NIH Council of Councils Working Group on Science of Science

Final Report

August 20, 2026

Table of Contents

Executive Summary 4

Overarching general recommendation: 4

Specific recommendations for applying science of science to NIH levers: 4

Specific recommendations for studying the system response: 5

Specific recommendations for studying outcomes: 5

Introduction 6

Working Group Findings 8

A Causal-Chain Framework 9

General recommendation: NIH should make greater use of rigorous empirical methods to study, evaluate, and improve its research funding policies, processes, and outcomes. 12

The Value of a Science of Science Approach to Decision-Making 16

Limitations to a Science of Science Approach to Decision-Making 17

Specific Recommendations 17

Applying the Science of Science to NIH Levers 17

Recommendation 1: Systematically study NIH grant mechanisms to inform adjustments that optimize project management and success. 18

Recommendation 2: Evaluate changes intended to strengthen NIH peer review. 18

Recommendation 3: Optimize NIH funding processes through experimentation, agent-based simulation, and data infrastructure. 19

Recommendation 4: Conduct science of science research to understand how review and funding processes influence the diversity of the scientific portfolio and transformative innovation. 20

Recommendation 5: Implement a core reproducibility and generalizability metadata framework. 21

Recommendation 6: Identify promising technologies and research methodologies to accelerate translation from research to commercialization. 22

Recommendation 7: Use science of science methods to measure, communicate, and maximize the societal impact of NIH-supported research. 22

Recommendation 8: Develop data, methods, and the evidence base to improve portfolio diversity and support paradigm-challenging science. 24

Recommendation 9: Build global science indices to guide strategic NIH investment and benchmark U.S. competitiveness. 25

Understanding the system response 25

Recommendation 10: Invest in datasets from university administrative systems to study scientists' behavior and career progression. 26

Recommendation 11: Use a human-centered approach to study the evolution of scientific fields. 26

Recommendation 12: Promote standards for data usage descriptors to support responsible data reuse and reproducibility. 27

Recommendation 13: Design data trust measures that can be used to support credible and replicable research. 28

Outcomes: Evaluating progress 29

Recommendation 14: Launch targeted living systematic reviews. 29

Recommendation 15: Develop a range of performance measures to capture outcomes of research. 30

Recommendation 16: Develop a range of performance measures to capture outcomes of training programs. 30

Recommendation 17: Improve and validate grant–patent linkages to better enable assessment of NIH's impact on health products. 31
Recommendation 18: Validate output measures against outcomes. 32
Recommendation 19: Support the development of measures for scientific openness, self-correction, and public trust. 32
Recommendation 20: Study K–12 levers that produce future scientists. 33
Recommendation 21: Study the effects of major policy changes in NIH funding priorities and procedures. 33

From Recommendations to Implementation: Strengthening the Science of Science Agenda at NIH 34

Acknowledgements 37

Glossary of Terms 38

Acronyms (Alphabetized) 42

Working Group Charge and Roster 43

Introduction 43

Charge 43

Roster 44

References 47

The editing process used Microsoft Copilot (Microsoft 365 Copilot; version current as of report finalization date). Edits produced by Copilot were limited to language refinement (brevity and clarity). All outputs were manually inspected, validated, and incorporated only when they aligned with the working group's scientific intent. Human reviewers maintained full responsibility for content integrity and compliance with NIH standards.

Executive Summary

In February 2025, the Director of the National Institutes of Health (NIH) and the Director of NIH's Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI) charged the Council of Councils with applying science of science principles to identify priority scientific opportunities that could strengthen NIH's programs, policies, and impacts in the biomedical research enterprise. The field of ***science of science***[1] examines how science itself works through systematic, empirical studies looking at how scientific research is organized, funded, conducted, and evaluated in order to improve its efficiency and impact.[1] Science of science can provide evidence-driven and data-linked insights to understand how NIH investments in research, training, and infrastructure, along with the policies and processes that guide them, could be optimized to lead to better scientific and societal outcomes. To study the effectiveness, transparency, and impact of biomedical research, content experts from academic, private non-commercial, and government agencies across the country were appointed to the Working Group on Science of Science (SoS WG) to advise NIH leadership on ways to leverage advances in science of science. This report presents the SoS WG's recommendations for NIH to consider that would enable data-driven decision-making, rigorous evaluation processes, and innovative policy reforms.

Overarching general recommendation:

The overarching recommendation of the SoS WG is that NIH should increase the use of rigorous empirical methods to study, evaluate, and improve its role as the principal funder of biomedical research in the U.S. To achieve this, NIH should:

a) Develop evaluation-ready policies and programs by clearly articulating objectives and framing them to the extent possible as testable interventions.
b) Strengthen its data infrastructure and increase access to NIH administrative data.
c) Expand its analytical and methodological assessment capacity.
d) Coordinate leadership and governance for science of science research.
e) Further invest in science of science research.

To accomplish the above, the Working Group proposes that NIH adopt a ***causal-chain framework*** to better understand how NIH policies influence scientific outcomes. The objective of this Working Group is to offer actionable recommendations on how to provide empirical evidence with the potential to improve NIH's ability to optimize resource allocation, accelerate innovation, enhance ***replication and reproducibility***, decrease health disparities, and increase societal impact of its investments and practices in biomedical research.

The Working Group created twenty-one recommendations and grouped them into three categories based on levers (tools and policies that directly shape research), system response (how the scientific community reacts) and outcomes (scientific discoveries and health improvements).

Specific recommendations for applying science of science to NIH levers:

1. Systematically study NIH grant mechanisms to inform adjustments that optimize project management and success.

[1] Here and throughout, bold italicized terms indicate that they are defined in the glossary.

2. Evaluate changes intended to strengthen NIH peer review.
3. Optimize NIH funding processes through experimentation, agent-based simulation, and data infrastructure.
4. Conduct science of science research to understand how review and funding processes influence the diversity of the scientific portfolio and transformative innovation.
5. Implement a core reproducibility and generalizability metadata framework.
6. Identify promising technologies and research methodologies to accelerate translation from research to commercialization.
7. Use science of science methods to measure, communicate, and maximize the societal impact of NIH-supported research.
8. Develop data, methods, and the evidence base to improve portfolio diversity and support paradigm-challenging science.
9. Build global science indices to guide strategic NIH investment and benchmark U.S. competitiveness.

Specific recommendations for studying the system response:

10. Invest in datasets from university administrative systems to study scientists' behavior and career progression.
11. Use a human-centered approach to study the evolution of scientific fields.
12. Promote standards for data usage descriptors to support responsible data reuse and reproducibility.
13. Design data trust measures that can be used to support credible and replicable research.

Specific recommendations for studying outcomes:

14. Launch targeted living systematic reviews.
15. Develop a range of performance measures to capture outcomes of research.
16. Develop a range of performance measures to capture outcomes of training programs.
17. Improve and validate grant–patent linkages to better enable assessment of NIH's impact on health products.
18. Validate output measures against outcomes.
19. Support the development of measures for scientific openness, self-correction, and public trust.
20. Study K–12 levers that produce future scientists.
21. Study the effects of major policy changes in NIH funding priorities and procedures.

Introduction

The National Institutes of Health is the principal funder of biomedical research in the United States.[2] Its investments support a broad range of activities, including basic research, scientific training, clinical translation, and the development of new medical products, technologies, and methodologies. NIH-funded research contributes to improved health outcomes, workforce development, and a range of economic activity associated with biomedical innovation.[3] [4]

The Director of the NIH has called for reforms[5] to ensure the agency can fully capitalize on emerging opportunities and address challenges such as the evolving nature of scientific research, increasing globalization and international competition, and the growing influence of artificial intelligence (AI). Advances in data and analytic capabilities now create unprecedented opportunities to strengthen how NIH evaluates, supports, and accelerates science. Leveraging these tools, the development and implementation of strategic reforms can enhance the impact and scientific merit of NIH-funded research, ultimately advancing disease prevention and treatment, improving health outcomes, and extending human life, and ensuring that these outcomes are applicable across populations and real-world settings.

NIH has a long history of working with researchers to generate evidence to inform improvements to its programs, policies, and processes. Through sharing access to NIH administrative data,[6] NIH has supported research on the biomedical workforce,[7] research productivity,[8] patents,[9] and drug approvals,[10] among other findings. These insights have directly informed policy innovations, including funding mechanisms that support research programs rather than individual projects, enhanced investigator stability and flexibility, and illustrated the value of NIH's data-driven approach.[11] Emerging evidence shows that who participates in science, what topics are studied, and how studies are designed work together to shape the direction, impact, and applicability of scientific knowledge.[12] Overall, this highlights the value of data-driven decision-making and supports continuing efforts to improve how NIH-funded science is conducted and advances human health.

NIH has taken important steps to build a robust evidence base to inform its own decision-making. Through the National Institute of General Medical Sciences' (NIGMS) Science of Science Approach to Analyzing and Innovating the Biomedical Research Enterprise (SoS: BIO) Program, NIH actively seeks to support research to understand the biomedical research system.[13] Answering to recommendations from the U.S. Government Accountability Office[14] and the Advisory Committee to the NIH Director to expand access to administrative data,[15] NIH is piloting a Science of Science Scholars Program that pairs researchers with NIH staff on studies to provide in-depth, policy-relevant analyses of NIH processes, programs, and outcomes.[16] Together, these efforts reflect a growing commitment to leveraging data, external expertise, and rigorous analysis to continuously improve the effectiveness and impact of NIH-supported research.

Building on these efforts and broader advances in metascience, NIH is increasingly positioned to deepen its capacity for evidence-informed decision-making. With a mission to advance science through cross-cutting, innovative strategies that foster collaboration and synergy across Institutes and Centers, the Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI) plays a central leadership role in coordinating agency-wide efforts to drive continuous

improvement. For example, DPCPSI has catalyzed engagement with the science of science research community through a range of initiatives designed to expand analytic capacity and bring new methods and perspectives to bear on NIH operations. To further strengthen NIH's ability to make research, resource, and training support decisions grounded in empirical evidence and the best available science, DPCPSI sought guidance from the Council of Councils to identify priority scientific opportunities where advances in the science of science could most effectively enhance the overall impact of NIH's programs and policies.

To that end, the Council of Councils officially established the Working Group on Science of Science at its April 21, 2025, meeting with the following charges:[17]

1. Review the efforts that NIH has undertaken to generate evidence about its programs, policies, and processes.
2. Identify significant gaps in knowledge where leveraging science of science and related research may better influence NIH decision-making.
3. Provide recommendations on potential approaches that NIH could consider to enhance its use of science of science.
4. Provide recommendations for how NIH can engage the science of science research community and communicate results of science of science research.

The Working Group brought together a deliberately multidisciplinary mix of expertise spanning economics and public policy, science-of-science, innovation, and technology-transfer research, open science and research evaluation, basic and experimental biomedical science, and clinical medicine, including infectious disease, neurology, and neuropsychology. Its members also represented universities, nonprofit organizations, philanthropy, and NIH leadership, providing perspectives from both the conduct of science and the funding, evaluation, translation, and administration of research. This breadth of expertise was selected because understanding how NIH investments produce scientific, economic, and health outcomes requires insight into not only biomedical science, but also incentives, workforce behavior, institutional dynamics, innovation pathways, and public value. The first meeting of the Working Group took place on September 4, 2025. A logic model consisting of five components was used to guide the initial deliberations of the Working Group:

1) Inputs: Resources and investments available for programs, projects and initiatives. This includes funding, agency staff, scientific infrastructure, data and information that NIH currently gathers from internal (e.g., portfolio analysis) and external (e.g., grant reviewers) sources for use in decision making.

2) Activities: Actions and processes that implement programmatic goals, including development of research concepts for funding, panel and study group deliberations, funding decisions, policies and oversight governing proposals and awards, scoring procedures and training mechanisms for program managers.

3) Outputs: Products and evidence generated from activities, including scientific publications, patents, preclinical findings, regulatory submissions, prototypes, clinical trial results, technology transfer artifacts, and expanding workforce capabilities.

4) Outcomes: Tangible results, such as acceleration of a scientific field and development of new interventions of outputs from NIH investments.

5) Impacts: How the outcomes changed public health measures such as reductions in mortality and morbidity, and changes in policies, and organizational practices.

Based on the themes laid out by the logic model, Working Group members were then divided into three subgroups:

1) Optimizing the deployment of NIH inputs.
2) Accelerating science through activities resulting in primary and secondary outputs.
3) Capturing outcomes and impacts to measure overall value.

Importantly, each subgroup framed their discussions in the context of priorities laid out by the NIH Director.[18] In addition, each subgroup considered:

1) Identifying the need for data and objective measurements to inform decision-making.
2) Establishing priorities for funding support mechanisms.
3) Accessing opportunities for engaging the science of science community.

Two major documents produced by NIH leadership over the past year provided context that guided WG deliberations. The first, "*Advancing NIH's Mission Through a Unified Strategy*" (August 2025),[19] identified the strategic context within which NIH could move to "fund the most meritorious science, address urgent health needs, and sustain a robust biomedical research workforce." The report identifies strategic priorities, including chronic health issues affecting Americans, chronic childhood diseases and nutrition, as well as next-generation tools such as AI, alternative (human biology-based) testing models, and real-world data platforms. The second, "*Leading in Gold Standard Science: An NIH Implementation Plan"* (August 2025),[20] identified nine principles of what has been termed gold-standard science. The approach developed in the Working Group's report aimed to align with these principles while extending them analytically. The Working Group focused on how each principle can be translated into measurable components of NIH policy and practice, and how evidence can be generated to assess whether these principles are being realized in practice.

The findings and recommendations to follow present the Working Group recommendations emerging from its discussions, beginning with a general recommendation that NIH adopt a process of continuous learning as a core operational principle to inform and improve NIH decision-making, followed by specific recommendations regarding what should be studied, what data will be needed, and the science of science approaches that could be employed.

Working Group Findings

The Working Group noted that NIH has made significant progress to date in developing the data and analytic infrastructure needed to support a science of science approach to improving the biomedical research enterprise. Building on longstanding investments in data systems and transparency, NIH has increasingly enabled both internal and external researchers to examine how research funding translates into scientific, workforce, and societal outcomes.

NIH has expanded access to administrative and portfolio data through platforms such as the NIH Research Portfolio Online Reporting Tools[21] and through controlled access to internal data resources. These efforts have enabled analyses of research productivity, workforce diversity, patenting, and drug development, contributing to a growing body of evidence on the scientific, economic, and public health impacts of NIH funding.[22] [23] [24] [25] [26]

NIH has largely relied on available conventional (e.g., bibliometric) measures to assess the impact of its research on advancing science and improving health. Research has been evaluated using a relatively limited set of measures of quality and productivity. Often the number of citations that articles receive from subsequent articles is used, which is known to be an inadequate measure of impact.[27] NIH has invested in efforts to overcome some of the challenges of counting the number of citations by developing more robust measures like the Relative Citation Ratio, which adjusts for differences across fields.[28] A harder challenge to overcome is that many of the outcomes readily captured focus on impact on the scientific literature and not on those outcomes related to translational impact and public health advancement, which are of greater interest to the public and better convey the value of NIH investments.

Fortunately, researchers in science of science policy[29] have developed a wider range of outcome measures that NIH can draw on to evaluate its work. These include richer measures based on citations, including the breadth of impact of research[30] and disruptiveness;[31] [32] measures based on text analysis;[33] [34] measures based on workforce development;[35] [36] [37] economic activity[38] [39] and, measures of impact on technology.[40] [41] [42] [43] NIH has also developed well-researched case studies[44] that illustrate the value of NIH investments in specific areas. Such work is informative for providing relatable narratives to convey the value of NIH activities. Outside researchers have used NIH data to conduct quantitative evaluations that link NIH investments to downstream outcomes, including patenting;[45] [46] health;[47] and drug development.[48]

NIH has also invested in data infrastructure to support reproducibility and transparency, including policies that promote data sharing and improved documentation of research outputs.[49] Emerging approaches, such as linking datasets to publications and tracking patterns of data use, are beginning to provide insight into how scientific data are generated, shared, and reused across the research ecosystem.[50]

Together, these and other efforts reflect a growing foundation for science of science research at NIH and support the adoption of science of science principles to inform NIH research and investment decisions. There is also growing interest across NIH Institutes, Centers, and Offices (ICOs) in strengthening the biomedical research ecosystem by supporting science of science activities that address issues such as sustaining a robust research workforce, expanding research capacity nationwide, fostering innovation, and improving rigor and reproducibility.[51] While the infrastructure continues to evolve, the evidence-based approach has been tested and provides a strong basis for expanding NIH's capacity to generate, integrate, and apply evidence to improve its programs, policies, and outcomes.

