REPORT

THE HISTORY AND POLITICAL ECONOMY OF NIH PEER REVIEW

Bhaven Sampat, *Professor, Department of Health Policy and Management, Columbia University*

Economic Studies at BROOKINGS

GOOD SCIENCE PROJECT

INSTITUTE FOR PROGRESS

Building a Better NIH is a project partnership between the USC-Brookings Schaeffer Initiative for Health Policy, the Hutchins Center on Fiscal and Monetary Policy, the Institute for Progress, and the Good Science Project.

The author would like to thank Pierre Azoulay, Larry Brown, Daryl Chubin, Robert Cook-Deegan, Richard Frank, Daniel Gross, Michael McGeary, Richard Nelson, Daniel Sarewitz, David Wessel, and Alexandra Witze for their comments and suggestions on an earlier draft. Heidi Williams, Caleb Watney, and participants in the Brookings/Institute for Progress/Good Science Project "Building a Better NIH" Conference also provided useful feedback. Nicholas Cosentino provided excellent research assistance.

The Brookings Institution is financed through the support of a diverse array of foundations, corporations, governments, individuals, as well as an endowment. A list of donors can be found in our annual reports published online. The findings, interpretations, and conclusions in this report are solely those of its author(s) and are not influenced by any donation.

The history and political economy of NIH peer review[1]

Bhaven N. Sampat, Columbia University and NBER

May 2023

Note: *This paper was prepared for the Building a Better NIH project, a partnership between the Brookings Institution, the Institute for Progress, and the Good Science Project. For more information about the project, please visit the Building a Better NIH project page.*

1 Introduction

The National Institutes of Health (NIH) is the world's largest public funder of biomedical research (Viergever and Hendriks 2016). The agency is currently composed of 27 Institutes and Centers with a collective annual budget of over $45 billion, most of which is used to support research at universities and academic medical centers across the U.S. ("Budget" 2014; "Institutes at NIH" 2023). The choices NIH makes in how it allocates this funding affect health and well-being globally.

Though its roots trace back to the 19th century, NIH assumed its modern form after World War II. It has a string of accomplishments to its credit. The agency has funded scientific research that helped shape clinical practice, illuminate health risks, and change health behaviors. NIH-supported research has contributed to developing hundreds of drugs and therapies, including antiretrovirals that reversed the course of the HIV-AIDS epidemic, cancer drugs, and, more recently, Covid-19 diagnostics, vaccines and treatments. It has supported the growth of academic medical centers across the country, helped train the U.S. scientific workforce, and contributed to the rise of the U.S. pharmaceutical, biotechnology, and medical device industries. For all these reasons, the NIH has long been known as the "crown jewel" of the federal government, historically attracting bipartisan political support unrivaled in health or science policy (Cook-Deegan and McGeary 2006). Economic research lends credence to this enthusiasm, suggesting a significant role of NIH-funded research in drug development and health improvements (Sampat and Lichtenberg 2011; Azoulay et al. 2019; Cutler and Kadiyala 2003).

Despite broad, high-level support, there are important disagreements and uncertainties about NIH research policy. A specific focus of current reform efforts is on the agency's peer review system, the machinery it uses to allocate the bulk of its funding. Through its unique "dual" peer

[1] I thank Pierre Azoulay, Larry Brown, Daryl Chubin, Robert Cook-Deegan, Richard Frank, Daniel Gross, Michael McGeary, Richard Nelson, Daniel Sarewitz, David Wessel, and Alexandra Witze for their comments and suggestions on an earlier draft. Heidi Williams, Caleb Watney, and participants in the Brookings/Institute for Progress/Good Science Project "Building a Better NIH" Conference also provided useful feedback. Nicholas Cosentino provided excellent research assistance.

The author did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this article. The author is not currently an officer, director, or board member of any organization with a financial or political interest in this article.

review system, applications for funding are evaluated first for scientific merit by "study sections" composed of scientists in the field, then funded by categorical and disease-oriented institutes and centers (e.g., the National Cancer Institute), which may also assess mission relevance. Today there are over 250 study sections, compared to 21 established after World War II, which review nearly 60,000 applications annually, compared to about 800 per year in 1946 ("CSR Data and Evaluations 2023; Van Slyke 1946). Since the end of the war, NIH funding, awarded primarily through these peer review panels, has increased 1,000-fold, from about $41 million to about $45 billion today (in 2020 dollars). The scale of the U.S. biomedical research system it supports has also grown: in 2022 it supported over 300,000 researchers at 2,500 unique institutions in all 50 states ("NIH Budget: What We Do" 2022). NIH accounts for over one-third of U.S. civilian (non-defense) federal R&D spending, more than three times NASA's R&D funding, and nearly six times that of the National Science Foundation (American Association for the Advancement of Science 2023).

Science and technology policy scholars have extolled NIH peer review as an effective way to balance scientific merit of research with considerations of utility and practical application (Stokes 1997; Mowery, Nelson, and Martin 2010). However, recently academics and other stakeholders have raised questions about the process: whether it effectively identifies and funds breakthrough research (Packalen and Bhattacharya 2020; Franzoni, Stephan, and Veugelers 2022) or the best science (Li and Agha 2015; Fang, Bowen, and Casadevall 2016), whether it favors established researchers (Fang and Casadevall 2009), whether there is racial/ethnic bias in evaluation (Ginther et al. 2011) or conflicts of interest (AAAS 2023), and whether its administrative burdens hinder scientific progress (Buck 2022). In addition to criticisms related to how well the process supports science, there are also questions about whether it effectively targets and promotes desired health outcomes (Gross, Anderson, and Powe 1999; Sarewitz 2016; Kaplan 2019; Callahan 2006; Crow 2011).[2]

Bubbling before the pandemic, criticisms of NIH peer review came to the surface with Covid-19. One set of concerns prompted by the outbreak is how well the agency supports breakthrough science and scientists, in light of the funding difficulties that crucial mRNA research and researchers faced at the NIH before the pandemic (Franzoni, Stephan, and Veugelers 2022). There were also allegations during the crisis that the NIH peer review process did not tackle applied Covid-19 research and development at the needed scale and scope, or with the necessary urgency (Gross and Emanuel 2022; Balaguru et al. 2022).

These and other claims about the NIH peer review process have prompted many reform efforts, including proposals to fund "people not projects" (Ioannidis 2011; Azoulay, Graff Zivin, and Manso 2013); to reduce administrative burdens (Buck 2022); to change the criteria and procedures used to evaluate applications; to move to lottery-based funding systems instead of peer review (Fang and Casadevall 2016); to try randomized trials of different funding approaches (Azoulay 2012); or even to create a new agency to more effectively fund research than the NIH (Sampat and Cook-Deegan 2021; Greenberg 1998).

This paper provides historical context for current reform efforts by chronicling the origins, history, and evolution of the NIH peer review process. I show that the basic peer review system landed almost accidentally at NIH near the end of World War II, and that many of the criticisms

[2] Still another concern is that the grants peer system has evolved to fuel hyper-competition in science, with a focus on grantsmanship over research quality and integrity (Stephan 2015; Alberts et al. 2014; Sarewitz, 2010).

today reflect concerns and tensions evident in science policy debates during and immediately after the war. Through review of evaluations of the NIH peer review process since the 1950s, I argue that a central tension, important for reform, is that the NIH peer review system has tried to accommodate sometimes competing goals: funding the best "basic" science and scientists, but also targeting specific health problems; promoting scientific freedom and autonomy, but also ensuring public accountability; and with different stakeholders valuing different goals. These tensions echo broad themes in the debates between scientist (and engineer) Vannevar Bush and Senator Harley Kilgore near the end of World War II about the very purpose of research funding and the best ways for the public to support research. I also show that one of the difficulties the current system faces—how to accurately and reliably evaluate the quality of a large number of applications for funding at scale—was foreseen by Bush and others at the end of the war. Other issues, including reducing administrative burden of the process on scientists and promoting non-conventional research, have also long been of concern to the NIH. Over decades, NIH has tried to overcome these difficulties through incremental reforms to the model developed in the 1940s, and continues to struggle with these issues today.

In Section 2, I describe how the World War II research effort provided the rationale and funds for expanding U.S. biomedical research funding, how the NIH became the major postwar funder, and how the wartime Committee on Medical Research's funding approach became the template for NIH's dual peer review system. This section also describes the key themes in the Bush-Kilgore debates during and after the war, which would resurface throughout the NIH's history, and early concerns about whether a peer review process could reliably evaluate scientific merit at scale. Section 3 provides an overview of the postwar growth of the agency, and the evolution of the peer review system. Section 4 discusses Bushian and Kilgorean perspectives on peer review performance, as evidenced in a series of external evaluations of the agency beginning in the 1950s. Section 5 describes internal evaluations and reform efforts, most of which are positive at a high level on its performance, but in their details raise questions about the ability of the peer review process to evaluate scientific merit at scale. Section 6 concludes by considering the implications of the historical perspectives for current reform efforts.

2 World War II and the Emergence of the Modern NIH

2.1 Prewar precedents

The roots of the prewar NIH extend to the Marine Hospital Service, which since the late 1700s had provided medical care for seamen. In 1887, the Service set up a laboratory in the Marine Hospital in Staten Island, NY, called the Hygienic Laboratory by its director, to better diagnose and manage infectious diseases. In 1891, the Service moved the Hygienic Laboratory to Washington DC (Harden n.d.). By the time the Marine Hospital Service was renamed the Public Health Service (PHS) in 1912, the Hygienic Laboratory had assumed an important role in epidemiological work on infectious diseases, water pollution, setting standards and helping with development of vaccines, and in general acting as a clearinghouse for biomedical research information used by clinicians and public health officials (Mandel 1996). The Hygienic Laboratory became the National Institute of Health in 1930 with the Ransdell Act, which authorized a research program including research beyond infectious diseases. Funding for these activities, however, was meager (Swain 1962). In 1937 the National Cancer Act created the National Cancer Institute (NCI), originally a PHS division parallel with NIH. The NCI was authorized to provide grants and fellowships to extramural research as well to set up the National Advisory Cancer Council to review applications and select awardees (Swain 1962). Under this authority, NCI funded a handful of extramural grants, fellowships, and traineeships in

the late 1930s and early 1940s (Munger 1960), though here again the numbers were small. While there was some movement in the direction of extramural research funding at NCI (which was not yet part of the NIH), before the war NIH had minimal experience with extramural grants, and no real peer review apparatus.

Other players in the biomedical innovation system were more experienced with evaluating and administering external grants. After industry, the biggest funders of biomedical research before World War II were private philanthropies (Ginzberg and Dutka 1989). The most significant was the Rockefeller Foundation. As William Schneider (2015) recounts, beginning in the early 20th century, Rockefeller had provided "block grants" to universities for general research funding, facilities, and the creation of new research institutes, as well as direct support of leading medical schools (Schneider 2015). But beginning with a 1928 re-organization, Rockefeller also began to fund fixed-term project grants. These small 3- to 5-year grants were attractive because they were lower commitment than large institutional grants and could be used to steer research to specific priorities the foundation wished to encourage. These grants were controversial. The head of Rockefeller's Medical Research Program, Alan Gregg, worried that they made the foundation "a huge dispensary of chicken feed" creating short-termism, timidity, and a focus on "square-in-program" research instead of deep exploration (Schneider 2015, 290). In addition to these concerns, Gregg was also frustrated about the difficulties of evaluating and administering large numbers of small grants, and administrative burden on the funders—a concern which would prove to be prescient in view of later difficulties that would confront the NIH.

Though these efforts anticipated some of the issues the postwar NIH program would encounter and may have familiarized biomedical researchers and administrators with grants, historians suggest a limited direct influence of the Rockefeller program or NCI peer review on the NIH (Schneider 2015; Mandel 1996). Instead, the main influence was the World War II Committee on Medical Research.

2.2 The Committee on Medical Research

In 1940, as the German armies advanced in Western Europe, U.S. involvement in World War II was imminent. Military leaders and policymakers knew that more so than in previous conflicts, this would be a technological war. The side that could best marshal the R&D needed by the military—for weapons development, communications and radar, jet propulsion, chemistry, optics, atomic fission, and medicine, among others—would have a major advantage. This was, on one hand, a source of optimism. In the period following World War I U.S. firms and universities had improved their scientific and technological capabilities in a range of fields. On the other hand, there was no serious federal policy for coordinating or supporting this research. Outside of agriculture, the U.S. government did not substantially fund extramural research before World War II, and as Vannevar Bush would later put it, had "no national policy for science" (Bush 1945).

Recognizing the gap, in June 1940 President Franklin D. Roosevelt authorized the creation of the National Defense Research Committee (NDRC). NDRC was led by Bush (former Dean of Engineering at MIT, and President of the Carnegie Institution of Washington) and included other leaders of the science establishment. NDRC was established to "coordinate, supervise, and conduct scientific research on the problems underlying the development, production, and use of mechanisms and devices of warfare" (Baxter 1968, 14). It funded research extramurally through contracts to leading firms and universities, which the government had not previously done on a large scale.

There was also a need for supporting medical research during the war. Major problems where research was needed included influenza, dysentery, malaria, wounds, venereal diseases, nutrition, physical hardships (such as sleep deprivation, coping with frigid temperatures, or oxygen deprivation), aviation medicine, and penicillin production at scale. While the NDRC did not initially include medical research, in June 1941 Roosevelt signed an Executive Order creating the Office of Scientific Research and Development (OSRD) composed of NDRC and a new Committee on Medical Research (CMR). CMR, chaired by University of Pennsylvania pharmacologist Alfred Newton Richards, was charged with mobilizing medical researchers and identifying "the need for and character of contracts to be entered into with universities, hospitals, and other agencies conducting medical research activities" (Executive Order 8807, 1940). The decision to create a new entity, CMR, to fund medical research reflected the limited research capacities of the NIH at the time, which had no real experience with extramural research funding.

Investigators—typically researchers at universities, hospitals, or foundations—would fill out simple "Proposals for Contract" forms. They were required to: (1) Describe the subject of the investigation with its background, present state of knowledge, significance in national defense, and plan of attack; (2) List its personnel, materials, and financial requirements; (3) State the investigative facilities available for the research; (4) Estimate its duration (Stewart 1948, 103).

