee on Science, 1998)

If indeed 73% of patent applicants cited public science, that would be a much more powerful number than the actual result, which was that 73% of the cited papers originated in publicly funded research. So an element of wishful thinking appears here, as it did with the Mansfield misquotes. The errors are clearly not random. The tendency to ignore reality and pretend numbers are more powerful than they are is one thing that makes scholars queasy and reluctant to interact with policymakers.

Nevertheless, most users did quote the result correctly, even 5d years later when the National Science Board (NSB) quoted the results in two documents:

An NSF-supported study found that 70 percent of the scientific papers cited in U.S. industry patents came from science supported by public funds and performed at universities, government labs, and other public agencies. (National Science Board, 2003, 2005)

Narin also briefed interested Congress members in a breakfast meeting organized by the NSF, as well as briefing the NSB. The NSB got interested and convened a subcommittee to write a report on Industry Reliance on Publicly-funded Research (IRPR). Caveats were a worry for the subcommittee, who found the topic to be more complex than anticipated. The minutes of a subsequent NSB meeting reported that

There are other indicators to account for ... It would be difficult to draw general conclusions, so the paper will contain a number of limited conclusions. Finally, there are issues of credibility to address. The Task Force was concerned that the paper not appear to be self-serving and that it be cautious about overstatement. Consequently, more study and discussion are needed as the Task Force's initial draft is revised. (Fannoney, 1997)

The chairman applauded the Task Force for its caution and urged them to continue their efforts, which resulted in an addendum to Science & Engineering Indicators 1998 entitled *Industry trends in research support and links to public research* (National Science Board, 1998).

Like Mansfield, Narin provided quantitative evidence that U.S. firms benefit from publicly funded research. Narin's new intermediate measure pointed out that firm patents increasingly referenced publicly funded research. This suggested that firms used, and therefore benefited from, public research. Narin's analysis was descriptive, serving to make an abstract entity, the national research system, visible and tangible.

Narin's paper did not focus on a single number. The author's summary of the paper would have been that references from U.S. patents to U.S.-authored research papers tripled over a 6-year period, from 1988 to 1994. Furthermore, the cited U.S. papers represented basic research in influential journals, authored at top research universities and laboratories, relatively recent to Narin's analysis, and heavily supported by public agencies. Intermediaries incorporating the result into overviews plucked the 73% number (73% of papers cited by U.S. industry patents are public sector science) out of the paper's introduction, and it was used repeatedly by those seeking to establish the value of research funding.

3.3. Case 3: Martin and Irvine's Gap (Britain)

In the mid-1980s, Ben Martin and John Irvine produced a series of commentaries in *Nature* arguing that British science was in decline as evidenced by trends in publication output and

government funding of research falling behind that of the Netherlands, France, Germany, Japan, and the United States (Irvine, Martin, et al., 1985; Irvine & Martin, 1986; Martin, Irvine, et al., 1987). The titles tell the story: “Charting the decline in British science”; “Is Britain spending enough ...”; and “The continuing decline ...” The first one was an analysis of trends in publication output and the second compared levels of research funding in the United Kingdom with those of competitors. Martin and Irvine disliked existing funding data and went around the world talking to agencies to collect proper funding data, reporting their analysis in the second commentary. The next year they updated the publication analysis in the third commentary.

The decline narrative attracted the government’s attention, and John Irvine was asked to meet with the responsible government minister, who wanted to know how big the funding gap was. Irvine offered £100 million, which was considered doable. OECD data suggest that the increase in the U.K. government funding of university research in 1987 at 13.4% was higher than in any other year between 1982 and 1994. The increase was £172 million; whether this was £100 million higher is difficult to see, because the increase the year before was £118 million and the year after was £115 million.

The constituency subject to Martin and Irvine’s analysis were public sector actors, scientists. A healthy public research sector is considered a public good broadly beneficial to society, and so a natural concern of national-level policymakers. Martin and Irvine’s paper on funding shortfalls was silent on exactly how much the United Kingdom was behind, which would have been the obvious focal number, and indeed, the minister requested this number at their meeting.

