Research Evaluation, 2023, **00**, 1–8
https://doi.org/10.1093/reseval/rvac046
Article

From intent to impact—The decline of broader impacts throughout an NSF project life cycle

Thomas Woodson [1,2,*], Sophia Boutilier[3]

[1]Department of Technology and Society, Stony Brook University, 1431 Computer Science, Stony Brook, NY 11794, USA
[2]InnoLab Research Center, University of Vaasa, Wolffintie 32, FI-65200, Finland
[3]Department of Sociology, Stony Brook University, Circle RD, SBS Building, Stony Brook, NY 11794, USA

*Corresponding author. E-mail: thomas.woodson@stonybrook.edu.

Abstract

It is important for funding agencies to evaluate if scientists accomplish their research goals. By comparing a representative sample of National Science Foundation abstracts and project outcome reports (PORs) from 2014 to 2017, this article investigates whether scientists attain the broader impacts they propose. We find that the number of broader impacts proposed in the abstracts is significantly higher than the number of broader impacts reported in the PORs. The trend is common across directorates and type of impact, except when impacts serve advantaged groups. Only the number of broader impacts for advantaged groups increases from the abstract to the POR. Despite the difference between proposed impact and reported impact, our study does not conclude that scientists are delinquent or disingenuous when they propose their research. Rather, we question the capacity of current frameworks to capture the quality of impacts and to weigh the relative importance of impacts that serve marginalized groups versus those that sustain the status quo.

Keywords: academic research; science policy; research evaluation; responsible research and innovation; broader impacts

1. Introduction

A central component of awarding a grant is ensuring the grant awardee completes a task. In some sectors, such as construction or product procurement, it is more straight forward to evaluate the grant's success due to the transactional nature of the award. In these cases, the granting agency can determine if the awardee accomplished the proposal because they will have the building or product in hand at the end of the contract. However, when it comes to science funding, such concrete measures are often incommensurate with the scientific process. First, it can be hard to measure whether an awardee accomplishes the task they propose because developing new scientific knowledge is subject to various technical challenges that may confound the outcomes. An inherent part of research is that the outcome cannot be known in advance. Second, it is common for scientists to propose one set of work, but in the process of conducting the experiment, they discover a new facet of the problem. These are typical parts of science that lead to important discoveries, rule out alternatives, and contribute to knowledge. Yet, these outcomes may not match an evaluation checklist. Scientific granting agencies struggle with the tension between the compliance of grant awardees and the inherent unpredictability of science (Bornmann 2017; Cozzens 2000).

In addition to generating scientific knowledge, it is common for research and development (R&D) funding organizations to require their projects have societal benefits. In the case of the National Science Foundation (NSF), a federal science funding body in the USA, applicants must explicitly discuss both the intellectual merit and the broader impact of the research in their funding proposals and final reports. However, requiring funded research to have a broad impact is not universally accepted. Scholars question whether scientists should be held to broader impacts given that research outcomes cannot be predicted *a priori* and requiring broader impacts may slow discovery (Polanyi 1962; Tretkoff 2007; Bozeman and Boardman 2009).

To contribute to the conversation on research evaluation and the beneficiaries of scientific funding, this study examines broader impacts throughout the funding cycle of NSF projects. Specifically, our research questions are: Do NSF-sponsored projects propose more broader impacts at the beginning of a project than they report at the end of the project? Second, do NSF-sponsored projects accomplish fewer broader impacts related to marginalized groups than advantaged groups? This study uses two evaluative frameworks, the Robert's classification strategy of broader impacts and Woodson's Inclusion-Immediacy Criterion (ICC), to classify the impacts in NSF grant abstracts and project reports (Roberts 2009; Woodson, Hoffman and Boutilier 2020; Woodson and Boutilier 2022). By comparing the broader impacts in the proposal and final project reports, we can better understand the nature of science funding and how it impacts society. This approach is unique because instead of relying on third-party evaluators to ascertain if the researchers accomplished their proposed task, this study uses the investigators' self-reported assessment to evaluate the research.

This article has five sections. In Section 2, the authors discuss the relevant literature on broader impact and research evaluation that inform this study. Section 3 outlines the methods used to generate the sample and compare the abstracts and project outcome reports (PORs). Section 4 discusses the article's main findings, and Section 5 highlights the implications of our findings.

2. Broader impacts

Since World War II, scientists, politicians, and the public have regularly emphasized that public funding for science should be accountable to taxpayers, the knowledge from discoveries should be shared, and science should be relevant to the public (Bush 1945; Penfield et al. 2014). This belief is reflected by research agencies around the world ranging from Europe's policies on responsible research and innovation (Wiwanitkit and Wiwanitkit 2014) to South Africa's policies using science to correct the horrors of apartheid (Marais et al. 2010). The US NSF mandated scientific impact through policies such as broader impacts. Though NSF broader impact policy has evolved over the last 70 years, it has always been an important part of the NSF's mission (National Science Foundation 2014). Since the 1980s, broader impact requirements have become increasingly integrated into the award process, and in 1997 the NSF codified that award applicants consider the intellectual merit and the broader impacts of their research.