CMR had to decide which of these proposals to fund. Beyond military need, there was a need to assess the scientific and technical feasibility of proposed projects. Lacking any real precedent, CMR's review process during the war—the predecessor of modern NIH peer review—relied on a committee structure created by the National Research Council's (NRC) Division on Medical Sciences (DMS) to liaise between the military and civilian experts on medical problems the military may face. There were 7 main NRC/DMS committees, and over 30 subcommittees, comprised of over 350 unpaid scientists and military personnel (Richards, 1946). Over the course of the war the committees met regularly "with no financial compensation beyond traveling and maintenance expenses" (Stewart 1948, 101). Through the committees, the CMR "gained the advice of several hundred men, who were specialists in their fields" (Stewart 1948, 101).

The NRC committees would evaluate scientific and technical feasibility, grading the applications as A, B, or C, or simply "disapprove." Then the CMR would evaluate them "from the point of view of their possible mediate or immediate effect in winning the war, their consistency with the program already in effect or projected, their personnel and budgets" (Stewart 1948, 103). CMR typically focused its attention on those proposals recommended by the NRC committee, making final decisions to approve or reject. CMR approval meant that an application was forwarded to Vannevar Bush for funding, and he followed the CMR recommendation in almost all cases.

During the war, the NRC committees reviewed 951 proposals, of which 638 were recommended to CMR, and 593 resulted in contracts. Mandel (1996) describes the NRC committees as "the first sustained large-scale exercise" in biomedical peer review (10). In relying on outside experts to assess scientific merit, followed by a second level of review for practical importance and then a recommendation for funding, the NRC/DMS committees and CMR served as a prototype for postwar peer review by the NIH (Mandel 1996).

In one crucial way the process differed from postwar peer review in medicine: there was little focus on basic research or on advancing breakthrough science for its own sake. Though Bush, Richards, and other OSRD and CMR leaders were strong advocates for basic research—among the strongest—all recognized that the time for basic research was before a crisis. During

the war there was also no interest in nurturing science or scientists over the long run. The focus was on solving problems, quickly.

2.3 The Bush-Kilgore Debates

The wartime research effort contributed to the development of a range of technologies that helped win the war, including, famously, radar and the atomic bomb (Baxter, 1946). Though less than one-tenth the size of NDRC in financial terms, CMR was also crucial to crisis resolution, helping support the mass production and clinical testing of penicillin, the development of antimalarial drugs, progress on numerous vaccines, and advancing work on blood substitutes, steroids, and other medical technologies that were crucial during the wartime effort. Chester Keefer, the "penicillin czar", later described it as "a novel experiment in American medicine, for planned and coordinated medical research had never been essayed on such a scale" (Keefer 1969, 62). So impressed were policymakers that even before the war was over, Roosevelt asked Bush to reflect on lessons from this experience for peacetime R&D policy. Bush's response, *Science, The Endless Frontier* (discussed in more detail below), is often considered the blueprint for postwar R&D policy.

However, there were disagreements on the lessons to be learned. Even during the war, New Deal liberals led by Senator Harley Kilgore (D-WV) raised concerns about aspects of OSRD. One issue was patent policy, which in many cases allowed funding recipients to retain title, which Kilgore viewed as a giveaway of publicly funded research to private firms that would exacerbate technological and economic inequality (Sampat 2020). A closely related concern was fairness. OSRD and CMR contracts were concentrated geographically and in certain institutions. While in Bush's view it was necessary to focus on the best researchers to resolve the crisis, Kilgore viewed this as self-serving to elite scientists like Bush. Throughout the war, he raised concerns that ideas from small independent inventors and institutions across the country were being ignored by the OSRD review process.

A series of bills introduced by Kilgore tried to reshape OSRD during the war but were unsuccessful. In 1944 he drafted legislation for peacetime funding of both basic and applied research through a new National Science Foundation. In contrast to what Kilgore regarded as elite and closed contracting procedures during the war, the foundation would be administered by a board with lay representatives and would sponsor research aimed (as OSRD was) at specific outcomes, "programmatically responsive, in a liberal fashion, to the political system" (Kevles 1977, 16).

In sharp contrast to Kilgore, academic scientists including OSRD leaders were critical of the government control of scientists during the war, and afterwards. CMR's director Richards viewed "regimentation" in science after the war as "abhorrent" and ultimately unproductive, on the view that the applied wartime successes were the "fruits" of previous basic research (Richards 1946, 578). Though the scientific community valued the prospect of increased government support after the war, it was cautious about government control of science, in medicine and beyond. While leaders of the scientific community understood the need for "bureaucratic interference with its professional autonomy" (Kevles 1977, 11) during the war, most scientists opposed this approach in peacetime. There was a feeling among many prominent scientists that Kilgore's approach would erode scientific autonomy—that it would socialize science. There were also some concerns from conservatives and industry that the Kilgore approach would represent an encroachment of the state into what were traditionally privately-run activities.

It was in this environment that Bush drafted *Science, The Endless Frontier.* The report was nominally a response from Bush to President Roosevelt to draw on lessons from OSRD for postwar support of science for national security, economic growth, and the war against disease. But in large part it was to seize the momentum from Kilgore, whose proposals for continuing government research funding in peacetime were gaining popularity. In the report, Bush made the case that basic research is necessary for economic, security, and health benefits, as the "pacemaker of technological progress." Anticipating economists' "market failure" rationale for public R&D funding, he argued that industry would not adequately support basic research, so government funding was required. He further argued that basic research was not plannable—many discoveries come from unexpected sources—thus making a pragmatic case for scientific autonomy. In the report's section on medical research, he asserted "Discoveries pertinent to medical progress have often come from remote and unexpected sources, and it is certain that this will be true in the future," emphasizing the need for scientific freedom (Bush 1945).

In a seminal treatment of the topic, Kevles (1977) summarized the situation as follows:

> The differences between Bush and Kilgore boiled down to a basic issue: Kilgore wanted a foundation responsive to lay control and prepared to support research for the advancement of the general welfare; Bush and his colleagues wanted an agency run by scientists mainly for the purpose of advancing science (16).

As the quote above hints, in *Endless Frontier* Bush, like Kilgore, advocated for a single research foundation. The foundation would have a broad remit, including medical research funding. That Bush did not envision NIH as the major funder may reflect its lack of serious experience in funding extramural biomedical research at the time. But he was also concerned that the NIH, as a mission-oriented agency (part of the Public Health Service), would be too distracted by applied endeavors to support the fundamental research that Bush favored:

> Operating agencies have immediate operating goals and are under constant pressure to produce in a tangible way, for that is the test of their value. None of these conditions is favorable to basic research. Research is the exploration of the unknown and is necessarily speculative … Basic scientific research should not, therefore, be placed under an operating agency whose paramount concern is anything other than research. Research will always suffer when put in competition with operations. The decision that there should be a new and independent agency was reached by each of the committees advising in these matters.

Bush wanted to start small, through a program of grants and fellowships for medical research, arguing for $5 million in the first year and that "[a]fter a program is under way perhaps 20 million dollars a year can be spent effectively." (That is about one-tenth of the NIH budget today, in real terms.)

Bush did not specify the process through which the funds would be allocated. He also appointed specific committees to reflect on the questions in the Roosevelt letter. The committee reports and recommendations are included as part of *Endless Frontier.* In Appendix C the Bowman Committee—tasked with the specific question what the government can do to aid research—provides more specifics than Bush did, proposing that to preserve scientific freedom "matching funds" be given to institutions to spend fluidly. This would preserve freedom but also reduce costs of administering the program. "[S]ince the grants are largely automatic in character, the board is freed from the burden of investigating intensively the large number of potential recipients and arriving at a decision in regard to the merits and defects of each," it read, adding

that “the experience of the private foundations demonstrates that judgements of this sort are extremely difficult and time consuming, even when pursued on a small scale” (69). Another report from a group explicitly focused on medical research, the Palmer Committee, recommended a brand new National Foundation for Medical Research run by eminent scientists, focused on largely unrestricted and long-term grants, and administered by universities, with limited political control or accountability. The Palmer Committee proposed even more autonomy than Bush's Foundation would have had. Bush rejected this proposal, recognizing that a medical research funding agency with so little accountability would be politically unrealistic (Fox 1987; Mandel 1996).

3 Peer Review at NIH: Origins and Overview

3.1 The Heir Unapparent

Bush, the influential architect of postwar innovation policy, and his nemesis Kilgore each proposed a single foundation for research funding that would include medical research. The Palmer Committee, representing the medical research establishment, favored a new autonomous foundation for medical research, independent of the new foundation Bush proposed, but also of the Public Health Service and the NIH. Donald Fredrickson (who would become NIH director in 1975) speculated that this choice may have reflected “a prevailing opinion that ‘public health research’ as represented by the NIH, was mainly sanitary engineering, vaccination, and vital statistics, and not serious ‘fundamental’ medical science” (Fredrickson 1993, 23).

But there were other currents as well. As described above, even before the war there were forays into extramural medical research funding by the government, mainly at NCI. Concurrent with the wartime effort and the Bush-Kilgore debates, the Public Health Service Act (Public Law 78-410, signed into law in March 1944) extended PHS's authority to make extramural grants under the National Cancer Act to all areas “to conduct and support research into the diseases and disabilities of man” (National Institutes of Health, 2022). It also made the NCI a division of the NIH and authorized the NIH to create other Institutes (Schneider, 2015; Strickland, 1988).

As Bush, Kilgore, and their allies debated competing legislation reflecting their respective views for the research foundation, a very practical question arose: what to do with active CMR contracts, most of which were with universities, as OSRD wound down. Bush wanted them at his proposed foundation, but that did not yet exist. Though few of the players in the debates about postwar science policy supported it, through deft political maneuvering by Surgeon General Thomas Parran and NIH Director Rolla Dyer (Strickland 1988; Fox 1987), including exaggerating NIH’s prewar experience with extramural funding, the NIH filled the vacuum under the new authorities created by PL 78-410. When it took them over in January 1946 the funds from 66 CMR contracts tripled the NIH’s budget and “positioned it to become the principal federal government vehicle for the performance and support of biomedical research for the foreseeable future and beyond” (Strickland 1988, 19).

That this was far from obvious for the time is evidenced by NIH’s appointment of the former head of the PHS’s Venereal Disease division, Cassius Van Slyke, to administer the program. Van Slyke was recovering from a heart attack, and in 1946 was given the job of administering the nascent NIH program—-as head of its new Research Grants Office—on guarantees it would not interfere with his recovery. It was described as to him as “an incidental, lower-left-hand-

drawer of the desk sort of activity" and that he "positively wouldn't have to work more than two hours a day and probably not more than four or five hours a week" (quoted in Strickland 1988, 22).

Things changed. Many of the inherited CMR grants were for the clinical study of penicillin, whose price dropped sharply in 1946, creating windfall funding for other research. In what has been called by one of its authors "the most naive letter ever to emanate from the national government in Washington," NIH leaders sent a note to deans of medical schools: "We have limited funds available for research purposes. If you have investigators who need these funds, let us hear by return mail" (Strickland 1988, 24). The response was overwhelming, and in the years that followed NIH requested and received funding from Congress to expand its funding for extramural research. Well before the funding agency Bush and Kilgore were deliberating was created (the National Science Foundation, established in 1950), medical research was already spoken for.

3.2 The Emergence of NIH Peer Review

The next question was how to handle the thousands of applications that came in response to the NIH's letter—how to make good "bets" on medical research (Strickland 1988). Van Slyke and colleagues experimented with various ad hoc approaches to gauging quality, including relying on bibliographies like Cattell's *American Men of Science* to identify scientists in adjacent areas and get their opinions on scientific merit (Strickland 1988, 24). The opinions were then taken to the National Advisory Health Council for final funding decisions, mimicking the dual review approach of CMR (Mandel 1996). By the end of 1946, this approach was superseded by formally asking scientists if they would be willing to serve on "study sections." The CMR Penicillin Panel was the first, renamed the Syphilis study section at NIH.

At the end of the first year, in a 1946 *Science* article Van Slyke advertised the program, now with 21 study sections and 250 leading scientists. He emphasized the need for government grants to encourage research of medical problems that scientists agree require research, a program "of scientists and by scientists" (Van Slyke 1946, 559). Echoing Bush and the Palmer Committee, he started by emphasizing "complete acceptance of a basic tenet of the philosophy upon which the scientific method rests: the integrity and independence of the research worker and his freedom from control, direction, regimentation, and outside interference" (559).

Like Bush, Van Slyke made not just an ideological but also a pragmatic case for academic freedom:

> During the war it frequently was necessary to sacrifice fundamental, not immediately applicable research in order to arrive at specific objectives promptly; promising bypaths often had to be by-passed. In the normal course of scientific investigation, however, the bypaths often lead to more important findings than do the roads from which they branch. Much of the most important research may not appear immediately to lend itself to clinical application, but it builds a large body of information, assembled parts from which may later have wide clinical applicability (559).

The article also described the mechanics of the arrangement. The final funding decisions would be made by the National Advisory Health Council of the NIH, the National Advisory Cancer Council for the NCI, and the National Mental Health Council for the newly created mental health Institute. Study sections would advise the three Councils. The study section recommendations

were to accept as submitted, reject, or defer decisions (requesting more information). Members were instructed to consider scientific merit, training, and facilitations for research. Then the applications were forwarded to the Council, where study section advice was usually accepted. The Surgeon General made the final, official funding decision (562).

Van Slyke emphasized other features distinguishing the program from the CMR. As a "peacetime program" there was no emphasis on "abnormal speed": 3- to 5-year projects were suggested as guidelines, with provisions for straightforward continuations (562-563). Bi-annual simple financial reporting was required to show the status of funds. In general, Van Slyke also wanted to acknowledge to investigators that the administrative burdens of the war—limited as they may seem from today's perspective—were not appropriate and would not be issues in peacetime.[3] There was a straightforward 4-page application form, to limit the burden on applicants.

3.3 Postwar Growth and Evolution

Van Slyke's missionary efforts were effective. Applications continued to pour in after the publication of his *Science* article. During the early post-World War II era there was also a dramatic increase in the size and scope of the NIH budget (see Figure 1). Two others who also helped shape the emergence of the modern NIH were disease advocates Mary Lasker and Florence Mahoney. Lasker was a New York socialite and wife of advertising magnate Albert Lasker, who brought techniques from advertising to sell the value of medical research to Congress (Drew 1967). (The oldest building on the NIH Campus, Building 60, was renamed the Lasker Building in 1984.[4]) Her chief ally was Mahoney, also wealthy and well-connected with journalists (with family ties to the powerful Cox national newspaper chain) and influential politicians (Robinson 2001). Both turned to support of medical research only after President Truman's national health insurance initiative had failed, opposed by the American Medical Association among other powerful interest groups (Robinson 2001). Medical research was a way to improve health with fewer political opponents. For the advocates, Lasker especially, the reason for NIH funding was always about health, rather than science for its own sake. Their applied focus sometimes put them in tension with NIH leadership and the scientific community, echoing themes in the Bush-Kilgore debates.

