

Chapter Title: Assessing the Benefits of Research

Book Title: Dangerous Science

Book Subtitle: Science Policy and Risk Analysis for Scientists and Engineers

Book Author(s): Daniel J. Rozell

Published by: Ubiquity Press. (2020)

Stable URL: <https://www.jstor.org/stable/j.ctv11cvx39.5>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



This book is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.



JSTOR

*Ubiquity Press* is collaborating with JSTOR to digitize, preserve and extend access to *Dangerous Science*

# Assessing the Benefits of Research

To answer the question of how we assess the risks and benefits of dangerous science, it helps to break down the problem. We will start with the assessment of benefits—a topic frequently revisited during budgetary debates over government funding of research. Trying to assess the benefits of research is a long-standing and contentious activity among science policy analysts and economists. Government-funded research constitutes less than one third of total research spending in the US, but public funding of research does not merely augment or even displace private investment (Czarnitzki & Lopes-Bento 2013). Rather, **public funding is critical to early-stage, high-risk research that the private sector is unwilling to fund**. The result is a disproportionate contribution of government funding to technological innovation (Mazzucato 2011). For example, while the private sector funds nearly all the pharmaceutical clinical trials in the US, public funding is still the largest source of pharmaceutical basic research.

## Methods of Valuation

So how do policymakers assess the benefits of publicly funded research? Let's look at some of the most common approaches.

### *Research as a jobs program*

In a simple input-output model of research, spending on salaries, equipment, and facilities associated with research has an analogous output of research jobs, manufacturing jobs, construction jobs, and so forth, but the impact of the actual research output is neglected (Lane 2009). While this is a simplistic way to

---

#### **How to cite this book chapter:**

Rozell, D. J. 2020. *Dangerous Science: Science Policy and Risk Analysis for Scientists and Engineers*. Pp. 9–27. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bci.b>. License: CC-BY 4.0

view research, it is popular for two reasons. First, it is relatively quantifiable and predictable compared to methods that focus on research output. For example, the STAR METRICS<sup>1</sup> program was started in 2009 to replace anecdotes with data that could be analyzed to inform the ‘science of science policy’ (Largent & Lane 2012). However, the first phase of STAR METRICS only attempted to measure job creation from federal spending (Lane & Bertuzzi 2011; Weinberg et al. 2014). A subsequent program, UMETRICS, designed to measure university research effects, used a similar approach by analyzing the same STAR METRICS data to determine the job placement and earnings of doctorate recipients (Zolas et al. 2015).

The second reason for the jobs-only approach is because job creation and retention is a primary focus of government policymakers. Elected officials may talk about the long-term implications of research spending, but the short-term impacts on jobs for their constituents are far more relevant to their bids for re-election. As US Representative George E. Brown Jr. noted, the unofficial science and technology funding policy of Congress is ‘Anything close to my district or state is better than something farther away’ (Brown 1999).

One outcome of viewing research in terms of immediate job creation is any research program may be seen as a benefit to society because all research creates jobs. However, this ignores that some research, through automation development or productivity improvements, will eventually eliminate jobs. Likewise, when research funding is focused on the desire to retain science and engineering jobs in a particular electoral district, it can diminish the perceived legitimacy of a research program. For example, there is a long-standing cynical perception of some US National Aeronautics and Space Administration (NASA) funding acting as a southern states jobs program (Berger 2013; Clark 2013).

### *Econometric valuation*

The jobs-only perspective is obviously narrow, so most serious attempts at measuring the benefits of research use broader economic indicators. Econometric methods have attempted to measure the value of research by either a microeconomic or macroeconomic approach.

The microeconomic approach attempts to estimate the direct and indirect benefits of a particular innovation, often by historical case study. The case study approach offers a depth of insight about particular technologies that is often underappreciated (Flyvbjerg 2006). However, it is time and resource intensive, and its detailed qualitative nature does not lend itself to decontextualized

---

<sup>1</sup> Science and Technology for America’s Reinvestment - Measuring the Effects of Research on Innovation, Competitiveness, and Science (a bit of a stretch for an acronym)

quantification.<sup>2</sup> Furthermore, actual benefit-cost ratios or rates of return for case studies tend to be valid only for the industry and the time period studied. As a result, they can be a poor source for forming generalizations about research activities. Additionally, innovation often comes from chance discovery (Ban 2006), which further complicates attempts to directly correlate specific research to economic productivity.

The macroeconomic approach attempts to relate past research investments to an economic indicator, such as gross domestic product (GDP).<sup>3</sup> This approach is more useful for evaluating a broader range of research activities. Using the macroeconomic approach, the value of research is the total output or productivity of an organization or economy based on past research investments. Three important factors have been noted when attempting a macroeconomic valuation of research (Griliches 1979):

1. The time lag between when research is conducted and when its results are used defines the timeframe of the analysis. Depending on the research, the time lag from investment to implementation may take years or decades.
2. The rate at which research becomes obsolete as it is replaced by newer technology and processes should be considered. The knowledge depreciation rate should be higher for a rapidly changing technology than for basic science research. For example, expertise in vacuum tubes became substantially less valuable after the invention of the transistor. Conversely, the value of a mathematical method commonly used in computer science might increase over time.
3. There is a spillover effect in research based on the amount of similar research being conducted by competing organizations that has an impact on the value of an organization's own research. This effect might be small for unique research that is unlikely to be used elsewhere. Further complicating this effect is the influence of an organization's 'absorptive capacity' or ability to make use of research output that was developed elsewhere (Cohen & Levinthal 1989). Even without performing substantial research on its own, by keeping at least a minimum level of research capability, an organization can reap the benefits of the publicly available research output in its field.