In 2010, the America Competes Reauthorization Act of 2010 mandated that the NSF continue to use broader impacts as one of the standards to judge grants (Public Law 111-274 2010). The goal of the law was to enhance US R&D competitiveness through a variety of initiatives such as more coordination of STEM education, prizes to incentivize R&D, regional innovation hubs, and increasing R&D funding by about $45.6 billion (Gonzalez 2011). One of the law's required the NSF continue applying the broader impact criterion to increase economic competitiveness, create a globally competitive workforce, improve STEM education, and increase national security (Public Law 111-274 2010). Because the broader impact criterion is law, there is intensified oversight and evaluation of the criterion. PIs need to give concrete attention to broader impacts and how their institutions support impacts.

However, the America Competes Reauthorization Act of 2010 gives the NSF flexibility in applying broader impacts. The NSF does not have a standardized list of broader impacts. Rather, the Foundation purposefully leaves broaden impacts vague so that PI's will be creative with impacts and will not simply apply the criterion as a check box exercise (Davis and Lass 2014). Because there is not a standard list, researchers often divide broader impacts into eight categories based on NSF statements. The categories (see Table 1) primarily identify types of activities carried out by research teams. Some of the categories are specific, such as K-12 (primary and secondary school) outreach, while others, such as potential societal benefits, give scientists considerable breadth to integrate impacts (Roberts 2009). A key factor in the broader impact criterion is that they are more than just output measurements. Rather the impact measurements align more closely with outcome measurements from the public policy/affairs literature (Dal Mas et al. 2019).

2.1 Dissent of broader impacts

Alongside the steady insistence that science have broader impact, other voices in the science community question the value of broader impacts. At the most fundamental level, some wonder whether broader impacts are necessary or even appropriate to require of scientists (Kelly and McNicoll 2011). These critiques challenge linking societal impacts with scientific discovery, arguing that researchers should be encouraged to explore free of constraints to deliver pre-determined results. Broader impact criteria hinder free exploration and are viewed as an assault on academic freedom (Chubb et al. 2017).

Others ask whether scientists are equipped to deliver social dividends and if these distract them from the research process for which they are trained (Burggren 2009). Broader impacts add extra burdens for scholars. Not only do they have to manage the main research project, but they must also keep track of broader impacts (Sivertsen and Meijer 2020). The extra burden could push scholars to do the minimal effort for broader impact activities.

A third problem is that many grants are awarded in peer review panels that are ill equipped to determine if the project has an impact (Derrick and Samuel 2016). Peer review is central to scientific funding, evaluation, and the self-regulation of science. In theory, a panel brings together different experts to fairly judge the potential or quality of work (Langfeldt 2004). Review panels have the most legitimacy if those charged with evaluating the work are experts in the field. This system quickly breaks down when the panel judges the merits of items not in their expertise. In evaluating the social impact of research, few scientific expert panels have the expertise to evaluate this area (Bozeman and Boardman 2009; Holbrook and Hrotic 2013).

Furthermore, scholars ask whether proposing impacts is just a formality to win a grant than a genuine intention to have impact (Mardis, Hoffman and McMartin 2012). The broader impact criterion could even incentivize scientists to exaggerate the impact of their research to secure funding. The scientists may believe that the best way to receive a competitive award is to aggressively hype the benefits with minimal evidence to justify their confidence (Intemann 2022). Even if scientist do not exaggerate the benefits of their research to be deceptive, they may be overly optimistic out of belief that their research will be impactful.

Although the NSF makes funding decisions based on broader impacts, it does not impose strict requirements or guidelines governing how scientists accomplish them. Once a

Table 1. Eight common categories for broader impact including an example for each impact (Roberts 2009)

Infrastructure for Science-Creating a user facility for scientists to operate a high-powered microscope	Broadening Participation-recruiting PhD students from underrepresented minority groups to participate in a conference
Training and Education-Training graduate students or incorporating research material in a course	Academic Collaboration-developing a cross institutional collaboration
K-12 (Primary and Secondary School) Outreach-Going to a primary school to discuss the life of a scientist	Potential Societal Benefits-creating a device to reduce greenhouse gases in the atmosphere
Outreach/Broad Dissemination-Sharing the results of the study with policymakers	Partnerships with Potential Users of Research Results-Working with industry to implement the research finding in a product

scientist receives funding, they have a great deal of freedom to conduct the research as they see fit, and there is little oversight directing them to fulfill the project goals. This funding strategy accounts for the realization that the scientific process can be unpredictable (Davis and Lass 2014; Penfield et al. 2014) but it could limit the broader impact of research. Despite some training for scientists on broader impacts, proposal abstracts can have aspirational and incomplete statements regarding impact (Mardis, Hoffman and McMartin 2012). Compared to the intellectual merit portion of a grant, broader impact strategies do not receive the same attention despite suggestions by the NSF to delineate broader impact rationale, approach, feasibility, and evaluation procedures in the grant proposal (Wilbert and Tedjasaputra 2021). Instead, broader impacts are often proposed through sweeping claims that are difficult to substantiate (Mardis, Hoffman and McMartin 2012).

Moreover, it is challenging connecting the claim to actual impact. A core part of the research evaluation literature evaluates and measures the claims of research. One study characterized 10 different assessment methods to highlight the differences in intellectual underpinnings of the framework and how that affects tools and measurements (Smit and Hessels 2021). For example, frameworks that emphasis the monetary value of impact may construct detailed cost–benefit analysis. Other frameworks, such as public values mappings, eschew only economic consideration and highlight non-economic aspects of R&D funding (Bozeman and Sarewitz 2011). Research evaluation struggles to measure causal claims of science, and in the case of broader impacts, linking a particular project to a long-term impact can be even more challenging (Penfield et al. 2014; Reale et al. 2018).