These "Noble Conspirators" befriended and lobbied influential members of Congress to expand funding (Drew 1967). They focused on powerful members of appropriations committees—the so-called "Cardinals" of Capitol Hill—especially chairs of the House and Senate Labor/Heath subcommittees responsible for the NIH budget. Their initial lobbying was to expand cancer research (Strickland 1972). In 1946, they helped push a bill to create a new National Institute of Mental Health (NIMH). In 1948 they also drafted and steered through Congress passage of legislation to create a new institute to secure long-term funding on heart disease, the National Heart Institute (now NHLBI, the National Heart, Lung, and Blood Institute; Strickland 1972).

---

[3] On the topic of financial reporting, Van Slyke emphasized "In order not to divert the time of the researcher unnecessarily from the actual conduct of research …[i]t is not desired that the preparation of these reports present any long, tedious burden" (563) and they should be brief and concise.

[4] Building 60 also now houses the Office of NIH History, whose staff and archivists have been indispensable in helping me locate and collect materials for this paper. I am especially thankful to Gabrielle Barr and Michele Lyons for their help and guidance.

Through the same act, in 1948 the agency was renamed the National Institutes of Health (McGeary and Smith 2002, Appendix A).

Other disease institutes followed (McGeary and Smith 2002; Varmus 2001), including the National Institute of Dental Research (now NIDCR) in 1948, the National Institute of Allergy and Infectious Diseases (NIAID) also in 1948 (originally the National Microbiological Institute), and in 1950 the National Institute of Arthritis and Metabolic Diseases (now NIDDK) and the National Institute of Neurological Diseases and Blindness (now NINDS).[5] As Table 1 shows, there are 27 Institutes and Centers (ICs) today, each with a different disease, organ, or professional focus, each with its own Director, Advisory Council, and Congressional appropriation.

About a ten-minute walk from the Lasker Building, Building One on NIH's Bethesda Campus is the James A. Shannon Building. Shannon, a physician and physiologist, had been a central member of the Committee on Medical Research's malaria effort. He joined NIH in 1949 as associate director of the new National Heart Institute and became NIH Director in 1955. The "Shannon Years" are often identified as the NIH's Golden Age (McGeary and Smith 2002). He worked closely with appropriations chairs Congressman John Fogarty (D-RI) on the House side, and Senator Lister Hill (D-AL), to increase the NIH budget.[6] Supporting the Congressional increases were "citizen witnesses" (typically representatives from disease groups or clinicians) supplied and coached by Lasker, Mahoney, and their fellow Noble Conspirators who made the case for more funding. Lasker and Mahoney worked to persuade powerful politicians to fund the NIH, through "reminding the lawmakers of their mortality," of the burden and cost of illnesses, the political popularity of medical research, and also by making targeted campaign contributions (Drew 1967, 79). The scientific community and universities, which had been cautious about federal funding until the end of World War II, emerged during the 1950s as another powerful interest group invested in increasing the budget.[7] Cook-Deegan and McGeary (2006) observe "[w]hile the two major health-related constituencies—scientists and disease research advocates—often differed over how and to what degree research should be planned and 'targeted,' they shared the common goal of increasing research funding" (179).[8]

Figure 1 shows the growth of the NIH budget over time. When Shannon became Director in 1955, the NIH budget was nearly $800 million (in 2020 dollars), 16 times what it was in 1945 in real terms. Under Shannon's leadership, between 1955 and 1968 the budget had increased another tenfold, and the agency grew into its modern form. During this expansion, there was

[5] NIH leaders and the scientific community were skeptical of new institute formation, instead preferring funding the best science irrespective of disease, a la Bush. But there was also recognition that creating entities with which disease groups could identify helped generate larger budgets, and thus that creation of new institutes to appeal to Kilgorean outcomes-oriented constituencies could be a necessary evil. These debates around new institute formation would continue along the same lines throughout the history of the agency (McGeary and Smith 2002; Varmus 2001).

[6] After they retired, the Lister Hill Center for Biomedical Communication, part of the National Library of Medicine, and the Fogarty International Center were established at the NIH, not far from the Shannon and Lasker buildings.

[7] McGeary and Smith (2002) observe that representatives of the four corners of this "quadrilateral" support for NIH—by the scientific community, disease groups, Congress, and the Executive Branch (including NIH)—"have a mutual interest in improving health through research … but they do not always agree on how, or how much relative to other national needs" (7).

[8] The authors also suggest that Bush did not foresee the potentially important role of advocates in boosting funding. This argument is similar to one made by Gruber and Johnson (2019) about the political economy of science funding more generally, that Kilgorean considerations (they highlight broad geographic funding) can help increase support for funding research.

some concern among some members of Congress and in the Executive Branch (particularly in the fiscally conservative Eisenhower administration) about "force-feeding" the NIH. Nevertheless, overall the focus was on growth (Greenberg 1967). Though there have been moments of belt-tightening and austerity in the decades since, overall the expansion has continued. Once wary of NIH funding, academic medical centers and biomedical researchers are now dependent on it (Alberts et al. 2014; Stephan 2015).

At the same time, the basic approach to reviewing and awarding grants has remained remarkably similar to the one adapted from OSRD. The first round of NIH's dual peer review is conducted by study sections comprised of volunteer external scientists from across the country, like the NRC/DMS committees that advised CMR. Peer reviewers provide priority score rankings to applications—a practice that goes back to one of the early policy decisions of NIH in the late 1940s. Today, applications are provided an overall impact score (averaged across eligible study section members' scores) and a set of specific scores by criteria (significance, investigator, innovation, approach, and environment). Raw overall scores are normalized and converted to percentiles.[9] With these scores in hand, together with summary statements with reviewer comments, Institute/Center Advisory Councils (or Boards) perform a second level of review, making recommendations to IC Directors who make the final funding decisions ("Peer Review at NIH" 2022). In theory the second layer of peer review by the ICs is meant to align funding with program priorities, including health and other mission-related considerations. In practice they have tended to follow the percentile rankings from study sections nearly in lockstep (McGeary and Smith 2002). [10] As we will see, the passive nature of second-stage review has long been part of a Kilgorean criticism: that the agency prioritizes scientific merit, as judged by external scientists, over actually targeting health or other applied outcomes. [11]

Since World War II, the NIH budget has grown 1,000-fold in real terms (Figure 1), and the number of categorical disease institutes has proliferated (Table 1). While the basic architecture of the peer review system has remained intact since the early postwar years, there has been considerable incremental evolution. The evolution of the process, described in more detail below, has reflected many external and internal evaluations of it, with influence from each stakeholder group that shaped overall growth in the scale and scope of the NIH: disease lobbies, scientists and universities, the Executive Branch, and Congress. The following two sections describe these evaluations, the issues they considered, and their effects on the NIH peer review system and process.

4 Evaluating Peer Review: The Major External Assessments

Since the initial wartime expansion, there have been numerous major external assessments of the NIH. The reviews have covered a range of issues, including the size of the NIH budget, new Institute and Center formation, the intramural program, indirect cost recovery, and others. In this

[9] While percentiling procedures have evolved (see below), they are currently based on the study sections' scoring previous two rounds.
[10] There is some heterogeneity across ICs in the extent to which they deviate from strict adherence to priority scores (Murrin 2021).
[11] While the vast majority of NIH funding goes through centralized peer review (through study sections organized by the Center for Scientific Review), about 20 percent of funding is reviewed by study sections run by the Institutes and Centers (IC review). IC review focuses on specific grant mechanisms, and in addition is used for targeted requests for applications (RFAs). See https://nexus.od.nih.gov/all/2012/03/30/does-it-matter-where-your-grant-application-is-reviewed.

section, I describe the main issues related to the grants peer review process raised in these assessments, showing the ongoing importance of issues in the Bush-Kilgore debates—scientific autonomy versus targeting, basic versus applied research, accountability versus bureaucratic interference—throughout the agency's history. I draw on the list of external reports helpfully compiled by Michael McGeary and Philip Smith (McGeary and Smith 2002). As will be evident, I have benefitted from their analyses of politics surrounding these reports as well.[12] I also consider several other external evaluations, not covered by McGeary and Smith, that discuss peer review or related topics.

4.1 The Early Reports: Scientific Freedom and Quality

The first major external assessment was conducted in the context of concerns about the NIH's exponential budget growth from the Eisenhower Administration. Among those worried was the first Secretary of the Department of Health, Education and Welfare (HEW), Olveta Culp Hobby. In 1955, Hobby asked the Director of the young National Science Foundation (NSF) to evaluate the nation's medical research effort, including the NIH. The committee assembled by NSF comprised eight academic scientists, chaired by Cyril Long, the Dean of Yale's Medical School (McGeary and Smith, 2002).

The Long Report (Long, 1955) voiced concerns that the NIH's public health orientation might interfere with universities' autonomy. Specifically, it expressed concerns about the categorical or disease focus of the NIH: that "the choice of educational and research opportunities may no longer rest in the hands of the educational institutions but … determined by funds available for those areas which others have decided to be of paramount importance" (46). Research fields at universities would be based on "popularity"—presumably the reference was to Congress and disease advocates—not scientific value. In general, the Committee recommended long project terms (up to 10 years), and what would today be called a people-not-projects approach ("the award of the grants should be based on the demonstrated or potential capacity of the investigator, rather than on a field of research"). It argued against Kilgorean ideas about targeting, and like Bush and Van Slyke dismissed "the widely held belief, no doubt fostered by certain wartime successes" targeting research funding to specific objectives.[13]

Like the Palmer Committee's Appendix to *Endless Frontier*, the Long Report expressed skepticism about the ability of a public health agency to secure broad scientific freedom and about the benefits of unplanned research. It recommended severing the extramural program from NIH and creating a new Office of Medical Research and Training to absorb its functions. Unsurprisingly, the recommendation to sever the extramural program was not met warmly by the NIH. Since Secretary Hobby retired from HEW in the summer of 1955 the report was essentially ignored (McGeary and Smith 2002).

The alarms Long sounded regarding scientific freedom would resurface in the later external reports, including in another major external assessment of the NIH, led by a former Yale Medical School Dean, Stanhope Bayne-Jones (Bayne-Jones 1958). The Bayne-Jones Report

[12] Though this list was prepared for a National Academy of Sciences evaluation of the organizational structure of the NIH (National Research Council 2003), and focuses on reports concerned with structure, the list is similar to other lists of external reports on peer review (e.g. National Institutes of Health GPRST 1976).

[13] This was mostly in the context of the intramural program. In that discussion the report claimed an "essential difference" between weapons development and disease eradication, where the former was viewed more amenable to "plan and design" and the latter characterized by "accidental discovery" (30).

was prepared amid continued budget growth throughout the 1950s—driven by the efforts of Director Shannon and disease advocates—and amid ongoing concerns from the Eisenhower Administration that NIH was “pushing biomedical research too quickly” (McGeary and Smith 2002, 130). The new HEW Secretary Marion Folsom, who joined in 1955 following Hobby's retirement, commissioned the study to assess a range of issues regarding the NIH, including budget size, staffing, the intramural program, indirect costs, Congressional earmarking, training grants, and the role the NIH should have in drug development.

The Bayne-Jones Report disagreed with the conclusion in the Long Report that the NIH overemphasized applied problems. It provided a strong endorsement of the NIH program as fulfilling both Bush’s and Kilgore's goals: “The twin dangers of bureaucratic interference with science, leading to loss of freedom by scientists and universities, and of bureaucratic lassitude, leading to failure to adapt programs to changing needs, have been avoided” (67). However, it echoed some of the concerns about freedom, recommending that the NIH grant system needed to provide “assured stability of support” and “freedom to shift resources and the substance” of projects (72). Like Bush, it also recommended institutional grants “provided under conditions which give the institutions a substantial degree of freedom in deciding how to use the funds” (73).[14]

A concurrent report was commissioned by the Senate Appropriations Committee’s Labor-HEW Subcommittee several years later, suggested by Republicans who continued to be uneasy about continuing large budget increases for the agency. A specific concern (echoing the Bowman Committee’s remarks in *Endless Frontier*) was reflected in President Eisenhower's statement in 1959: that the budget was increasing faster than the procedure of “careful appraisal” of applications was adapting, thus lowering the quality of funded research. The Committee of Consultants on Medical Research, led by Emory's Vice President for Health Services Boisfeuillet “Bo” Jones, issued a report in 1960 (Jones 1960). The Jones Report went deeper into the administrative machinery of the grants process than the previous reports. The headline conclusion of the report was that the NIH peer review approach was “extremely successful” in allocating funds for medical research and “has assured consistently high standards for the research supported, gained the confidence of the scientific community and maintained the traditional freedom of both institutions and investigators” (xiv). Based on an assessment of the quality of study sections, the standards, and application success rates, it argued in its “subjective, but very strongly held opinion” that the quality of review had, if anything, increased over time (29). It did suggest the need for additional study sections and to divide study sections, to keep up with expected growth in scale, and new administrative staff to manage the program.

Following its overall assessment of high quality, and because there were many new opportunities for scientific advances and the value of medical research was so high, the Committee viewed current funding as inadequate, and recommended rapid expansion as soon as possible. Like the earlier reports, the Jones Report emphasized that this funding “must be

[14] The Bayne-Jones Report also raised some Kilgorean concerns, including around how the focus on scientific quality may lead to geographic inequities, and recommended that NIH take into consideration the value of "a geographically dispersed system of first-rate, non-Federal research institutions" (70). Further, in a section on "Research on Problems of Social Importance" it emphasized that as a public institution, the NIH should "support research not only because of its inherent scientific worth but also because of its value to the solution of urgent social problems" (74). Though (beyond generally supporting the categorical disease structure) it did not specify how this balancing act should occur.

provided to an increasing degree under terms and conditions which will permit scientists still broader latitude to pursue their own ideas and assure them greater stability of support" (8).

4.2 The Fountain Report and Accountability

In their discussions of peer review the early reports were mainly focused on intellectual freedom, in line with arguments Bush had made. There was little focus on the Kilgorean theme of accountability for how funds are spent. This was to change in the early 1960s with a series of evaluations spearheaded by Congressman Lawrence Fountain (D-NC) from the Intergovernmental Relations Subcommittee of the House Committee on Government Operations.