---

<sup>2</sup> This has not stopped big-data enthusiasts from trying. For example, keyword text-mining was performed on a 7,000 case study audit of research impacts in the United Kingdom. Ironically, the point of the audit was to add complementary context to a quantitative assessment (Van Noorden 2015).

<sup>3</sup> While GDP is a popular economic indicator, detractors dislike its simplicity which ignores many factors, including inequality, quality-of-life, happiness, and environmental health (Masood 2016; Graham, Laffan & Pinto 2018).

In general, quantifying any of the above factors is easier for applied research than for basic research. Likewise, it is easier to quantify private benefit to a particular organization than public benefit. Another factor that prevents easy identification of the economic value of research is the general lack of data variability. Research funding rarely changes abruptly over time, so it is difficult to measure the lag between research investments and results (Lach & Schanker-man 1989).

The most common approach for determining the economic rate of return for research is growth accounting where research is assumed to produce all economic growth not accounted for by other inputs, such as labor and capital. Economists often refer to this unaccounted growth as the Solow residual (Solow 1957). A comprehensive review (Hall, Mairesse & Mohnen 2009) of 147 prior research studies that used either an individual business, an industry, a region, or a country to estimate the rate of return of research found a variety of results. The majority of studies found rates of return ranging from 0 to 50 percent, but a dozen studies showed rates over 100 percent<sup>4</sup>—a wide interval that portrays the difficulty of quantifying the benefits of research.

Not surprisingly, the return on research is not constant across fields, countries, or time, so any estimates from one study should be used cautiously elsewhere. Likewise, it is important to distinguish general technological progress from research. While technological progress may account for most of the Solow residual, a non-negligible amount of innovation occurs outside of funded research programs (Kranzberg 1967, 1968). Ultimately, due to the many potential confounding factors, such as broader economic conditions or political decisions, it is difficult to show a causal relationship for any correlation of productivity or profit with research.

Unfortunately, some advocacy groups have issued reports that imply a simple direct relationship between scientific investment and economic growth. Such statements are unsupported by historical data (Lane 2009). For example, Japan spends a higher proportion of GDP on research than most countries but has not experienced the expected commensurate economic growth for the past two decades. Likewise, at the beginning of this century, research spending in the US was about 10 times higher than in China. As a result, US contribution to the global scientific literature was also about 10 times higher than China's. However, in the following decade, China's economy expanded 10 times faster than the US economy.<sup>5</sup> The exact relationship between research spending and economic growth remains unclear; the only consensus is that research is beneficial.

---

<sup>4</sup> One outlier study conservatively (or humbly) estimated rates between -606 and 734 percent.

<sup>5</sup> Ironically, China's robust economy allowed it to dramatically increase its research spending and eventually surpass the US in total number of science publications according to NSF statistics.

### *Valuation by knowledge output*

Given the difficulties of using econometric methods to assess research, economists have explored other methods that avoid monetizing research benefits and the private versus social benefit distinction. One popular alternative is to use academic publications. Despite its relative simplicity compared to economic growth, publications are still problematic. First, comparisons are complicated because scientific publication is not equally valued among all fields and organizations. Publication in prestigious journals is often essential to career advancement in academia but is relatively unimportant for industrial research scientists. Likewise, an ever-increasing proportion of research is being disseminated outside standard academic journals via the internet—open access pre-print archives, research data repositories, and code-sharing sites have all become common. It is unclear how these new modes of information sharing should be measured. Second, the method does not assess the relative value or visibility of an individual publication. This issue is partially addressed by using the number of citations rather than the number of publications. However, citations are not a clear sign of quality research. For example, citations are commonly made to provide basic background information (Werner 2015). Similarly, a journal's impact factor—the average number of citations for a journal's articles in the past year—is widely derided as a proxy for research output quality (yet the practice is still shamefully common). Conversely, a lack of citations does not necessarily indicate a lack of social benefit. The information in an academic research article may be widely used in non-academic publications—reports, maps, websites, and so forth—without ever generating a citation that can be easily found. Likewise, online papers that have been repeatedly viewed and downloaded, but never cited, may have more public value than a cited paper with less internet traffic. Additional shortcomings are similar to traditional econometric approaches: what is the appropriate lag time for publications and what time window should be considered for counting publications based on the depreciation rate of scientific knowledge (Adams & Sveikauskas 1993)?