The literature recommends proposing a solution rather than identifying a problem; Martin and Irvine identified a problem. However, in line with recommendations, a solution existed— increase spending on U.K. science. There was some leeway in putting a number on the size of the funding gap depending on which countries were included in the comparison group. The estimate of £100 million that Irvine provided to the minister met the criteria of policy feasibility, and so had an impact. However, in a later update of the work, Martin and Irvine wrote that £500 million would have been required in 1987 to attain the “European” mean (Irvine, Martin, & Isard, 1990).

3.4. Case 4: Butler’s Perverse Incentives (Australia)

As a result of a 1992 policy that linked publications in indexed journals to university research funding, Linda Butler (2003) found that the Australian share of world publication output grew, but the citation performance of Australia fell from number 6 in 1981 to number 10 in 1999. Butler argued that this was because once the policy took effect, authors prioritized producing more papers, publishing in lower impact factor journals after the policy was introduced.

In later reflections, Butler concluded “Australia’s research evaluation policy had become a disincentive to research excellence” (Butler, 2003). A series of reports between 2002 and 2004 by Australia’s Department of Education, Science and Training (DEST) in many places incorporated data from a bibliometric study by Donovan and Butler (Donovan & Butler, 2003) and used the finding of declining citation impact.

Between 2002 and 2004, a series of consultative white papers published by DEST took stock of the Australian university research system. The reports incorporated data from a bibliometric study by Donovan and Butler in many places. In addition, the white papers reported the finding of declining citation impact. These reports progress from highlighting the positive,

to reporting the result exactly as the paper reports it, to summarizing the key point, to elaborating on the policy context and causes.

Australia's relative citation impact is falling behind most other comparable OECD countries, but that in the science disciplines, a few universities stand out in their overall performance. (Australia Department of Education, Science, and Training, 2002)

Australian academic publishing has increased since funding authorities have started to link allocation of research funds to number of publications. However, Figure 2.7 shows that the fastest increases in publication have been in Quartile 3 and 4 journals (i.e. those with a below-median impact). (Australia Department of Education, Science, and Training, 2003a)

[the publication component of the funding formula] rewards quantity rather than quality and has led to an increase in Australian articles appearing in lower-impact journals. (Australia Department of Education Science, and Training, 2003b)

The Research Quantum which rewards universities for publication output is likely to have boosted publications, however, the absence of a strong quality criterion to the Research Quantum publications measure may adversely affect the impact of journals produced by the sector [sic]. (Australia Department of Education, Science, and Training, 2003c)

In essence, these arise from the difficulties inherent in using a simple numerical measure as a proxy for the highly diverse and complex outcomes that are desired of the research system. Many stakeholders cite the findings of Dr Claire Donovan and Linda Butler (K30) that a rise in the number of publications has been accompanied by a significant decline in citation impact. Stakeholders appreciate that reliance on proxies can induce aberrant behaviour, both on the part of university administrators as they seek to optimise their institution's positions, and on the part of individual researchers. (Australia Department of Education, Science, and Training, 2004)

As a result, Australia changed its university evaluation system to incorporate two to four weighted categories of journals, a feature directly responding to the conclusions of Butler's analysis.

Universities were the constituency analyzed by Butler. A healthy public research sector is considered a public good broadly beneficial to society, so weakening universities' overall research competitiveness would be salient for national policymakers. The analysis illustrated unintended negative consequences arising from the perverse incentives built into the evaluation system and so was highly salient to those in charge.

The Butler case never focused on a single number. There was a number to highlight, namely that Australia fell from the sixth to 10th ranked country in citation share/publication share, which was in the second sentence of the paper's abstract. Nevertheless, perhaps because a simplistic focus on a single number was the cause of the problem Butler highlighted (universities could put a dollar value on a paper indexed in the Web of Science), intermediaries including Butler's result in their overviews explained that Australia's relative citation impact was falling behind most other comparable OECD countries (Australia Department of Education, Science, and

Training, 2002), that the fastest publication increases were in below median impact journals (Australia Department of Education, Science, and Training, 2003b), or that policies have led to an increase in Australian articles appearing in lower impact journals (Australia Department of Education, Science, and Training, 2003a). Despite the lack of “a number” the message was quite clear and tangible and was acted upon.

Like Martin and Irvine, Butler identified a problem with a straightforward solution—revise the evaluation process for Australian universities—that was within the scope of the agency involved—DEST.