A first Fountain Report (Intergovernmental Relations Subcommittee 1961) began by noting the enormous growth in the scale of the NIH budget, and the role of Congress in expanding it beyond the agency's request (in 8 of 10 years between 1950 and 1960), to a budget of $3.5 billion (2020 dollars) by the end of the decade (Figure 1). The question was whether these funds were spent effectively. The report raised concerns about quality, taking a deeper dive into the administration of the grants than the earlier evaluations. It observed that the share of applications in the top priority score cohort had been decreasing, which suggested diminished quality. It also observed that the second layer of review by Advisory Councils, meant to add "program objectives" and "policy considerations"—above and beyond scientific merit determinations by study sections—was increasingly a rubber stamp of study section assessments, as the volume of applications grew and detailed second stage review became impracticable. The implication was that the major lever in the system to achieve Kilgorean targeting was broken.

Fountain's primary focus, however, was on accountability of the funds. Throughout the growth period of the 1950s there had been at NIH a philosophy of "brains management" over "book management"— "helping the investigator get on with the job ... with a minimum of trauma and disregard, if necessary, of administrative protocol" (Mandel 1996, 91). This was by design, to promote freedom. However, as scale grew, it became difficult not only to ensure scientific quality, but also monitor how the funds were spent. The thoroughness of early peer review described in Van Slyke's *Science* article was no longer possible. It was not feasible to review budgets for reasonableness, and the NIH had no machinery to do so. In sharp contrast to previous reports, and to Van Slyke's explicit attempts to limit burden, Fountain complained that "reporting requirements are at a minimum" (27). Fountain was particularly concerned about the inadequacy of post-award management: limited review (including of budgets) after the initial grant. His report argued that the NIH policy "permitting the investigator unlimited freedom to change an approved research project once the grant has been awarded" while appealing to scientists "is not fully protective of public funds" (34). It suggested that changes in the project should be evaluated for importance, significance, and approach just as the original application had been. While it allowed that changes may be needed, the report argued for prompt reporting of these changes and more intensive scrutiny, and raised questions about the ability of the NIH to provide such oversight.[15]

[15] Fountain also questioned whether, under the prevailing approach, the NIH was really funding projects or surreptitiously funding people: "Although a sharp distinction is not possible in many instances, it might be desirable to have a clearer identification of the intent of Federal support in the individual case—whether it is intended principally to support the research efforts of the man or the specific project" (35). The report also asked questions about the appropriateness of charging salaries to NIH grants

The NIH paid lip service to the Fountain Report but was generally dismissive (McGeary and Smith 2002). Shannon dismissed concerns around post-award monitoring as “essentially trivial” and emphasized the freedom was built in, necessary for the agency's success—a feature, not a bug (126). He observed the concerns may have been appropriate for a contract system—like that in World War II—but not for a system of scientific grants. Like Bush and Van Slyke before him, Shannon made a case for freedom as a necessary condition for progress, testifying that “Freedom is defended not on abstract grounds, or as an inherent right of scientists, but as the prime condition for assuring maximum yield on the taxpayer's investment” (“Health Research and Training Hearings” 1961, 15). While in 1962 the NIH officially replied that it was “in accord” with the recommendations, and NIH administrators “definitely intend eventually to make all desirable changes,” the Fountain Committee felt it was being ignored (Committee on Government Operations 1962, 10). According to one NIH official interviewed by Daniel Greenberg they were correct: “there was a feeling that time would pass and the whole thing would be forgotten” (Greenberg 1963, 1076).

A second Fountain Report in 1962 noted the agency had “done relatively little” to implement the changes, and affirmed that publicly funded grants, like contracts, demanded public accountability (Committee on Government Operations 1962). It concluded that Congress was “overzealous” in funding medical research and that “the pressure for spending increasingly large appropriations has kept NIH from giving adequate attention to basic management problems” (26).

This had serious repercussions. Rather than large increases to the agency’s request, Congress reduced the budget in FY1964, for the first time in years (McGeary and Smith 2002; Strickland 1988). NIH responded with “a rapid and complex evolution of administrative mechanisms” in its extramural peer review program (Mandel 1996, 97). The NIH changes included creating new regulations, codified in the 1963 NIH grants manual, as “a massive synthesis of procedures currently used by 22 PHS Institutes and program divisions, to which new regulations responsive to the Fountain Committee requirements were continually added” (Mandel 1996, 103). One NIH administrator observed: “all of a sudden a new document was born called ‘the regulations’ where a whole series of ‘thou shalts’ and ‘thou shalt nots’ were written down for the first time as regulations which had the thrust of the law” (Goldstein 1986, 8). The new provisions included, among others, requirements that grantees not be allowed to alter objectives without permission, “effort reporting” detailing the share of time researchers spend on their NIH grants versus other activities, and restrictions on shifting funds across budgetary categories.

Since NIH could not itself monitor compliance details for thousands of grants disbursed to hundreds of universities, the burden was shifted to universities and PIs. Many in the scientific community objected to the new administrative tasks. A 1963 *Science* editorial “More Paperwork, Less Research” remarked on new processes for continuation grants that “Grantees report they must spend from 1 to 7 days in obtaining information and filling out the forms” and the collective “work on this form will cost the nation millions of dollars in lost time from research” P.H.A. 1963, 725). The editorial also warned of the bureaucratization of the NIH, replacing scientists running the program with “administrative types” who “can only run scared, go by the book, and introduce

(specifically, concerns that universities were paying more than regular salary for those on NIH grants), prompting several subsequent inquiries of this issue.

all kinds of excuses for delay" (725).[16] There was another flurry of responses in *Science* several months later (Edsall et al., 1963), including around whether the provisions requiring notifications in cases of changes to research were appropriate for grants (as opposed to contracts), and claims they would limit exploration of "unexpected leads" that were core to how research progresses. The push for accountability and responsible budgeting of public funds—similar to some of Kilgore's concerns about how wartime R&D funds were being spent—led to the regulations and administrative burdens that Bush, Van Slyke, and the medical establishment had feared, and accompanying claims that this would stifle medical progress.[17]

4.3 Wooldridge and the Quality of Peer Review

In the aftermath of the Fountain imbroglio, President Kennedy commissioned yet another review, which President Johnson set up in 1964, chaired by aerospace engineer Dean E. Wooldridge. The Wooldridge Committee interpreted its charge broadly: "to judge whether the American people are getting their money's worth" from the billion dollars spent on NIH and "to recommend any changes in organization or procedure" to improve effectiveness (Wooldridge 1965, xv). The Committee interviewed scientists and institutional administrators from a range of fields and universities, conducted site visits, and reviewed case studies of successful and unsuccessful grant applications.

The 1965 Wooldridge Report (Wooldridge 1965) provided a strong endorsement of the NIH and continued expansion of funding overall. More so than previous reports, it acknowledged the growing difficulties of evaluating the quality of applications at scale, and what has been called the "workload crisis" at the NIH (Mandel 1996):

> A major difficulty of the present arrangement relates to the time that the Study Section members must devote to their advisory work. The cost today is considerable in terms of the manhours of scarce scientific effort required. Unfortunately, most pressures are in the direction of increased demands on the Study Sections (Woolridge 1965, 21).

It noted the desire among extramural investigators for "more site visits … ways of improving research plans … more explanation of the reasons why proposals are not approved"—in other words, a return to the hands-on review activities of the 1940s and early 1950s (21).

Wooldridge implicitly acknowledged the managerial and accountability issues raised by Fountain. However, the report noted, monitoring the progress of the thousands of grants the agency awarded across the country would be infeasible. Accordingly, it pushed NIH to shift responsibilities for grant administration to the grantee institutions, including by strengthening the grant administrative capacity of weak institutions. If such responsibilities were accompanied with funding (through institutional grants), "a decrease in the red tape harassments that currently

[16] Another editorial complained that the grant system would increasingly be run by "former scientists who for a variety of reasons become involved in the regulation of science rather than contributing to it creatively" (P.H.A., 1963).

[17] Beyond accountability for funds, other regulations were added over the years, including in the mid-1960s new requirements for institutions to comply with human subjects protection and the Civil Rights Act. By the mid 1980s, Mandel (1996) writes that the traditional role of peer review in judging scientific merit was being undermined by "activist congressional efforts" to use it "as a managerial tool for cost control, fraud monitoring, risk assessment, and other ancillary oversight roles" infringing on the "fundamental guarantee" of scientific freedom in the Bush Report.

annoy many scientific investigators, along with an increase in the quality of institutional accountability of government funds, should simultaneously be available" (1).

The Wooldridge Report made the case for NIH to focus on "basic science" while recognizing the "practical" benefits of categorical and disease orientation.[18] It also raised concerns about the cost, efficacy, and quality of review of new, large, coordinated and often applied programs that had arisen at NIH, relative to funding untargeted research. It was particularly critical of the Cancer Chemotherapy Program, which was created in 1955 to test agents against cancer.[19] Overall, however, the Wooldridge Report provided a strong endorsement of the NIH approach, including peer review, and helped buoy funding (McGeary and Smith 2002).

4.4 Peer Review, Targeting, and Priority Setting

Beginning in the mid-1950s NIH increasingly relied on contracts to achieve Kilgorean goals of planning, including "filling gaps" in the portfolio and "prodding efforts in particular areas" where, presumably, grant peer review alone could not (Strickland 1988, 68). Contracts were common instruments for supporting the large coordinated programs discussed by Wooldridge. Following the Wooldridge Report's criticism of the Cancer Chemotherapy Program, disease advocates pushed for review of cancer contracts by the National Advisory Cancer Council. The NIH strongly objected, and the issue was referred by the HEW Secretary to a new committee, chaired by Jack Ruina, a consultant who had experience with defense contracting, for evaluation. Beyond the specific instigating issue—the authority of the Council to review contracts—the Ruina Report (Ruina 1966) also discussed more general questions of basic versus directed research, and difficulties in supporting the latter through traditional peer review mechanisms:

> Biomedical research is now yielding, with increasing frequency, results which promise major benefits in health and longevity, but which call for large-scale *directed research* or for *development* before they can be put into use. The National Institutes of Health, which in the past two decades has selected and supported, in major part, the research leading to these promising activities, now faces the problem of exploiting them … New methods of organization and procedure may be needed (2).

[18] At least one critic asked, "Is it proper to assume the role of disease chaser in the popular prints while playing the more correct role of fundamental researcher in scientific circles?" (Cooper 1965, 1435).
[19] According to Wooldridge's assessments, this program had not justified the $200 million that had been spent on it to that date. Wooldridge raised issues regarding mismanagement of this and other coordinated programs, and difficulties in "managing" coordinated programs by the NIH more generally, and also regarding lack of thorough scientific review of such programs. On the Chemotherapy Program, the Report also observed "there is no evidence that the initiation of the large scale program was preceded by any very extensive attempts, involving either pilot tests or analyses, to compare the value of the probable results with the value of the results that a similar expenditure might have been expected to produce in other biomedical research: by just lifting the overall level of the traditional research grant program" (40). (Though it did not make this point, there was no evidence of the opposite either, i.e. that traditional grants had higher returns than the coordinated programs.) Another large coordinated program flagged in the Wooldridge Report was the Framingham Heart Study, an epidemiological study whose results would have an important role in advances against cardiovascular disease (Cutler 2004). See Patel (2012) regarding controversy regarding the Framingham Heart study at NIH during the 1960s.

The Ruina Report highlighted the difficulties of steering research through the peer-reviewed grant system, recommending broader use of contracts by the NIH to direct research. Mary Lasker and other disease advocates took the broader issue, of the NIH's limited machinery to support applied research and development, to President Johnson. They continued to push it when advocating for the National Cancer Act in 1971 (Rettig 1977). A big theme in the Congressional and public debate around the Act was whether the NIH peer review process was able to target a specific disease and make progress rapidly, or whether it "boxed in" program officials who wanted to achieve specific goals (Strickland 1972).[20] Advocates argued (unsuccessfully) for establishing a new organization separate from NIH that would move more quickly, and for more direct accountability of researchers to politicians and disease groups (Rettig 1977; Strickland 1972).

The debate around the National Cancer Act re-surfaced Kilgorean questions about the ability of the NIH review process to support applied research. In its wake, there were pushes to review the entire NIH (Rettig 1977). The President's Biomedical Research Panel (BRP) was convened to do so. The Panel Report (Murphy 1976) noted "increasing and frequently changing public demand for allocation of resources according to public perceptions of important health goals, rather than on the basis of scientific opportunities. In response to these factors, there is public pressure for the Congress to play a strong role in setting program direction, emphasis, and budget levels" (5). The Panel also raised questions about whether the NIH was the proper locus for these applied activities. Unlike the Long Report, when the concern was that applied public health activities would crowd out fundamental research, for BRP the concern was the precise opposite. The BRP report highlighted "differences of opinion regarding the proper role of the NIH" as the central tension:

> Many in the science community prefer that the NIH revert to a "pure" research institution. Others within this same community and elsewhere feel that this new responsibility is appropriate and that the mission of the NIH encompasses knowledge applications in the interest of imperative healthcare and public well-being (7).

The BRP also discussed a range of other issues related to peer review—among others, the impact of the recently passed Sunshine Laws (the Freedom of Information Act of 1967, the Federal Advisory Committee Act of 1972, and the Privacy Act of 1974), topics that would also be the focus of internal NIH reports, discussed below. Though it raised concern about peer review and targeting, like many previous reports its review of peer review of science was generally positive.[21]

---

[20] In his testimony to the Fountain Committee fifteen years prior, Director Shannon described the fundamentals of the investigator-initiated grant peer review system, observing "While extensive, planned, coordinated research centrally initiated … is part of the system, the keystone is response to needs and opportunities as expressed by the scientific community. This is not support at random, but support of research in response to the internal logic of science." (Fountain 1962, 14). While he did not say so explicitly, such a system would make targeting to specific applied goals—those of Congress, disease advocates, or NIH officials—difficult.

[21] A concurrent report commissioned by Paul Rogers (D-FL), chair of the House Subcommittee on Health and the Environment, was also supportive of the NIH and included a chapter on peer review. Despite its generally positive assessment, the Rogers Report summarized contemporary criticisms of the process, including conflicts of interest and "in-group" behavior, and the lack of broad geographic representation and of young investigators on study sections. Another criticism it noted is that the process was "reactive to the initiative, interests, and whims of individual researchers," not suitable for targeting research, and had a "strong bias for basic scientific inquiry and a disdain for applied research" (43). It also recounted concerns about the "*managerial choice* to strictly follow descending project priority in funding" (italics in

Questions about peer review and targeting also arose in several subsequent evaluations. In 1984 the Institute of Medicine (IOM) report "Responding to Health Needs and Scientific Opportunity: The Organizational Structure of the National Institutes of Health" (Institute of Medicine 1984) described the push for new Institutes as part of the push for "practical applications." In arguing against the proliferation of new Institutes as a way to support practical applications (curing specific diseases), the report acknowledged the lack of existing mechanisms (within peer review) for achieving programmatic goals, and the limited role of councils in overturning study section recommendations focused on science. The report suggested the need for better responses to health needs than through the blunt tool of new Institute formation, including solicitation of applications for grants and potentially "high relevance scores" for specific areas (19).[22]

A subsequent IOM report, in 1998, studied the question of how NIH set priorities across disease areas, and the related concern that NIH focused more on scientific opportunities than public health considerations in its priority setting and allocation decisions (Institute of Medicine 1998). The context was the belief by some members of Congress, in the scientific community, and among some disease advocates that AIDS, breast cancer and other high-profile diseases got more funding than they should given their societal burden, because of the strength of their disease lobbies (Sampat 2012).