Scientists love to create models for complex problems, so it should be no surprise that a model was created that estimates the ultimate number of citations for a particular article (Wang, Song & Barabási 2013). The model included the following characteristics: citations accrue faster to papers that already have many citations, a log-normal decay rate for future citations, and a general factor that accounts for the novelty and importance of a paper. However, the model required 5 to 10 years of citation history to make projections, and the difficulty of properly calibrating the model limited its utility (Wang et al. 2014; Wang, Mei & Hicks 2014). Conversely, an extensive study that looked at 22 million papers published over the timespan of a century in the natural and social sciences found that citation histories are mixed and unpredictable (Ke et al. 2015). Some extreme papers, labeled 'sleeping beauties,' accumulated few citations for

decades and then suddenly peaked—presumably because an important application for the research occurred at a much later date. Likewise, some of the most novel papers tend to languish for years in less prestigious journals but are eventually recognized by other fields for their original contributions and eventually become highly cited (Wang, Veugelers & Stephan 2016). Generally speaking, using short-term citations as a metric for assessing research is a bad idea.

A similar non-monetary approach for measuring research benefits is to count the number of patent citations in a particular field (Griliches 1979; Jaffe, Trajtenberg & Henderson 1993; Ahmadpoor & Jones 2017). This method has the benefit of better assessing the practical value of research activities and capturing the technological innovation component of research that is likely to have high social benefit. However, this method also shares some of the drawbacks of the publication approach as well as a few unique drawbacks of its own. The economist Zvi Griliches observed that a US productivity peak in the late 1960s was followed by a decline in patents granted in the early 1970s and that both events were preceded by a decline in the proportion of GDP devoted to industrial research spending in the mid-1960s. Whether productivity and patents followed a 5- to 10-year lag behind research spending was difficult to determine given that among other factors, the number of patents per research dollar also declined during that time period, an energy crisis occurred during that time period, and other countries suffered similar productivity losses without the drop in research funding (Griliches 1994).

Fluctuations in patent generation may also be due to the national patent office itself. For example, the 2011 Leahy-Smith America Invents Act, which took effect in 2013, changed the unique US first to invent patent system to a more standard first to file system. This makes comparisons before and after the new system more difficult. Likewise, as Griliches noted, stagnant or declining funding for a patent office could limit the throughput of the department or prevent it from keeping up with growing patent application submissions. This very phenomenon appears to have occurred in the US since the innovation boom of the Internet age (Wyatt 2011).

Ultimately, patents remain a limited and non-representative measure of research benefits. There is poor correlation between patents and public benefit because most benefits come from a small subset of all patents and only about half of all patents are ever used and fewer are ever renewed (Scotchmer 2004). Also, not all organizations patent their inventions at the same rate because the value of a patent is distinct from the value of the invention (Bessen 2008). Pharmaceutical patents can be extremely valuable; whereas, software-related innovations are more difficult to defensibly patent and are often obsolete before the patent is even awarded. A review of the 100 most important inventions each year from 1977 to 2004, as judged by the journal *Research and Development* ('R&D 100 Awards'), found that only one-tenth were actually patented (Fontana et al. 2013). Most companies relied on trade secrets or first-to-market advantages

rather than patents. Patents allow a holder to litigate against infringement, but this legal right is often too expensive and time-consuming for all but the largest organizations to carry out. Alternatively, a large collection of related patents can create a 'patent thicket,' where its primary value is rent-seeking and slowing competitors, not social benefit. A CDC list of the most important public health achievements of the 20th century contained no patented innovations (Boldrin & Levine 2008), suggesting patents are indeed a very poor measure of research social benefit. Nonetheless, patents are still widely used as a measure of research value for lack of a convincing alternative.

While economists view the measurement of knowledge output to be problematic but possible, others believe the problem is intractable or at least not quantifiable in any honest way. Philosopher Paul Feyerabend argued that a careful study of the history of science shows the truth or usefulness of any particular scientific theory or line of research may not be appreciated for decades or even centuries (Feyerabend 2011). He gave one extreme example of the theory proposed by the Greek philosopher Parmenides of Elea (5th century BCE) that all matter has the same fundamental nature. The theory was abandoned for over 2,000 years before being revived by particle physicists in the 20th century. A more recent example is the theory of continental drift, first proposed in 1596 by Flemish cartographer Abraham Ortelius. The theory was revived in 1912 by meteorologist Alfred Wegener, who unsuccessfully championed the idea for two decades.<sup>6</sup> After the steady accumulation of supporting evidence, the idea was eventually incorporated into the theory of plate tectonics in the 1960s, which now serves as a cornerstone of modern geoscience. Perhaps the most relevant example is the theory that fossil fuels cause global warming, which was first proposed by Swedish scientist Svante Arrhenius in 1896. His work was inspired by British scientist John Tyndall's 1859 work on the radiative heat absorption properties of carbon dioxide and water vapor and their likely effects on the planet's surface temperature. Despite winning the 1903 Nobel Prize in Chemistry for foundational work in electrochemistry, Arrhenius' work detailing the correct mechanism and mathematical relationship between infrared absorption and atmospheric carbon dioxide concentrations was largely ignored for almost a century before anthropogenic climate change was realized to be an unprecedented threat to humanity.

Even though economists are generally trying to measure the short-term societal benefits of more tangible and immediate research, selecting a lag time is merely a choice of analytical convenience. There were decades between the development of quantum physics and technologies based on quantum theory: transistors, lasers, magnetic resonance imaging, and so on. The theory is over a century old and yet new technologies, such as quantum computers, are still in

---

<sup>6</sup> His failure was partly scientific, his observations had no good explanatory mechanism, and partly social, he was an outsider to the geology community and a German World War I veteran.



development. It would be hard to argue that these were impractical or unimportant benefits that could be left out of a realistic benefits assessment. It would seem even a field of research that has yet to yield useful results—such as string theory (Castelvecchi 2015)—should not be dismissed as long as it still has intellectual inspirational value; one never knows what is yet to transpire. Likewise, how does one measure the benefits of long-term research that may require decades to yield significant findings (Owens 2013b).