The challenge of connecting declaration and impact is especially pronounced for impacts concerning underserved groups, such as low-income populations and ethnic minorities (Mardis, Hoffman and McMartin 2012). Previous studies found that impacts directed at underrepresented groups were the least likely to be proposed or achieved (Watts, George and Levey, 2015). The same study finds scientists over-exaggerated expected impacts for underrepresented groups by 110% and the failure to deliver on these impacts drove the overall drop in impacts from the proposal to completion stages. It is unclear from these findings whether scientists were disingenuous in their commitment to underrepresented groups or if the reductions in reported impacts were simply due to the challenge of achieving the goals, lenient reporting requirements, or the over-optimism of the PI.

De Jong and Muhonen (2020) developed a measure of societal impact capacity to capture the 'ability of academic researchers to realize benefits to society based on academic research'. The measure focuses 'on the processes and conditions that lead to societal impact' and not assessing the actual outcome (De Jong and Muhonen 2020). They find that high performing European countries, such as France and Germany, have more capacity to do research that impacts society than low performing European countries like Estonia and Romania. De Jong and Muhonen's analysis directs scholars' attention to country level aspects of inequality in broader impacts. This study differs because it does not analyze the capacity of scholars to achieve broader impacts, but rather, who benefits from the impact.

Given the debates over scientist's capacity to deliver broader impacts, the degree they detract from the research process, and the disproportionate attrition of impacts geared toward marginalized groups, we analyze grant abstracts and reports submitted by NSF grant winners. These data allow us to examine the life cycle of expected impacts and to capture if PIs completed their proposed broader impacts.

3 Data and methods

To answer our research questions, our study evaluates winning NSF awards to determine the types and rates at which impacts are achieved. Following Watts, George and Levey (2015), we evaluate both abstracts, submitted at the proposal stage, and Project Outcome Reports (PORs), submitted at the end of the funding period. This comparison allows us to explore whether scientists over-propose impacts. It also allows us to compare broader impact completion across NSF directorates. The seven research NSF directorates roughly relate to scientific fields (see Table 4 for list of directorates). We investigate impacts proposed and reported using two evaluative frameworks, the Robert's classification of NSF's Broader Impact and the Immediacy Inclusion Criterion (IIC). The Robert's classification categorizes impact based on activity while the IIC observes who is targeted by the impacts and how impacts are integrated into the research process (see coding scheme section below for more details).

We recognize there is a relatively short time, about 2–5 years, between a grant's abstract and POR. However, the POR is the last report required by the NSF. Once the grant is finished and the POR is submitted, the grantee does not need to keep pursuing the broader impacts discussed in the grant, nor are they required to update the Foundation on further research developments. While the structure of reporting limits a longer view, it does allow scholars to report immediate broader impacts closely related to their work. As time goes on, it is more challenging to attribute impact to a particular intervention or discovery due to the temporal nature of evaluation (Penfield et al. 2014; Spaapen and van Drooge 2011). Therefore, we consider the POR to be valuable data to measure impact because it is tied closely to the project and written by the PI.

3.1 Sample

The data for this study are a random sample of 400 pairs of proposal abstracts and PORs that received NSF funding from 2014 to 2017. The abstracts are a short summary statement of the proposed research, and the PORs are follow-up reports submitted at the conclusion of funding where scientists describe their activities during the award period. The NSF specifies that both the abstract and the POR should address the project's task, findings, intellectual merit, and broader impacts (National Science Foundation 2021). The abstracts and PORs are public-facing documents, and consequently, the NSF requires PIs to write them so the public can understand them. As result, the abstracts and PORs are short and easier to code than technical documents or research articles. However, the nature of the abstracts and PORs limit the amount of detail the coders can extract from the documents.

Information on awarded grants is openly available on the NSF website. We downloaded all the NSF grants from 2014 to 2017 and then randomly selected 400 grants for analysis. We chose to limit the sample to grants awarded between 2014 and 2017 because, in 2013, the NSF updated their broader impact criterion (National Science Foundation

2014). Moreover, it was not possible to analyze grants after 2017 because many of NSF's research awards are 3 years, and consequently, at the time of the analysis POR data was not available for awards disbursed after 2017.

The NSF awarded a total number of 45 312 grants from 2014 to 2017. A random sample of 400 grants gives the team sufficient power to determine the significance of our results at the 95% confidence level. In addition to the raw data, the team added the grants' directorates by matching a grant's recorded program officer to the directorate where the program officer is based. Directorate data allow us to track how broader impacts compare across fields within the NSF.

3.2 Coding schemes

To analyze the data, we applied two coding schemes to all 400 pairs of abstracts and PORs, the scheme developed by Roberts (2009) and the Immediacy Inclusion Criterion (IIC) developed by Woodson (Woodson, Hoffman and Boutilier 2020; Woodson and Boutilier 2022). The Robert's classification is widely used by scholars studying NSF broad impact criterion and it groups broader impacts into common activity types (see Table 1) (Roberts 2009; Mardis, Hoffman and McMartin 2012; Kamenetzky 2013; Wiley 2014). While the Roberts' categories capture useful information about what kinds of impacts are practiced by scientists, previous research finds that impacts for people from underrepresented groups in STEM fields, such as women and ethnic minorities, are less likely to be achieved (Watts, George and Levey 2015), and that equal access to scientific dividends cannot be assumed (Bozeman 2020).