Related to this, other disease organizations had begun to lobby Congress aggressively for "soft" earmarks in the mid-1990s. NIH is atypical among research funding agencies in that Congressional appropriators do not typically include "hard" earmarks requiring specific amounts of funding for particular projects or diseases. According to former Director Harold Varmus (2009), this reflects "Congressional confidence in NIH's system of peer review" (150). However, "soft" earmarks—in the form of report language attached to appropriations bills, often aiming to get NIH to issue requests for applications (RFAs) for particular diseases—grew in the 1990s. Soft earmarks, though non-binding, were also criticized as interfering with the process, according to one legislator "turning the floor of Congress into a scientific peer review panel" (IOM 1998, 26). But their defenders argued that they were the only real way to introduce health-related considerations and objectives (e.g. we need to spend more on disease X, we want to support technology Y or build new scientific field Z) into a peer review process otherwise narrowly focused on scientific opportunity and merit. Like the 1984 report, the 1998 IOM report noted the potentially important role of RFAs—often from Congressional directives—in priority setting and directing research.

---

original) which, according to critics, "tends to assure that applied work or work of greater program interest to the institutes does not get funded" (43). It raised a range of issues with advisory councils, including the need for a more active role in priority setting, and that certain disciplines that could help the councils judge relevance and social impact ("such as law or sociology") were not represented (47). There were also allegations, hearkening back to Fountain, about lack of enforcement of requirements (e.g., for delinquent progress reports). See Rogers (1976).

[22] There are related questions about the ability of the NIH peer review process to respond quickly to immediate public health crises. An Institute of Medicine Report, *The AIDS Research Program of the National Institutes of Health* (1991), suggests the early NIH response to HIV/AIDS was hindered by "standard operating procedures … oriented toward support of long-range basic research rather than swift, coordinated action" (21). Eventually, the NIH responded more quickly through instruments outside of the standard investigator-initiated peer review toolkit, including through its intramural program, contracts, use of RFAs, and funding applications with low-peer review scores "out of order." I thank Mike McGeary for bringing this to my attention. The Covid-19 pandemic also raised questions about the ability of the NIH to pivot quickly to respond to a crisis.

Another National Research Council/IOM report was also motivated by the growth in NIH Institutes and Centers and was nominally about NIH's organizational structure (National Research Council 2003). It was written on the tail end of the doubling of the NIH's budget over the 1997-2002 period. Like previous assessments, the report was positive towards the agency, and of peer review as "the best guarantee we have overall that scientists will carry out research that is of high quality and high potential for scientific progress" (ix). It acknowledged, as previous reports had, the limited ability of the NIH peer review process to target specific diseases or health problems, a Kilgorean concern. It also raised concerns that a peer review process focused on accountability and quality may discriminate against high-risk, high-reward research requiring "immediate reaction," and suggesting adopting elements of the very different Defense Advanced Research Projects Agency (DARPA) approach in the NIH.[23]

The major external evaluations surveyed above reveal the persistence of many of the Bush-Kilgore tensions throughout the history of the NIH: basic versus applied, science versus application, serendipity versus steering, freedom versus accountability. Another theme discussed in many of the external evaluations was the question of whether the process is an old boys' club, biased toward specific institutions or investigators.[24] This concern relates to questions raised by Kilgore about the distribution of funding, but Bush too appears to have appreciated potential inequities that may result from funding the best science.[25]

5 Internal Assessments of the Peer Review System

5.1 GPRST, Privacy, and Quality

In addition to the major external reports, there have also been scores of internal studies of NIH peer review. The first comprehensive "self-examination" was by the 1975 Grants Peer Review Study Team (National Institutes of Health 1976) led by Ruth Kirschstein, who would become acting NIH Director in 1993 and again from 2000-2002. The impetus for GPRST was another set of imperatives from outside the peer review system. Congress had recently enacted the "Sunshine" Laws providing public access to records and meetings by executive agencies, including the Freedom of Information Act (FOIA), the Privacy Act, the Federal Advisory Committee Act (FACA), and the Government in Sunshine Act, aimed at helping the public understand and inform the activities of executive agencies.

FOIA raised questions about whether information about funded grants and grant applications would be available to the public, and concerns at the NIH about whether this might undermine the intellectual property of investigators and discourage good applications. Some interpretations of FACA would have made study section meetings open, and beyond intellectual property, there were concerns about reviewer privacy and the ability to attract good reviewers, the key input to high-quality review. There were also changes being proposed to FACA to incorporate more lay

[23] The DARPA model includes active program management to achieve specific objectives, among other features. See Cook-Deegan (1996) and Azoulay et al. (2019).
[24] This is different from, though may correlate with, the claim discussed above that peer review may be biased against specific kinds of research (e.g., clinical research, applied research, or even novel research).
[25] This may have been another reason—beyond scientific freedom considerations and difficulties in evaluating large number of project grants—why Bush preferred institutional funding to individual grants. Such grants could be used to lift up lower-ranked institutions, as was proposed in a number of the early external evaluations reviewed above.

membership in federal advisory groups (a Kilgorean idea), and questions on whether this might undermine the scientific rigor of peer review.[26] The Privacy Act could potentially allow PIs and others to collect information about their applications: one question was what information from the review would be released, and to whom.

GPRST was formed in April 1975 to assess the legal constraints of these laws and "devise procedural alternatives" (Mandel 1996, 148). Beyond the Sunshine Act issues, the GPRST also took a broader mandate: to "[e]xamine in critical detail the entire process of peer review and make, where necessary, recommendations for modifications or change" (GPRST Phase I, V).[27] The Committee was composed of NIH officials, and over a two-year study it administered surveys of study section members and advisory council members, mailed questionnaires to grantees, and held public hearings on the strengths and weaknesses of the peer review system.

The GPRST reported that its various respondents expressed "general satisfaction" with the review system, and that there was a lack of enthusiasm for changes "arising from Congressional and public concerns with the openness of Government processes" (GPRST Phase I, 58). On the potential changes from the Sunshine Laws, it recommended making information (including priority scores) available to applicants, and opening to the public the process for nominating study section members and advisory council members, changes that were adopted (National Institutes of Health, 1978). The GRPST raised concerns that broad release of information (including research designs and protocols) would make scientists less willing to include "new" ideas in their applications if they were available to the public, undermining the quality of review and discouraging new ideas, and recommended changes to FOIA to provide statutory exemption for such materials.[28] It also speculated on potentially negative effects of the Sunshine Laws—in particular whether they would lead to under-disclosure of important

---

[26] The Department of Defense's Congressionally-directed Breast Cancer Research Program, which began in the 1990s, is a potentially interesting point of comparison. The program, explicitly focused on funding high-risk, high-impact research not fundable through NIH, has a dual peer review structure similar to NIH. Unlike NIH, it includes "consumers" (breast cancer patients and survivors) as voting members in both the scientific and programmatic stages of peer review. See IOM (1997). At NIH today, public representatives participate in the second stage of peer review (advisory councils) but not the first (study sections).

[27] As Kirschstein put it to the NIH Director in an interim report: "How can a system devised in an era of elitism, of secrecy, and of economic growth … be adopted to an era in which stress is on equal opportunity, openness, and limited availability of funds?" (quote reproduced from Mandel 1996, 148).

[28] Part of the GPRST was devoted to ongoing litigation in the Washington Research Project case (*Washington Research Project, Inc. v. Department of Health, Education and Welfare*), where a non-profit group challenged the NIH/DHEW denial of its FOIA request for information on 11 NIMH grants on the effects of psychotropic drugs on children with learning disabilities. In September 1974 the U.S. Court of Appeals for the District of Columbia ruled that grant applications and their continuation/renewal/supplemental applications that were approved or pending were subject to disclosure under FOIA. In the 1980s, researcher George Kurzon prevailed in FOIA litigation in getting a list of successful and unsuccessful applications from another NIH Institute, the NCI, arguing that without such data it is difficult to evaluate the NIH's effectiveness, and against the government's claim that there are significant privacy interests (including stigma to unfunded investigators) served by withholding this information. Kurzon wanted the data to examine whether NIH was biased against innovative applications. The agency ultimately released to Kurzon a printout of 7,609 successful and unsuccessful applicants. See Chubin and Hackett (1990). A 2001 article (Reporters Committee 2001) describes a later victory by Kurzon for similar NIMH data. In general, the NIH has been reluctant to release information on unfunded applications, and/or priority scores, to outside researchers to evaluate its effectiveness, citing privacy concerns like those discussed in the GPRST report.

information in proposals, and whether this may disproportionately hurt funding chances for new investigators without track records—and recommended periodic evaluation of this new set of regulations on innovation and scientific progress.

Among the broader set of peer review issues that were considered, issues related to the "workload crisis"—the high burden of a growing number of applications on peer reviewers and administrative staff—were also on the agenda (Mandel 1996). GPRST recommended that the NIH Director have the authority to create new study sections and split them, and to set workload limits, among other approaches to handling the problem of quality evaluation at scale. It also brought attention to a specific aspect of the quality problem: difficulties faced by unorthodox and innovative research in the peer review process. Though it considered a range of factors (membership of peer review groups, the review criteria, the fact that the process funds projects not people), it ultimately laid the blame on limited funding, writing "[A]s the dollars available for research become more and more limited, there is a tendency (almost inevitable and often uncontrollable) to invest in the "safe bet."" (GPRST Phase 1, 170). It recommended various changes, including that the applicant more precisely identify innovative aspects of their research, that innovativeness be discussed in summary statements, and "that the NIH consider the feasibility of developing an experiment involving limited support" for high-risk, innovative proposals, "part of a large, much-needed effort to examine the processes of decision-making in allocating research support" (GPRST Phase 1, 171).

5.2 Subsequent evaluations and the evolution of peer review

Many later reports reiterated some version of the high-level assessment of GPRST: that most scientists it contacted endorsed the existing system, and that the quality of NIH peer review was high (Mandel 1996). Yet the questions they asked and their analyses reveal continued concern about the ability to do quality review at scale, and the high burden on the NIH staff and the external scientific community that would be required to do so (Mandel 1996). Mandel (1996) observes that the NIH study sections "undertook what Bush insisted could not be done" in evaluating the scientific merit of large numbers of small projects at scale (45). The internal evaluations and reports—often coming back to the very same issues every few years—emphasize the difficulties of doing so.

A specific quality-related issue that received considerable attention in internal reports is the priority scores of applications by study sections. Priority score data have been analyzed in external and internal analyses to look at a range of issues, including how the quality of applications changed over time (whether quality kept up with increased funding), the validity and reliability of scores as a measure of scientific quality, and whether priority scores were effectively determinative of funding, i.e. whether the Institutes and Centers (ICs) were outsourcing to study sections "decisions that by right belonged to Institute directors" (Mandel 1996, 117). Another perennial topic has been heterogeneity in scoring by study sections. In discussing whether priority scores should be used to make funding choices, a 1968 report (authored by former National Heart Institute administrator Franklin Yeager) suggested no, given "study section bias" and difficulties in comparing across study sections; it advocated for cuts to program grants and large grants instead, as a means of rationing funds (Mandel 1996, 117). The sources of bias include gaming of the process by some study sections, to enhance the probability of getting "their" science funded in the second stage of peer review (Mandel 1996; National Institutes of Health 1976). The difficulty of "interdigitating" raw scores from different study sections posed challenges for ICs throughout the 1960s (NIH 1981).

In the face of this confusion, in 1971 the NIH introduced normalized priority scores (Mandel 1996). Study sections reported both raw and normalized scores, and different ICs used different scores in funding choices, creating confusion for applicants, Advisory Councils, and all around (National Institutes of Health 1976). The GPRST suggested a single score would be useful but did not indicate which one (raw or normalized). Throughout the 1970s there was heterogeneity in how these scores were used.[29] In 1980, NIH Director Donald Fredrickson instituted an NIH-wide convention to use raw scores, leaving the "unique task [of] determining how to relate ratings of one IRG [study section] to another" to the Institutes (NIH 1981). A 1984 Priority Score Working Group Report reiterated that if priority scores are not comparable, they are questionable as measures of merit. This Group eschewed statistical correction, instead focusing on orientation and training of reviewers to promote standardization. This Working Group and others noted related concerns about priority score creep, in which a large and growing share of applications garnered high priority scores—perhaps because applications were improving, perhaps because of study section gaming to advantage their area of science (NIH 1984). In 1987, the Working Group on the Movement of Priority Scores ("Wigmops") suggested improving comparability by using percentile rankings (following a practice that was being used by NHLBI) which, together with other changes it recommended to the process, became the standard across ICs by the late 1980s (Mandel 1996).[30]

Scores were also prominent in various internal assessments of peer review that were launched as part of the NIH's response to Clinton-Gore's "reinventing government" efforts in the early 1990s. In 1994 the NIH created the Rating of Grant Applications (RGA) Committee (technically a subcommittee of the Committee on Improving Peer Review; NIH Office of Extramural Research, 1999). Among the RGA's ten recommendations, several focused on the rating scale, including a suggestion that there be an eight-point integer scale, with higher numbers reflecting better scores, replacing the existing 41-point scale (1.0 to 5.0 in 0.1-unit increments), to avoid false precision. There was also discussion of the specific review criteria should be, and whether they should be individually scored and aggregated. RGA recommended three criteria for assessing grant applications (significance, approach, and feasibility), individually scored, and that the overall "global" score for an application be mathematically derived from the individual scores.

RGA recommendations were also reviewed by a new study group, the Peer Review Oversight Group (PROG), which was to be a centralized NIH body, created by NIH Director Varmus in 1996, charged with coordinating, evaluating, and suggesting changes to NIH peer review (NIH *Peer Review Notes*, October 1996). Based on information from pilot studies and input from the scientific community, PROG found support for some of RGA's recommendations, but reservations about basing overall scores on individual criterion scores (NIH Office of Extramural Research, 2002). Varmus suggested that there was "anxiety" in the reviewer community that under such an approach scientists would lose their autonomy, and criticized the RGA approach as trying to "mechanize what had previously been a largely intuitive process" (NIH *Peer Review Notes,* June 1997, 2).