Selecting a lag time by a cutoff function that is designed to capture most of the citations, patents, or economic growth based on past research is based on the questionable assumption that only the intended outcome of applied research is of interest. However, the history of technology suggests secondary unintended discoveries, both good and bad, are important. For example, in the pharmaceutical industry, drugs are commonly repurposed when they are unexpectedly found to treat a disease other than their intended target. Thus, selecting a time period for the evaluation of research may capture some of the intended outcomes but miss the secondary serendipitous discoveries (Yaqub 2018).

### *Valuation by multiple metrics*

The various metrics discussed so far appear to be poor measures of the social benefits of research. They are popular primarily because they make use of the available data, not because they necessarily measure the desired outcomes. Metrics are frequently pursued with the noble intention of improving accountability and transparency but do not often accomplish either because they tend to oversimplify complex processes and create perverse incentives to game the system when metrics are used to reward or punish individuals.<sup>7</sup>

For example, if patents become a preferred metric of research productivity, some researchers will knowingly generate patents that are of questionable licensing value to improve their likelihood of securing future funding. Likewise, the frequent practice of using the number of publications as a metric has led to academic complaints about ‘salami-slicing’ research and jokes about the ‘least publishable unit.’ Quantitative assessments of research output in the United Kingdom, Australia, and New Zealand may have created the unintended consequence of pushing researchers away from high-risk basic research and toward more conventional, short-term, applied projects to improve their rankings (Owens 2013a; McGilvray 2014). History suggests abandoning basic research in favor of seemingly more predictable short-term applied research is

<sup>7</sup> The metrics-focused system analysis approach of Secretary of Defense Robert McNamara is often blamed for the tragic poor decision-making surrounding the war in Vietnam as well as his later missteps as president of the World Bank. Proponents of metrics often point to successes in far simpler and more quantifiable human endeavors, such as baseball (Muller 2018).

probably counterproductive. For example, how could one predict that the germ theory of disease developed in the 19th century would be the impetus for the modern sanitation techniques responsible for much of the increase in average life expectancy in the 20th century? A review of almost 30 years of biomedical research grants found that basic and applied research were equally likely to be cited in patents (Li, Azoulay & Sampat 2017). Of course, the underlying observation is not new. Abraham Flexner first made the argument that basic research yields important social benefits in his 1939 essay, *The Usefulness of Useless Knowledge*. It appears the message requires frequent repetition.

Despite these critiques, there has been some hope that using a family of complementary metrics would yield an improved estimate over individual research measurements. For example, a combination of publication citations to capture basic research and patents to capture technology development might appear to be a complementary set of measurements. The STAR METRICS program was created to measure the impact of US federally funded research using a multi-dimensional approach. Some of the proposed indicators included (Federal Demonstration Partnership 2013):

- number of patents;
- number of start-up companies;
- economic value of start-up companies over time;
- future employment of student researchers;
- impacts on industry from research;
- the number of researchers employed;
- publications and citations; and
- long-term health and environmental impacts.

While the STAR METRICS approach avoided some of the limitations of individual metrics previously discussed, it was questionable how many of the proposed metrics could be measured in practice or how representative the final set of metrics would be. Given the difficulty of the task, it was not surprising when the full implementation of STAR METRICS program was abandoned in 2015.

A more successful program, the Innovation Union Scoreboard, has been evaluating the research efforts of European Union member states since 2007 (Hollanders & Es-Sadki 2014). It encompasses 25 metrics, including multiple indicators for educational outcomes, scientific publications, patents, public research investments, private research investments, employment, and other economic indicators. As with similar programs, the Innovation Union Scoreboard is by necessity restricted to indicators for which there are data. As such, unquantifiable benefits are missed.

Despite the difficulties of quantitatively valuing research, the era of big-data has inspired an entire alphabet soup of research assessment systems, none of which can be easily compared to each other. Detractors have argued that these broad quantitative measurement tools are just as non-representative and easily

gamed as the many popular, but widely derided, college ranking schemes. It has yet to be seen if any of these multi-metric systems will improve research or how—outside their own definition—success will be determined. However, the rush to quantitative assessment is not universal. The Chinese Academy of Sciences moved away from an existing 24 indicator multi-metric research ranking system to a qualitative system based on peer review (Kun 2015). The motivation was a desire to place emphasis on the real social value of research rather than on easily measured surrogates.

### *Value-of-information analysis*

Econometric methods would appear to be the obvious choice for performing a research cost-benefit analysis. However, as previously discussed, this is a difficult task even for research that has already been conducted. Estimating the value of future research is even more uncertain as it requires the questionable assumption that the future will be much like the past. This is a difficult assumption to defend because history shows the progress of technology to be inconsistent and unpredictable. Computer technology has exceeded most predictions made in the 20th century, yet utilities powered by nuclear fusion have stubbornly remained a technology of the future. Unfortunately, there is no consistent set of criteria that will predict whether a particular research project will succeed. The list of contributing factors is extensive, and there is even disagreement among studies regarding the magnitude and direction of influence of each factor (Balachandra & Friar 1997).