To address these issues, we also evaluate the abstracts and PORs using the IIC which focuses on who benefits from the broader impact and how the impact relates to the underlying research (see Table 2). The IIC was developed by Woodson and it examines broader impacts based on the immediacy of the broader impact to the underlying research and whom will benefit from the impact. Rather than classifying the activities like in Roberts, the IIC hopes to understand inequality that could arise in the broader impacts of research.

In the IIC, each of the dimensions has three levels. Inclusion refers to who benefits from the impact and its three levels are advantaged, universal, and inclusive. Advantaged groups are wealthy or privileged communities, such as other academics or consumers with high-purchasing power. The category *universal* refers to impacts that help everyone regardless of status. Innovations with universal impact often solve problems related to public goods, like reducing pollution or improving the electrical grid. Lastly, *inclusive* impacts are targeted toward marginalized groups. These groups are generally the last to benefit from research and can face barriers to accessing innovations. Since this study was conducted on grants from the USA, we defined marginalize groups as low-income populations, LGBTQ communities, people with disabilities, and women and racial minorities underrepresented in STEM fields.

Immediacy refers to the incorporation of the impact in relation to the research project. If the impact is *intrinsic* to the research project, then the impact and the research are closely aligned. A good example of an intrinsic impact is developing a smart grid. The research, developing a smart grid, has clear and immediate benefit to society because it makes energy more sustainable and reliable (Tuballa and Abundo 2016). It is nearly impossible to separate the research from the broader impact. If the impact is *direct*, then the impact is central to the research project but not the project's goal. Training graduate students while conducting research is the most common direct impact. Training a graduate student is not the goal of most science research funding, but to accomplish the research goals, the PIs must train them. Finally, impacts that are extrinsic are additions to the research and have no clear connection to the main intellectual thrust of the project. If a high-energy physicist speaks to a secondary school class about becoming a scientist, the impact is extrinsic to the research.

The three levels of inclusion and immediacy intersect with each other to generate nine categories of broader impacts (see Table 2).

3.3 Coding reliability and approach

The data were coded using content analysis. The codebook consists of 18 codes as defined by each criterion (9 for the IIC, 8 for the BIC, and 1 additional code for 'no BI'). The team coded the data using a manifest coding technique. Manifest coding labels only explicit mentions of the topics of interest as identified in the codebook. Manifest coding avoids making inferences from the text or looking for deeper meaning written between the lines (Kleinheksel et al. 2020). This approach was necessary given the breadth of scientific research discussed by grantees and is also consistent with previous scholarship (Watts, George and Levey 2015). In addition, manifest coding is the best approach to minimize coding biases that could occur for scientific disciplines familiar to the authors. It is easy for the coders' backgrounds to give them undue advantage in judging the impacts of the research in their own fields.

Before the full sample was coded, the team tested for intercoder reliability. A sample of 30 PORs were coded by both authors to check for consistency, and during this process, we discussed ambiguous cases and refined the codebook. Initially, our intercoder reliability Krippendorff alpha score was 0.453 which signifies moderate agreement (Hallgren 2012). When we investigated the reliability of each code, we

Table 2. Inclusive-immediacy criterion (author generated)

		Immediacy of Broader Impacts		
		Intrinsic	Direct	Extrinsic
Inclusivity of broader impacts	Universal (everyone)	Developing smart grid technology	Collaborating with public sector utility	Visiting primary school class
	Advantaged (status quo)	Creating smart watches/ fabrics,	Training graduate research assistant	Developing new graduate course
	Inclusive (marginalized group)	Developing new malaria medicines	Training underrepresented groups in STEM	Discussing STEM careers at a low-income high school

discovered that we had only had a moderate disagreement in two of the 18 codes, potential societal benefits and universal intrinsic. These categories have been noted by other researchers to be particularly difficult to code (Von Schomberg 2013; Watts, George and Levey 2015). When we removed the codes with low agreement our alpha increased to 0.637, suggesting substantial intercoder agreement (Hallgren 2012).

3.4 Methodological challenges

The coders faced numerous challenges analyzing the data. First, some of the abstracts and PORs used vague wording or aspirational language that did not clearly demonstrate how the research achieved the broader impacts. Another challenge the coders faced is that they had to understand the proposed impact. Even though the abstracts and PORs should be written for a general audience, some of them are use very specific and technical language. Consequently, the grant awardees think they are conveying an obvious impact, especially to their peers in the field, but to an outside observer, it was not clear there is an impact.