4. DISCUSSION

Relevance, legitimacy, and accessibility are the crucial characteristics of knowledge used by decision-makers (Contandriopoulos et al., 2010). The above discussion explored how each result was made more accessible to policymakers. The cases also shared several characteristics salient for decision-makers. The research in all four cases highlighted important differences between fields, thereby showing disparities and distribution of burden and benefits. Geography is important to elected officials, who represent a constituency and care about comparisons with peers. Mansfield’s work, a typically economic view “from nowhere,” surveyed U.S. firms, so the result concerned the return from U.S. research to American society. Narin, Martin and Irvine, and Butler built their cases through international comparison. Science policy is a national-level concern, with links between innovation and economic prosperity, so policymakers do not want to fall behind other nations. Martin and Irvine and Butler were arguing precisely that Britain and Australia were falling behind. Narin argued that U.S. firms were leaders in building on public science.

The credibility of the messenger and the evidence were also evident. Three authors were academics well known in their fields at respected universities. Narin owned a consulting firm that had a long record of publishing in the top journals. The reputations of *Research Policy* and *Nature* vouched for the scientific adequacy of the evidence in the cases. In addition, the Congressional Research Service and the NSF closely examined the Mansfield and Narin work for technical adequacy (National Science Board, 1998; Webre, 1993), furthering the credibility attribution. Finally, none of these results rested on *p*-values to support the significance of their findings. Three were entirely descriptive, and it was the magnitude of the effect that attracted attention.

Sarkki, Niemela, et al. (2014) identified four trade-offs between the three key attributes that afflict the science–policy interface: a time trade-off between doing research (legitimacy) and interfacing with policymakers (accessibility) a clarity–complexity trade-off between simple (accessible) and nuanced (legitimate) communication of results, a time–quality trade-off between timely (relevant) versus in-depth (legitimate) analysis, and a push–pull trade-off between responding to policy demands (relevant) or identifying emerging issues (legitimacy). These cases did not suffer time–quality conflict, because they set the policy agenda by originating issues in investigator initiated work (i.e., identifying emerging issues on the pull side of the push–pull trade-off). Thus there was no pressure for fast, and perhaps lower quality, responses to opportunities presented by a short-lived policy window.

However, the cases did make visible the considerable time invested in interfacing with policymakers beyond the production of the journal article (i.e., the time trade-off). The four journal articles were not the main vehicle for communication with policymakers. Rather, their results were incorporated in broader summaries written by agencies, advocacy groups, and

other intermediaries. Intermediaries were focused on policy work and so had the time to develop documents for policymakers. Intermediaries also had the freedom to drop the caveats and clarify and focus the result, while borrowing the credibility of the original result to enhance the credibility of their own documents. The resulting *New York Times* article and agency white papers took the results closer to the heart of policymaking through clearer, contextualized, and more targeted messaging.

The clarity–complexity trade-off was also present. In each case more complexity was on offer. Mansfield gave a speech at the AAAS detailing methods and limitations in the analysis of social returns from R&D (Mansfield, 1991b), and the CBO reviewed the work (Webre, 1993). Meyer (2000) took issue with Narin’s interpretation of papers cited in patents, and an NSB task force reviewed the work (Fannoney, 1997). Leydesdorff, as well as Braun, Glänzel, and Shubert, disagreed with Martin and Irvine’s bibliometric portrayal of decline, pointing out that the conclusion depended on a series of methodological choices: whole versus fractional author counts, fixed versus dynamic/expanding journal set, and types of publications included, and *Scientometrics* (1991), volume 20 issue 2 was devoted to the debate. Belatedly, Butler’s analysis was revisited in a debate in a special section of the *Journal of Informetrics* (2017), volume 11, issue 3. The high visibility attending the policy impact of the four cases likely prompted the reviews and academic debates. Paradoxically, Porter’s analysis would suggest that the debates may have served to enhance the perceived objectivity and factuality of the numbers. Porter argued that openness to possible refutation by other experts reduces the demands on personal credibility to vouch for impersonal numbers (Porter, 1995, p. 214).