A Causal-Chain Framework

In considering a science of science approach to better position NIH's role in the biomedical research ecosystem, the Working Group formulated a ***causal-chain framework***. This framework

views NIH as an active participant in a dynamic research system, in which policy and process choices influence the behavior of scientists, institutions, and research communities to work together to ultimately determine the pace, direction, and impact of scientific progress. This perspective expands the current focus to embrace a broader, evidence-based understanding of how and why NIH actions produce particular outcomes and impacts. This framework expands the full chain linking NIH investments in research, resources, and training to downstream outcomes and impacts, including basic science discoveries, workforce and economic development, translation into clinical practice, and improvements in human health.

The framework includes three key elements: levers, system responses, and outcomes (see Figure 1).

1. **Levers — What NIH Controls.**

Levers are the instruments directly under NIH's authority: statements of research priorities, funding mechanisms, administrative policies and infrastructure, and research resources. Science of science treats each lever as a testable intervention with a specified ***theory of change***. This is the entry point of the causal chain. Because NIH controls the science funding levers directly,[52] evidence from science of science analyses and assessments about their effects and interactions can translate immediately into action. This allows the identification of how any given policy, process, or structure is intended to achieve its goals and what improvement might be possible. Examples of NIH's processes, structures, and policies that could be considered include the scale and duration of funding mechanisms; design, administration, and criteria of the peer review process; requirements for relevance and generalizability across populations; the reporting of rigor, reproducibility, and transparency; and organizational and digital infrastructure.

2. **How the System Responds.**

When NIH adjusts one of its levers, the scientific community responds. Responses include applications, formation of collaborative teams or consortia, the expansion or contraction of specific scientific fields, distribution of research topics, institutional responses, scholarly publications, patent applications, development of new technologies, expansion of training opportunities, among others.

Science can be analyzed as labor performed under incentives and constraints. Incentives are motivators—monetary or non-monetary rewards, benefits, or costs—that encourage individuals or organizations to change behavior, work harder, or take specific actions. Constraints are limitations, restrictions, or controlling factors that define the boundaries of what is possible, restricting actions, behavior, or choices. Scientists can change career choices and embark on new research fields, teams and collaborations can form and dissolve, and the institutional dynamics by which knowledge is produced and transmitted can change. Of course, there are confounding factors that the field of science of science can consider to guide decisions that increase the likelihood of achieving the desired result.

Studying the system response can help NIH understand how the manipulation of levers may or may not produce the intended effects and to inform adjustments in the levers accordingly. This requires collecting data and developing empirical measures to better understand scientists' choice of fields, careers, and topic selection, of the structure of scientific teams, and the

organization and responses of scientific institutions. The impact of changes in scientific workforce incentives and constraints could then be assessed in terms of their impact on scientific discovery, workforce outcomes, translation and innovation, and health outcomes.

3. Outcomes — Evaluating Effects of the Response.

The outcomes of responses to levers are the scientific discoveries and the pace at which they occur; their relevance to populations, including those disproportionately affected by poor health and increased mortality; robustness, performance, and composition of the biomedical workforce; the speed and effectiveness of translating research into clinical practice and commercial products; reduction in health disparities and improved health outcomes, and other responses that ultimately impact human health and longevity.

Critically, outcomes should not be assessed in isolation but whenever possible relative to the ***counterfactual*** use of resources. For example, a finding that NIH funding produced "X" number of publications is important but insufficient alone in informing the need for future NIH investment. Instead, knowing what outcomes the same dollars would have produced under a different allocation would provide additional, useful information to guide future decision-making.

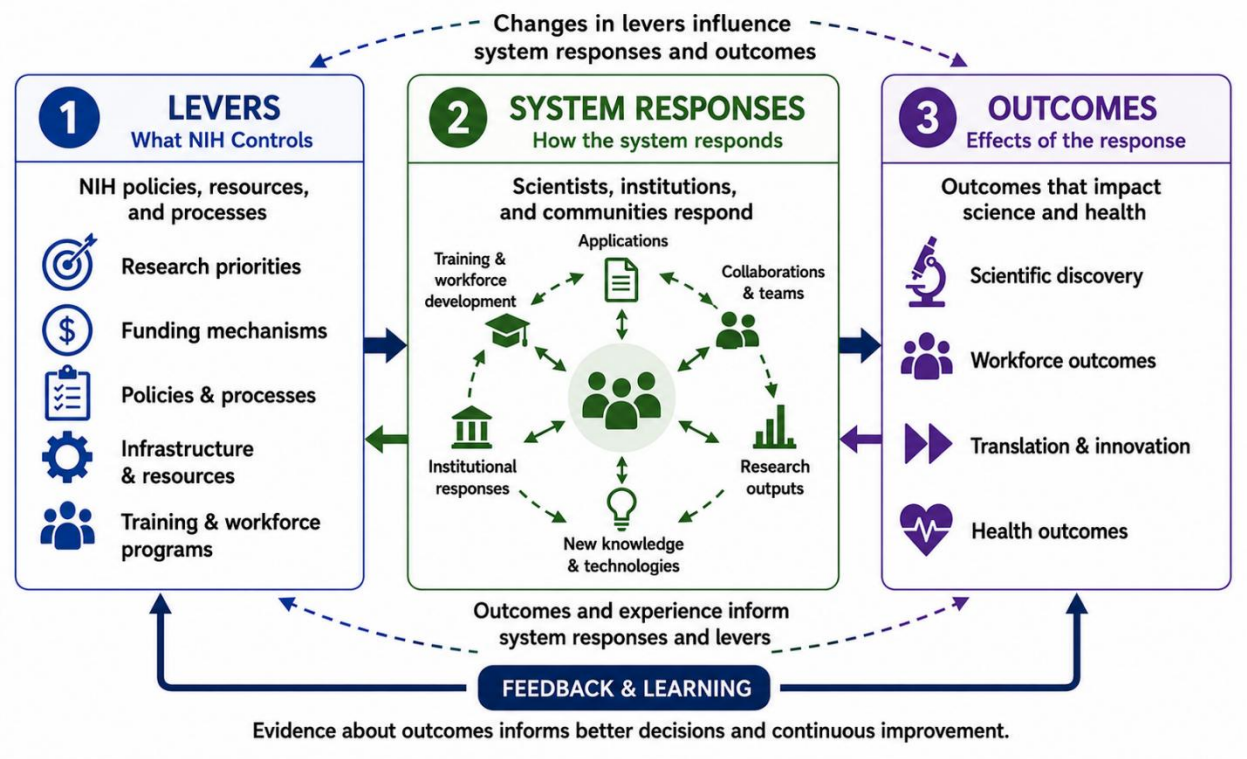


Figure 1. NIH causal-chain framework.
NIH levers shape research-system responses, which lead to outcomes in discovery, workforce development, translation, and health impact. Evidence (e.g., observations, measured data, etc.) from these outcomes feeds back to improve how NIH uses its levers.
This figure was made with GPT-5.5 and reviewed by the SOS WG for accuracy and alignment with WG findings and recommendations.

Importantly, these components are not independent stages but parts of an integrated system. Changes in one element can propagate throughout the system in complex and sometimes unintended ways. For example, modifications to peer review criteria (levers) will alter investigator behavior and reshape collaboration patterns (system responses), that ultimately influences the types of discoveries that emerge (outcomes), which in turn accumulate experiences and insights that cyclically inform levers and system responses. Understanding these interdependencies is essential for designing policies that improve overall system performance rather than optimizing individual components in isolation. This is precisely what science of science seeks to accomplish.

By organizing the analysis of NIH investments and activities within this causal-chain framework, science of science provides a systematic way to assess NIH policies as testable interventions within a complex system. This approach enables NIH to move beyond intuition and retrospective assessment of traditional metrics and towards a more prospective, outcomes-based model of decision-making. This outcomes-based approach could then inform iterative improvements across the biomedical research enterprise.

The Working Group used this approach to identify gaps in current processes and clarify where evidence is most needed. Improvements in outcomes, such as faster translation or a more robust workforce, depend not only on NIH investments and resources, but on how use of those investments and resources shape behavior of the scientific community within the research system. As a result, evaluating NIH's effectiveness requires understanding both the direct effects of policy and operational choices and the system responses through which those effects are realized. The sections that follow apply this framework to examine and recommend how NIH can more effectively deploy its levers, understand system responses, and evaluate outcomes. In doing so, this report provides a roadmap for how NIH can take a more systematic, outcomes-based approach to improving decision-making processes to shape the biomedical research enterprise.

General recommendation: NIH should make greater use of rigorous empirical methods to study, evaluate, and improve its research funding policies, processes, and outcomes.

A central contribution of the science of science is the recognition that research funding systems can themselves be studied, evaluated, and improved using rigorous empirical methods. To fully realize this potential, NIH should consider itself to be not only a funder of science but a continuous learning organization that systematically evaluates evidence as to how its decisions shape scientific behavior and downstream outcomes, constantly seeks to improve its processes, and uses evidence-based outcomes to make appropriate adjustments to informed decision-making.

Many U.S. science agencies implement policies and reforms through iterative inputs such as informed judgment, stakeholder engagement, and lessons from prior experience.[53] While these are valuable inputs, they are often episodic, retrospective, and insufficient on their own for optimizing a system as complex and consequential as the U.S. biomedical research enterprise.[54] Embedding continuous learning requires a shift from relying primarily on periodic feedback and accumulated experience to explicitly studying and where possible testing whether policy and process changes achieve their intended effects. The resulting evidence can then be used not only

to inform decision-making but to create an ongoing cycle of measurement, evaluation, adaptation, and refinement.

NIH has undertaken multiple efforts that reflect elements of this approach, but these efforts have not yet been organized in a systematic or comprehensive way. Implementing a more structured science of science approach at NIH would build on existing strengths while requiring further development of several complementary capabilities. Taken together, these capabilities would support a more continuous systematic approach to understanding how NIH policies and actions influence scientific behavior and outcomes, and how those policies might be refined over time.

The Working Group elaborated on its general recommendation by addressing A) study topics, B) data needs, C) methods, D) governance, and E) funding, with the following sub-recommendations:

A. NIH should develop evaluation-ready policies and programs by clearly articulating objectives and framing them to the extent possible as testable interventions.

Building on current practices, there may be opportunities to regularly incorporate evaluation plans that clearly specify intended objectives, key outcome measures, data sources, and timelines for assessment into policies and actions undertaken by NIH. For example, funding mechanisms, peer review criteria, and reporting requirements should be explicitly conceptualized as measurable interventions within a complex system.

Each policy and action should be associated with a theory of change that specifies how it is expected to influence scientific behavior and outcomes. This framing enables systematic assessment, and where possible comparison of alternative designs, allowing for accumulation of institutional knowledge over time. By defining success criteria in advance of actual implementation of a policy or action, NIH could ensure that reforms and actions are evaluated more consistently and transparently, rather than retrospectively or selectively. If possible, evaluation should be grounded in credible counterfactuals that allow an assessment of what would have happened in the absence of the intervention or under an alternative design. Wherever feasible, and recognizing that it might not always be possible, NIH should incorporate randomized trials, phased rollouts, or ***quasi-experimental designs*** that allow for causal inference.

B. To address data needs, NIH should strengthen its data infrastructure and increase access to administrative data resources.

The impacts of NIH funding and policy changes often unfold over extended time horizons. Accordingly, a data infrastructure that links proposal, award, and outcome data including scientific productivity, scientific workforce composition and training trajectories, translation into clinical practice, and broader societal impacts on health will allow NIH to assess not only short-term outputs, but long-term returns on investment. By creating new mechanisms while maintaining privacy and protecting intellectual property, NIH could allow qualified extramural researchers access to administrative data using principles of science of science, where expertise could be further expanded and new approaches towards data linkages exploited. Linkages to de-identified ***person-centered data*** (i.e., capturing workforce dynamics and demographic composition, team formation, and career trajectories) offer opportunities to better understand how NIH actions shape the research enterprise over time. Advances in longitudinal data systems and the use of standardized ***metadata*** related to research design, data use, and reproducibility

could further strengthen this foundation. Improving access to NIH administrative data is in keeping with the *Evidence-Based Policymaking Act of 2018.*

While there is limited funding for science of science research, some opportunities (e.g., the National Science Foundation's (NSF) Science of Science: Discovery, Communication and Impact,[55] NIGMS-funded SoS: BIO program,[56] Sloan Foundation,[57] and the Abdul Latif Jameel Poverty Action Lab Science for Progress)[58] exist where consistent pathways to access data can be created. As the science of science community coalesces,[59] it has produced a rich body of findings, considerable data, methodologically rigorous study designs, and a core of experts who are eager to use data, particularly pre-decisional data, to generate insights that might improve NIH performance. NIH should consider implementing a multi-pronged strategy to harness this interest and expertise to provide transparency and improve NIH activities. For example, the Science of Science Scholars Pilot Program provides an opportunity for experienced science of science researchers to access NIH internal administrative data to conduct in-depth analyses.[60] To scale and further improve access to data, NIH should consider conducting studies to identify data (e.g., administrative, peer review, and research data) that can provide evidence for relevant science of science questions.

The conditions required for accessing the various data types of public or restricted data should be explicitly defined and a publicly accessible catalog of metadata created. The catalog can also detail conditions for conducting analyses, reporting outcomes, and specifying whether findings should be kept internal for use within NIH or reported publicly. To ensure NIH's conditions for accessing data are successfully met, NIH should also consider implementing and/or augmenting current application processes to access data. If enacted, such processes should respect the sensitivity of NIH's administrative data and ensure researchers can conduct studies.

C. NIH should expand its analytical and methodological assessment capacity.

Those studying NIH both internally and externally have employed approaches and methodologies such as traditional evaluation and bibliometrics to assess its programs, policies, and operations. NIH could expand on these methods using experimental and quasi-experimental approaches where appropriate and feasible to further strengthen its assessments. In addition, scientific workforce metrics could be developed to better capture translation, workforce size, training pathways, development of expertise, and broader impacts. Methodological refinement could be achieved through a coordinated effort combining internal analytic capacity and partnerships with the broader science of science research community.

D. NIH should identify processes to better coordinate leadership and governance for science of science research

Central to the effort to better systematize NIH's ability to study itself is the development of coordinated leadership and governance structures to support these activities. Establishing a focal point for analytic coordination[61] could help prioritize evaluation efforts, promote consistency in methods, strengthen collaborations with the external science of science community, and facilitate the integration of evidence into decision-making across ICOs.

For the principles of science of science to be integrated systemwide into policies, guidance, and operations across the NIH, the Working Group recommends that science of science activities be coordinated by the newly proposed NIH Office of Research Economics, Planning, and Analysis (OREPA) within DPCPSI in the Office of the Director. This placement enables strategic coordination of science of science activities across all NIH ICOs and facilitates priority setting, planning and funding for multi-ICO initiatives. A centralized office could help ensure that analytic insights and assessments are fully integrated into policy and decision-making processes by NIH leadership. This function could also support the development of evaluation standards, oversee investments in data and analytic infrastructure, and facilitate collaboration with external researchers and organizations engaged in science of science research. This office could also establish safeguards and analytic standards that promote the independence, rigor, reproducibility, and objectiveness of research findings used to inform NIH policies and decisions. With funding within the Office of the Director, OREPA could catalyze efforts across NIH to further develop science of science methodologies, enhance and increase availability of data, establish and refine policies and procedures to embed science of science principles in NIH operations, and design new strategies for funding science of science activities both internally and by the extramural community. As with other DPCPSI offices, the Council of Councils should play an important advisory role.

E. NIH should make greater use of science of science research

Implementing the Working Group's recommendations on science of science will require deliberate and sustained investment through funding mechanisms that are both flexible enough to support novel initiatives and durable enough to outlast any single budget cycle. The NIH Director's Office is uniquely positioned to lead this effort, drawing on existing NIH-wide funding instruments that operate above the level of individual ICOs. Chief among these is the Common Fund, which was established by Congress precisely to support high-priority, cross-cutting research that no single ICO would or could champion alone. Common Fund programs are time-limited by design. Typically five to ten years in duration, Common Fund programs are intended to catalyze fields and establish infrastructure before transitioning sustained activities elsewhere. This makes the Common Fund an ideal vehicle for launching new priority areas during their formative phase, while longer-term stewardship structures are developed. From the moment any such program is launched, however, evaluation must be built in as a core structural requirement, not as an afterthought. The evaluation criteria should include clearly defined metrics, timelines, and accountability mechanisms established at the outset and revisited at regular intervals throughout the program's life. Beyond the Common Fund, the NIH Director's Office has access to several other trans-agency mechanisms that could be considered.

For initiatives that involve extramural grant awards, dedicated Notice of Funding Opportunities should be issued under trans-NIH authority, using established award mechanisms based on the scale and collaborative nature of the work. Cooperative agreements in particular (the U-series mechanisms) offer NIH program staff a more active role in shaping and coordinating funded projects, which is appropriate when the goal is to build a coherent, cumulative body of evidence rather than simply funding independent investigators working in parallel. NIH should also be open to contracts to support long-term analyses and products that have direct implications for decision-making.