A third problem is that for some awards there was a large mismatch between the stated broader impact and the likelihood that the broader impacts could come to fruition. For example, many innovations such as novel materials, high-end computing, or medicines can be very expensive and only available to powerful groups, like the military, or rich individuals. At some point, the innovative could become cheap and mainstream, but the coders could not reasonably jump to those conclusions without violating the manifest coding protocol. As a result, PORs must explicitly say the research helps marginalized groups or reduces the cost to be classified as inclusive. Otherwise, we categorized it as helping advantage groups/maintaining the status quo or helping everyone universally. Also, if a POR made wildly broads claims such as, 'our research will cure cancer', we considered these types of broader impacts aspirational.[1]

A similar challenge is weighting the scope of an impact against another one. How does improved concrete compare with more efficient wind energy? Both have potential universal societal benefit but are they equal in their impact? The current frameworks cannot capture magnitude or relative importance. Given the tools currently available to us, we simply code each instance of an impact category per grant. The sums of these are reflected in the descriptive statistics provided in Section 4.

4 Results

Below is a list of the main results from the analysis. Our first analysis compares the difference in the number of broader impacts in the abstract versus the PORs across the 400 grants (POR-abstract=difference). When we plot the histogram of the differences, the distribution is largely normal with a slight skew to the left (see Figure 1). The mean difference is −0.5125.

We then conducted a simple *t*-test comparing the number of broader impacts from the abstract to the POR (see Table 3). The analysis shows there are significantly fewer broader impacts in the PORs than the abstracts. This is an important result and confirms observations made by other scholars like Watts, George and Levey (2015). This team found in the NSF's Division of Biology, PI's reported fewer BIs in the PORs than the abstracts in three out of the five categories they explored.

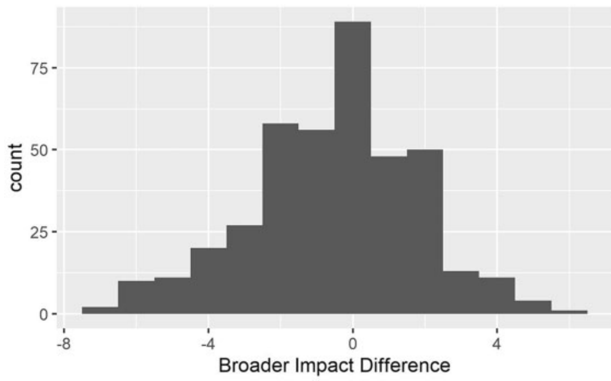

Figure 1. Histogram of the difference between the number of broader impacts in the abstract and POR (POR-abstract=difference).

Table 3. T-test comparing the number of broader impacts in the abstracts and PORs

	Mean	95% Confidence interval		*t*	df	Sig. (two-tailed)
POR—Abstract	−0.5125	−0.74	−0.284	−4.4095	399	1.34E−05

Next, we analyze the broader impacts by directorate (see Table 4). The largest directorates in our study are Engineering (ENG) and Mathematical and Physical Sciences (MPS) with 93 awards each in the study, while the smallest directorate is Education and Human Resources (EHR) with only 26 grants. The number of grants in our sample roughly correlates with the overall size of the directorates.

One important observation from Table 4 is that despite variation in the number of awards and amount of money granted by directorate, there is a general decline in broader impacts from the abstract to the project outcome report. Five out of the seven NSF directorates had more broader impacts expected in the abstracts than reported in the PORs. Only grants in EHR and MPS achieve as many impacts as they propose.

When we examine the number of broader impacts by types of impacts, the results show a similar decline between the abstract and PORs (see Tables 5 and 6). Note that in this part of the analysis, we are not showing counts within the same grant. Rather, we are reporting the total number of abstract and PORs with those types of impacts. The most common broader impacts from Roberts' categorization are training and education and potential social benefit. Overall, there is a clear pattern that abstracts have more expected impacts than are reported in the PORs. However, in two categories, partnership with potential users and training and education, PIs discuss more broader impacts in the PORs than abstract. It is not clear why these categories increase, but it could be that it is easier to achieve these impacts, so they are reported more often. It is also important to note that grants associated with no broader impacts increase 167% from the abstracts to the PORs. This is a large increase, and it further shows that PIs propose more impacts than they accomplish.

Applying the IIC to the data, Table 6 shows the number of impacts in each immediacy and inclusion category at the two

Table 4. Descriptive statistics of grants awarded by directorate from 2014 to 2017

NSF directorates	No. of awards	Impacts in abstract	Impacts in PORs	% Change	Total funds awarded	Median award	SD
Biological Sciences (BIO)	34	175	114	−35	$14,478,405	$306,081	$315,457
Computer & Information Science & Engineering (CISE)	35	130	110	−15	$8,100,957	$150,000	$212,652
Education and Human Resources (EHR)	26	86	86	0	$9,125,311	$249,885	$293,942
Engineering (ENG)	93	315	290	−8	$24,100,058	$200,000	$261,399
Geosciences (GEO)	63	270	209	−23	$15,962,529	$204,120	$193,387
Mathematical and Physical Sciences (MPS)	93	312	313	0	$27,570,328	$223,482	$269,518
Social, Behavioral and Economic Sciences (SBE)	53	162	108	−33	$7,562,766	$99,772	$144,850

Table 5. Number and percent change of proposed and reported broader impact, by Roberts' classification

Broader impact classification	Abstract	POR	% Change
Academic Collaborations	96	62	−35
Broaden Participation	81	69	−15
Infrastructure for Science	104	91	−13
K-12 (primary and secondary school) Outreach	54	47	−13
Outreach Broad Dissemination	39	16	−59
Partnerships with Potential Users of Research Results	49	58	18
Potential Societal Benefits	163	115	−29
Training and Education	211	224	6
No broader impact	15	40	167
Total[a]	797	682	−14

[a] Totals are the sum of the columns less those coded as No BI.