The complexity introduced in the debates did not filter through to the policy discussion. Rather, in these cases a single number played an outsized role, capturing the essence of the scholarly analyses and facilitating its communication. In only one case did the study’s author highlight the single number. In two other cases, intermediaries or the decision-maker extracted it. Credible academics are wary of the single number and its potential for misuse—witness the caveats surrounding the 28% number that was offered and its subsequent evolution. On the other side of the equation, it is unreasonable to expect policymakers to invest in understanding the complexities unique to every study. Intermediaries know this and enhance accessibility by offering a story with numbers that is uncomplicated, clear, and stripped of unnecessary content except the credibility of the author of the number. This is why white papers and policy documents contain references. The number is never truly alone; even at its most accessible, it is accompanied by a halo of credibility derived from its referenced source. Thus staged, the number conveys fairness and impartiality, shielding decision-makers from accusations of arbitrariness and bias, and so motivating change (Porter, 1995, p. 8).

Alternatively, intermediaries may be in the business of providing reductive and one-sided arguments in favor of policy goals. The exaggerations we found in the U.S. cases are not unique. Elson, Ferguson, et al. (2019) found that exaggeration characterized 79.2% of 24 policy statements about media effects on behavior made by U.S. organizations representing scientists. They concluded that “in the majority of policy statements we reviewed, the approach of the organization appeared to be reductive in that complexities and inconsistencies in research were ignored in favor of a narrative supportive of the organization’s policy agenda.” Stirling (2010) has argued that such pressures to simplify should be resisted. Stirling’s argument concerns environmental policy advice in areas where putting numbers on risk is inadequate because uncertainties loom large (toxic exposures, fish stocks, etc.), uncertainties being nonquantifiable unknowns. Such radical uncertainty is absent here. Nevertheless, the tension

between simpler (more accessible) and more nuanced (more legitimate) messaging, is pervasive when research engages with policy.

5. CONCLUSION

We have attempted to learn from historical exemplars of quantitative analysis that influenced science policy. We strengthened our analysis using comparisons with best practices identified in a review of literature on use of evidence in health policy. The first lesson was that quantitative studies of science have the potential to make an important contribution to improving the governance of research by providing convincing evidence of system-level problems and societal contributions. Researchers who wish to engage with policymakers should be mindful of the perspective of decision-makers, for example their interest in geographically defined constituencies, in sizeable effects, and in the distribution of benefits and harms from policy adjustments. They should also not shy away from drawing clear conclusions, even to the extent of identifying a single number that best encapsulates their main finding. Researchers, especially in the larger U.S. system, should look for opportunities to engage with intermediaries who draw on scholarly literature in their advocacy, although this will create tension if distortions are introduced. Paradoxically, although publishing in high-credibility journals is important for establishing the legitimacy of results, intermediary organizations may not have access to such paywalled articles, so providing access in another way may be necessary. Technical debate seems often to accompany policy impact, so researchers should not be surprised if they are required to defend their analyses. We hope that these lessons provide useful guidance to the next generation of scientometricians aspiring to improve the governance of research.

ACKNOWLEDGMENTS

We would like to thank Stephanie Noble for her work on creating an annotated bibliography for this project. Her work was indispensable to the synthesis in this paper.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

Early work on the framework synthesis was funded by a seed grant from the Executive Vice President for Research at Georgia Tech and the Quick Wins program to the primary author. The authors were able to more fully develop these concepts and do the empirical work through the generosity of NSF award 51366FE.

REFERENCES

- Alsalam, N., Beider, P., Gramp, K., & Webre, P. (1998). *The economic effects of federal spending on infrastructure and other investments*. Washington, DC: Congressional Budget Office.
- American Society for Cell Biology. (2012). *San Francisco declaration on research assessment*. <https://sfedora.org/read/> (accessed June 19, 2020).
- Atkins, D., Siegel, J., & Slutsky, J. (2005). Making policy when the evidence is in dispute. *Health Aff (Millwood)*, 24(1), 102–113. <https://doi.org/10.1377/hlthaff.24.1.102>
- Australia Department of Education, Science, and Training. (2002). *Varieties of excellence: Diversity, specialisation and regional engagement* (DEST issues paper, DEST). Canberra, Australia: Department of Education, Science, and Training.
- Australia Department of Education, Science, and Training. (2003a). *Evaluation of knowledge and innovation reforms: Issues paper*. Canberra, Australia: Department of Education, Science, and Training.
- Australia Department of Education, Science, and Training. (2003b). *Mapping Australian science & innovation: Main report*. Canberra, Australia: Department of Education, Science, and Training.