The Value of a Science of Science Approach to Decision-Making

Strengthening NIH's approach to science of science by employing a causal-chain framework and systematizing and expanding analysis and evaluation to better inform decision-making will contribute to advancing NIH Leadership priorities in numerous ways. Following are examples where the Working Group identified as ways in which a science of science approach could assist leadership decision-making.

Portfolio optimization and resource allocation. A science of science approach can provide evidence on whether NIH's current allocation matches its stated priorities in specific areas of science, and the potential benefit or loss to health and longevity. Simply put, it is a question about whether the money is going to the right projects at the right times by considering portfolio composition and stated research, resource, and training priorities. This also pertains to supporting a diversity of scientific and capacity building approaches, applicability across populations, and effectiveness in real world settings.

Workforce pipeline and training efficiency. Science of science data and methodology can provide actionable information to NIH leadership regarding the extent to which NIH funding mechanisms and program opportunities in support of scientific training is achieving the goal of invigorating a robust and highly skilled biomedical workforce across educational, scientific, and demographic backgrounds and career stages.

Speed of translation. How long does it take for an NIH-funded discovery to translate into improvements in patient care? Science of science can map the Levers-to-Response-to-Outcomes chain specifically for analyzing and assessing the effectiveness of clinical translation, identifying bottlenecks, hurdles, and hinderances and suggesting ways to address and mitigate them. It can also assist in identifying promising innovations. Understanding which parts of the chain are bottlenecks and what adjustments to those levers are most likely to address it, requires the kind of empirical tracing the causal-chain framework is designed to support.

Optimizing funding decisions. Science funding requires decisions under substantial uncertainty. NIH has to choose among many meritorious proposals whose outcomes may be difficult to predict, emerge only over long-time horizons, and may produce benefits that are simultaneously scientific, clinical, societal, and economic. Although this resembles financial portfolio decision-making, the analogy is necessarily limited. Scientific outcomes are not readily expressed through a common price or return metric, and the value of one project may depend on the knowledge, methods, or infrastructure generated by others.[62]

Decision science can provide useful tools for making funding choices more systematic and transparent. Relevant considerations may include the potential scientific or public-health value of a project, the uncertainty surrounding its likely outcomes, its relationship to other funded work, and its contribution to the overall balance of the research portfolio. These factors should not be reduced to a single formula or used to rank proposals mechanically. Rather, they can supplement expert judgment by helping NIH evaluate tradeoffs among near- and long-term goals, incremental and transformative research, complementary and overlapping projects, and investments whose broader value may extend beyond their immediate results. NIH's current two-stage peer review system partly follows this logic: study sections assess scientific validity and value and likelihood for success while advisory councils review for mission relevance and

programmatic priorities. NIH should consider pilot studies that apply this decision-making framework by testing different approaches towards assessing risk, return, and diversification.

Like all NIH investments, science of science activities are subject to opportunity costs. Every dollar devoted to evaluation, data infrastructure, or methodological research is a dollar that is not available for direct support of biomedical research, training, or research resources. Accordingly, investments in science of science should themselves be viewed as testable interventions and evaluated against appropriate counterfactuals. The objective is to identify the level and types of investment that generate improvements in decision-making sufficient to outweigh the resources they consume.

Limitations to a Science of Science Approach to Decision-Making

While these approaches offer a promising path for strengthening NIH's evidence-driven decision-making, several limitations must be acknowledged. Many recommendations depend on integrating administrative, bibliometric, clinical, and workforce datasets that remain fragmented, inconsistently structured, or incomplete. For example, linkages between NIH grants and downstream patents or products are hampered by underreporting in patents and the lack of publicly available patent-product linkages for biologics, diagnostics, and other health technologies. Similarly, constructing longitudinal, person-centered data requires merging university human resource sponsored-project, procurement, publication, and patent records, yet such data are currently available only through specific initiatives and are not uniform across institutions. Efforts that rely on text analysis and AI also confront constraints: metadata describing study design, reproducibility, data provenance, and data usage are not standardized across NIH-funded research, limiting the reliability of automated classification or evidence-synthesis tools. Finally, outcomes such as clinical translation or population health impact require long-horizon data that are difficult to assemble and validate, complicating attempts to generate credible counterfactuals or evaluate policies in a timely manner. Together, these limitations indicate that many proposed innovations will require substantial data infrastructure development before they can be fully realized. The Working Group recommends that NIH undertake an analysis of the needed infrastructure to overcome gaps and deficiencies that until addressed will limit its ability to fully apply science of science approaches to decision-making.

Specific Recommendations

Building on its recommendation for NIH to study its processes and activities, the Working Group developed specific recommendations covering support mechanisms, priority study areas, data needs, and methods across the framework's stages: levers, system responses, and outcomes. Throughout, the recommendations are presented as considerations for exploration, with the goal of informing a broader, evolving science of science research and policy agenda.

Applying the Science of Science to NIH Levers

Levers are the tools, resources, and mechanisms directly controlled by NIH, such as funding priorities, grant mechanisms, peer review criteria, research resources and infrastructure, and others. They represent the entry points for policy interventions that can shape the direction and

outcomes of biomedical research. The following recommendations suggest specific topic areas for study (grant mechanisms, peer review), data needs, and approaches.

Recommendation 1: Systematically study NIH grant mechanisms to inform adjustments that optimize project management and success.

NIH's extramural program includes 91 distinct grant mechanisms that differ in a variety of ways, such as grant type, eligibility, or institution type (e.g., small business). The range of these mechanisms are difficult for researchers to understand and for program officers to manage. NIH should consider using science of science methodology to reorganize mechanisms around the scientific goals they are intended to address. This would change investigator behavior, reshape collaboration patterns, impact institutions, and ultimately influence the types of discoveries that emerge. In addition, NIH should study whether new, targeted mechanisms are needed where no such opportunities currently exist. These might include support for long-term research, investigator mobility, renewable people-centered funding, and replication studies. Finally, NIH should analyze peer review criteria to better and more explicitly align with each mechanism's stated purpose, and to develop program-level evaluation metrics to assess how effectively each mechanism addresses its intended goal.

Recommendation 2: Evaluate changes intended to strengthen NIH peer review.

Recent changes to NIH's peer review process[63] have fully separated scientific evaluation from funding decisions: the Center for Scientific Review now manages peer review while the ICOs make funding decisions based on a combination of review scores and programmatic priorities. This separation was intended to promote impartiality and transparency, but it may be vulnerable to weaknesses introduced through dominant frameworks defining what counts as valid science, disadvantaging novel approaches, and narrowing of expertise as study section boundaries shape who reviews whom. Major reforms, including the 2008 Enhancing Peer Review initiative and 2025 changes instituted by NIH (i.e., Simplified Review Framework and Implementing a Unified NIH Funding Strategy to Guide Consistent and Clearer Award Decisions), have been adopted with limited evaluation or randomized evaluation. Other changes include those to promote early career researchers and spread resources more widely. These include promoting the success of Early Stage Investigators through preferential funding policies and career development mechanisms such as K awards, the NIH Director's Early Independence Award, the NIH Director's New Innovator Award, and the NIGMS Maximizing Investigators' Research Award for new and early-stage investigators, all of which are designed to accelerate investigator independence and improve funding stability.

Adjustments that intend to minimize and impose limits on the important advisory role of peer-review have been proposed[64] and will undoubtedly be proposed in the future. These deserve further study to provide evidence of their effectiveness and to guide future decision-making. For example, a large study[65] from 2015 that examined the ability of peer review to select the highest impact science that will proceed to improve human health showed that peer review is the single most important predictor in identifying transformative science, science with applied potential, and to screen out low-quality research. In light of continued changes to peer review processes, studies such as this should be repeated.

The Working Group recommends NIH study proposed policy changes. For instance, the working group recommends using existing administrative data on reviewer scores and outcomes to

simulate the effects of the proposed changes in peer review to evaluate the likely impact on funding decisions. Simulations could also be performed to evaluate other possible changes in scoring. Work could simulate the effects of simplified review on the tradeoff between importance and innovation and between rigor and feasibility. Other work should evaluate the split in authority between the Center for Scientific Review and ICOs. Work could also evaluate the effects of historic changes in peer review (e.g., to promote early career researchers).

Another way a robust science of science infrastructure could contribute to peer review prospectively is to enable systematic assessment of effectiveness by experimenting with alternative approaches prior to making changes. While international funders have piloted alternatives (e.g., lotteries, golden tickets, multi-mechanism models), none have tested them against traditional approaches in a controlled setting. The result is that institutional learning is limited. NIH could commit to gathering evidence and testing alternative funding strategies before fully implementing them. This could be operationalized through an annual innovation challenge prize competition to design, prioritize, and evaluate experiments in peer review and grant allocation. To test the impact of alternative funding strategies on transformative science and science with applied potential, NIH should consider funding projects that apply science of science methodology to test alternative approaches to assess peer review and grant allocation. Additionally, funded projects could build longitudinal data infrastructure linking proposals, review processes, funding decisions, and long-term outcomes to assess how outcomes of grants and grant portfolios could be improved and to facilitate randomized or quasi-experimental designs.

Consideration should also be given for providing secure access to this longitudinal data infrastructure for approved extramural researchers and requiring an evaluation plan with defined counterfactuals, outcome measures, and learning timelines for proposed policy and process changes to peer review.

Recommendation 3: Optimize NIH funding processes through experimentation, agent-based simulation, and data infrastructure.

NIH's funding processes have evolved largely without systematic experimentation, limiting the ability to identify which policies and practices most effectively support or hinder high-impact science. The effect of key design choices, such as proposal requirements, award mechanisms (e.g. grants, contracts, prizes, etc.), funding allocation rules, review criteria, page limits, and others, are rarely tested in a rigorous, comparative way. At the same time, growing administrative burden and mixed evidence on funding concentration, investigator incentives, and targeted funding opportunities offer opportunities for analysis and assessment.

NIH should pilot means to manage its funding processes as an object of continuous learning through experiments. A structured program of randomized and quasi-experimental studies would allow NIH to evaluate alternative approaches to application requirements, review processes, and funding allocations. In parallel, technological advances make it possible to simulate peer review using "***in silico***" committees trained on historical data, enabling low-risk testing of alternative policies at scale before real-world implementation.

NIH should consider the following:

- Testing alternative pathways in funding processes as a core NIH practice. Priority areas should include introducing pre-proposal stages, varying reliance on Notice of Funding

Opportunities and program announcements, testing allocation rules (e.g., lotteries among meritorious proposals or ***out-of-order funding***), and experimenting with award size, duration (e.g., multi-year funding), and structure. Additional areas to consider for study include funding of investigators versus projects, adjusting limits on total investigator funding, reducing soft-money dependence, modifying budget reporting, and evaluating tradeoffs between administrative burden and research quality.

- Piloting bold alternatives to current systems, including simplified or randomized funding allocation among proposals meeting baseline standards, to assess effects on return on investment.
- Using existing data on scores and outcomes to simulate the effects of the recent changes in peer review to evaluate the likely impact on funding decisions. Simulations could also be performed for other possible changes in scoring. This work should include changes to scoring to prioritize disease relevance, scientific rigor and feasibility.
- Developing and evaluating the effectiveness of ***in silico review systems*** (i.e., AI-based multi agent systems to model applications and peer review committees trained on historical grant data), to replicate and study reviewer biases and scores. Once validated, these systems could test alternative allocation mechanisms (e.g., lotteries, modified scoring, different panel compositions) rapidly and without risk to applicants.
- Providing access to data infrastructure linking applications, review outcomes, and funding decisions enabling both simulation and external research.
- Pilot and study approaches to maximize dissemination and transparency of research results.

Recommendation 4: Conduct science of science research to understand how review and funding processes influence the diversity of the scientific portfolio and transformative innovation.

Peer review systems are generally effective at identifying technically strong science but are less reliable at distinguishing incremental advances from transformative breakthroughs.[66] [67] [68] Dense intellectual networks characterized by repeated co-authorships, high citation clustering, and institutional concentration can constrain diversity of thought and reinforce dominant paradigms. They may also focus attention on approaches that fall neatly within disciplinary lines.

Epistemic narrowing occurs when a scientific field or review panel becomes intellectually dominated by a single framework, set of hypotheses, or methodology, crowding out alternative approaches. Emerging bibliometric tools now allow measurement of "disruptiveness," conceptual novelty, inter- / trans- / multi-disciplinarity, and long-horizon citation impact, providing feasible mechanisms for evaluating whether review processes preferentially fund incremental versus field-redefining work and how each contributes to improvements in health.[69] [70] [71] [72] [73] [74]

NIH has implemented safeguards including rotating membership on study sections, conflict-of-interest rules, and standardized scoring criteria. Special emphasis panels and *ad hoc* reviewers are used to target reviewers with the expertise needed to assess interdisciplinary, novel, or niche proposals. However, NIH does not currently appear to employ systematic, quantitative approaches to characterize the range of scientific perspectives within study sections.

The Working Group did not identify an existing formal mechanism to assess patterns such as the concentration of approaches, the breadth of methodological or theoretical representation, or how these features evolve over time. NIH should consider science of science approaches to better understand these dynamics and their relationship to scientific outcomes. Potential areas for examination include:

- Funding patterns within study sections, to characterize the distribution of supported hypotheses, models, methods, and disciplines to better understand how these reflect the scope of applications received. Such analyses could help contextualize how study sections engage with established and emerging areas of research.
- Collaboration and citation patterns among reviewers and funded investigators, as one of several ways to describe the structure of scientific communities represented in review.
- Longitudinal examination of research outputs within study section portfolios, including downstream translational activity where measurable, to better understand how funding patterns relate to trajectories of scientific progress.
- Characterization of scoring patterns and their association with subsequent outcomes, including variability in scores across criteria (e.g., innovation) and their relationship to longer-term scientific impact. These analyses could help inform understanding of how review criteria are applied across different types of science.

Recommendation 5: Implement a core reproducibility and generalizability metadata framework.

Key information on study design, analytic transparency, population characteristics, and replication status is not consistently reported or easily aggregated across NIH-funded research. This deficiency limits the ability to assess research robustness, detect redundancy, and track progress toward reliable knowledge. At the same time, evidence across multiple fields shows that repeatability is often difficult to establish and that uncertainty in the literature may be underappreciated.[75] [76] [77] [78] [79] [80]

Programs like the Defense Advanced Research Projects Agency's Systematizing Confidence in Open Research and Evidence program[81] demonstrate the value of combining AI and human evaluation to systematically assess research credibility and generate scalable indicators of reliability.[82] NIH has prioritized reproducibility, transparency, and rigor through initiatives aligned with its "Gold Standard Science" framework.[83] However, NIH lacks a comprehensive, enterprise-wide metadata system to capture reproducibility and ***generalizability*** across funded research. Existing efforts highlight potential value of such data.[84] [85] [86] [87] [88] [89] A minimal, interoperable core metadata set that captures study design (e.g., recruitment, randomization, blinding, power justification, subgroup analytic strategies), transparency indicators (e.g., ***preregistration***, data and code availability), population characteristics, and replication status could enable cross-project comparison, portfolio-level analysis, and machine-assisted evidence synthesis, consistent with ***Findable, Accessible, Interoperable, and Reusable (FAIR)*** data principles.[90] [91]

NIH should consider assessing the feasibility and value of extending approaches similar to the Systematizing Confidence in Open Research and Evidence program[92] to biomedical research, focusing on systematic evaluation of repeatability across key stages of the research pipeline and development of scalable indicators of credibility. A potential outcome would be implementation

of a core metadata framework integrated with existing systems and designed to minimize burden on investigators. Standardizing and aggregating these data would improve monitoring of research quality and the accumulation of reliable knowledge, supporting measures such as replication, methodological diversity, and time to high-confidence findings.

Recommendation 6: Identify promising technologies and research methodologies to accelerate translation from research to commercialization.

Commercialization is an important pathway for translation,[93] [94] [95] but valuable research findings are often lost between the bench and the bedside because of a translation gap or valley of death between government funding and industry support.[96] [97] [98] [99] [100] NIH has long recognized the importance of translation, creating the National Center for Advancing Translational Sciences to rigorously engineer improvements in translation.[101] And despite efforts such as the Small Business Innovation Research and the Small Business Technology Transfer programs at NIH, funding gaps persist. Part of the problem is that NIH has limited ability to systematically identify and track potentially valuable technologies, which means promising discoveries are not always identified and brought to market. This makes it difficult to analyze and engineer improvements to the translation process itself.