Table 6. Number and percent change of proposed and reported broader impacts, by the IIC classification

	Abstract	POR	% Change
Immediacy			
Intrinsic	353	222	−37
Direct	218	283	30
Extrinsic	74	48	−35
Total	645	553	−14
Inclusion			
Advantaged	349	350	0
Universal	187	122	−35
Inclusive	109	81	−26
Total	645	553	−14

time points—when the project was proposed (abstract) and when its final report was submitted (PoR). The first observation is that about 353/645 (53%) of the expected broader impacts are classified as intrinsic. This result directly addresses a common critique that broader impacts distract scientists from doing research, causing them to spend time doing tasks extrinsic to their research, like primary school outreach. The immediacy data should assuage the concern that scientists will be distracted from research to do broader impact activities. The majority of reported broader impacts (91%) are intrinsic or directly related to the research process. Indeed, by the end of the funding cycle, researchers report *more* direct impacts than they expect at the proposal stage.

The results from the analysis of the inclusion categories indicate that advantaged groups are insulated from diminishing impacts throughout a project life cycle (see Table 6). Expected and reported impacts for advantaged groups are stable over time (349 and 350, respectively). This result directly addresses the second research question. NSF-sponsored projects accomplish fewer broader impacts related to marginalized groups than advantaged groups, in both absolute and relative terms. Decreases in universal and inclusive impacts account for the general decline in impacts. Inclusive impacts are the smallest category in absolute terms but do not experience as much relative decline as universal impacts.

5 Discussion and conclusion

Our findings indicate a systematic decline in impacts from the proposal stage to the conclusion of the funding period. Impacts decline in all but four categories (13 of 17) and all but two directorates (5 out of 7). Across the sample, impacts decline by 14% from the abstract to the POR, but the decline is most pronounced when the impacts are intrinsic to the research or targeted at marginalized groups. This finding is troublesome given the NSF's commitment to broaden participation by engaging underrepresented groups. Furthermore, the Foundation's goals of prosperity, security, health, and growth (National Science Foundation 2020) can be better accomplished in a more equal society (Intemann 2009) which behooves the NSF to prioritize inclusive impacts. Yet, as other studies find, grants with inclusive impacts have less funding and prove more difficult to achieve (Watts, George and Levey 2015; Woodson and Boutilier 2022). By contrast, grants report more impact for advantaged groups from the abstract to the POR. From this finding, it appears the interpretation and enforcement of broader impact policy is not serving marginalized groups. This deficit not only maintains the status quo but may hinder the development of scientific thought due to an absence of diversity (Intemann 2009). Impacts that serve the general population suffer less attrition than inclusive impacts, but these still make up less than a quarter of impacts achieved by the end of the research period.

Like most research, our study raises more questions and there are various other explanations for the decline in impacts that cannot be directly measured. One major explanation for the decline in reported impacts is scientists overpromise impact to get funded. Some of the overpromising could be deliberate, but a part of the hype may be due to the natural optimism of a scientists. They often think their research is more impactful than it really is, and that they are skilled enough to solve major challenges (Holton 1976; Hochschild and Sen, 2015). Without optimism, the scientific enterprise would grind to halt because fewer people would try to find solutions to challenging, if not impossible, problems. Other scholars will need to study if optimism is driving the decline in impacts, but we cannot rule out this factor in influencing these results.

A second issue affecting the results is the challenge of tracking and tracing impact throughout a project. Even if a project has impact, for example mentoring students, these impacts can easily go untracked and uncounted. As a result, at the end of the project, it might appear that there was little impact, but many of the impacts were just invisible.

Third, unclear reporting guidelines could reduce reported impact. The NSF tries to balance giving clear guidance on broader impacts and leaving room for novel approaches to broader impacts. However, a reoccurring complaint of scientists applying for grants that the broader impact criterion is still unclear (Tretkoff 2007).

Fourth, there could be weak enforcement mechanisms to require reporting. This is certainly a weakness of the NSF's broader impact strategy. The foundation requires PIs to discuss broader impacts in their abstracts and PORs, but each directorate and panel review has different norms in assessing the standards. Moreover, there is no oversight in whether the PIs accomplish their proposed impacts and report it back correctly. The system gives a lot of freedom and trust to the PI's. The lack of enforcement can cause a difference in reporting.

In addition to the factors outside of scope of the study that affect broader impacts, our study has several limitations. First, it is difficult to judge the magnitude and depth of an impact using the current tools and the descriptions scientists routinely provide. The lack of granular and comparable information means that all impacts reported are counted as 'one' regardless of their capacity for social transformation. Planning a conference and discovering a cure for cancer are both considered impacts, albeit directed at different target groups, but clearly not equivalent in their effects. Our evaluative approach cannot capture this distinction.

Second, as mentioned in the Section 2, understanding the attribution and longevity of impacts are related challenges. As Mardis, Hoffman and McMartin (2012) point out, impacts are more likely to occur long after the research project. Unfortunately, the longer the time horizon, the more difficult it is to determine the cause of the impact. Because NSF awardees are only required to report impacts completed during the funding period, there may be an unintended emphasis on low hanging fruit, such as training graduate students. This could help to explain the high rates of impacts in the advantaged and direct categories. While such an approach may be pragmatic for scientists receiving NSF funding, it may discourage efforts to pursue broader impacts across the population that take longer to accomplish. As required by law, the NSF's needs to judge grants by its broader impact, but the reporting policies may reinforce impacts for advantaged groups because they are easier to achieve.