- Australia Department of Education, Science, and Training. (2003c). *National report on higher education in Australia: 1991–2001*. Canberra, Australia: Department of Education, Science, and Training.
- Australia Department of Education, Science and Training. (2004). *Evaluation of knowledge and innovation reforms consultation report*. Canberra, Australia: Department of Education, Science, and Training.
- Brownson, R. C., Chriqui, J. F., & Stamatakis, K. A. (2009). Understanding evidence-based public health policy. *American Journal of Public Health, 99*(9), 1576–1583. <https://doi.org/10.2105/AJPH.2008.156224>
- Brownson, R. C., Dodson, E. A., Kerner, J. F., & Moreland-Russell, S. (2016). Framing research for state policymakers who place a priority on cancer. *Cancer Causes & Control, 27*(8), 1035–1041. <https://doi.org/10.1007/s10552-016-0771-0>
- Brownson, R. C., Fielding, J. E., & Maylahn, C. M. (2009). Evidence-based public health: A fundamental concept for public health practice. *Annual Review of Public Health, 30*, 175–201. <https://doi.org/10.1146/annurev.publhealth.031308.100134>
- Burris, S., Wagenaar, A. C., Swanson, J., Ibrahim, J. K., Wood, J., & Mello, M. M. (2010). Making the case for laws that improve health: A framework for public health law research. *Milbank Quarterly, 88*(2), 169–210. <https://doi.org/10.1111/j.1468-0009.2010.00595.x>
- Butler, L. (2003). Explaining Australia’s increased share of ISI publications—The effects of a funding formula based on publication counts. *Research Policy, 32*(1), 143–155.
- Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., ... Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences, 100*(14), 8086–8091.
- Coffman, J. M., Hong, M. K., Aubry, W. M., Luft, H. S., & Yelin, E. (2009). Translating medical effectiveness research into policy: Lessons from the California Health Benefits Review Program. *Milbank Quarterly, 87*(4), 863–902. <https://doi.org/10.1111/j.1468-0009.2009.00582.x>
- Committee on Science, United States House of Representatives, One Hundred and Fifth Congress. (1998). *Unlocking our future: Towards a new national science policy*. Washington, DC: United States GPO.
- Contandriopoulos, D., Lemire, M., Denis, J. L., & Tremblay, É. (2010). Knowledge exchange processes in organizations and policy arenas: A narrative systematic review of the literature. *Milbank Quarterly, 88*(4), 444–483.
- Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Malden, MA: Blackwell.
- Dagenais, C., Laurendeau, M. C., & Briand-Lamarche, M. (2015). Knowledge brokering in public health: A critical analysis of the results of a qualitative evaluation. *Evaluation and Program Planning, 53*, 10–17. <https://doi.org/10.1016/j.evalprogplan.2015.07.003>
- Dodson, E. A., Geary, N. A., & Brownson, R. C. (2015). State legislators’ sources and use of information: Bridging the gap between research and policy. *Health Education Research, 30*(6), 840–848. <https://doi.org/10.1093/her/cyv044>
- Donovan, C., & Butler, L. (2003). Submission to the Knowledge and Innovation Review. In *Research Evaluation and Policy Project*. Australian National University.
- Dutton, J. E. (1997). Strategic agenda building in organizations. In Z. Shapira (Ed.), *Organizational decision making* (pp. 81–107). Cambridge: Cambridge University Press.