For future research decisions, an alternative to traditional econometric or knowledge output approaches is to use value-of-information (VOI) analysis. In VOI, the value of the research is measured by estimating its expected value to a particular decision and weighing it against the cost of obtaining that information (Morgan, Henrion & Small 1990; Fischhoff 2000).<sup>8</sup> For example, knowing the transmissibility of a particular pathogen has value for public health officials in their decision of how to prepare for future pandemics. This value can be measured in any agreeable units—money, lives saved, response time, and so on. The primary strength of this approach is that it deals directly with the value of research to the decision maker (Claxton & Sculpher 2006). By comparison, high quality research, as measured by knowledge output methods, has no clear correlation to societal benefit only an assumed link. Because VOI is a forward-looking predictive method of valuation rather than a backward-looking

---

<sup>8</sup> VOI literature often uses the term ‘expected value of perfect information,’ which is simply the difference between the value of the decision made with complete information compared to existing information. Restated, this is the value of removing uncertainty from the decision process.

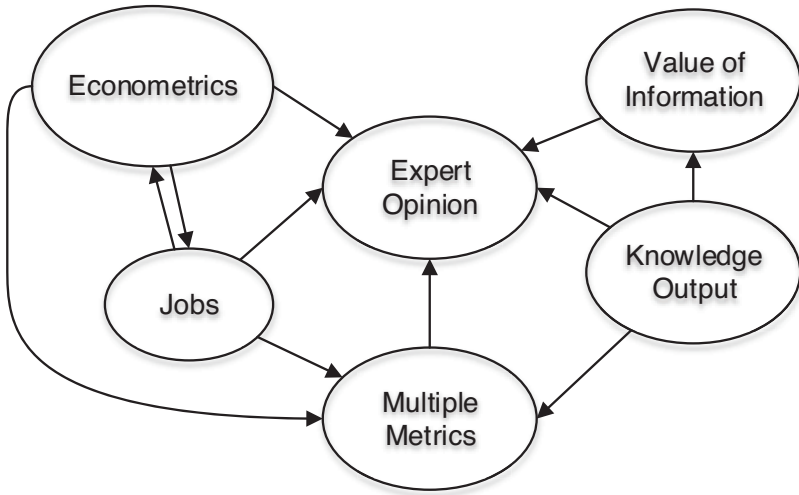
reflective method, it sidesteps the issue of making comparisons between past and future research.

Another strength is that VOI analysis is a more theoretically complete and consistent method of research valuation. Performing a cost-benefit analysis using a family of economic, knowledge, and social metrics can use collected data, but that data will generally be an incomplete measure of the total value of research and will often consist of proxies for the characteristics we would prefer to measure. Conversely, a VOI approach can place a direct value on factors that are difficult to monetize: aesthetic, intellectual, or even the cultural significance of a scientific discovery. Thus, VOI is complete in the sense that any recognized benefit can be included in the analysis.

However, the thoroughness of the VOI approach comes at the price of subjective estimates and value judgments. VOI is a productive decision tool only when one can reasonably estimate the value of obtaining the information. For that reason, VOI is often applied to business, engineering, and applied science decisions (Keisler et al. 2013). For example, VOI would be useful for estimating whether a particular medical test has value for decisions about patient treatment. However, it is harder to use VOI for estimating the value of highly uncertain basic research. VOI is subjective when it measures subjective things. It cannot create certainty out of uncertainty.

The thoroughness of the VOI approach also complicates analysis due to the broader array of potential social benefits that might be considered. While VOI is simple in concept, it can be quite complex in practice. For this reason, the VOI approach is often used in conjunction with an influence diagram—a visual representation of a decision process that represents variables as nodes and interactions between variables as arrows (Howard & Matheson 2005). The influence diagram serves as a visual aid to elucidate and organize the often complex interaction of factors that can affect the value of basic research. However, an influence diagram with more than a dozen or so nodes and arrows tends to become an unreadable labyrinth that provides little insight.

As an example, Figure 1 shows the relation among the various ways in which research can be valued as an influence diagram. Each form of valuation is represented as a node with arrows indicating if the method informs another method. For example, job creation is often concurrent with economic growth (but not always), so we would expect these two research valuation methods to be closely related. Likewise, both jobs and economic growth can be used in a multi-metric approach or in an expert opinion approach. Knowledge output, in the form of citations and patents, can also be used in a multi-metric approach and is similar to the VOI approach in that both are non-monetary and can more easily characterize the value of basic research with no immediate practical applications. Expert opinion, discussed in the next section, is the most comprehensive approach in that it can make use of all the other methods of valuation. However, in practice, expert opinion can range from superficial to comprehensive.



**Figure 1:** Methods of valuing research.

Although less formal than the VOI approach, a similar process can be used to reconcile the supply and demand for science research (Sarewitz & Pielke Jr 2007). This is done by collecting the information required by policymakers (the demand) through workshops, surveys, interviews, and committees. Using the same process, regular assessments are made regarding whether the research (the supply) is actually being used. Rather than placing a common value on the information, the intent is only to re-align research priorities to maximize social benefit. In theory, this is a great idea because useful science can happen by accident, but more useful science will happen when it is done with purpose. The priority re-alignment process is much less subjective than VOI in the sense that it does not attempt to quantitatively compare research programs. However, it is also time consuming in that it seeks input from all stakeholder groups and can be difficult to complete when contentious issues preclude a consensus on the demand for science. Furthermore, it is difficult to predict what research will actually yield the most social benefit; focusing only on immediate applied research would miss important basic research that eventually yields important technology.