That same visibility problem extends further upstream, to the scientific methods and tools that make discoveries possible in the first place. Many biomedical advances depend not only on new hypotheses, targets, or disease areas, but on new ways of observing, measuring, manipulating, modeling, and experimenting. Science of science research suggests that scientific methods and instruments extend researchers' cognitive and sensory capacities.[102] Major discoveries often follow from new methods and tools that make previously inaccessible phenomena measurable or experimentally tractable.[103] New scientific fields are often triggered by powerful new methods.[104] [105] Yet NIH has little systematic way to track which methods and tools are in use, where they are creating bottlenecks, or where a new tool could unlock many downstream discoveries. This is a blind spot that mirrors, and compounds, the lack of visibility into which discoveries are ready to translate.

NIH should consider developing and refining mechanisms to identify promising technologies and the researchers conducting them to support translation. NIH should also consider developing tools to track the methods, instruments, software, data resources, protocols, models, assays, and computational workflows that make new biomedical knowledge possible to identify methodological bottlenecks and technologies that could unlock many downstream discoveries. Drawing on portfolio analysis, expert input, and AI-enabled text and network analysis, NIH could build a shared evidence base spanning both promising technologies and the methods that enable them, and use it to analyze and engineer improvements across the translation pipeline.

Recommendation 7: Use science of science methods to measure, communicate, and maximize the societal impact of NIH-supported research.

Every year, NIH invests billions of dollars in biomedical research,[106] yet links between funded science and downstream societal outcomes remain poorly understood. Evaluation relies largely on bibliometric indicators (e.g., publications, citations, and patents), which reflect scientific influence but not whether research improves health, changes clinical practice, or reduces disease burden. At the same time, many unfunded proposals might contain ideas with substantial

unrealized societal value. These gaps reflect the inherent difficulty of predicting, at the time of funding, which investments will ultimately matter most.

Advances in AI agents create new opportunities to study these pathways at scale to better assess and maximize societal returns. AI agents can now integrate and reason across diverse data sources, including grant applications, peer review, investigator histories, publications, patents, clinical trials, and health outcomes, at unprecedented scale. Such analyses would spark new questions: Where has NIH funding produced outsized societal returns? Where has strong scientific output failed to translate into meaningful downstream benefit? Which promising directions were overlooked? Importantly, NIH can learn from funded and unfunded applications, helping to identify systematic blind spots such as promising topics, unconventional approaches, overlooked combinations of expertise, and areas where impact may be left unrealized. These tools could augment, not replace, expert judgment and be evaluated for validity, robustness, and fairness.

NIH should consider launching a science of science initiative focused on societal-impact discovery and portfolio learning. The initiative could have two closely connected components that would enable NIH to learn from its full set of opportunities and use that knowledge to direct resources toward science with the greatest potential for societal impact:

First, NIH should consider developing secure infrastructure and rigorous methods for linking research investments to downstream outcomes. This would include creating appropriately protected datasets that connect funded and unfunded applications with subsequent scientific, technological, clinical, economic, and public-health outcomes. NIH should support research to develop more meaningful measures of societal impact, trace the pathways through which research produces real-world benefits, and deploy AI tools for impact assessment, opportunity mapping, and portfolio analysis. These tools should augment rather than replace expert judgment and should be evaluated for validity, robustness, transparency, fairness, and potential bias.

Second, NIH should treat access to ***large language models*** (LLMs) and related computational resources as shared research infrastructure for the science-of-science community. Text analysis is central to this field, including the classification of grants and publications and the measurement of novelty, diffusion, convergence, and scientific consensus. LLMs have dramatically expanded what is possible, with technical feasibility no longer the greatest challenge for many text-analysis tasks. LLM Application Programming Interface expenses in the form of ***token costs*** are now the limiting factor for conducting many cutting-edge, science of science studies. NIH, and especially the National Library of Medicine, should treat LLM access as shared research infrastructure rather than leaving each team to solve the problem independently. Negotiated token pricing with LLM providers for NIH related science of science, subsidized model access, benchmark tasks, and evaluation tools would enable science of science scholars to ask questions about the evolution of fields, the structure of knowledge, and the meaning of textual novelty at a scale that is currently out of reach. Shared infrastructure would reduce duplication, broaden participation, and allow researchers to investigate the evolution and impact of biomedical science at substantially greater scale than currently possible.

These tools could also create a major opportunity to make science more accessible to the public. For example, AI-enabled tools that better interface with PubMed database could help users identify and prioritize publications most relevant to their concerns, understand where consensus exists, and interpret what findings do and do not mean. On the public side, AI-enabled database tools could move beyond keyword search toward natural-language querying, evidence synthesis, consensus mapping, uncertainty explanation, and plain-language interpretation.

Because public support for science depends in part on whether people can understand and navigate scientific literature, these tools could ultimately create both research and civic value. NIH, through the National Library of Medicine, should conduct a set of carefully controlled pilots to develop and evaluate public-facing AI tools grounded directly in the underlying scientific literature. These tools should provide traceable sources, clearly distinguish objective evidence from interpretation and opinion, communicate uncertainty, and resist overstating preliminary or contested findings. Evaluation should address accuracy, bias, user comprehension, accessibility, and effects on public understanding and trust.

Together, these investments would enable NIH to learn more systematically from its complete portfolio of opportunities, direct resources toward research with the greatest potential for societal benefit, and communicate more clearly how publicly funded science contributes to health and society. The result would be a biomedical knowledge system that is not merely searchable but increasingly measurable, understandable, and useful.

Recommendation 8: Develop data, methods, and the evidence base to improve portfolio diversity and support paradigm-challenging science.

Science of science can inform ways to broaden the NIH portfolio to increase support of paradigm-challenging and ***adversarial science***. The NIH portfolio can be thought of as analogous to an investment portfolio, much of ***modern portfolio theory*** shows that diverse, even uncertain, investments reduce overall risk.

NIH should consider the following:

- Support analyses to actively study and manage the NIH portfolio to increase the diversity of funded ideas, approaches, and methods.
- Use of relationships between scores and proposal content to identify intellectual similarity and concentration.
- Identify projects where a small number of reviewers view the project as promising even without widespread consensus.
- Use of citation analysis, ***natural language processing***, and/or validated generative AI to identify and characterize core conceptual underpinnings and methods in proposals, flagging projects that depart from the most common underpinnings.
- Develop and validate methods to evaluate portfolios at the program level.
- Apply science of science to explore best practices for an adversarial science initiative that promotes constructive disagreement, scientific dissent, and viewpoint diversity, including collaborative projects between teams with competing theoretical frameworks.
- Develop and track clear success criteria and evaluation plans to accompany all investments in portfolio diversity initiatives.

- Identify projects that cross disciplinary boundaries or paradigms or otherwise appear risky to assess how they were evaluated, and what their outcomes may be.
- Use validated generative AI to infer reviewers' preferences on research and potential biases in reviews.

A particular focus of this work should be to assess how to prioritize higher-return on investment work over lower-return work, including work that fits disciplinary and/or paradigmatic research and researchers. For instance, there is evidence that interdisciplinary research face review impediments that may limit the funding of high risk-high-reward research.[107]

Recommendation 9: Build global science indices to guide strategic NIH investment and benchmark U.S. competitiveness.

Concerns are growing that scientific progress has slowed in areas critical to human welfare. One indicator is stagnating life expectancy improvements for some populations.[108] At the same time, U.S. leadership at the scientific frontier is increasingly contested, with other global players (e.g., China) now leading or co-leading in multiple domains.[109] [110] This changing landscape is directly relevant to NIH's interest in strengthening biomedical research competition. Effective competition requires benchmarking U.S. science not only internally but against global peers.

NIH should consider supporting a measurement infrastructure to track scientific progress across fields, institutions, and countries. Such indices would include novelty, citation influence, patents, clinical trials, translation, reproducibility, and health impact. This evidence base could identify where stronger domestic competition is most needed and clarify predictors and methods for determining when cutting-edge work or collaboration abroad yields greater returns for U.S. health than strictly domestic investment.

NIH also should consider establishing a science of science program that develops open, continuously updated global indices of scientific progress relative to research investment across biomedical fields. Priorities could be informed by portfolio analysis and external input by scientific experts. The program could also support research comparing ***marginal productivity*** of NIH funding across domestic and international settings, and pilot carefully governed funding experiments and/or partnerships guided by these indices. The aim is to provide a stronger empirical basis for deciding when domestic investment and/or global engagement will best advance U.S. health outcomes.

Understanding the system response

Knowledge is generated by academic scientists, industry researchers, government scientists, clinicians, and interdisciplinary teams. NIH policy and process changes alter incentives and generate responses that are sometimes unexpected and may lead to unexpected outcomes. Science of science can begin to systematically investigate how individuals and organizations respond and make recommendations based on evidence. The Working Group centered their recommendations on data and methods to better understand scientists' behavior, the impact of scientific fields, and the required data that will be needed to measure outcomes and impacts.

Recommendation 10: Invest in datasets from university administrative systems to study scientists' behavior and career progression.

NIH tracks publications and patents but lacks a data infrastructure to understand how its funding shapes scientist behavior, team structure, scientific communities, and institutional incentives.[111] Data on researchers and teams are needed to assess how policies and processes drive innovation,[112] workforce dynamics,[113] [114] [115] translation, dissemination of results, access gaps, and training outcomes.[116] [117] Without such data, NIH cannot fully connect its funding policies and processes to subsequent scientific, workforce, economic and health impacts.

NIH should consider negotiating agreements with funded institutions to provide access to deidentified administrative data, including information from human resources, sponsored grants and contracts, cooperative agreements, and procurement. These data could then be linked to scientific outputs (publications, patents, products, and other resources) at the individual, organizational, and award levels. With appropriate safeguards to protect personal privacy, linkages could also be made to external sources such as state employment and workforce records. The resulting datasets could be used to evaluate how NIH investments impact researchers' career trajectories and to assess differences in the returns on those investments across institutions, research environments, and researcher career stages and other characteristics.

One such example is the Universities Measuring the EffecTs of Research on Innovation, Competitiveness, and Science (UMETRICS).[118] [119] [120] [121] Developed following recommendations made in the early 2000s of the White House Office of Science and Technology Policy[122] and supported by NIH, the Alfred P. Sloan Foundation, the Ewing Marion Kauffman Foundation and the National Science Foundation (NSF). UMETRICS is a people-based data infrastructure that links information from participating university administrative systems, including human resources, and sponsored projects to publications, patents, products and dissertations. Recent NSF support has expanded the scope by integrating UMETRICS data with state employer and workforce data records.[123] [124] [125] [126] [127]

Datasets such as UMETRICS could be used to assess how emerging technologies, including AI, affect the structure, organization, and operation of scientific teams. NIH should consider establishing pilot projects potentially in collaboration with NSF that enable researchers to use university administrative data to examine the returns on research investments in terms of employment, earnings, economic impact, scientific productivity, and career development.[128] Most importantly these projects should help clarify how people, teams, and institutions propagate and translate scientific ideas into advances in health outcomes and longevity.

Recommendation 11: Use a human-centered approach to study the evolution of scientific fields.

NIH has invested in strong biomedical classification systems, including Medical Subject Headings (MeSH), for organizing and retrieving literature and Research, Condition, and Disease Categorization for standardized reporting of research expenditures. While effective for tracking publications, topics, diseases, spending, and reporting, these systems are less well equipped for analyzing how scientific fields emerge, grow, interact, and innovate.[129] [130]

A large science of science literature provides a complementary approach to characterizing scientific fields and understanding innovation that could expand measurement and reporting.[131] [132] Researchers have argued that scientific fields are really networks of people whose movements and relationships govern how knowledge develops, diffuses and gets used.[133] Others have also described a people-centric approach to understanding how commercial innovation evolved from university research.[134] [135] [136] [137] More recent evidence demonstrates the importance of a people-centric approach to characterizing research fields.[138] [139] [140] [141] [142] [143]

This literature has informed the development of an extensive science of science conceptual and empirical framework for characterizing scientific fields as a network of invisible colleges of people whose movements and relationships govern how knowledge develops, diffuses and is applied.[144] Some well-known examples in health sciences include molecular biology[145], biotechnology[146] and stem cell research.[147]

New technologies like LLMs and ***network analysis***, when validated, can enable a human-centered approach to the classification of science. Using a science of science framework, fields can be defined by patterns of collaboration, shared models, common data, the transmission of ideas from academia to commerce[148] and the source of high-growth startups.[149] Network analysis groups researchers by co-authorship, shared grants, or topic similarity, while LLMs help identify common models and datasets.[150]

NIH should develop and apply human-centered classification and network analysis to provide empirical evidence to study the evolution of scientific fields and evaluate how NIH investments contribute to their emergence, growth, and impact. This human-centered network information could also complement existing NIH classification systems. For example, MeSH could incorporate network signals (co-authorship patterns) to improve indexing accuracy by relying on collaborative network signals and changes in researchers' interests over time to refine document classification, entity resolution and relation extraction for ambiguous or changing categories. Integrating non-linguistic document to document linkages has been shown to improve supervised classification of scientific documents into MeSH categories and the addition of similar socially defined connections among documents also yields promising results for unsupervised classification of grants and publications suggesting that such extensions may improve indexing in cases where gold standard data are present and in emerging fields that lack them.[151] A human-centered approach could also improve the classification of diseases and conditions used in NIH spending reports by allowing categories to evolve as scientific communities, concepts, and research priorities change. This could provide NIH leadership and the public with more timely and informative evidence about how scientific fields develop and how NIH investments contribute to their progress.

Recommendation 12: Promote standards for data usage descriptors to support responsible data reuse and reproducibility.

NIH prioritizes collaboration and data sharing and has invested in major infrastructures like the Protein Data Bank[152] and the Knockout Mouse Phenotyping Project.[153] But access to data alone is not enough: researchers need structured metadata that link datasets to publications, authors, and institutions, as well as data trustworthiness to support reproducibility and public confidence. Policies like the *2016 Commission on Evidence-based Policymaking*[154] and the *Foundations for Evidence-Based Policymaking Act of 2018*[155] emphasize using data to inform policy and require agencies to produce data inventories, track and report data use, and engage with the public.

NIH should consider using new tools that can generate data usage statistics that demonstrate how datasets are used by employing validated LLMs to mine scientific publications.[156] Such statistics could support an institutional framework that recognizes data as an output and develops norms in data citations, metadata, classifications, and persistence. Results could be reported through platforms like the National Data Platform and Data Commons.[157] Such an incentive-based framework could ensure that data information is persistently referenceable, navigable and contextual. Application of these descriptors could strengthen NIH investments by showing which data are used most, by whom, and where gaps or risks exist.[158] They can also reveal how quickly repositories drive data use, affects speed discovery and translation, indicate quality and reliability of research, and the return to taxpayers. They can also identify dataset experts and recognize their contributions. This would help to better align data investments with scientific impact and public value.[159] [160]

Recommendation 13: Design data trust measures that can be used to support credible and replicable research.

Data reuse across fields creates a need for contextual trust measures, not simple binary judgments.[161] Data trust measures can systematically document how data are produced and used for specific purposes.[162] They support data utility (whether data are fit for specific purposes, including appropriate and inappropriate use), and data quality (how data are produced, processed, and validated, using traceable evidence).[163] Trustworthiness is a broader concept than reproducibility and replicability. It reflects the extent to which data and findings are credible and can support valid and reliable conclusions.

All data carry risks when reused outside their original context. Data quality signals make these risks explicit, including known biases, temporal or spatial limitations, and potential failure modes in automated or high-stakes applications. Ultimately, the measures can serve to strengthen data trust through community validation, as researchers and other users test and refine these descriptors over time.[164]

Developing such measures is operationally difficult because of (at least) three challenges. The first is quantifying the elements of data quality that underpin data trust: producing a scalable set of signals or a rubric.[165] The second is identifying the different purposes for which data can be used and applying the rubric to assess data quality in that context. Data quality is not a characteristic of any specific dataset, but rather a "context-dependent attribute that must be assessed relative to specific goals and users."[166] The third challenge is engaging a user community to define the use to which the data will be put, to ensure that data quality signals are sound, and to establish trust in data use. This last challenge is particularly important, since trust is not a static construct, but evolves over time based on a series of observations and interaction.[167]

NIH should consider developing data trust measures based on (i) identifying and translating the elements of data quality into a rubric that fits the NIH context, (ii) capturing those elements as metadata for specific NIH use cases of interest, and (iii) disseminating the results in data repositories in a manner that that can be readily understood, tested, and certified by the relevant NIH communities.[168] Such data trust measures could help NIH scientists and NIH program staff

make informed decisions and align data use with reliable, real-world performance. Data-trust measures for both data quality in production (accuracy, reliability, objectivity, producer reputation, and data production) and use have been developed in other contexts.

Outcomes: Evaluating progress

For NIH, the key outcomes are the real-world, long-term results that the investments are intended to produce (not necessarily the short-term products such as publications). Outcomes include accelerating the pace of biological discovery, strengthening the quality and capacity of the biomedical workforce, expanding the translation of research into clinical benefits and commercial applications, improving rigor, transparency, and reproducibility, and ultimately advancing human health and longevity.