- Elson, M., Ferguson, C. J., Gregerson, M., Hogg, J. L., Ivory, J., Klisanin, D., ... & Wilson, J. (2019). Do policy statements on media effects faithfully represent the science? *Advances in Methods and Practices in Psychological Science, 2*(1), 12–25.
- Fannoney, S.E. (1997). AGENDA ITEM 18: NSB Occasional Paper “Publicly-Funded Research.” In *Approved Minutes Open Session 346th Meeting National Science Board November 13–14, 1997*, National Science Foundation, Arlington, VA. Retrieved from: <https://www.nsf.gov/nsb/meetings/1997/nov/opensess.htm>, January 2, 2020.
- Fielding, J. E., & Briss, P. A. (2006). Promoting evidence-based public health policy: Can we have better evidence and more action? *Health Affairs (Millwood), 25*(4), 969–978. <https://doi.org/10.1377/hlthaff.25.4.969>
- Fielding, J. E., & Frieden, T. R. (2004). Local knowledge to enable local action. *American Journal of Preventive Medicine, 27*(2), 183–184. <https://doi.org/10.1016/j.amepre.2004.04.010>
- Gamble, V. N., & Stone, D. (2006). U.S. policy on health inequities: The interplay of politics and research. *Journal of Health Politics, Policy and Law, 31*(1), 93–126. <https://doi.org/10.1215/03616878-31-1-93>
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusions of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly, 82*(4), 581–629.
- Hanney, S. R., Gonzalez-Block, M. A., Buxton, M. J., & Kogan, M. (2003). The utilisation of health research in policy-making: Concepts, examples and methods of assessment. *Health Research Policy and Systems, 1*(1), 2.
- Hicks, D., (2012). Performance-based university research funding systems *Research Policy, 41*, 251–261. <http://dx.doi.org/10.1016/j.respol.2011.09.007>
- Hicks, D., Wouters, P., Waltman, L., de Rijke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics: Use these 10 principles to guide research evaluation *Nature, 520*, 429–431. <https://doi.org/10.1038/520429a>
- Irvine, J., Martin, B., Peacock, T., & Turner, R. (1985). Charting the decline in British science. *Nature, 316*, 587–590.
- Irvine, J., & Martin, B. R. (1986). Is Britain spending enough on science? *Nature, 323*, 591–594.
- Irvine, J., Martin, B. R., & Isard, P. A. (1990). *Investing in the future*. Cheltenham, UK: Edwin Elgar.
- Jewell, C. J., & Bero, L. A. (2008). “Developing good taste in evidence”: Facilitators of and hindrances to evidence-informed health policy-making in state government. *Milbank Quarterly, 86*(2), 177–208. <https://doi.org/10.1111/j.1468-0009.2008.00519.x>
- Laugesen, M. J., & Isett, K. R. (2013). Evidence use in New York City public health policymaking. *Frontiers in Public Health Services and Systems Research, 2*(7), 2.
- Lavis, J. N., Oxman, A. D., Moynihan, R., & Paulsen, E. J. (2008). Evidence-informed health policy 1—Synthesis of findings from a multi-method study of organizations that support the use of research evidence. *Implementation Science, 3*, 53. <https://doi.org/10.1186/1748-5908-3-53>
- Lavis, J. N., Robertson, D., Woodside, J. M., McLeod, C. B., & Abelson, J. (2003). How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Quarterly, 81*(2), 221–248. <https://doi.org/10.1111/1468-0009.t01-1-00052>
- Lemay, M. A., & Sá, C. (2014). The use of academic research in public health policy and practice. *Research Evaluation, 23*(1), 79–88.
- Mansfield, E. (1991a). Academic research and industrial innovation. *Research Policy, 20*(1), 1–12.
- Mansfield, E. (1991b). Social returns from R&D: Findings, methods and limitations. *Research Technology Management, 34*(6), 24–27.