In standard VOI literature, benefit is derived from additional knowledge, and it is assumed that the value of information can never be negative because a decision maker can always choose to ignore low-value information (Blackwell 1951). However, experiments suggest decision makers are often unable to ignore unhelpful information once it is known due to a ‘curse of knowledge’

(Camerer, Loewenstein & Weber 1989).<sup>9</sup> Furthermore, decision makers are often unaware when information is unhelpful based on their surprising willingness to pay for unhelpful information (Loewenstein, Moore & Weber, 2003). This questions the basic assumption that the value of information is never negative because it can be ignored without cost.

We can extend this concept of negative value of information to include research that may yield knowledge that has potential public harm, such as dual-use research that has obvious use by military, terrorists, or criminals. Without the negative VOI concept, research cannot be any worse than wasted effort. With the idea of negative VOI, some research programs may yield information we might prefer not to know or find morally objectionable (Kass 2009). Likewise, some research might harm the public because it is erroneous. For example, a 1998 *Lancet* paper linked the MMR vaccine with autism. Although later discredited and retracted, it fueled suspicion regarding the safety of childhood vaccination; subsequent outbreaks of preventable diseases and multiple fatalities occurred in communities that disproportionately avoided vaccination (Gross 2009).

### *Qualitative assessment by expert opinion*

A 1986 US Office of Technology Assessment report reviewed a variety of quantitative methods for determining the value of research and the prevalence of such methods in industry and government. The report found that the majority of managers preferred 'the judgment of mature, experienced managers' as the best method for assessing the value of research (OTA 1986). **Formal quantitative models were perceived to be misleading due to their simplistic nature, which missed the complexity and uncertainty inherent in the decision-making process.**

Given the issues with various quantitative methods as previously described, it is not surprising that expert opinion is still the gold standard in estimating the value of research. However, qualitative expert review is also problematic. Two fundamental difficulties with using expert opinion are conflicts of interest and unavoidable bias. Specialists are usually employed within their field of expertise, which leads to a weak, but pervasive, financial conflict of interest. Likewise, people tend to attach the most value to activities on which they have spent the most time. This phenomenon, referred to as effort justification

---

<sup>9</sup> An example of the curse of knowledge occurs in teaching. It is extremely difficult to imagine one's own state of mind before a concept was understood. This leads to teachers often overestimating the clarity of their instruction and the comprehension in their students (Weiman 2007).

(Festinger 1957) or the IKEA effect (Norton, Mochon & Ariely 2012)—because people tend to value an object more when they assemble it themselves—can lead experts to unintentionally overestimate the value of the research with which they have been most involved. Even the appearance of conflict between what is in the best interest for the general public versus the experts themselves decreases credibility and can make research assessment discussions look like special interest lobbying.

One way to partially compensate for potential expert bias is to actively seek competing views. Philosopher Philip Kitcher recommends an ‘enlightened democracy’ where well-informed individuals selected to broadly represent society set science research agendas (Kitcher 2001). This ideal is set as a middle ground between a ‘vulgar democracy’, where science suffers from the ‘tyranny of the ignorant’, and the existing system, where a struggle for control over the research agenda is waged between scientists (internal elitism) and a privileged group of research funders (external elitism). Some influences on the science research agenda, such as focused lobbying by well-informed advocates, defy this idealized distinction between a scientific elite and an uninformed public. Nonetheless, the struggle to maintain a balanced and representative set of research policymakers is real.

One example of this struggle was the President’s Science Advisory Committee created by US President Eisenhower to provide cautious science policy analysis during the American pro-science panic that occurred after the launch of Sputnik in October 1957. The Committee’s criticism of President Kennedy’s manned space program and President Johnson’s and President Nixon’s military programs led to its ultimate demise in 1973 (Wang 2008). The subsequent Office of Technology Assessment served in a similar role for the US Congress but fared only marginally better lasting from 1972 to 1995. It attempted to maintain neutrality by only explaining policy options without making explicit recommendations. However, its general critique of President Reagan’s Strategic Defense Initiative—mockingly called Star Wars—created conservative antipathy that eventually led to its demise. Suggestions have been made on how to make such science advisory bodies more resilient (Tyler & Akerlof 2019), but these anecdotes suggest that balanced counsel on science policy can be difficult to maintain.

Compared to quantitative methods, assessment by expert opinion is time-consuming and expensive. The tradeoff is supposedly a better assessment. Unfortunately, the historical record is less than convincing. For example, the National Science Foundation (NSF) uses peer review panels to assess research proposals based on the significance of goals, feasibility, the investigator’s track record, and so on, but the process may not be capable of predicting even relative rankings of future research impact. For a study of 41 NSF projects funded a decade prior, panelists’ predictions of future success were found to have no significant correlation with the actual number of publications and citations coming from each funded project (Scheiner & Bouchie 2013).