Using a science of science approach, progress toward these outcomes should be evaluated not in isolation, but relative to what might have occurred under alternative uses of the same resources. The Working Group therefore focused its recommendations on methods and processes that can more effectively measure and assess how scientific communities respond to actions taken through NIH policy, funding, programmatic, and other institutional levers by better capturing outcomes.

Recommendation 14: Launch targeted living systematic reviews.

In fast-moving scientific fields, conventional literature reviews that summarize published studies and other scientific evidence at a single point in time can quickly become outdated. As new data, methods, and findings emerge, these fixed summaries may no longer represent the current state of knowledge. This makes it harder to determine which conclusions are well supported, where uncertainty remains, and whether the field is moving toward a shared interpretation of the evidence. Without that shared understanding, researchers and decision-makers may struggle to reach consensus, which in turn slows confident, evidence-based decision-making.

Living systematic reviews (LSRs) can address this by continuously updating evidence while maintaining rigor to reduce the lag between discovery and synthesis of new and innovative ideas.[169] [170] LSRs have been used in areas such as oncology and infectious disease.[171] A related model, the Pop-Up Journal, focuses on curating research that could help to bridge the science–policy gap on a single key research question and gather evidence needed for decision-making.[172] NIH has longstanding investments in evidence synthesis capacity, including systematic review support through the NIH Library.[173] The Agency for Healthcare Research and Quality has operationalized living systematic reviews within its Evidence-based Practice Center program, providing a federal precedent.[174] However, NIH has not yet institutionalized living evidence synthesis as a core, enterprise-wide function.

The NIH should consider launching a limited number of targeted living evidence synthesis pilots in domains characterized by rapid evidence generation, high decision relevance, and demonstrable uncertainty. Examples of appropriate pilot areas include rapidly evolving fields in which new evidence accumulates continuously, and decisions have important implications for research priorities or clinical care. Potential domains include Alzheimer's disease and related dementias, long COVID, obesity therapeutics, antimicrobial resistance, rare disease therapeutics, and environmental health. These areas combine rapid evidence generation, high practitioner decision relevance, and persistent uncertainty.

Each pilot should require a preregistered protocol specifying search cadence, updating triggers, inclusion criteria, strength-of-evidence grading, and transparent versioning. Ideally, living evidence synthesis could provide a practitioner infrastructure to shorten the interval between discovery, cumulative assessment and clinical use. For each pilot, links should be made between study design characteristics, population attributes, and replication status to the synthesis platform to enable machine-assisted evidence surveillance, reducing burden and redundancy. Each pilot could include from its inception an evaluation plan with key performance indicators including time from study publication to incorporation in the synthesis, influence on funding or policy decisions, the cost per each update cycle, and user uptake.

Recommendation 15: Develop a range of performance measures to capture outcomes of research.

Rigorous evaluation depends on having endpoints that accurately measure the desired outcomes. Using a range of performance measures is important for NIH because its outcomes are broad, long-term, and often difficult to capture with a single metric. Measures should reflect not only immediate outputs, such as publications, grants awarded, and clinical trials launched, but also longer-term impacts, including scientific breakthroughs, improved health outcomes, workforce development, innovation, and public trust. A balanced set of measures will help NIH demonstrate accountability while recognizing that biomedical research often produces benefits over many years and through complex pathways. It also supports better decision-making by showing where investments are producing progress, where gaps remain, and how NIH's work contributes to public health and scientific advancement.

Measures should be adaptable and capture major policy changes in NIH funding priorities no matter whether the changes occur gradually or on a rapid timetable. For example, recent policies resulting in grant terminations or narrowing of already-awarded funds should be evaluated using common outcome metrics that will assess both negative and beneficial effects of such changes. Metrics should inform rather than replace expert judgment.

NIH should work with the science of science community to develop and validate different metrics to evaluate NIH funding, including distal and near endpoints, such as publications, clinical translation, public health advances and other measures to evaluate the performance of its portfolio. Near endpoints should include measures that address the trustworthiness of the science, including the extent to which the research is accountable, evaluable (transparent and shared openly), well-formulated, reviewed, controlling for bias, reducing error, inclusive of both positive and negative findings, statistically sound, and whether the research claims are based in facts and warranted by the evidence. Once developed and validated, the measures could be used to provide evidence to rigorously and transparently optimize funding and used to communicate the value of NIH investments to the research community and to the broader public. This analysis should be conducted independently of NIH.

Recommendation 16: Develop a range of performance measures to capture outcomes of training programs.

In improving upon NIH performance measures, one area of particular importance is the outcomes of trainees. Evidence indicates that NIH training programs have large impacts on future scientific productivity and a considerable economic footprint.[175] [176] [177] [178] Moreover, the relationship between the research outcomes of labs and their training outcomes are far from one-

to-one,[179] so that a focus on improving research outcomes does not necessarily improve upon training outcomes.

At the same time, NIH has often trained more researchers than the biomedical research enterprise is funded to effectively absorb,[180] [181] [182] so that increasing the quality and impact of training should be a priority over arbitrarily set numbers of trainees. Much of the work on measuring research outcomes should expand beyond publications to examine how research contributes to technological and medical advances, while also developing better measures of the capabilities and career trajectories of trained researchers, including their movement into businesses, occupations, and earnings.

Conventionally, NIH has largely treated research and training separately. That is, there is a set of funding mechanisms that support trainees and early career researchers (e.g., "T"s and "F"s and "K"s) and a separate set of mechanisms that support research (e.g., "R"s). Yet most trainees are supported on research projects and not training grants.[183] Moreover, there is evidence that the emphasis on mechanisms to support early career researchers, while intended to ease the transition to independent funding, has kept early career researchers from applying for research projects.[184]

Given the importance of capturing training outcomes, the NIH should develop and validate a series of measures specifically related to the impact of training supported on all funding mechanisms, including research projects, fellowships, and training grants. This could include partnering with NSF in the expansion of the Industries of Ideas, in which NSF piloted a novel people-centric approach to assess the impact of its investments on regional firms and jobs in key technology areas in Ohio. NIH could develop additional impacts such as changes in the firm demand for education and skills - to include NIH research investments.

Recommendation 17: Improve and validate grant–patent linkages to better enable assessment of NIH's impact on health products.

A large body of research attempts to link NIH funding to patents and drugs using administrative and citation data.[185] [186] [187] [188] [189] While promising, this approach has key limitations including underreporting of patents in iEdison / Research Portfolio Online Reporting Tools Expenditures and Results (RePORTER)[190] [191] and public data on patents linked to specific products only exist for small-molecule drugs, not biologics or other health products.[192] However, patents and drugs are widely recognized as providing a narrow and incomplete lens on outcomes, and existing ways to link grants to patents and patents to drugs (including through citations) suffer from missing data and lack of serious validation. Limited information on the value of NIH-supported patents constrains science of science analyses and might lead to underestimation of the broader outcomes of NIH investment.

NIH should consider generating a more robust evidence base on the outcomes of its investments by improving existing linkages, validating commonly used measures, and supporting the development and validation of new/improved links to health outcomes. This could be done, for example, by partnering with the U.S. Patent and Trademark Office (USPTO) to reconcile grant support data across RePORTER, iEdison, and patent records, including those recorded in USPTO certificates of correction to issued patents.

More complete information on government-funded patents and better patent-product linkages can help generate more reliable and comprehensive information on health products directly linked to NIH research. As an example, NIH could support validating in-text and front-page citations in patents as measures of intellectual influence and knowledge flows, and in non-citation-based ***natural language processing*** approaches linking patents to NIH grants. NIH could seed studies that validate citation-based and other bibliometric links from NIH-funded research to policy documents, clinical guidelines, pivotal trials, and other outcomes beyond patents. Additionally, NIH could explore increasing transparency by making iEdison data on licensing and revenues available to researchers, requiring grantee institutions to report back any U.S. Food and Drug Administration approved products and associated patents resulting from NIH grants and making data broadly available. This would allow links to health products to go beyond small molecule pharmaceuticals (where patent-product linkages are publicly available using the Orange Book[193]).

Recommendation 18: Validate output measures against outcomes.

NIH already provides a significant body of data linking its research funding to patents, publications and clinical trials through RePORTER. These conventional science of science measures (e.g. citation-based measures of patent and publication value; text-based measures of novelty and disruptiveness) measure the output of NIH research.

Whether these and other measures are associated with public health outcomes (e.g. quality-adjusted life years, disability-adjusted life years, mortality) is less well-known. NIH should validate or promote the validation of these measures against measures of health impact and health outcomes.

NIH engagement with stakeholders across ICOs can help identify the health outcomes most relevant to each of their missions and research portfolios. This process can also clarify which conventional science of science indicators are most informative for assessing progress toward those outcomes, how the usefulness of established indicators varies across ICOs, and where improved or additional indicators are needed.

Recommendation 19: Support the development of measures for scientific openness, self-correction, and public trust.

NIH leadership has made restoring public trust a priority.[194] Practices like data sharing, disclosure of uncertainty, and reproducibility are intended to strengthen trust. Trust is typically measured as a multidimensional construct (both self-reported attitudes and observed behaviors). Current assessments of science applications and publications do not include institutional measures of scientific freedom, openness to dissent, willingness to publish null or unpopular results, peer-review and editorial gatekeeping, self-censorship, and the capacity for self-correction, or communication of results with community-based partners.

The Working Group understands that not all these characteristics are under the control of the scientists themselves. The decisions by publishers, conference organizers, are important. Further, these concepts are harder to measure than bibliometrics, but they are not beyond study. Surveys, archival work, confidential administrative data, text analysis, and public-opinion research can be

combined to develop and validate credible indicators of whether scientific communities are open, and self-correcting, and suggest ways to increase public confidence**.**

In addition, NIH should consider incentivizing research specifically aimed at measuring the institutional conditions that sustain trustworthy science. NIH should consider support for work on community engagement, openness of debate, replication climates, censorship and self-censorship, the treatment of ***heterodox*** ideas, and the relationship between these factors and public trust. It should include retrospective studies as well as prospective work on institutional designs that better protect contestation and self-correction. The aim is not to relitigate past controversies but to build a durable evidence base for a more open, self-correcting, and trusted scientific enterprise.

Recommendation 20: Study K–12 levers that produce future scientists.

The U.S. scientific workforce has become increasingly reliant on immigration,[195] which has provided a major resource. But as global competition for scientific talent intensifies, it is risky to rely too heavily on immigration of scientists. Federal efforts to discourage and reduce immigration of highly qualified foreign researchers compounds the issue.

At the same time, there is surprisingly little rigorous work on the levers in domestic K–12 education that affect the number and quality of students who become scientists. Work on how K–12 education affects future scientists and Science, Technology, Engineering, and Mathematics (STEM) participation has been funded primarily by NSF, through its STEM education and workforce programs. Other federal agencies have contributed as well, including NIH through its Science Education Partnership Award,[196] Department of Education, Department of Energy, NASA, and other private and philanthropic funders. Existing measures of school quality prioritize test scores, graduation rates, and college attendance and do not focus on the production of future scientists. NIH has a direct interest in understanding the upstream determinants of participation in the biomedical workforce.

NIH and the American biomedical enterprise would benefit from a dedicated research initiative on the domestic scientist pipeline, with emphasis on linked longitudinal datasets. NIH should use a science of science methodology to link early educational environments and investments to later outcomes such as STEM persistence, PhD entry, biomedical research employment, and NIH grant participation and subsequent publication, patents and career outcomes. This would make it possible to identify levers, such as advanced science and math access, teacher quality, laboratory exposure, mentorship, research experiences, science competitions, or advising, that have long-run career effects, and for which demographic groups. The output should be an evidence base to study the mechanisms and upstream investments that most reliably increase the future supply and quality of U.S.-based scientists.

Recommendation 21: Study the effects of major policy changes in NIH funding priorities and procedures.

Recent federal actions have introduced broad disruptions across the NIH-supported biomedical research enterprise, affecting awarded grants, workforce capacity, funding opportunities, peer review, funding decisions, and the scope of supported scientific domains. These changes include large-scale grant terminations, reductions in workforce and funding mechanisms, modifications

to funded project terms, targeted cuts to diversity, equity, and inclusion (DEI) research, altered review and funding procedures, and expanded scrutiny of scientific alignment with agency priorities. Like most previous funding and policy changes, these were implemented in the absence of systematic evidence evaluating their likely effects on research processes, funding outcomes, and the broader biomedical research enterprise. Significant changes to NIH policies, programs, or funding priorities, regardless of their origin, policy rationale, or timing, should be treated as opportunities for rigorous evaluation using science of science methods.

Going forward, NIH should treat all policy shifts as testable, system-level interventions, and rigorously evaluate their effects and use a science of science approach to capture how changes in funding, workforce composition, and topic legitimacy propagate through the research ecosystem. Specifically, NIH should consider launching independent science of science evaluations of major policy changes in NIH funding priorities. Evaluations should account for NIH's legal obligations and address empirical questions as to the effects of the policy changes on NIH outcomes and create longitudinal datasets that link affected and unaffected awards to publications, workforce exits, health outcomes, and other measures. At the same time, the evaluations should explicitly test whether new policies produce beneficial portfolio reallocations. Importantly, this evaluation should prioritize outcome measures that extend beyond traditional bibliometrics to include health-relevant impacts, such as the generation, dissemination, and implementation of evidence needed to reduce health disparities, respond to public health threats, and improve population health.

More broadly, NIH should reaffirm that future changes to funding policy, portfolio priorities, and administrative processes are introduced only alongside empirical evidence, or structured and transparent evaluation plans capable of generating such evidence, that they will improve scientific productivity, workforce resilience, and, ultimately, health outcomes. Embedding this standard within a continuous learning framework would align NIH policy with core science of science principles, ensuring that decisions are not only responsive to external directives but are also accountable to evidence on their effects across the full research–to–health causal chain.

From Recommendations to Implementation: Strengthening the Science of Science Agenda at NIH

The Working Group's report underscores that NIH is uniquely positioned to strengthen the biomedical research enterprise by adopting a systematic, evidence-driven, science of science approach to understanding how its actions and policies shape scientific behavior and outcomes. By applying a causal-chain framework, investing in robust data infrastructure, and embedding continuous learning across its programs, NIH can more effectively evaluate the impact of its decisions at-scale and adapt its strategies to maximize scientific progress and public health benefit. The recommendations presented here offer a diversity of options and pathways for NIH to build a roadmap for continuing excellence and success. Though beyond the scope of this Working Group's charge, it is important to note that NIH would benefit from piloting alternative approaches and assure that the results of the proposed approaches in these recommendations be used to improve upon NIH programs, policies, and operations. If so, the Working Group believes that these actions will allow NIH not only to steward biomedical research more effectively, but

also to accelerate discovery, enhance transparency and rigor, and ensure that the Nation’s biomedical research investments achieve their greatest possible impact on human health.

Undertaking any of these recommendations will take strategic commitments of time and funding. The Working Group does not advocate funding of its recommendations by assessing a proportional "tax" or set-aside against existing ICO budgets. While such mechanisms have occasionally been used at NIH, the current fiscal environment characterized by sustained pressure on discretionary spending and flat or declining appropriations in real terms means that any mandatory assessment would come directly at the expense of investigator-initiated research. The Working Group's credibility depends on proposing solutions that strengthen the overall enterprise, not ones that pit NIH-wide priorities against the scientific communities that individual ICOs serve.

Instead, the Working Group suggests a dedicated, line-item funding through the NIH Office of the Director's budget. The selection of topics and themes to be funded under these initiatives must follow a rigorous, transparent, and reproducible process, grounded in systematic review of the scientific literature, identification of critical knowledge gaps, and expert input from credentialed scientists with relevant domain expertise. This process should be explicitly and vigorously protected from distortion by political considerations and institutional conflicts of interest. Conscious and unconscious bias in priority-setting, including the tendency to favor fashionable research areas, well-networked investigators, or institutional incumbents over genuinely high-impact but underexplored questions, must likewise be actively identified and mitigated through structured decision-making frameworks and diverse and balanced external advisory bodies selected based on scientific merit and objectivity.

Finally, and perhaps most importantly, the Working Group recommends the establishment of a formal, recurring evaluation framework that operates at both the internal and external levels, on a defined and predictable schedule. Internally, NIH program officers and initiative leads should conduct structured annual reviews assessing whether science of science funded activities are on track to meet their stated objectives, whether the evidence base has shifted in ways that warrant course correction, and whether administrative processes are functioning efficiently and equitably. Externally, independent scientific advisory bodies composed of researchers, clinicians, and public health experts with no financial or institutional conflicts relative to the programs under review should conduct comprehensive evaluations at regular multi-year intervals, with findings made publicly available in full. These external reviews must be empowered to recommend reallocation, restructuring, or discontinuation of programs that are not demonstrably advancing decision-making at NIH.