A final goal of the broader impact criterion is that it forces scientists to discuss the societal value of their work which can reduce the perception that science is wasting public funding (Kamenetzky 2013). However, what is 'wasted' science funding and how do evaluators account for 'failed' research projects? Reporting on failure is near taboo in the research community (Reale et al. 2018), yet it is not necessarily tantamount to waste. This article cannot determine if NSF funding is wasteful. Ultimately, politicians and the public will need to debate whether they believe funding science, even science that has no clear impact, is important.

Funding

This work was supported by the National Science Foundation (Grant No. 1926494) and by BusinessFinland (Grant No. 33346/31/2020).

Conflict of interest statement. None declared.

Note

1. The BIs coded as aspirational did not yield significant insights based on directorate or target group. Therefore, this code is not included in this analysis.

References

Bornmann, L. (2017) 'Measuring Impact in Research Evaluations: A Thorough Discussion of Methods for, Effects of and Problems with Impact Measurements', *Higher Education*, 73: 775–87.

Bozeman, B., and Sarewitz, D. (2011) 'Public Value Mapping and Science Policy Evaluation', *Minerva*, 49: 1–23.

Bozeman, B. (2020) 'Public Value Science', in *Issues in Science and Technology*. <https://issues.org/public-value-science-innovation-equity-bozeman/> accessed 27 Dec 2022.

Bozeman, B., and Boardman, C. (2009) 'Broad Impacts and Narrow Perspectives: Passing the Buck on Science and Social Impacts', *Social Epistemology*, 23: 183–98.

Burggren, W. W. (2009) 'Implementation of the National Science Foundation's "Broader Impacts": Efficiency Considerations and Alternative Approaches', *Social Epistemology*, 23: 221–37.

Bush, V. (1945) *Science the Endless Frontier a Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development*. Washington DC: United States Government Printing Office.

Chubb, J., Watermeyer, R., and Wakeling, P. (2017) 'Fear and Loathing in the Academy? the Role of Emotion in Response to an Impact Agenda in the UK AND Australia', *Higher Education Research & Development*, 36: 555–68.

Cozzens, S. (2000) 'Assessing Federally-Supported Academic Research in the United States', *Research Evaluation*, 9: 5–10.

Dal Mas, F., Massaro, M., Lombardi, R., and Garlatti, A. (2019) 'From Output to Outcome Measures in the Public Sector: A Structured Literature Review', *International Journal of Organizational Analysis*, 27: 1631–56.

Davis, M., and Laas, K. (2014) 'Broader Impacts" or "Responsible Research and Innovation"? a Comparison of Two Criteria for Funding Research in Science and Engineering', *Science and Engineering Ethics*, 20: 963–83.

De Jong, S. P. L., and Muhonen, R. (2020) 'Who Benefits from Ex Ante Societal Impact Evaluation in the European Funding Arena? A Cross-Country Comparison of Societal Impact Capacity in the Social Sciences and Humanities', *Research Evaluation*, 29: 22–33.

Derrick, G. E., and Samuel, G. N. (2016) 'The Evaluation Scale: Exploring Decisions about Societal Impact in Peer Review Panels', *Minerva*, 54: 75–97.

Gonzalez, H. B. (2011) *Reauthorization of the America COMPETES Act: Selected Policy Provisions, Funding, and Implementation Issues Summary*. Washington, D.C.: Congressional Research Service. <www.crs.gov> accessed 27 Dec 2022.

Hallgren, K. A. (2012) 'Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial', *Tutorials in Quantitative Methods for Psychology*, 8: 23–34.

Hochschild, J., and Sen, M. (2015) 'Technology Optimism or Pessimism about Genomic Science: Variation among Experts and Scholarly Disciplines', *Annals of the American Academy of Political and Social Science*, 658: 236–52.

Holbrook, J. B., and Hrotic, S. (2013) 'Blue Skies, Impacts, and Peer Review', *RT: A Journal on Research Policy & Evaluation*, 1: 1–24. doi: 10.13130/2282-5398/2914.

Holton, G. (1976) 'Scientific Optimism and Societal Concerns: A Note on the Psychology of Scientists*', *Annals of the New York Academy of Sciences*, 265: 82–101.

Intemann, K. (2009) 'Why Diversity Matters: Understanding and Applying the Diversity Component of the National Science Foundation's Broader Impacts Criterion', *Social Epistemology*, 23: 249–66.

Intemann, K. (2022) 'Understanding the Problem of "Hype": Exaggeration, Values, and Trust in Science', *Canadian Journal of Philosophy*, 52: 279–94.

Kamenetzky, J. R. (2013) 'Opportunities for Impact: Statistical Analysis of the National Science Foundation's Broader Impacts Criterion', *Science and Public Policy*, 40: 72–84.

Kelly, U., and McNicoll, I. (2011) *Through a Glass, Darkly: Measuring the Social Value of Universities*. <https://www.publicengagement.ac.uk/sites/default/files/80096%20NCCPE%20Social%20Value%20Report.pdf> accessed 27 Dec 2022.