---

[29] By the late 1970s, several ICs used normalized scores (NIA, NIAID, NICHHD, NCI, NEI, NHBLI) while others did not. This had implications for not just individual grants but IC level funding: when the Congress in 1980 decided to not fund grants scoring above 212, the NHI (which used normalized scores) claims it effectively lost $12 million in funding (Mandel 1996), about 2 percent of its budget that year.

[30] Under percentiling, the priority scores for the current and previous two meetings of a study section were used to normalize. NIH *Peer Review Notes* (May 1988) observed this change would "emphasize the importance of relative rank and provide compensation for widely differing scoring practices" across study sections.

Another major focus of PROG, as with GPRST, was on whether NIH's peer-review system was too conservative and resulted in not enough high-risk, high-reward science. Following its deliberations, a new set of review criteria was implemented in 1997: Significance, Approach, Innovation, Investigator, and Environment (NIH, 1997). The addition of "Innovation" was controversial, both in terms of whether it was something reviewers could actually evaluate but also fears that it could penalize certain types of important (but not especially innovative) research (Marshall 1997).[31] The process would continue to rely on an overall score, not (as per RGA) an average of individual criterion scores, but based on reviewers' subjective assessment of the application overall.

Fifty years after the original study sections were formed, NIH officials and the extramural research community were also worried that the organization of the study sections, which evolved haphazardly for administrative reasons, was no longer serving the goal of identifying the best science. The list of concerns were long: that new fields had no study sections, that clinical research was not fairly reviewed, that too many of the "best" proposals ended up in a few study sections (causing the best science to compete with itself), among others (Alberts et al. 1999). In April 1998 a new Panel on the Scientific Boundaries of Peer Review was created, chaired by Bruce Alberts, the president of the National Academy of Sciences. The major divisions of the Center for Scientific Review, their Integrated Review Groups (IRGs), and the study sections they include were re-organized over the early 2000s, resulting in the basic structure in place today ("Center for Scientific Review" 2015).[32]

The most recent major set of reforms occurred under the auspices of Elias Zerhouni, who took over as NIH Director in 2002.[33] The "Trans-NIH Effort to Enhance Peer Review" which began in 2007, suggested a range of changes to reduce administrative burden, to enhance consistency in the scoring system, to improve review quality, to improve success rates of early-career investigators, and to promote transformative research. The result was the implementation in 2009 of "the most sweeping enhancements to peer review in its 60-year history" (NIH *Peer Review Notes*, September 2009, 1). These changes implemented the basic scoring system (scoring on a 9-point whole number scale; applications scored on five core criteria, plus an independent overall score) and review criteria that are used today (NIH Peer Review 2023).

Table 2 summarizes these and other incremental changes in the peer review process, based on information compiled from the 1996-2010 editions of *Peer Review Notes*, a publication used to update reviewers and the extramural community about policy developments. In many cases, not all, these reflected recommendations from the various reports and study groups described above. In general, the changes focused on improving review quality, especially given the constantly expanding scale of the enterprise, but also on reducing administrative burden, promoting equity, reducing bias and conflict of interest, supporting early-career investigators, and promoting innovation, among others. Since these are many of the same goals of current

[31] Claude Lenfant, a former head of NHLBI, worried this would penalize clinical research. The question of whether the peer review process discriminates against applied and clinical research is another long-standing tension at the NIH, and also divides along Bush-Kilgore lines (Kastor 2010, Chapter 2).
[32] Yet another centralized approach to peer review reform was the Peer Review Advisory Committee (PRAC), replacing PROG and others, created in 2005. PRAC recommended changes to shorten the application to relieve burden on both applicants and reviewers, and continued work to improve identification of innovative and high-impact work, including through new mechanisms (*NIH Peer Review Notes, September* 2006) and through changing scoring approaches (*NIH Peer Review Notes, September* 2007).
[33] Varmus retired from this post in 1999. From 2000 to 2002, Ruth Kirschstein was Acting Director.

reform efforts, revisiting the effects of these earlier changes (or barriers to implementation, why they didn't succeed) could be informative.[34]

6 Conclusions

The basic architecture of NIH's peer review process was built on the dual review structure created at the end of World War II. Donald Stokes (1997) and other science policy scholars have praised this model as an effective approach to funding research, balancing scientific opportunity and use considerations. However, there has been growing criticism of the peer review process, including that it is too conservative, administratively burdensome, biased, not efficient at choosing the highest quality proposals, and insufficiently responsive to the nation's actual health problems and priorities. These criticisms have intensified since the Covid-19 pandemic and have spurred a range of recent reform proposals, most focused on peer review.

How, then, can a 75-year overview of NIH peer review inform current reform efforts and help improve the NIH for the future? In summary, there are five main implications. First, the history reveals the complexities of NIH reform, including the difficulties of simultaneously investing in discovery for pure science's sake while also addressing societal goals—which are themes that trace back to Vannevar Bush and Harvey Kilgore in the 1940s and 1950s. Paraphrasing Stokes (1997), Americans fund research not for what it is, but what it is for. When it comes to the NIH, there are different perspectives on this among different stakeholders. Chubin and Hackett (1990) identify a number of "desiderata" of peer review, including what I have called Bushian goals of scientific advancement and innovation, but also Kilgorean goals of responsiveness to social needs, equity, and accountability. Traditionally high levels of bipartisan support for the NIH may require peer review to respond to numerous objectives, not just one. Current reform efforts need to be specific about the endpoints they are targeting and to explicitly consider potential tradeoffs between different outcomes. As Azoulay et al. (2013) observed in their discussion of incorporating aspects of the Howard Hughes Medical Institute (HHMI) "people not projects" approach at NIH, to promote incentives of exploration, a public agency does not have the same latitude to focus only on breakthrough science or exploration, since the system has other goals and thus political constraints.[35]

---

[34] The internal evaluations also triggered "experiments" within the NIH. Beginning with the GPRST, internal reports emphasized the importance of testing changes to the process before implementing them. Jerome G. Green, the Director of the NIH's Division of Research Grants (which later became the Center for Scientific Review) from 1986 to 1995, was a strong proponent of experiments, remarking "It is regrettable that many who go from conducting research to administering research lose their fondness for careful experimentation and tend to accept their intuitions about the process of review" (McManus 1989, 9). Many of the major changes in peer review described above, including those emanating from the recommendations of internal reports, were first piloted by NIH (used by some study sections not others, some types of applications not others, and some Institutes not others). For example, "triage" of applications—where reviewers are asked to identify non-competitive applications ex ante, and these are neither discussed or scored—was evaluated initially on proposals resulting from RFAs, finding "little likelihood that a highly competitive application would be wrongly disqualified" (199), and then for all applications (Mandel 1996). Under Varmus, NIH also experimented with "just in time" data (detailed budgets only required after first round of review) to reduce administrative burden, small "chunk" grants for fixed amounts (limiting budget details), and approaches to rewarding innovative ideas (Marshall 1994b).

[35] On the longstanding tensions between the "health" and "science" aspects of the NIH mission (Strickland 1972; Rettig 1977; Cook-Deegan and McGeary 2006; Sampat 2012), it is possible that finding and funding the best science is indeed the most effective route to improving health—as Bush, Van Slyke

Second, the history shows that many of the changes that are central to current policy reform efforts—identifying innovation and breakthroughs, reducing administrative burden, improving quality of review and informational content of scores, supporting early-career researchers, and countering biases—have been of central concern to NIH itself for decades, and targets of incremental reforms. Going forward, reform efforts need to better understand why prior attempts have not solved the problems. Were they too modest in scope, did they fail in implementation or to follow through, did internal politics stymie them, were more resources needed, was the culture of NIH or the extramural community too hard to change? Or something else? Third, and related, revisiting "natural" experiments—reforms tried by some Institutes or study sections or for some grants but not others, or with staggered implementation—could be a useful input to future reform.

A fourth thing the history makes clear is that there have been many empirical evaluations of the peer-review process (and changes to it), but mainly by internal analysts. The agency's reluctance to involve external analysts in analyses of peer review data may reflect fear that full transparency may reveal cracks in the armor and threaten NIH budgets.[36] But it is also rooted in historical concerns about privacy, based on concerns about getting high-quality applications and high-quality reviewers to assess them. External analyses can bring objectivity, as well as data and empirical approaches beyond that available internally. Congress could help overcome both constraints by supporting the agency in creating infrastructure to make available the data needed for objective external research and evaluation using NIH peer review data (including scores and unfunded applications), with appropriate safeguards for confidentiality and privacy (Stein 2022; Buck and Kadakia 2022).

A fifth implication relates to the specific issue of administrative burdens and regulations facing NIH applicants. The historical review suggests since the beginning, the NIH's primary focus was limiting administrative burden on applicants and grantees. Many of the rules and regulations now associated with NIH bureaucracy come not internally from the NIH, but externally from Congress. If these no longer serve their intended purposes, it would be useful for legislators to endorse and provide support for reviewing and reducing administrative burden. For example, Congress might usefully ask the agency, through report language or other levers, to document the range of regulations facing applicants and their historical rationale, and assess whether the original rationale continue to make sense. Or it could commission external assessments (e.g. through the National Academies or a special commission) to do so. As with reform efforts, evaluation of the costs and benefits of regulations should consider the broad range of objectives the agency must consider, and the myriad stakeholders who have historically sustained large NIH budgets.

There is another, more radical, implication of the observation that the NIH peer review process struggles to balance numerous competing aims. Chubin and Hackett (1990) asked in their important study of peer review (which had a broader focus than just the NIH) whether "perhaps the peer review process has been pressured to serve so many distinct purposes that it serves none well" (2). Partly because of the difficulties of doing "all of the above" within the NIH,

---

and others have asserted. Health versus science may be a false dichotomy. But the evidence base on this remains limited and would need to be strengthened to get buy-in from stakeholders whose support is predicated on the expectation that NIH research will yield health improvements, and who wish to steer the agency towards specific health priorities.

[36] Richard Klausner, NCI Director from 1980-1988, describes peer review as a sacred cow: "People say these things [criticisms of peer review] in private, not public, because they're concerned that such comments may discourage the public and Congress from supporting the NIH" (Kastor 2010, 110).

reformers since the 1950s have suggested replacing or supplementing the NIH with other research funding agencies—from the Long Report's suggestion to sever the extramural program from the NIH (to protect basic research) to cancer advocates' proposal to separate the NCI from NIH (so the NCI could focus on targeted investigations). Veteran science journalist Dan Greenberg called for an "NIH Two" far from Bethesda (Greenberg 1998) which would be "unencumbered by tradition or prior commitments" (Greenberg 2007). These debates have been rehashed in modern proposals that the newly created ARPA-H, meant to catalyze breakthrough science and technology, be housed in a separate entity from the NIH (Sampat and Cook-Deegan 2021).[37] Thus far none of these separatist attempts has been successful. Correcting perceived deficiencies in the biomedical research system—whether with respect to Bushian ends of supporting science or Kilgorean goals of technology development and targeting outcomes—could ultimately require not just going through the 75-year-old NIH peer review structure, but also around it.

[37] Drawing inspiration from the DARPA model (see the end of Section 4, above), disease advocates proposed ARPA-H to allow the federal government to better fund breakthrough research, develop technology platforms, do big projects, and otherwise fund research that NIH peer review process does not. Advocates argued for an independent agency, on the theory that NIH cannot be fixed from within. NIH officials countered that there were complementarities between what the proposed ARPA-H would do and existing NIH activities (Sampat and Cook-Deegan 2021). In 2022, ARPA-H was created as an "independent entity" within NIH.

7 References

Alberts, Bruce M., Francisco J. Ayala, David Botstein, Ellen Frank, Edward W. Holmes, Ronald D. Lee, Eduardo R. Macagno, et al. 1999. "Proposed Changes for NIH's Center for Scientific Review." Science 285 (5428): 666–67. https://doi.org/10.1126/science.285.5428.666.

Alberts, B., M. W. Kirschner, S. Tilghman, and H. Varmus. 2014. "Rescuing US Biomedical Research from Its Systemic Flaws." *Proceedings of the National Academy of Sciences* 111 (16): 5773–77. https://doi.org/10.1073/pnas.1404402111.

American Association for the Advancement of Science (AAAS). 2023. "Historical Trends in Federal R&D" Accessed March 8, 2023. https://www.aaas.org/programs/r-d-budget-and-policy/historical-trends-federal-rd.

Azoulay, Pierre. 2012. "Turn the Scientific Method on Ourselves." *Nature* 484 (7392): 31–32. https://doi.org/10.1038/484031a.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat. 2019. "Public R&D Investments and Private-Sector Patenting: Evidence from NIH Funding Rules." *The Review of Economic Studies* 86 (1): 117–52. https://doi.org/10.1093/restud/rdy034.

Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso. 2013. "National Institutes of Health Peer Review: Challenges and Avenues for Reform." *Innovation Policy and the Economy* 13 (January): 1–22. https://doi.org/10.1086/668237.

Balaguru, Logesvar, Chen Dun, Andrea Meyer, Sanuri Hennayake, Christi Walsh, Christopher Kung, Brittany Cary, et al. 2022. "NIH Funding of COVID-19 Research in 2020: A Cross-Sectional Study." *BMJ Open* 12 (5): e059041. https://doi.org/10.1136/bmjopen-2021-059041.

Baxter, J.P. 1946. *Scientists Against Time*. Science in World War II, v. 1. Little, Brown. https://books.google.com/books?id=57lgAAAAIAAJ.

Bayne-Jones, Stanhope. 1958. "The Advancement of Medical Research and Education through the Department of Health, Education, and Welfare: Final Report of the Secretary's Consultants on Medical Research and Education." Washington, D.C.: U.S. Government Printing Office.

Buck, Stuart. 2022. "How To Actually Reduce Administrative Burden on Research." Substack newsletter. *The Good Science Project* (blog). October 13, 2022. https://goodscience.substack.com/p/how-to-actually-reduce-administrative.

Buck, Stuart, and Kushal T. Kadakia. 2022. "Investing In The Science Of Science: What Medicare Can Teach The NIH About Experimentation." *Health Affairs Forefront*. https://doi.org/10.1377/forefront.20221020.721179 .

"Budget." 2014. National Institutes of Health (NIH). October 31, 2014. https://www.nih.gov/about-nih/what-we-do/budget.