The authority and independence of these evaluative bodies must be protected by charter, and their findings must feed directly and transparently into subsequent funding decisions. Progress reports and assessments by internal and external advisory bodies should be reported back at regular intervals to the NIH Council of Councils. An NIH that holds itself to this standard of evidence-driven, bias-conscious, and continuously evaluated stewardship will not only make better use of public resources, but it will also model for the broader scientific community the integrity that the enterprise of human health research demands. When, as the largest funder of biomedical research in the world, NIH harnesses the full power of science by better

understanding the data and processes that fuel it, it will unlock discoveries that can cure disease, extend human life, and transform what is possible for generations to come.

Acknowledgements

The Working Group is grateful to Dr. Kaye Husbands Fealing for reviewing and providing feedback on this report.

Glossary of Terms

This glossary provides plain-language descriptions of technical terms and concepts used in the Working Group's Science of Science report that may be unfamiliar to readers outside of economics and the social sciences.

Adversarial Science Adversarial science is a method where scientists with opposing views work together to design and conduct studies, aiming to resolve disputes and enhance research rigor. This approach brings rivals together to agree on methods, fostering transparency and reducing confirmation bias.

Causal-Chain Framework A structured way of thinking about cause and effect in a complex system. In this report, it refers to tracing how NIH policy decisions (inputs/levers) lead through a series of intermediate responses (mediators) to ultimately produce outcomes (scientific discoveries, health improvements). The framework insists on understanding not just whether something happened, but why and through what pathways, rather than simply correlating inputs with outputs.

Counterfactual / Counterfactual Reasoning In evaluation, a counterfactual is a hypothetical question, such as "What would have happened if the policy or intervention had NOT been put in place?" It is the comparison baseline needed to assess whether a policy truly caused an observed result. For example, if NIH-funded research produced 100 new drugs, the counterfactual would ask "how many of these drugs would have been developed *without* that funding?" Without counterfactual reasoning, it is impossible to know whether an outcome was caused by the policy or would have occurred anyway.

Epistemic Narrowing / Paradigm Capture A situation in which a scientific field or review panel becomes intellectually dominated by a single framework, set of hypotheses, or methodology, crowding out alternative approaches. Paradigm capture specifically refers to a peer review study section becoming so entrenched in one theoretical model that proposals using different approaches are systematically disadvantaged, even if scientifically valid. At NIH, this can slow scientific progress by suppressing potentially transformative heterodox ideas.

FAIR Data Principles A framework for data management and sharing that stands for: Findable, Accessible, Interoperable, and Reusable.[197] FAIR data principles are widely adopted in the scientific community to ensure that research data can be discovered by others, accessed under clear conditions, used across different systems and software, and reused for future research. NIH has increasingly required data sharing policies aligned with FAIR principles.

Generalizability The degree to which findings from a specific study apply to other populations, settings, or conditions beyond those directly studied. A treatment that works in a clinical trial conducted in a specific hospital with a specific patient population may or may not generalize to the broader population. This is particularly relevant in biomedical research, where study populations have historically not been representative of all demographic groups.

Heterodox Science Heterodox science refers to research that challenges or departs from the dominant theoretical framework in a field. Like the concept of adversarial science, a heterodox

approach is a deliberate strategy of funding competing research teams with opposing theoretical frameworks to test ideas more rigorously through constructive scientific disagreement. This concept aims to prevent groupthink and encourage exploration of alternative explanations.

In Silico Review / Multi-Agent AI System Computer-based simulation of processes that would otherwise require human deliberation. An *in silico* review committee would be an AI system trained on historical grant data to mimic the decisions of human peer review panels, allowing researchers to test how different funding rules would perform without disrupting the live grant system. Multi-agent means that the AI system would include multiple interacting components, analogous to multiple reviewers on a traditional review group panel.

Large Language Model (LLM) / Token Costs Large language models (LLMs) are AI systems trained on vast amounts of text that can read, summarize, classify, and generate language at high capability. Accessing these systems via an API (application programming interface) typically incurs costs based on 'tokens', meaning the number of words or word-pieces processed. At the scale of analyzing millions of scientific papers or grant applications, these token costs become a significant financial constraint on research.

Living Systematic Review (LSR) / Evidence Synthesis A systematic review is a rigorous, structured summary of all available evidence on a research question. A living systematic review is continuously updated as new studies are published, in contrast to traditional static reviews that quickly become outdated. Evidence synthesis is the broader practice of combining findings from multiple studies to arrive at more reliable conclusions than any single study can provide.

Marginal Productivity Economic concepts describing how the value of additional investment changes as the amount invested increases. Marginal productivity refers to the additional output gained from one more unit of investment. Diminishing returns describes the phenomenon where each additional dollar (or grant) produces less benefit than the previous one. This concept is relevant to debates about whether concentrating multiple large grants in a single laboratory is optimal.

Metadata / Metadata Set Data that describes other data. In research contexts, metadata includes structured information about a study's design, methods, sample characteristics, data sources, and outcomes rather than the underlying data itself. A core metadata set for reproducibility would be a standardized set of fields that all funded studies must report, enabling systematic comparisons across research and automated monitoring of research quality.

Modern Portfolio Theory A framework from financial economics showing that a diversified portfolio of investments with varying risk and return profiles can achieve better outcomes than concentrating resources in the highest-expected-return opportunities alone. Applied to research funding, it suggests that NIH should support a diverse mix of research approaches and risk levels, including some speculative, high-risk/high-reward projects, rather than concentrating only on the safest or most conventional science.

Natural Language Processing A branch of artificial intelligence that enables computers to understand, interpret, and generate human language text. In the science of science context,

natural language processing tools can analyze grant applications, scientific papers, and reviewer comments at large scale, identifying topics, measuring novelty, detecting intellectual similarity across proposals, and extracting structured information from unstructured text.

Network Analysis / Citation Network / Co-Authorship Network Mathematical methods for studying the structure of relationships among interconnected entities. A citation network maps how papers cite one another, revealing intellectual lineages and the spread of ideas. A co-authorship network maps collaborations among researchers. Clustering coefficients measure how tightly interconnected a group is. For example, high clustering in a study section's reviewers could signal intellectual insularity or paradigm capture.

Out-of-Order Funding When an NIH Institute, Center, or Office (ICO) funds a grant application that did not achieve a sufficiently high priority score to be within the normal payline prior to the 2026 policy relative to its peer review ranking. This discretion allows institutes to pursue specific programmatic priorities but requires transparency and justification to prevent arbitrary or opaque decision-making.

Person-Centered Data Infrastructure Data systems organized around individual researchers, tracking the people who conduct science, their roles, career trajectories, training, collaborations, and economic contributions, rather than focusing solely on grants or publications. This approach enables a more complete picture of how NIH investments shape the scientific workforce over time.

Portfolio Optimization / Portfolio Analysis Applying concepts from financial investment portfolio management to the allocation of research funding. A financial portfolio should be diversified across assets of different risk and return profiles. Adapting this principle to scientific investment, a research portfolio should balance investments across different scientific questions, methodologies, and risk levels. Portfolio correlation analysis would thus examine whether funded projects are too similar to one another, which might reduce the portfolio's overall value relative to its cost.

Preregistration The practice of publicly registering a study's hypotheses, methods, and analysis plan before data collection begins. Preregistration makes it possible to distinguish confirmatory findings (testing a hypothesis stated in advance) from exploratory findings (patterns discovered after the fact). It reduces 'p-hacking,' selective reporting, and publication bias which have been associated with the reproducibility crisis in science.

Quasi-Experimental Design / Quasi-Experiments Research methods that approximate a true controlled experiment without randomly assigning subjects to groups. Because it is often ethically or practically impossible to randomly assign who receives research funding, scientists instead seek to exploit plausibly exogenous natural variations, such as comparing researchers who scored just above or just below a funding cutoff, to estimate the causal effect of a policy. These simple 'natural experiments' potentially allow researchers to draw more precise causal conclusions than simple observational comparisons. The downsides of such quasi-experimental analyses include: 1) the obtained estimates are local in nature (i.e. the estimates are only applicable to a very limited types of individuals such as those likely to score near the cutoff), 2)

the estimates do not reveal what happens if policy is adopted on a larger scale (i.e. the estimates ignore so called general equilibrium effects), and 3) the incentive to focus on less important research questions due to unavailability of plausible exogenous variation for studies on the most important questions (such as impact of large research funding shocks on health, or effect of country-wide school closures on health outcomes).

Reproducibility vs. Replicability Related but distinct concepts central to scientific rigor. Reproducibility is the ability of independent researchers to test a hypothesis through multiple methods and consistently achieve results that confirm or refute it, ensuring findings are generalizable and robust across different approaches. Replicability is the ability to perform the same experiment or study using the same methods and conditions to achieve the same result. The reproducibility crisis refers to the widely documented finding that many published scientific results fail to hold up when re-tested, raising concerns about publication bias, p-hacking, and inadequate methods reporting.

Science of Science (SoS) / Metascience The systematic, empirical study of how science itself works, examining how scientific research is organized, funded, conducted, and evaluated in order to improve its efficiency and impact.[198] [199] Just as scientists study diseases or materials, metascientists study the scientific enterprise as a system. It draws on economics, sociology, computer science, and data analytics to understand what makes research productive and how funding policies shape scientific progress.

Theory of Change An explicit, testable statement of the logic behind a policy or program, describing what it is intended to achieve, what assumptions underlie it, and through what specific pathway it is expected to produce its intended effects. It is a roadmap from “here is what we are doing” to “here is why we believe it will work”. Making theories of change explicit allows them to be tested, refined, and compared.

Acronyms (Alphabetized)

- AI — Artificial Intelligence
- DPCPSI — Division of Program Coordination, Planning, and Strategic Initiatives
- FAIR — Findable, Accessible, Interoperable, and Reusable
- ICO — Institute, Center, or Office
- LLM — Large Language Model
- LSR — Living Systematic Review
- MeSH — Medical Subject Headings
- NIGMS — National Institute of General Medical Sciences
- NIH — National Institutes of Health
- NSF — National Science Foundation
- RePORTER — Research Portfolio Online Reporting Tools Expenditures and Results
- SoS — Science of Science
- SoS: BIO — Science of Science Approach to Analyzing and Innovating the Biomedical Research Enterprise
- STEM — Science, Technology, Engineering, and Mathematics
- UMETRICS — Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science
- USPTO — U.S. Patent and Trademark Office
- WG — Working Group

Working Group Charge and Roster

Introduction

As the largest public funder of biomedical research in the world, NIH is the driving force behind decades of advances that have expanded scientific knowledge and improved health. As stewards of public funds, NIH continues to seek new ways to engage the research community, including science of science, which applies rigorous and empirical analysis to study how science is done, to help us understand the impacts of NIH policies and programs. NIH has a long history of engaging with researchers to generate evidence to inform improvements to its programs, policies, and processes. This includes making a wide array of grants-related data and statistics publicly available through the NIH Research Portfolio Online Reporting Tools (RePORT) website and providing outside researchers access to NIH administrative data. NIH also supports research to strengthen NIH decision making through the NIGMS Science of Science Approach to Analyzing and Innovating the Biomedical Research Enterprise Program. Recently, the Office of Extramural Research and the Office of Evaluation, Performance, and Reporting launched a pilot NIH Science of Science Scholars Program that will pair science of science researchers with NIH staff to conduct in-depth analyses on NIH processes, programs and outcomes of NIH-supported research. In light of these efforts and other advancements in biomedical research, NIH is seeking continual improvement in generating evidence to inform its work and to drive more discoveries to improve health.

Charge

To maximize NIH's use of evidence in decision-making, DPCPSI is seeking advice from experts to identify scientific areas with the greatest potential for improving upon NIH's programs, policies, and operations. Specifically, the Council of Councils Working Group on Science of Science is charged with:

1. Reviewing the efforts that NIH has undertaken to generate evidence about its programs, policies, and processes.
2. Identifying significant gaps in knowledge where leveraging science of science and related research may better influence NIH decision-making.
3. Providing recommendations on potential approaches that NIH could consider to enhance its use of the science of science and ensuing results.
4. Providing recommendations for how NIH can engage the science of science research community and communicate results of science of science research.

Roster

Co-Chairs

Kevin C. Kent Lloyd, DVM, PhD
Council of Councils Member
Distinguished Professor, Department of Surgery
School of Medicine
Director, Mouse Biology Program
Associate Director, Comprehensive Cancer Center
University of California, Davis

Marina Volkov, PhD
Director
Office of Evaluation, Performance, and Reporting
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health

Executive Secretary

Ira Kuhn, PhD
Science Advisor to the Director
Office of Evaluation, Performance, and Reporting
Division of Program Coordination, Planning, and Strategic Initiatives
Office of the Director
National Institutes of Health

Working Group Members

Pierre Azoulay, PhD
International Programs Professor of Management
Professor, Technological Innovation, Entrepreneurship, and Strategic Management
Sloan School of Management
Massachusetts Institute of Technology

Maryann Feldman, PhD
Watts Professor of Public Policy and Management
Watts College of Public Service and Community Solutions
Professor
School of Public Affairs
Research Director
Global Center for Technology Transfer
Arizona State University

Monica Gandhi MD, MPH
Director, UCSF-Bay Area Center for AIDS Research
Professor of Medicine and Associate Chief
Division of HIV, Infectious Diseases, and Global Medicine
University of California San Francisco

Daniel Goroff, PhD
Vice President and Program Director
Alfred P. Sloan Foundation

Karen C. Johnston, MD, MSc
Harrison Distinguished Professor of Neurology, Department of Neurology
Associate Vice President for Clinical & Translational Research
Office of the Vice President for Research
University of Virginia

Julia Lane, PhD
Professor Emerita
Wagner Graduate School of Public Service
Provostial Fellow for Innovation Analytics
New York University Wagner

Jennifer J. Manly, PhD
Professor of Neuropsychology
Taub Institute for Research on Alzheimer's Disease & the Aging Brain
Department of Neurology
Columbia University Irving Medical Center

Brian Nosek, PhD
Co-Founder and Executive Director
Center for Open Science
Professor
University of Virginia

Mikko Packalen, PhD
Associate Professor
Department of Economics
University of Waterloo

Bhaven Sampat, PhD
Professor
School of Government and Policy
Carey Business School
Johns Hopkins University

Dashun Wang, PhD
Kellogg Chair of Technology
Professor of Management & Organizations
Professor of Industrial Engineering & Management Sciences
Director, Center for Science of Science and Innovation
Co-Director, Ryan Institute on Complexity

Director, Northwestern Innovation Institute
Northwestern University

Caleb Watney, MA
Co-Founder and Co-CEO
Institute for Progress

Bruce Weinberg, PhD
Eric Byron Fix-Monda Endowed Chair of Economics and Public Affairs
Ohio State University