- Mansfield, E. (1998). Academic research and industrial innovation: An update of empirical findings. *Research Policy*, 26(7), 773–776.
- Martin, B. R., Irvine, J., Narin, F., & Sterritt, C. (1987). The continuing decline of British science. *Nature*, 330(6144), 123–126.
- Martin, B., & Irvine, J. (1985). Charting the decline in British science. Commentary. *Nature*, 316, 587–590.
- Martin, B., & Irvine, J. (1986). Is Britain spending enough on science? Commentary. *Nature*, 323, 591–594.
- Martin, B., & Irvine, J. (1987). The continuing decline of British science. Commentary. *Nature*, 330, 123–126.
- Meagher, L., & Lyall, C. (2013). The invisible made visible: Using impact evaluations to illuminate and inform the role of knowledge intermediaries. *Evidence & Policy*, 9(3), 409–418. <https://doi.org/10.1332/174426413x14818994998468>
- Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy*, 29(3), 409–434.
- Murthy, L., Shepperd, S., Clarke, M. J., Garner, S. E., Lavis, J. N., Perrier, L., ... Straus, S. E. (2012). Interventions to improve the use of systematic reviews in decision-making by health system managers, policy makers and clinicians. *Cochrane Database Systematic Reviews*, 9, CD009401. <https://doi.org/10.1002/14651858.CD009401.pub2>
- Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between US technology and public science. *Research Policy*, 26(3), 317–330.
- National Science Board. (1998). *Industry trends in research support and links to public research*. Washington, DC: National Science Board.
- National Science Board. (2003). *Fulfilling the promise: A report to Congress on the budgetary and programmatic expansion of the National Science Foundation*. Washington, DC: National Science Board.
- National Science Board. (2005). *2020 Vision for the National Science Foundation* (NSB 05-142). Washington, DC: National Science Board.
- Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research*, 14, 2. <https://doi.org/10.1186/1472-6963-14-2>
- Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential Address, American Political Science Association. *American Political Science Review*, 92(1), 1–22.
- Pisano, M. (2016). How research can drive policy: Econometrics and the future of California's infrastructure. *Public Administration Review*, 76(4), 538–539.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Redman, S., Turner, T., Davies, H., Williamson, A., Haynes, A., Brennan, S., ... Green, S. (2015). The SPIRIT Action Framework: A structured approach to selecting and testing strategies to increase the use of research in policy. *Social Science & Medicine*, 136, 147–155. <https://doi.org/10.1016/j.socscimed.2015.05.009>
- Role of Public Investment in Promoting Economic Growth: Hearing before the Committee on Financial Services, U.S. House of Representatives, 110th Cong. 39. (2007). Retrieved from: <https://www.govinfo.gov/content/pkg/CHRG-110hhrg35409/pdf/CHRG-110hhrg35409.pdf>, January 6, 2020.
- Sabatier, P. A., & Jenkins-Smith, H. C. (1993). *Policy change and learning: An advocacy coalition approach*. Boulder, CO: Westview.
- Sarkki, S., Niemela, J., Tinch, R., van den Hove, S., Watt, A., & Young, J. (2014). Balancing credibility, relevance and legitimacy: A critical assessment of trade-offs in science-policy interfaces. *Science and Public Policy*, 41(2), 194–206. <https://doi.org/10.1093/scipol/sct046>
- Science policy: The candidates' response. (1992). *Science*, 258(5081), 384–385.
- Stamatakis, K. A., McBride, T. D., & Brownson, R. C. (2010). Communicating prevention messages to policy makers: The role of stories in promoting physical activity. *Journal of Physical Activity and Health*, 7(suppl. 1), S99–107.
- Stirling, A. (2010). Keep it complex. *Nature*, 468(7327), 1029.
- Stone, D. (1989). Causal stories and the formation of policy agendas. *Political Science Quarterly*, 104(2), 281–300.
- Task Force on the Future of American Innovation. (2006). *Measuring the moment, Benchmarks of our Innovation Future II*. Retrieved from <http://innovationtaskforce.org/docs/Measuring%20the%20Moment%20Report.pdf>
- Tricco, A. C., Cardoso, R., Thomas, S. M., Motiwala, S., Sullivan, S., Kealey, M. R., ... Straus, S. E. (2016). Barriers and facilitators to uptake of systematic reviews by policy makers and health care managers: a scoping review. *Implementation Science*, 11(4). <https://doi.org/10.1186/s13012-016-0370-1>
- Troy, L. M., & Kietzman, K. G. (2016). Enhancing evidence-based public health policy developing and using policy narratives. *Journal of Gerontological Nursing*, 42(6), 11–17. <https://doi.org/10.3928/00989134-20160516-04>
- van de Goor, I., Hamalainen, R. M., Syed, A., Lau, C. J., Sandu, P., Spitters, H., ... REPOPA Consortium. (2017). Determinants of evidence use in public health policy making: Results from a study across six EU countries. *Health Policy*, 121(3), 273–281. <https://doi.org/10.1016/j.healthpol.2017.01.003>
- VanLandingham, G., & Silloway, T. (2016). Bridging the gap between evidence and policy makers: A case study of the Pew-MacArthur results first initiative. *Public Administration Review*, 76(4), 542–546. <https://doi.org/10.1111/puar.12603>
- Webre, P. (1993). *A review of Edwin Mansfield's estimate of the rate of return from academic research and its relevance to the federal budget process*. Washington, DC: CBO Staff Memorandum.