While expert panels are frequently used with the idea that group decisions are better than individual reviews, scientists are not immune to social dynamics that hinder good decision-making. Non-academics or other outsiders can be sidelined, dominant personalities or senior scientists may expect deference, or the panel may engage in groupthink. Larger studies of the NIH peer review process have found that there is no appreciable difference between high- and low-ranked grant proposals in their eventual number of publications per grant, number of citations adjusted for grant size, or time to publication (Mervis 2014). However, a study of 137,215 NIH grants awarded between 1980 and 2008 found that the highest-rated grant proposals yielded the most publications, citations, and patents such that a proposal with a review score one standard deviation above another generated 8 percent more publications on average (Li & Agha 2015). Critics have questioned the cause of this correlation considering journal publications are also based on peer-review; thus, any correlation may only indicate measurement of the same reputational system.

The journal peer-review system was the subject of another study that followed the publication history of 1,008 submissions to 3 top medical journals (Siler, Lee & Bero 2014). Of the 808 manuscripts that were eventually published, the lowest-rated submissions tended to receive the least eventual citations. However, the top 14 papers were all rejected at least once, which suggests the most innovative high-impact work is often unappreciated by the peer-review process.<sup>10</sup>

Perhaps the most damning critique of expert opinion comes from the many examples throughout history of substantial scientific research that went unappreciated by experts to an extent that is almost comical in hindsight. For example, biologist Lynn Margulis' paper proposing that mitochondria and chloroplasts in eukaryotic cells evolved from bacteria (Sagan 1967) was originally rejected by over a dozen journals. Over a decade later, DNA evidence confirmed the theory and Dr. Margulis was eventually elected to the National Academy of Sciences and given various awards, including the National Medal of Science. In another example, materials engineer Dan Shechtman needed two years to get his paper identifying the existence of quasicrystals published (Shechtman et al. 1984). He was met with ridicule from the scientific community and was even asked to leave a research group. This work eventually earned Dr. Shechtman the Nobel Prize in Chemistry in 2011.

The imprecision of peer review should come as no surprise to anyone who has published in the academic literature for some time. It is not uncommon to receive multiple reviewer comments that make contradictory assessments of a manuscript's quality or request mutually exclusive changes. Critiques of peer review have been common since its inception (Csiszar 2016), and there have

---

<sup>10</sup> While most rejections were desk rejections (that is rejections by the journal editors), this is, in practice, part of the peer-review process. This conservatism may be a sign of 'normal science' in action (cf. Kuhn 1962).

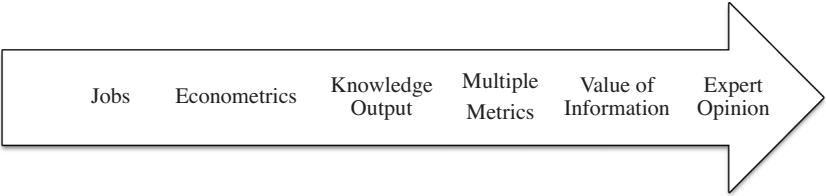


been many attempts to improve the process: abandoning anonymous reviews to improve accountability, publishing reviews to improve transparency, using double-blind reviewing to remove bias for or against the author’s reputation or publication history, or even awarding grants by random lottery to proposals that meet established quality standards. Some of the calls for reform are in fundamental conflict with each other—some want to fund projects not pedigrees, while others want to fund people rather than projects—and each side has a plausible argument. While these changes may make the process fairer, it is unclear if they also improve the ability of experts to assess the long-term merit of research. Ultimately, we are left with the likelihood that expert opinion is the worst way to assess the benefits of research, except for all the other methods.

Implications for Assessing the Benefits of Research

A comparison of the most common ways policymakers assess the benefits of research provides some insight into science policy. Figure 2 shows the various approaches previously discussed ordered from the narrowest to the broadest conception of benefits. This also corresponds to ordering from the most objective to most subjective. That is, assessing the job creation potential of a research program is comparatively objective and data-driven, while expert opinion requires considerable use of subjective estimates and value judgments. The choice of approach depends on the purpose of the assessment. One can use these various methods to obtain answers that are either objective and incomplete or comprehensive and subjective but not objective and comprehensive.

For example, in the H5N1 virus case study in the previous chapter, the benefits of research using potential pandemic pathogens are highly influenced by social factors suggesting that a broader conception of benefits is more appropriate for assessment. Specifically, influenza research is most useful for regions that have functional public health systems. Given the uneven distribution of basic public health services in the world, any research benefits are far more limited in extent than in an ideal world. The problem is further exacerbated by



**Figure 2:** Ways of assessing the benefits of research ordered by increasing completeness of benefits that can be considered and also increasing uncertainty and subjectivity of estimates.

the frequent regression of public health services in regions experiencing war and failed governments. The sad reality is most people in the world have no access to an influenza vaccine of any kind. Since past influenza pandemics were not recognized in their early stages, the likelihood that tailored vaccines can be quickly distributed worldwide is small. These factors undermine the immediate practical health benefits of this research. While this nuanced view of science research benefits is useful, it increases the difficulty of quantification and the uncertainty of the assessment.