References

[1] Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, Radicchi F, Sinatra R, Uzzi B, Vespignani A, Waltman L, Wang D, Barabási AL. Science of science. Science. 2018 Mar 2;359(6379):eaao0185. doi: 10.1126/science.aao0185. PMID: 29496846; PMCID: PMC5949209.
[2] National Academy of Medicine. The State of the U.S. Biomedical and Health Research Enterprise: Strategies for Achieving a Healthier America. Washington, DC: National Academies Press; 2024. doi:10.17226/27588. PMID: 39680696.
[3] https://www.unitedformedicalresearch.org/annual-economic-report/
[4] Dzau VJ, Yamamoto KR. Government-Funded Health and Biomedical Research Is Irreplaceable. NAM Perspect. 2025 Aug 18;2025:10.31478/202508b. doi: 10.31478/202508b. PMID: 41970625; PMCID: PMC13064907.
[5] https://www.nih.gov/about-nih/nih-director/statements/advancing-nihs-mission-through-unified-strategy
[6] nexus.od.nih.gov/all/2019/03/04/seeking-input-on-the-need-to-enhance-access-to-nih-grants-data/
[7] journals.lww.com/academicmedicine/fulltext/2016/08000/gender, race ethnicity, and national institutes of.23.aspx
[8] www.ncbi.nlm.nih.gov/pmc/articles/PMC4769156/
[9] pubmed.ncbi.nlm.nih.gov/28360137/
[10] pubmed.ncbi.nlm.nih.gov/29440428/
[11] https://www.nigms.nih.gov/Research/mechanisms/MIRA
[12] Kozlowski D, Larivière V, Sugimoto CR, Monroe-White T. Intersectional inequalities in science. Proc Natl Acad Sci U S A. 2022 Jan 11;119(2):e2113067119. doi: 10.1073/pnas.2113067119. PMID: 34983876; PMCID: PMC8764684.
[13] www.nigms.nih.gov/Research/specificareas/Pages/Science-of-Science-Policy-Approach-to-Analyzing-and-Innovating-the-Biomedical-Research-Enterprise-(SCISIPBIO).aspx
[14] www.gao.gov/products/gao-23-105656.
[15] acd.od.nih.gov/documents/presentations/12132018NextGen_report.pdf
[16] https://dpcpsi.nih.gov/oepr/pilot-nih-science-science-scholars-program
[17] https://dpcpsi.nih.gov/council/april-21-2025-agenda
[18] https://www.nih.gov/about-nih/nih-director/statements/advancing-nihs-mission-through-unified-strategy
[19] https://www.nih.gov/about-nih/nih-director/statements/advancing-nihs-mission-through-unified-strategy
[20] https://www.nih.gov/sites/default/files/2025-08/2025-gss.pdf
[21] https://report.nih.gov/
[22] Ginther DK, Schaffer WT, Schnell J, Masimore B, Liu F, Haak LL, Kington R. Race, ethnicity, and NIH research awards. Science. 2011;333(6045):1015–1019. doi:10.1126/science.1196783.
Supports: workforce diversity and analyses using NIH administrative data.
[23] Jacob BA, Lefgren L. The Impact of Research Grant Funding on Scientific Productivity. Journal of Public Economics. 2011;95(9–10):1168–1177. doi:10.1016/j.jpubeco.2011.05.005.
Supports: research productivity and bibliometric analyses of NIH-funded investigators.
[24] Azoulay P, Graff Zivin JS, Li D, Sampat BN. Public R&D Investments and Private-sector Patenting. Review of Economic Studies. 2019;86(1):117–152. doi:10.1093/restud/rdy034.
Supports: patenting and innovation resulting from NIH funding.
[25] Cleary EG, Jackson MJ, Zhou EW, Ledley FD. Comparison of Research Funding for the Development of New Drug Approvals from Public and Private Sources. JAMA Health Forum. 2023;4(4):e230511.
Supports: NIH funding and drug development.
[26] Sampat BN, Lichtenberg FR. What Are the Respective Roles of the Public and Private Sectors in Pharmaceutical Innovation? Health Affairs. 2011;30(2):332–339.
[27] Fong, Eric A., Wilhite, Allen W. 2017. "Authorship and citation manipulation in academic research." PLOS ONE, 12(12), e0187394. doi:10.1371/journal.pone.0187394
[28] Hutchins, B. Ian, Xin Yuan, James M. Anderson, George M. Santangelo. 2016. "Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level." PLOS ONE. September 6, 2016. https://doi.org/10.1371/journal.pbio.1002541
[29] Husbands Fealing, Kaye Julia I. Lane, John H. Marburger, and Stephanie S. Shipp, eds. *The science of science policy: A handbook*. Stanford University Press, 2020.
[30] Staudt, Joseph, Huifeng Yu, Robert P. Light, Katy Borner, Gerald Marschke, and Bruce A. Weinberg. 2018. "High-impact and transformative science (HITS) metrics: definition, exemplification, and comparison." PLoS One 13(7, July 19): e0200597 https://doi.org/10.1371/journal.pone.0200597.research

[31] Funk, Russel J. and Jason Owen-Smith. 2016. "A Dynamic Network Measure of Technological Change." Management Science 63 (No. 3, March): 791-817. https://doi.org/10.1287/mnsc.2015.2366
[32] Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large teams develop and small teams disrupt science and technology."Nature 566 (13 February): 378–382. https://www.nature.com/articles/s41586-019-0941-9
[33] Packalen, Mikko, and Jayanta Bhattacharya. 2020. "NIH Funding and the Pursuit of Edge Science." PNAS 117(22):12011–16. https://www.pnas.org/doi/10.1073/pnas.1910160117
[34] Packalen, Mikko. 2019. Edge Factors: Scientific Frontier Positions of Nations" Scientometrics. 188(3) 787-808. https://pmc.ncbi.nlm.nih.gov/articles/PMC6818715/
[35] Weinberg, Bruce A., Jason Owen-Smith, Rebecca Rosen, Lou Schwarz, Barbara McFadden Allen, Roy Weiss, Julia Lane. 2014 "Science Funding and Short-Term Economic Activity." Science 304 (April 4): 41-43. DOI:10.1126/science.1250055
[36] Azoulay, Pierre, Wesley H. Greenblatt, and Misty L. Heggeness. 2021. "Long-term Effects from Early Exposure to Research: Evidence from the NIH "Yellow Berets." Research Policy, 50 (Issue 9, November): 104332. https://doi.org/10.1016/j.respol.2021.104332
[37] Sattari, Reza, Jung Bae, Enrico Berkes, and Bruce A. Weinberg. 2022. "The Ripple Effects of Funding on Researchers and Output." Science Advances April 22. Vol 8, Issue 16. DOI: 10.1126/sciadv.abb7348
[38] Lane, Julia, Jason Owen-Smith, and Bruce A. Weinberg. "How to track the economic impact of public investments in AI." *Nature* 630, no. 8016 (2024): 302-304.
[39] Zolas, Nikolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen Smith, Rebecca Rosen, Barbara McFadden Allen, Bruce Weinberg, and Julia Lane. 2015. "Wrapping it up in a person: Examining employment and earnings outcomes for PhD recipients." Science 350 (December 11, No. 6266): 1367-1371. http://www.sciencemag.org/content/350/6266/1367
[40] Azoulay, Pierre, Danielle Li, and Bhaven Sampat. 2017. "The Applied Value of Public Investments in Biomedical Research." Science 356 (Issue 6333), pp. 78-81, 2017. DOI: 10.1126/science.aal0010
[41] Azoulay, Pierre, Joshua Graff Zivin, Danielle Li, and Bhaven Sampat. 2019. "Public R&D Investments and Private Sector Patenting: Evidence from NIH Funding Rules." Review of Economic Studies 86 (Issue 1): 117–152, 2019. https://doi.org/10.1093/restud/rdy034
[42] Marx, Matt and Aaron Fuegi, 2020. "Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles." Strategic Management Journal 41(9):1572-1594. https://onlinelibrary.wiley.com/doi/full/10.1002/smj.3145)
[43] Marx, Matt and Aaron Fuegi, 2022. "Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations." Journal of Economics and Management Strategy 31 (Issue 2, Summer): 369-392.
[44] https://www.nih.gov/about-nih/impact-nih-research/our-stories
[45] Azoulay, Pierre, Danielle Li, and Bhaven Sampat. 2017. "The Applied Value of Public Investments in Biomedical Research." Science 356 (Issue 6333), pp. 78-81, 2017. DOI: 10.1126/science.aal0010
[46] Azoulay, Pierre, Joshua Graff Zivin, Danielle Li, and Bhaven Sampat. 2019. "Public R&D Investments and Private Sector Patenting: Evidence from NIH Funding Rules." Review of Economic Studies 86 (Issue 1): 117–152, 2019. https://doi.org/10.1093/restud/rdy034
[47] McKibbin, Rebecca and Bruce A. Weinberg. 2021. "Does Research Save Lives? The Local Spillovers of Biomedical Research on Mortality." NBER Working Paper No. 29420. DOI 10.3386/w29420
[48] Blume-Kohout, M.E. (2012), Does Targeted, Disease-Specific Public Research Funding Influence Pharmaceutical Innovation?. J. Pol. Anal. Manage., 31: 641-660. https://doi.org/10.1002/pam.21640
[49] Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature. 2014 Jan 30;505(7485):612-3. doi: 10.1038/505612a. PMID: 24482835; PMCID: PMC4058759.
[50] Fenner M, Crosas M, Grethe JS, Kennedy D, Hermjakob H, Rocca-Serra P, Durand G, Berjon R, Karcher S, Martone M, Clark T. A data citation roadmap for scholarly data repositories. Sci Data. 2019 Apr 10;6(1):28. doi: 10.1038/s41597-019-0031-8. PMID: 30971690; PMCID: PMC6472386.
[51] https://grants.nih.gov/funding/find-a-fit-for-your-research/highlighted-topics/54
[52] As a funder, NIH has authority over where and how it's appropriated funds are spent, but the dollar amount is set by Congress.
[53] Office of Science and Technology Policy. Biennial Report on the Implementation Status of Federal Scientific Integrity Policy and Practice. Washington, DC: OSTP; 2024.
[54] Valdez, Bill, and Julia Lane. "The science of science policy: a federal research roadmap." *Office of Science and Technology Policy. Washington, DC, Executive Office of the President* (2008).
[55] https://www.nsf.gov/funding/opportunities/sosdci-science-science-discovery-communication-impact

[56] https://loop.nigms.nih.gov/2019/02/apply-to-scisipbio-a-joint-initiative-between-nigms-and-nsf-to-support-research-on-the-science-of-science-and-innovation-policy/
[57] https://sloan.org/programs/research/economics-2
[58] https://www.povertyactionlab.org/initiative/science-progress-initiative
[59] Fortunato, S., et al. (2018). Science of science. Science, 359(6379), eaao0185.
[60] https://dpcpsi.nih.gov/oepr/pilot-nih-science-science-scholars-program
[61] For example, through a central function or designated leadership role.
[62] arXiv:2608.18175v1 [physics.soc-ph]. https://doi.org/10.48550/arXiv.2608.18175
[63] https://www.nih.gov/news-events/news-releases/nih-centralizes-peer-review-improve-efficiency-strengthen-integrity
[64] https://www.regulations.gov/document/OMB-2026-0034-0001
[65] Li D, Agha L. Research funding. Big names or big ideas: do peer-review panels select the best science proposals? Science. 2015 Apr 24;348(6233):434-8. doi: 10.1126/science.aaa0185. Epub 2015 Apr 23. PMID: 25908820.
[66] Narasimhan Danthi, Colin O. Wu, Peibei Shi, and Michael Lauer, 2014. "Percentile Ranking and Citation Impact of a Large Cohort of National Heart, Lung, and Blood Institute–Funded Cardiovascular R01 Grants." Circulation Research Volume 114, Number 4. https://doi.org/10.1161/CIRCRESAHA.114.302656
[67] Ferric C Fang, Anthony Bowen, Arturo Casadevall. NIH peer review percentile scores are poorly predictive of grant productivity. Elife. 2016 Feb 16;5:e13323. doi: 10.7554/eLife.13323. PMID: 26880623; PMCID: PMC4769156.
[68] Li, Danielle and Leila Agha. 2015. "Big Names or Big Ideas: Do Peer Review Panels Select the Best Science Proposals?" Science, 24 April 2015: Vol. 348, No. 6233.
[69] Funk, Russel J. and Jason Owen-Smith. 2016. "A Dynamic Network Measure of Technological Change." *Management Science* 63 (No. 3, March): 791-817. https://doi.org/10.1287/mnsc.2015.2366
[70] Staudt, Joseph, Huifeng Yu, Robert P. Light, Katy Borner, Gerald Marschke, and Bruce A. Weinberg. 2018. "High-impact and transformative science (HITS) metrics: definition, exemplification, and comparison." *PLoS One* 13(7, July 19): e0200597https://doi.org/10.1371/journal.pone.0200597.research
[71] Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large teams develop and small teams disrupt science and technology."*Nature* 566 (13 February): 378–382. https://www.nature.com/articles/s41586-019-0941-9
[72] Packalen, Mikko, and Jayanta Bhattacharya. 2020. "NIH Funding and the Pursuit of Edge Science." PNAS 117(22):12011–16. https://www.pnas.org/doi/10.1073/pnas.1910160117
[73] Huifeng Yu, Gerald Marschke, Joseph Staudt, Matthew B. Ross, and Bruce A. Weinberg. 2023. "Publish or Perish: Selective Attrition as an Unifying Explanation for Lifecycle Patterns in Innovation." *Journal of Human Resources* 58 (July, Issue 4): 1307-1346. DOI: https://doi.org/10.3368/jhr.59.2.1219-10630R1.
[74] Packalen, Mikko and Bhattacharya, Jay. 2017. Neophilia Ranking of Scientific Journals. Scientometrics. 110(1) 43-64. https://pubmed.ncbi.nlm.nih.gov/28713181/
[75] Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.
[76] Camerer, C. F., Dreber, A., Holzmeister, F., et al. (2016). Evaluating replicability of laboratory experiments in economics. Science, 351(6280), 1433–1436.
[77] Errington, T. M., Iorns, E., Gunn, W., et al.; Reproducibility Project: Cancer Biology. (2021). Investigating the replicability of preclinical cancer biology. eLife, 10, e71601.
[78] Miske, O., Abatayo, A. L., Daley, M., Dirzo, M., Fox, N., Haber, N., Hahn, K. M., Struhl, M. K., Mawhinney, B., Silverstein, P., Stankov, T., Tyner, A. H., Adamkovič, M., Alzahawi, S., Anafinova, S., Awtrey, E., Axxe, E., Bailey, J., … Errington, T. M. (2026). Investigating the reproducibility of the social and behavioural sciences. Nature, 652, 126–134. https://doi.org/10.1038/s41586-026-10203-5
[79] Tyner, A. H., Abatayo, A. L., Daley, M., Field, S., Fox, N., Haber, N. A., Hahn, K. M., Struhl, M. K., Mawhinney, B., Miske, O., Silverstein, P., Soderberg, C. K., Stankov, T., Abbasi, A., Aberson, C. L., Aczel, B., Adamkovič, M., Albayrak, N., Allen, P. J., … Errington, T. M. (2026). Investigating the replicability of the social and behavioural sciences. Nature, 652, 143–150. https://doi.org/10.1038/s41586-025-10078-y
[80] Aczel, B., Szaszi, B., Clelland, H. T., et al. (2026). Investigating the analytical robustness of the social and behavioural sciences. Nature, 652, 135–142. https://doi.org/10.1038/s41586-025-09844-9
[81] https://www.cos.io/score-evidence
[82] https://www.darpa.mil/research/programs/systematizing-confidence-in-open-research-and-evidence
[83] https://www.whitehouse.gov/presidential-actions/2025/05/restoring-gold-standard-science/#:~:text=3.,results%20as%20positive%20outcomes;%20and
[84] https://iris.isr.umich.edu/

[85] Lane, Julia I., Jason Owen-Smith, Rebecca Rosen, and Bruce A. Weinberg. 2015 "New linked data on science investments, the scientific workforce and the economic and scientific results of science" Research Policy 44 (Issue 4, November): 1659-71. http://www.sciencedirect.com/science/article/pii/S0048733315000025
[86] Lin, Zihang, Yian Yin, Li Liu, and Dashun Wang. 2023. "SciSciNet: A large-scale open data lake for the science of science research." Scientific Data 10 (Article Number 315). https://doi.org/10.1038/s41597-023-02198-9
[87] https://relianceonscience.org/
[88] Marx, Matt and Aaron Fuegi, 2020. "Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles." Strategic Management Journal 41(9):1572-1594. https://onlinelibrary.wiley.com/doi/full/10.1002/smj.3145)
[89] Marx, Matt and Aaron Fuegi, 2022. "Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations." Journal of Economics and Management Strategy 31 (Issue 2, Summer): 369-392.
[90] Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: fair enough? Eur J Hum Genet. 2018 Jul;26(7):931-936. doi: 10.1038/s41431-018-0160-0. Epub 2018 May 17. PMID: 29777206; PMCID: PMC6018669.
[91] https://datascience.nih.gov/data-ecosystem
[92] https://www.cos.io/score
[93] Williams, R. Sanders and Susan Desmond-Hellmann. 2011. "Making Translation Work." *Science* **332** 1359-1359. DOI:10.1126/science.1209338
[94] Institute of Medicine (IOM). 2012. *Accelerating the Development of New Drugs and Diagnostics: Maximizing the Impact of the Cures Acceleration Network: Workshop Summary*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13452.
[95] Institute of Medicine (IOM). 2013. *The CTSA Program at NIH: Opportunities for advancing clinical and translational research.* Washington, DC: The National Academies Press.
[96] Butler, Declan. 2008. "Translational research: Crossing the valley of death." Nature 453 (7197):840–842. doi: 10.1038/453840a.
[97] Roberts SF, Fischhoff MA, Sakowski SA, Feldman EL. 2012. 'Perspective: transforming science into medicine: how clinician-scientists can build bridges across research's "valley of death"'. Academic Medicine. 87:266–70.
[98] Fishburn C. Simone. 2013. "Translational research: The changing landscape of drug discovery." Drug Discovery Today 18(9–10):487–494. doi: 10.1016/j.drudis.2012.12.002.
[99] Seyhan, Attilia A. 2019. "Lost in translation: The valley of death across preclinical and clinical divide–identification of problems and overcoming obstacles." Translational Medicine Communications 4 (1): 1–19. doi: 10.1186/s41231-019-0050-7.
[100] Park Sang-Min and Nicholas Vonortas. 2023. "Translational research: from basic research to regional biomedical entrepreneurship." Small Business Economics 60(4):1761–83. doi: 10.1007/s11187-022-00676-9.
[101] Collins, F. S. (2011). Reengineering Translational Science: The Time Is Right. Science Translational Medicine, 3(90), 90cm17. https://doi.org/10.1126/scitranslmed.3002747
[102] Krauss, Alexander. 2024. "Redefining the Scientific Method: As the Use of Sophisticated Scientific Methods That Extend Our Mind." PNAS Nexus 3:pgae112. doi:10.1093/pnasnexus/pgae112.]
[103] Krauss, Alexander. 2026. "New Tools Drive Scientific Discovery: Evidence from All Nobel-Prize and Major Non-Nobel Breakthroughs." Humanities and Social Sciences Communications, accepted/forthcoming.]
[104] Krauss, Alexander. 2025. "New Scientific Fields Are Triggered by Powerful New Methods." Humanities and Social Sciences Communications 12:1590. doi:10.1057/s41599-025-05797-6.
[105] Krauss, Alexander, Ariel Rosenfeld, and Lutz Bornmann. 2026. "The Rising Dominance of Methods across Science: Evidence from over Three Million Papers across Science." Manuscript under review, PNAS.
[106] https://www.nih.gov/about-nih/organization/budget
[107] Enrico Berkes, Monica Marion, Staša Milojević, Bruce A. Weinberg. 2024. "Slow convergence: Career impediments to interdisciplinary biomedical research." *Proceedings of the National Academy of the Sciences* 121 (32) e2402646121 https://doi.org/10.1073/pnas.2402646121.
[108] Dowd, J.B., Polizzi, A. and Tilstra, A.M. (2025), Progress Stalled? The Uncertain Future of Mortality in High-Income Countries. Population and Development Review, 51: 257-293. https://doi.org/10.1111/padr.12687
[109] R. Wu, C. Esposito, & J. Evans, Shifting power asymmetries in scientific teams reveal China's rising leadership in global science, Proc. Natl. Acad. Sci. U.S.A. 122 (44) e2414893122, https://doi.org/10.1073/pnas.2414893122 (2025).
[110] Packalen, Mikko. 2019. Edge Factors: Scientific Frontier Positions of Nations" Scientometrics. 188(3) 787-808. https://pmc.ncbi.nlm.nih.gov/articles/PMC6818715/