Bush, Vannevar. 1945. "Science, the Endless Frontier." 1945. https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm.

Callahan, Daniel. 2006. What Price Better Health?: Hazards of the Research Imperative. Berkeley, Calif.: University of California Press. https://www.ucpress.edu/book/9780520246645/what-price-better-health

"Center for Scientific Review (CSR)." 2015. National Institutes of Health (NIH). July 10, 2015. https://www.nih.gov/about-nih/what-we-do/nih-almanac/center-scientific-review-csr.

Chubin, Daryl E., and Edward J. Hackett. 1990. *Peerless Science: Peer Review and U. S. Science Policy*. Illustrated edition. Albany, N.Y: SUNY Press. https://sunypress.edu/Books/P/Peerless-Science2

Committee on Government Operations. 1962. "Administration of Grants by the National Institutes of Health: Reexamination of Management Difficulties. Twenty-First Report by

the Committee on Government Operations, 87th Congress, Second Session. House Rept. No. 1958." Washington, D.C.: U.S. Government Printing Office.
Cook-Deegan, Robert, and Michael McGeary. 2006. "The Jewel in the Federal Crown?: History, Politics, and the National Institutes of Health." In *History and Health Policy in the United States*, 9780813539874:176–201. Rutgers University Press. http://www.scopus.com/inward/record.url?scp=36749015940&partnerID=8YFLogxK.
Cooper, J. D. 1965. "Onward the Management of Science: The Wooldridge Report: NIH Was Not Cleared on All Counts by the Wooldridge Committee, Which Itself Is Rated Low on Methodology." *Science* 148 (3676): 1433–39. https://doi.org/10.1126/science.148.3676.1433.
Crow, Michael M. 2011. "Time to Rethink the NIH." *Nature* 471 (7340): 569–71. https://doi.org/10.1038/471569a.
"CSR Data & Evaluations | NIH Center for Scientific Review." n.d. Accessed May 10, 2023. https://public.csr.nih.gov/AboutCSR/Evaluations.
Cutler, David, and Srikanth Kadiyala. 2003. *Measuring the Gains from Medical Research*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/M/bo3643529.html
Cutler, David M. 2004. Your Money or Your Life: Strong Medicine for America's Health Care System. Oxford New York: Oxford University Press. https://scholar.harvard.edu/cutler/publications/your-money-or-your-life-strong-medicine-americas-health-care-system
Drew, Elizabeth. 1967. "The Health Syndicate: Washington's Noble Conspirators." *Atlantic Monthly* 220.
Edsall, John T., Robert P. Fischelis, L. H. Fountain, Fritz Lipmann, Ernst A. Scharrer, Helen Wendler Deane, Walter J. Burdette, et al. 1963. "More on Paper Work." *Science* 140 (3567): 721–34.
Executive Order 8807 - Establishing the Office of Scientific Research and Development. (1940). https://www.presidency.ucsb.edu/documents/executive-order-8807-establishing-the-office-scientific-research-and-development
Fang, Ferric C., Anthony Bowen, and Arturo Casadevall. 2016. "NIH Peer Review Percentile Scores Are Poorly Predictive of Grant Productivity." *ELife* 5 (February): e13323. https://doi.org/10.7554/eLife.13323.
Fang, Ferric C., and Arturo Casadevall. 2009. "NIH Peer Review Reform—Change We Need, or Lipstick on a Pig?" *Infection and Immunity* 77 (3): 929–32. https://doi.org/10.1128/IAI.01567-08.
———. 2016. "Research Funding: The Case for a Modified Lottery." *MBio* 7 (2): e00422-16. https://doi.org/10.1128/mBio.00422-16.
Fountain, L.H. 1962. *The Administration of Grants by the National Institutes of Health.* Washington: U. S. Govt. Print. Off. https://catalog.hathitrust.org/Record/001558338.
Fox, Daniel M. 1987. "The Politics of the NIH Extramural Program, 1937-1950." *J Hist Med Allied Sci*, 20. https://doi.org/doi:10.1093/jhmas/42.4.447.
Franzoni, Chiara, Paula Stephan, and Reinhilde Veugelers. 2022. "Funding Risky Research." *Entrepreneurship and Innovation Policy and the Economy* 1 (January): 103–33. https://doi.org/10.1086/719252.
Fredrickson, Donald S. 1993. "Biomedical Science and the Culture Warp." In *Emerging Policies for Biomedical Research*, edited by William Kelley. Washington, D.C.: Association of Academic Health Centers. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101584939X148-doc.
Ginther, Donna K., Walter T. Schaffer, Joshua Schnell, Beth Masimore, Faye Liu, Laurel L. Haak, and Raynard Kington. 2011. "Race, Ethnicity, and NIH Research Awards." *Science* 333 (6045): 1015–19. https://doi.org/10.1126/science.1196783.

Ginzberg, Eli, and Anna B. Dutka. 1989. *The Financing of Biomedical Research*. Baltimore: Johns Hopkins University Press.
Goldstein, Murray. 1986. An Interview with Dr. Murray Goldstein by Steven P. Strickland Ph.D. on the Occasion of the 100th Anniversary in 1987 of the National Institutes of Health.
Greenberg, D. S. 1963. "NIH and Fountain: Part of Problem Is That an Atmosphere of Suspicion Has Enveloped the Relationship." *Science* 140 (3571): 1076–78.
———. 1967. "Representative Fogarty Dies at 53." *Science* 155 (3759): 180–180. https://doi.org/10.1126/science.155.3759.180.
———. 1998. "Creaky NIH Is Past Its Prime – Baltimore Sun." 1998. https://www.baltimoresun.com/news/bs-xpm-1998-07-12-1998193026-story.html.
———. 2007. "We Need a Second NIH." *APS Observer* 20 (June). https://www.psychologicalscience.org/observer/we-need-a-second-nih.
Gross, C. P., G. F. Anderson, and N. R. Powe. 1999. "The Relation between Funding by the National Institutes of Health and the Burden of Disease." *The New England Journal of Medicine* 340 (24): 1881–87. https://doi.org/10.1056/NEJM199906173402406.
Gross, Cary P. and Emanuel, Ezekiel J. 2022. "The Missing Part of America's Pandemic Response." The Atlantic. June 5, 2022. https://www.theatlantic.com/ideas/archive/2022/06/nih-covid-vaccine-research-studies/661182/.
Gruber, Jonathan, and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. First edition. New York: PublicAffairs. https://www.hachettebookgroup.com/titles/jonathan-gruber/jump-starting-america/9781541762480/?lens=publicaffairs
Harden, Victoria. n.d. "A Short History of the National Institutes of Health - Office of NIH History and Stetten Museum." Accessed May 10, 2023. https://history.nih.gov/display/history/A+Short+History+of+the+National+Institutes+of+Health.
Institute of Medicine (US) Committee to Review the Department of Defense's Breast Cancer Research Program. 1997. *A Review of the Department of Defense's Program for Breast Cancer Research*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/5808.
Institute of Medicine. 1984. *Responding to Health Needs and Scientific Opportunity: The Organizational Structure of the National Institutes of Health*. https://doi.org/10.17226/762.
———. 1998. *Scientific Opportunities and Public Needs: Improving Priority Setting and Public Input at the National Institutes of Health*. The National Academies Collection: Reports Funded by National Institutes of Health. Washington (DC): National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK45365/.
"Institutes at NIH." n.d. National Institutes of Health (NIH). Accessed May 10, 2023. https://www.nih.gov/institutes-nih.
Intergovernmental Relations Subcommittee. 1961. *Health Research and Training; the Administration of Grants and Awards by the National Institutes of Health*. 87th Cong. 1st Sess. House. Report No. 321. Washington: U.S. Govt. Print. Off. https://catalog.hathitrust.org/Record/101721139.
Ioannidis, John P. A. 2011. "Fund People Not Projects." *Nature* 477 (7366): 529–31. https://doi.org/10.1038/477529a.
Jones, Boisfeuillet. 1960. "Federal Support of Medical Research: Report of the Committee of Consultants on Medical Research to the Subcommittee on Departments of Labor and Health, Education, and Welfare of the Committee on Appropriations, United States Senate, 86th Congress, Second Session." Washington, D.C.: U.S. Government Printing Office.

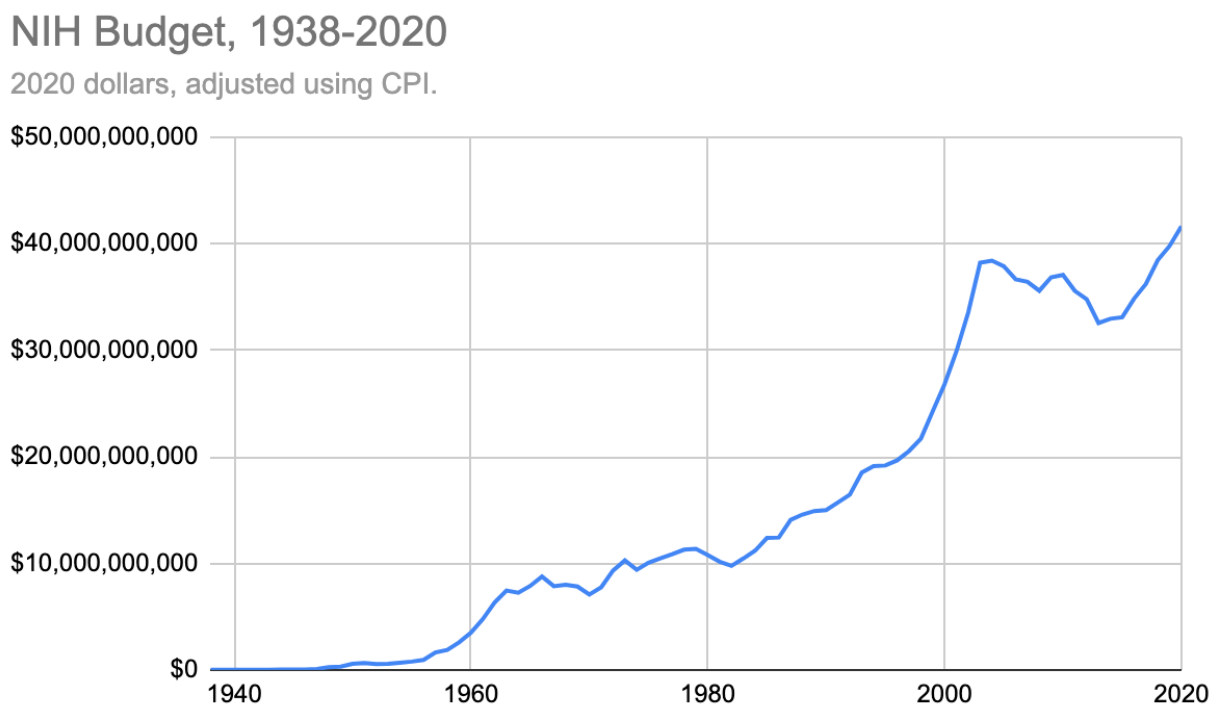
Kaplan, Robert M. 2019. *More than Medicine: The Broken Promise of American Health*. Cambridge, MA: Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674975903
Kastor, John. 2010. *The National Institutes of Health: 1991-2008*. Oxford, New York: Oxford University Press. https://global.oup.com/academic/product/the-national-institutes-of-health-9780199737994?cc=us&lang=en&
Keefer, C. S. 1969. "Dr. Richards as Chairman of the Committee on Medical Research." *Annals of Internal Medicine*, November, Suppl 8:61-70.
Kevles, Daniel J. 1977. "The National Science Foundation and the Debate over Postwar Research Policy, 1942-1945: A Political Interpretation of Science--The Endless Frontier." *Isis* 68 (1): 5–26. https://www.jstor.org/stable/230370
Li, Danielle, and Leila Agha. 2015. "Big Names or Big Ideas: Do Peer-Review Panels Select the Best Science Proposals?" Science 348 (6233): 434–38. https://doi.org/10.1126/science.aaa0185.
Long, Cecil. 1955. "Medical Research Activities of the Department of Health, Education and Welfare: Report of the Special Committee on Medical Research Appointed by the National Science Foundation at the Request of the Secretary of Health, Education and Welfare." National Science Foundation.
Mandel, Richard. 1996. *A Half Century of Peer Review, 1946-1996.* Division of Research Grants, National Institutes of Health.
Marshall, Eliot. 1994 "Peer Review Reforms Get Good Review." *Science* 265 (5171): 467–68.
———. 1997. "NIH Plans Peer-Review Overhaul." *Science* 276 (5314): 888–89. https://doi.org/10.1126/science.276.5314.888.
McGeary, Michael, and Philip M. Smith. 2002. "Organizational Structure of the National Institutes of Health," 187.
McManus, Rich. 1989. "Panel Finds Peer Review Nearly Peerless" *The NIH Record*, April 18.
Mowery, David C., Richard R. Nelson, and Ben R. Martin. 2010. "Technology Policy and Global Warming: Why New Policy Models Are Needed (or Why Putting New Wine in Old Bottles Won't Work)." *Research Policy* 39 (8): 1011–23. https://doi.org/10.1016/j.respol.2010.05.008.
Munger, Mary. 1960. "Growth of the Extramural Programs of the National Institutes of Health, FY1946-FY1958."
Murphy, Franklin. 1976. "Report of the President's Biomedical Research Panel." Washington, D.C.: DHEW.
Murrin, Suzanne. 2021. "Selected NIH Institutes Met Requirements for Documenting Peer Review But Could Do More To Track and Explain Funding Decisions." U.S. Department of Health and Human Services: Office of the Inspector General. https://oig.hhs.gov/oei/reports/OEI-01-19-00140.pdf.

National Institutes of Health GPRST. 1976. *Grants Peer Review: Report to the Director, NIH, Phase I*. Bethesda, Md.: National Institutes of Health. https://catalog.hathitrust.org/Record/000705747.
National Institutes of Health. 1978. *Decisions by Director, NIH, on Recommendations of Grants Peer Review Study Team*. U.S. Department of Health and Human Services.
National Institutes of Health. "Legislative Chronology." 2022. http://www.nih.gov/about/almanac/historical/legislative_chronology.htm#nineteenforty.
National Institutes of Health. Peer Review Notes (May 1988, June 1996, September 2006, September 2007, September 2009).
National Institutes of Health (NIH). 1997. "Review Criteria for and Rating of Unsolicited Research Grant and Other Applications." NIH Guide, Volume 26, Number 22. https://grants.nih.gov/grants/guide/notice-files/not97-010.html.