Upon reflection, we can see that some types of research are more amenable to particular forms of assessment. This suggests scientists involved in ‘blue skies’ basic research that has only job creation as an immediate quantifiable benefit should avoid getting locked into an econometric valuation debate. When basic science is treated as a mere economic engine, the weaknesses rather than the strengths of curiosity-driven research are emphasized, resulting in weak justifications. Rather, basic science should be honestly argued on intellectual, aesthetic, and even moral grounds if support from the general public is expected.

For example, in 1970, Ernst Stuhlinger, a scientist and NASA administrator, responded to a letter from Sister Mary Jucunda. Given the plight of starving children in Africa, she questioned the expenditure of billions of dollars for manned space flight (Usher 2013). Stuhlinger’s response is an eloquent defense of the value of research in general but a rather weak defense of space exploration based on several proposed practical benefits—none of which are actually dependent on manned space flight: satellite data to improve agricultural output, encouraging science careers, increasing international cooperation, and serving as a more benign outlet for Cold War competition. However, Stuhlinger wisely closes the letter with a reference to an enclosed photograph of the Earth from the Moon and hints at its worldview changing implications. The 1968 picture, now referred to as ‘Earthrise,’ was later described by nature photographer Galen Rowell as ‘the most influential environmental photograph ever taken’ (Henry & Taylor 2009). Sometimes the greatest benefits cannot be quantified.

### Implications for Research Allocation

There appears to be no method for assessing the benefits of research that is comprehensive, objective, and quantitative. This can make any research assessment process rather contentious if all the stakeholders are not already in agreement. Some science policy experts have suggested that the best science funding strategy is simply stable investment over time (Press 2013). The NSF estimated that over \$1 billion was spent over a 40-year timespan on the search for gravitational waves. The result was a technical and intellectual achievement that yielded a Nobel Prize and a new sub-field of astronomy.

However, without the benefit of hindsight, it is hard to present a clear justification of what constitutes optimal research support. And without

justification, proposed funding goals can appear arbitrary and claims of shortages or impending crises may be met with skepticism (Teitelbaum 2014; National Science Board 2016). While this advice rightly acknowledges that research budgets should not be based on the perceived viability of individual projects, it fails to resolve the question of selection. Should policymakers treat and fund all research requests equally?

Clearly, the general public does have science research priorities. A quick internet search of charities operating in the US yields dozens of charities that include cancer research as part of their mission but none for particle physics. The intellectual pleasures of discovering the Higgs boson in 2013 were real, but medical science, with its more immediate application to human health, attracts considerably more public attention. This exact allocation issue was recognized 50 years ago by philosopher Stephen Toulmin who wrote ‘the choice between particle physics and cancer research becomes a decision whether to allocate more funds (a) to the patronage of the intellect or (b) to improving the nation’s health. This is not a technical choice, but a political one’ (Toulmin 1964). The purpose here is not to argue over whether medical research is more worthwhile than particle physics. Rather, it is to highlight how different methods of valuing research have ethical and pragmatic dimensions that effect science policy. A jobs-only valuation approach might prefer funding particle physics research for the many construction and engineering jobs it supports. Meanwhile, an econometric approach might prefer medical research based on historical growth rates in the pharmaceutical sector. Finally, a knowledge output approach might be ambivalent between the two options.

Of course, even with explicit consideration, the expression of public values<sup>11</sup> in science policy is not assured in the near term. For example, if a nation chose to scale back on ‘curiosity’ science, it is not clear that displaced scientists and engineers would necessarily start working on applied projects that would more directly minimize human suffering. Scientists and engineers are not fungible commodities, neither are they devoid of personal preferences regarding how they spend their time—research is not simply a zero-sum game. Likewise, public research funding is generally small compared to many other government expenditures, which may have considerably less societal benefit. In this century, the US federal budget for science research has been approximately one tenth of military spending. One can only imagine the benefits to humanity if those numbers were reversed.

William Press, president of the American Association for the Advancement of Science, stated ‘[a] skeptical and stressed Congress is entitled to wonder

<sup>11</sup> Public values can be defined as the ethical consensus of society on what constitutes the rights, freedoms, and duties of individuals, organizations, and society. This definition also acknowledges that public values are not necessarily fixed, monolithic, or entirely compatible with each other (for example, valuing both liberty and security) (Bozeman 2007).

whether scientists are the geese that lay golden eggs or just another group of pigs at the trough' (Press 2013). Questioning the social value of science was prevalent in the early 20th century (Bernal 1939), but this skeptical attitude about the US science community fell out of favor for several decades after Vannevar Bush rather successfully argued that science should be insulated from the political process (Bush 1945).<sup>12</sup> Nonetheless, research assessment and funding decisions have always been predicated on an expectation of societal benefit. The predominant belief has been that research funding directly translates into knowledge and innovation. The problem is determining exactly what those benefits are and what they are worth.

In summary, there is no universally acceptable method for assessing the benefits of research. This does not mean that assessing the benefits of science research is impossible or uninformative, only that formal quantitative benefits assessments should be used with extreme caution. Quantitative results that appear objective may be hiding a great deal of subjectivity. Failure to consider the limitations of each method risks letting the chosen method shape the goal of the assessment—the reverse of what constitutes good policymaking.

---

<sup>12</sup> See Guston (2000) for a more detailed discussion of the history of changing expectations of science.