Kleinheksel, A. J., Rockich-Winston, N., Tawfik, H., and Wyatt, T. R. (2020) 'Demystifying Content Analysis', *American Journal of Pharmaceutical Education*, 84: 7113–37.

Langfeldt, L. (2004) 'Research Evaluation Judging Quality Expert Panels Evaluating Research: Decision-Making and Sources of Bias', *Research Evaluation*, 13: 51–62.

Marais, H. C., Pienaar, M., and Plalliling, S. (2010) 'Evolution of the South African Science, Technology and Innovation System 1994-2010 Evolution of the South African Science', *Technology, Innovation and Development*, 2: 89–102.

Mardis, M. A., Hoffman, E. S., and McMartin, F. P. (2012) 'Toward Broader Impacts: Making Sense of NSF's Merit Review Criteria in the Context of the National Science Digital Library', *Journal of the American Society for Information Science and Technology*, 63: 1758–72.

National Science Foundation. (2014) *Perspectives on Broader Impacts*. <https://www.nsf.gov/news/news_summ.jsp?cntn_id=133319> accessed 27 Dec 2022.

National Science Foundation. (2020) *About the National Science Foundation*. <https://www.nsf.gov/about/#:~:text=The National Science Foundation> accessed 27 Dec 2022.

National Science Foundation. (2021) *Proposal & Award Policies & Procedures Guide*. <https://www.nsf.gov/pubs/policydocs/pappg22_1/nsf22_1.pdf> accessed 1 Jul 2022.

Penfield, T., Baker, M. J., Scoble, R., and Wykes, M. C. (2014) 'Assessment, Evaluations, and Definitions of Research Impact: A Review', *Research Evaluation*, 23: 21–32.

Polanyi, M. (1962) 'The Republic of Science: Its Political and Economic Theory', *Minerva*, 1: 54–73.

Public Law 111-274 (2010).

Reale, E., Avramov, D., Canhial, K., Donovan, C., Flecha, R., Holm, P., Larkin, C., Lepori, B., Mosoni-Fried, J., Oliver, E., Primeri, E., Puigvert, L., Scharnhorst, A., Schubert, A., Soler, M., Soòs, S., Sordé, T., Travis, C., and Van Horik, R. (2018) 'A Review of Literature on Evaluating the Scientific, Social and Political Impact of Social Sciences and Humanities Research', *Research Evaluation*, 27: 298–308.

Roberts, M. R. (2009) 'Realizing Societal Benefit from Academic Research: Analysis of the National Science Foundation's Broader Impacts Criterion', *Social Epistemology*, 23: 199–219.

Sivertsen, G., and Meijer, I. (2020) 'Normal versus Extraordinary Societal Impact: How to Understand, Evaluate, and Improve Research Activities in Their Relations to Society?', *Research Evaluation*, 29: 66–70.

Smit, J. P., and Hessels, L. K. (2021) 'The Production of Scientific and Societal Value in Research Evaluation: A Review of Societal Impact Assessment Methods', *Research Evaluation*, 30: 323–35.

Spaapen, J., and van Drooge, L. (2011) 'Introducing "Productive Interactions" in Social Impact Assessment', *Research Evaluation*, 20: 211–8.

Tretkoff, E. (2007) 'NSF's "Broader Impacts" Criterion Gets Mixed Reviews', *American Physical Society News*, 16. <https://www.aps.org/publications/apsnews/200706/nsf.cfm> accessed 28 Dec 2022.

Tuballa, M. L., and Abundo, M. L. (2016) 'A Review of the Development of Smart Grid Technologies', in *Renewable and Sustainable Energy Reviews*, Vol. 59, pp. 710–725. London: Elsevier Ltd.

Von Schomberg, R. (2013) 'A Vision of Responsible Research and Innovation', in *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, pp. 51–74. West Sussex, UK: Wiley. doi: 10.1002/9781118551424.ch3.

Watts, S. M., George, M. D., and Levey, D. J. (2015) 'Achieving Broader Impacts in the National Science Foundation, Division of Environmental Biology', *BioScience*, 65: 397–407.

Wilbert, T., and Tedjasaputra, V. (2021) *NSF 101: Five Tips for Your Broader Impacts Statement*. National Science Foundation Website. <https://beta.nsf.gov/science-matters/nsf-101-five-tips-your-broader-impacts-statement>.

Wiley, S. L. (2014) 'Doing Broader Impacts? The National Science Foundation (NSF) Broader Impacts Criterion and Communication-Based Activities', in *Iowa State University Digital Repository*, Vol. 1, pp. 12–26. doi: 10.1017/CBO9781107415324.004.

Wiwanitkit, S., and Wiwanitkit, V. (2014) 'Broader Impacts' or 'Responsible Research and Innovation'?', *Science and Engineering Ethics*, 20: 1149–50.

Woodson, T., and Boutilier, S. (2022) 'Impacts for Whom? Assessing Inequalities in NSF-Funded Broader Impacts Using the Inclusion-Immediacy Criterion', *Science and Public Policy*, 49: 168–78.

Woodson, T. S., Hoffmann, E., and Boutilier, S. (2020) 'Evaluating the NSF Broader Impacts with the Inclusion-Immediacy Criterion: A Retrospective Analysis of Nanotechnology Grants', *Technovation*, 101: 1–9.