National Research Council. 2003. *The Organizational Structure of the National Institutes of Health*. *Enhancing the Vitality of the National Institutes of Health: Organizational Change to Meet New Challenges*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK43491/.
NIH Office of Extramural Research. 1999. "Overview Of Report On Rating of Grant Applications." January 17, 1999. https://web.archive.org/web/19990117092208/http://www.nih.gov/grants/peer/rga.htm#download.
———. 2002. "PROG on the Rating of Grant Applications: Deliberations and Decisions." June 12, 2002. https://web.archive.org/web/20020612135138/http://grants.nih.gov/grants/peer/prog_on_rga.htm
NIH. 1981. "Presentation on the History of Priority Scores and Their Use In Funding Decisions, Office of NIH History: Box 3 (Extramural, DRG Cost-DRG Liason)."
"NIH Budget: What We Do." 2022. National Institutes of Health (NIH). 2022. https://www.nih.gov/about-nih/what-we-do/budget.
"NIH Data Book." 2022. 2022. https://report.nih.gov/nihdatabook/.
"NIH Peer Review." n.d. NIH Central Resource for Grants and Funding Information. Accessed May 11, 2023. https://grants.nih.gov/grants/peer-review.htm.
NIH Priority Score Working Group. 1984. "NIH Priority Score Working Group: Report and Recommendations."
Packalen, Mikko, and Jay Bhattacharya. 2020. "NIH Funding and the Pursuit of Edge Science." *Proceedings of the National Academy of Sciences* 117 (22): 12011–16. https://doi.org/10.1073/pnas.1910160117.
Patel, Sejal S. 2012. "Methods and Management: NIH Administrators, Federal Oversight, and the Framingham Heart Study." *Bulletin of the History of Medicine* 86 (1): 94–121. https://doi.org/10.1353/bhm.2012.0017.
P. H. A. (1963). More Paper Work, Less Research. *Science*, *139*(3556), 725–725. http://www.jstor.org/stable/1710199
Reporters Committee. 2001. "Doctor Wins Access to Names of Rejected Grant Applicants." *The Reporters Committee for Freedom of the Press* (blog). 2001. https://www.rcfp.org/journals/the-news-media-and-the-law-fall-2001/doctor-wins-access-names-reje/.
Rettig, Richard A. 1977. *Cancer Crusade: The Story of the National Cancer Act of 1971*. Princeton, NJ: Princeton University Press.
Richards, A. N. 1946. "The Impact of the War on Medicine." *Science* 103 (2680): 575–78. https://doi.org/10.1126/science.103.2680.575.
Robinson, Judith. 2001. *Noble Conspirator: Florence S. Mahoney and the Rise of the National Institutes of Health*. 1st ed. Washington, D.C: Francis Press.
Rogers, Paul. 1976. "Investigation of the National Institutes of Health, Prepared by the Staff, Committee on Interstate and Foreign Commerce, U. S. House of Representatives and Its Subcommittee on Health and the Environment, 94th Congress, 2nd Session." Washington, D.C.: U.S. Government Printing Office.
Ruina, Jack. 1966. "Report of the Secretary's Advisory Committee on the Management of National Institutes of Health Research Contracts and Grants." Washington, D.C.: U.S. Government Printing Office.
Sampat, Bhaven N. 2012. "Mission-Oriented Biomedical Research at the NIH." *Research Policy* 41 (10): 1729–41. https://doi.org/10.1016/j.respol.2012.05.013.
———. 2020. "Whose Drugs Are These?" *Issues in Science and Technology*, July. https://issues.org/drug-pricing-and-taxpayer-funded-research/.

Sampat, Bhaven N., and Robert Cook-Deegan. 2021. “An ARPA for Health Research?” *Milbank Memorial Fund*. https://www.milbank.org/quarterly/opinions/an-arpa-for-health-research/.
Sampat, Bhaven N., and Frank R. Lichtenberg. 2011. “What Are The Respective Roles Of The Public And Private Sectors In Pharmaceutical Innovation?” *Health Affairs* 30 (2): 332–39. https://doi.org/10.1377/hlthaff.2009.0917.
Sarewitz, Daniel. 2016. “Saving Science.” *The New Atlantis* (blog). 2016. https://www.thenewatlantis.com/publications/saving-science.
Schneider, W. H. 2015. “The Origin of the Medical Research Grant in the United States: The Rockefeller Foundation and the NIH Extramural Funding Program.” *Journal of the History of Medicine and Allied Sciences* 70 (2): 279–311. https://doi.org/10.1093/jhmas/jrt074.
Stein, Carolyn. 2022. “Unlocking Federal Grant Data To Inform Evidence-Based Science Funding.” *Day One Project* (blog). 2022. https://www.dayoneproject.org/ideas/unlocking-federal-grant-data-to-inform-evidence-based-science-funding/.
Stephan, Paula E. 2015. *How Economics Shapes Science*. First Harvard University Press paperback edition. Cambridge, Mass.: Harvard Univ. Press.
Stewart, Irvin. 1948. *Organizing Scientific Research for War: The Administrative History of the Office of Scientific Research and Development*. Boston: Little, Brown.
Stokes, Donald E. 1997. *Pasteur’s Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press.
Strickland, Stephen P. 1972. *Politics, Science, and Dread Disease: A Short History of United States Medical Research Policy*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674594890.
———. 1988. *The Story of the NIH Grants Programs*. Lanham, MD: University Press of America.
Swain, Donald C. 1962. “The Rise of a Research Empire: NIH, 1930 to 1950: The Spectacular Growth of NIH Came after Two Decades of Careful Planning by the Public Health Service.” *Science* 138 (3546): 1233–37. https://doi.org/10.1126/science.138.3546.1233.
“The AIDS Research Program of the National Institutes of Health.” 1991. Institute of Medicine. https://doi.org/10.17226/1769.
Van Slyke, C. J. 1946. “New Horizons in Medical Research.” *Science* 104 (2711): 559–67.
Varmus, Harold. 2001. “Proliferation of National Institutes of Health.” *Science* 291 (5510): 1903–5. https://doi.org/10.1126/science.1059063.
Varmus, Harold. 2009. The Art and Politics of Science. New York: W.W. Norton & Company. http://www.ncbi.nlm.nih.gov/books/NBK190622/.
Viergever, Roderik F., and Thom C. C. Hendriks. 2016. “The 10 Largest Public and Philanthropic Funders of Health Research in the World: What They Fund and How They Distribute Their Funds.” *Health Research Policy and Systems* 14 (1): 12. https://doi.org/10.1186/s12961-015-0074-z.
Wooldridge, Dean. 1965. “Biomedical Science and Its Administration: A Study of the National Institutes of Health.” Washington D.C.: The White House.

Figure 1: NIH Budget Growth



Note: Based on annual Congressional appropriations (https://www.nih.gov/about-nih/what-we-do/nih-almanac/appropriations-section-1). Converted to real (2020) dollars using the Consumer Price Index (CPI).

Table 1: NIH Institute and Center History

Institute/Center Name	Abbreviation	Origin Date
National Cancer Institute	NCI	1937
National Institute of Mental Health	NIMH	1946
Center for Scientific Review	CSR	1946
National Institute of Allergy and Infectious Diseases	NIAID	1948
National Heart, Lung, and Blood Institute	NHLBI	1948
National Institute of Dental and Craniofacial Research	NIDCR	1948
National Institute of Diabetes and Digestive and Kidney Diseases	NIDDK	1948
National Institute of Neurological Disorders and Stroke	NINDS	1950
Warren Grant Magnuson Clinical Center	CC	1953
Eunice Kennedy Shriver National Institute of Child Health and Human Development	NICHD	1962
National Institute of General Medical Sciences	NIGMS	1962
National Center for Research Resources	NCRR	1962
Center for Information Technology	CIT	1964
National Library of Medicine	NLM	1968
National Eye Institute	NEI	1968
John E. Fogarty International Center	FIC	1968
National Institute of Environmental Health Sciences	NIEHS	1969
National Institute on Aging	NIA	1974
National Institute for Arthritis and Musculoskeletal Diseases	NIAMS	1986
National Institute of Nursing Research	NINR	1986
National Institute on Deafness and Other Communication Disorders	NIDCD	1988
National Human Genome Research Institute	NHGRI	1989
National Institute on Alcoholism and Alcohol Abuse	NIAAA	1992
National Institute on Drug Abuse	NIDA	1992
National Center for Complementary and Integrative Health	NCCIH	1992
National Institute on Minority Health and Health Disparities	NIMHD	1993
National Institute of Biomedical Imaging and Bioengineering	NIBIB	2000
National Center for Advancing Translational Sciences	NCATS	2011

Notes: Compiled from information provde
"Origin Date" refers to date the entity originally joined the NIH, whether or not originally an Institute.

Table 2: Changes to Peer Review, 1996-2010, Excerpted from *Peer Review Notes*

***Peer Review Notes* Edition**	**Entry Title**	**Notes**	**Focus**
June 1996	Recalibration of Scoring in DRG Study Sections	Response to concerns about score compression and diminished discriminatory power	Quality of review
	Committee Management Policies	Howard Hughes Medical Institute (HHMI) investigators no longer in conflict for reviewing applications of other HHMI investigators	Quality of review; conflict of interest
	Study Section Membership and Voting Rules	"Temporary" members of study sections allowed to vote	Quality of review
	Just-in-Time Submission of R29 and K-Series Grant Applications	R29 and K awards submitted using "just in time" approach, limiting administrative and budgetary information required at time of submission	Applicant burden
	Research Enhancement Awards Program (REAP)	To fund applications which are outside the payline but of interest to ORWH, OBSSR, OAM, ODS	Equity; Programmatic considerations
October 1996	NIAID Accelerated Council Review		Speed
	New NIH Policy On Submission Of Revised (Amended) Applications	Limits number of amended applications to two	Reviewer burden
June 1997	New Instructions for Scoring Fellowship Applications	Aim to spread scores and develop consistency across study sections	Review quality
	DRG to Review Applications Solicited By NIGMS for High Risk/High Impact Research	R21 mechanism to provide pilot supprt for groundbreaking research	Innovation
October 1997	DRG renamed the Center for Scientific Review (CSR)		
	NIH Adopts Explicit Statements of Review Criteria	Five explicit criteria to structure critiques and discussions: significance, approach, innovation, investigator, environment	Review quality; Innovation
	Area (R15) Program Update	AREA (Academic Research Enhancement Award) program, intended to support non-research intensive institution, reorganized from seed grants to funding new research	Equity

Table 2 (continued):

***Peer Review Notes* Edition**	**Entry Title**	**Notes**	**Focus**
June 1998	Change in NIH Policy of Supporting New Investigators	FIRST and R29 mechanisms eliminated; new investigators encouraged to submit R01; R01s from new investigators flagged for reviewers, and reviewers encouraged to adjust review criteria (less preliminary data; less reliance on track record)	Supporting younger investigators; countering biases
October 1998	Consistent Scoring	Scores recalibrated using 3.0 as target median score; aims to spread scores and decrease inconsistent scoring	Review quality
February 1999	NIH Implements Modular Research Grants	Direct costs in $25,000 modules, limited budget detail, enabling reviewers to focus on science. Applies to specific grant types only	Applicant burden; Reviewer burden; Quality of review
January 2000	Conflict of Interest: A Timely and Necessary Liberalization	Blanket conflict of interest waiver for multi-component institutions (e.g. University of California campuses) for purposes of review	Conflict of interest; Review quality
May 2000	Revised IRB Policy	IRB approval no longer required prior to submission (instead, only after initial peer review)	Administrative burden
January 2002	New Instructions for Evaluating Grant Applications Involving Human Subjects	In response to concerns raised by a GAO study, reviewers of clinical research must assess protection of human subjects, data safety and monitoring plan, inclusion of women plan, inclusion of minorities plan, inclusion of children plan	Research ethics and integrity
May 2003	Study Section Reorganization Update	"Design Phase" Complete; Implementation In Progress; First Meetings of New Study Sections Occur	Review quality
	NIH Data Sharing Policy	Investigators seeking >$500,000 in direct costs in any year must include a data sharing plan	Research ethics and integrity; Open science

Table 2 (continued):

Peer Review Notes Edition	Entry Title	Notes	Focus
May 2004	CSR's Reorganization Nears Completion	Final implemention for Integrated Review Groups (IRGs) suggested by Panel on Scientific Boundaries of Peer Review (PSBR)	Review quality
May 2005	Updated Review Criteria	For investigator-initiated applications. Each of the 5 criteria (significance, approach, innovation, investigators, environment) considered in determing an "overall" global score, weighing "as appropriate" given nature of application. Follows from NIH "Roadmap" efforts to better accomodate interdisciplinary work, translational research, clinical research	Review quality
May 2005	Model Organism Sharing Plans	Sharing plans required for research proposing new model organisms	Research ethics and integrity; Open science
May 2006	Multiple Principal Investigators	Implements a government policy to recognize >1 PI on a grant	Equity; Team science
September 2007	CSR to Abolish Application Deadlines for Chartered Reviewers	For R01/R21. To reward reviewers. Will be reviewed by "Special Emphasis Panels"	Review quality
September 2007	Big Step in Shortening the Review Cycle	Shorter review cycles for new investigator R01 applications; can be resubmitted next review cycle. Includes accelerated release of scores and critiques from review.	Speed; New investigators
January 2009	New Year Brings Changes to Peer Review	Implementing recommendations from the Trans-NIH Effort to Enhance Peer Review. Scoring on a 9-point scale (1 for exceptional; 9 poor); applications scored on five core criteria, and (for applications that are discussed) overall impact; clustering of new investigator applications; limits to one resubmission. New formatted reviewer critiques to focus review.	Quality of review; new investigators; reviewer burden; administrative burden

Table 2 (continued):

***Peer Review Notes* Edition**	**Entry Title**	**Notes**	**Focus**
January 2009	New Reviewers Can Lighten Load with Extended Terms	New reviewers can choose four- or six-year terms, the latter allowing to limit number of annual meetings	Reviewer burden; quality of review
January 2010	College of CSR Reviewers to Readily Match Reviewers and Applications	Pool of special reviewers to provide rapid online review of applications for specialized translational and complex reseaerch	Review quality; reviewer burden
January 2010	More Reviewers Can Now Submit Applications Anytime	Extended 2007/2008 changes in appliaction deadlines to more reviewers, to reduce burden reviewers may have in simultaneously preparing and reviewing applications	Reviewer burden; quality of review
Notes: Excerpted from archival issues of *Peer Review Notes*, 1996-2010. Final column is based on my own categorization of the focus of the reform.			