[111] Myers, Kyle. 2020. "The elasticity of science." American Economic Journal: Applied Economics 12 (4): 103-134.).
[112] Lane, J., J. Owen-Smith, and B. Weinberg, How to track the economic impact of public investments in AI. Nature 2024. 630: p. 302-304.
[113] Ross, M.B., et al., Women are credited less in science than men. Nature, 2022. 608(7921): p. 135-145.
[114] Benjamin Jones and Bruce A. Weinberg. "Age Dynamics in Scientific Creativity." *Proceedings of the National Academy of the Sciences* 108 (November 22, 2011): 18,910-14. doi: 10.1073/pnas.1102895108.
[115] David M. Blau and Bruce A. Weinberg. 2017. "Why the United States Science and Engineering Workforce Is Aging Rapidly." *Proceedings of the National Academy of the Sciences* 117 (No. 15): 3879-3884. doi:10.1073/pnas.1611748114.
[116] Azoulay, Pierre, Danielle Li, and Bhaven Sampat. 2017. "The Applied Value of Public Investments in Biomedical Research." Science 356 (Issue 6333), pp. 78-81, 2017. DOI: 10.1126/science.aal0010
[117] Azoulay, Pierre, Joshua Graff Zivin, Danielle Li, and Bhaven Sampat. 2019. "Public R&D Investments and Private Sector Patenting: Evidence from NIH Funding Rules." Review of Economic Studies 86 (Issue 1): 117–152, 2019. https://doi.org/10.1093/restud/rdy034
[118] https://iris.isr.umich.edu/ The UMETRICS program at the Institute for Research on Innovation and Science now includes transaction level information on over $192 billion in sponsored project spending on over 580,000 funding awards from all sources at over 100 campuses. It covers wages to about 985,000 employees
[119] Lane, Julia, and Stefano Bertuzzi. "Measuring the results of science investments." Science 331.6018 (2011): 678-680.
[120] Lane, Julia, and Stefano Bertuzzi. "The STAR METRICS project: current and future uses for S&E workforce data." Science of Science Measurement Workshop, held Washington DC. Vol. 12. 2010.
[121] Largent, Mark A., and Julia I. Lane. "STAR METRICS and the Science of Science Policy." Review of Policy Research 29.3 (2012): 431-438.
[122] Valdez, B. and J. Lane, The science of science policy: a federal research roadmap. Report to The Subcommittee on Social, Behavioral and Economics Sciences, Committee on Science, National Science and Technology Council. Washington, DC: Office of Science and Technology Policy, Executive Office of the President, 2008.
[123] Industries of Ideas. 2026; Available from: https://industriesofideas.ai/.
[124] Bruce A. Weinberg, Jason Owen-Smith, Rebecca Rosen, Lou Schwarz, Barbara McFadden Allen, Roy Weiss, Julia Lane. 2014 "Science Funding and Short-Term Economic Activity." Science 304 (April 4): 41-43. DOI:10.1126/science.1250055.
[125] Nikolas Zolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen Smith, Rebecca Rosen, Barbara McFadden Allen, Bruce Weinberg (Corresponding Author) and Julia Lane. 2015. "Wrapping it up in a person: Examining employment and earnings outcomes for PhD recipients." Science 350 (December 11, No. 6266): 1367-1371. http://www.sciencemag.org/content/350/6266/1367
[126] https://iris.isr.umich.edu/
[127] Julia I. Lane, Jason Owen-Smith, Bruce A. Weinberg. 2024. "How to track the economic impact of public investments in AI." Nature 630 (June 13): 302-304. https://www.nature.com/articles/d41586-024-01721-1
[128] https://industriesofideas.ai
[129] Hadfield, R.M., Delay and bias in PubMed medical subject heading (MeSH®) indexing of respiratory journals. medRxiv, 2020: p. 2020.10.01.20205476.
[130] Torrey, E.F., et al., Using the NIH Research, Condition and Disease Categorization Database for research advocacy: Schizophrenia research at NIMH as an example. Plos one, 2020. 15(11): p. e0241062.
[131] Hadfield, R.M., Delay and bias in PubMed medical subject heading (MeSH®) indexing of respiratory journals. medRxiv, 2020: p. 2020.10.01.20205476.
[132] Torrey, E.F., et al., Using the NIH Research, Condition and Disease Categorization Database for research advocacy: Schizophrenia research at NIMH as an example. Plos one, 2020. 15(11): p. e0241062.
[133] Diana Crane's Invisible Colleges: Diffusion of Knowledge in Scientific Communities (1972) University of Chicago Press
[134] Annalee Saxenian 1994 Regional Advantage: Culture and Competition in Silicon Valley and Route 128,
[135] Nicholas Mullins. 1972 "The Development of A Scientific Specialty: The Phage Group and the Origins of Molecular Biology." Minerva. 10(1) 51-82
[136] Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. "Intellectual human capital and the birth of US biotechnology enterprises." The American economic review 88.1 (1998): 290.
[137] Zucker, Lynne G., Michael R. Darby, and Maximo Torero. "Labor mobility from academe to commerce." Journal of Labor Economics 20.3 (2002): 629-660.

[138] Fleming, Lee, Santiago Mingo, and David Chen. "Collaborative brokerage, generative creativity, and creative success." Administrative science quarterly 52.3 (2007): 443-475; Balsmeier, Benjamin, Lee Fleming, Matt Marx, and Seungryul Ryan Shin. "Startups, unicorns, and the local influx of inventors." Review of Economics and Statistics (2025): 1-44.
[139] Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. Science, 308(5722), 697-702.
[140] Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. "Superstar extinction." The Quarterly Journal of Economics 125.2 (2010): 549-589.; Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. "Does science advance one funeral at a time?." American Economic Review 109, no. 8 (2019): 2889-2920.
[141] Scott, C. T., McCormick, J. B., DeRouen, M. C., & Owen-Smith, J. (2011). Democracy derived? New trajectories in pluripotent stem cell research. Cell, 145(6), 820-826.
[142] Zolas, Nikolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen Smith, Rebecca F. Rosen, Barbara McFadden Allen, Bruce A. Weinberg, and Julia I. Lane. "Wrapping it up in a person: Examining employment and earnings outcomes for Ph. D. recipients." Science 350, no. 6266 (2015): 1367-1371.
[143] Cui, H., Lin, Y., Wu, L., & Evans, J. A. (2026). Aging and the narrowing of scientific innovation. Science, 392(6798), 588-591.
[144] Diana Crane *Invisible Colleges:Diffusion of Knowledge in Scientific Communities* (1972) University of Chicago Press
[145] Nicholas Mullins. 1972 "The Development of A Scientific Specialty: The Phage Group and the Origins of Molecular Biology." *Minerva. 10(1) 51-82*
[146] Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. "Intellectual Human Capital and the Birth of US Biotechnology Enterprises." *The American Economic Review* 88, no. 1 (1998): 290-306.
[147] Scott, Christopher Thomas, Jennifer B. McCormick, Mindy C. DeRouen, and Jason Owen-Smith. "Democracy derived? New trajectories in pluripotent stem cell research." *Cell* 145, no. 6 (2011): 820-826.
[148] Zucker, Lynne G., Michael R. Darby, and Maximo Torero. "Labor mobility from academe to commerce." *Journal of Labor Economics* 20, no. 3 (2002): 629-660.
[149] Balsmeier, Benjamin, Lee Fleming, Matt Marx, and Seungryul Ryan Shin. "Startups, unicorns, and the local influx of inventors." *Review of Economics and Statistics* (2025): 1-44.
[150] Owen-Smith. Jason (2026). Person-Centric AI Research Community Identification: Initial Seed Set Validation. Zenodo. https://doi.org/10.5281/zenodo.20425418; Owen-Smith, Jason. (2026). Research Domain Clustering & Labeling: Overview & Initial Validation. Zenodo. https://doi.org/10.5281/zenodo.20434423
[151] Singh, Amanpreet, Mike D'Arcy, Arman Cohan, Doug Downey, and Sergey Feldman. "Scirepeval: A multi-format benchmark for scientific document representations." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 5548-5566. 2023.
[152] https://www.rcsb.org/
[153] Tudor Groza, Federico Lopez Gomez, Hamed Haseli Mashhadi, Violeta Muñoz-Fuentes, Osman Gunes, Robert Wilson, Pilar Cacheiro, Anthony Frost, Piia Keskivali-Bond, Bora Vardal, Aaron McCoy, Tsz Kwan Cheng, Luis Santos, Sara Wells, Damian Smedley, Ann-Marie Mallon, Helen Parkinson, The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease, Nucleic Acids Research, Volume 51, Issue D1, 6 January 2023, Pages D1038–D1045
[154] https://www.congress.gov/bill/114th-congress/house-bill/1831
[155] https://www.congress.gov/bill/115th-congress/house-bill/4174
[156] Harvard Data Science Review Special Issue, 2024 Democratizing Data: Discovering Data Use For Research and Policy https://hdsr.mitpress.mit.edu/specialissue4; https://democratizingdata.ai/events/2024/harvard-data-science-review/
[157] National Data Platform. *Pediatric Asthma Dashboard*. Available from: https://pediatric-asthma.di.ndp.utah.edu
[158] Chenarides, L., Ladislau, R., Noy, N., Parashar, M., Porter, S., & Lane, J. (2026). Data-usage descriptors as search metadata: the case of food security data and the National Data Platform. Nature: Scientific Data. DOI : 10.1038/s41597-026-07753-8.
[159] https://www.nature.com/articles/508312a
[160] Lane, Julia, Alfred Spector, and Michael Stebbins. "An invisible hand for creating public value from data." *Harvard Data Science Review* 4 (2024): 19804.
[161] Liang, W., et al., Advances, challenges and opportunities in creating data for trustworthy AI. Nature Machine Intelligence, 2022. 4(8): p. 669-677.
[162] Hendra, R., & Lane, J. (2026). Elements of Data Quality: A Foundation for Data Trust (V1). https://doi.org/10.5281/zenodo.20478444

[163] Guillen-Aguinaga, M., et al., *Data quality in the age of AI: A review of governance, ethics, and the FAIR principles.* Data, 2025. **10**(12): p. 201
[164] Mayer, R.C., J.H. Davis, and F.D. Schoorman, An integrative model of organizational trust. Academy of management review, 1995. 20(3): p. 709-734.
[165] Gualo, F., et al., Data quality certification using ISO/IEC 25012: Industrial experiences. Journal of Systems and Software, 2021. 176: p. 110938.
[166] Guillen-Aguinaga, M., et al., Data quality in the age of AI: A review of governance, ethics, and the FAIR principles. Data, 2025. 10(12): p. 201
[167] Mayer, R.C., J.H. Davis, and F.D. Schoorman, An integrative model of organizational trust. Academy of management review, 1995. 20(3): p. 709-734
[168] Hendra, Richard, and Julia Lane. "Elements of Data Quality: A Foundation for Data Trust (V1)". Zenodo, May 31, 2026. https://doi.org/10.5281/zenodo.20478444.
[169] Elliott, J. H., et al. (2014). Living systematic reviews: An emerging opportunity. PLoS Medicine, 11(2), e1001603.
[170] Elliott, J. H., et al. (2017). Living systematic review: 1. Introduction. Journal of Clinical Epidemiology, 91, 23–30.
[171] https://pmc.ncbi.nlm.nih.gov/articles/PMC7328560/
[172] Dalmeet Singh Chawla "Pop-up journals for policy research: can temporary titles deliver answers?" Nature Index, February 25, 2026. https://www.nature.com/articles/d41586-026-00446-7
[173] https://www.nihlibrary.nih.gov/services/evidence-synthesis
[174] https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/epc-evidence-reviews-federal-agencies.pdf
[175] Weinberg, Bruce A., Jason Owen-Smith, Rebecca Rosen, Lou Schwarz, Barbara McFadden Allen, Roy Weiss, Julia Lane. 2014 "Science Funding and Short-Term Economic Activity." Science 304 (April 4): 41-43. DOI:10.1126/science.1250055
[176] Zolas, Nikolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen Smith, Rebecca Rosen, Barbara McFadden Allen, Bruce Weinberg, and Julia Lane. 2015. "Wrapping it up in a person: Examining employment and earnings outcomes for PhD recipients." Science 350 (December 11, No. 6266): 1367-1371. http://www.sciencemag.org/content/350/6266/1367
[177] Azoulay, Pierre, Wesley H. Greenblatt, and Misty L. Heggeness. 2021. "Long-term Effects from Early Exposure to Research: Evidence from the NIH "Yellow Berets." Research Policy, 50 (Issue 9, November): 104332. https://doi.org/10.1016/j.respol.2021.104332
[178] Sattari, Reza, Jung Bae, Enrico Berkes, and Bruce A. Weinberg. 2022. "The Ripple Effects of Funding on Researchers and Output." Science Advances April 22. Vol 8, Issue 16.
[179] Xuan Jiang, Joseph Staudt, and Bruce A. Weinberg. "Publish and Train or Perish: Valuing the Early Career Outcomes of STEM PhD Recipients." Working Paper.
[180] NIH. 2012. "Biomedical Research Workforce Working Group Report." https://acd.od.nih.gov/documents/reports/Biomedical_research_wgreport.pdf. Accessed March 5, 2025.
[181] Alberts, Bruce, Marc W. Kirschner, Shirley Tilghman, and Harold Varmus. 2014. "Rescuing US biomedical research from its systemic flaws." PNAS 111 (Issue 16, April 14): 5773-5777. https://doi.org/10.1073/pnas.1404402111
[182] Xuan Jiang, Joseph Staudt, and Bruce A. Weinberg. "A Tale of Two Markets? Knowledge, Skill, and STEM Career Outcomes." Working Paper.
[183] Lauer, Mike. 2023. "Number of Postdoctoral Researchers Supported by NIH Grant Awards FY 2017-FY 2022." https://nexus.od.nih.gov/all/2023/03/02/number-of-postdoctoral-researchers-supported-by-nih-grant-awards-fy-2017-fy-2022/ Accessed on February 22, 2025.
[184] Heggeness, Misty L., Frances Carter-Johnson, Walter T. Schaffer, and Sally J. Rockey. (2016). "Policy implications of aging in the NIH-funded workforce." Cell Stem Cell, 19(1), 15-18.
[185] Azoulay, Pierre, Joshua Graff Zivin, Danielle Li, and Bhaven Sampat. 2019. "Public R&D Investments and Private Sector Patenting: Evidence from NIH Funding Rules." Review of Economic Studies 86 (Issue 1): 117–152, 2019. https://doi.org/10.1093/restud/rdy034
[186] Ouellette LL, Sampat BN. Using Bayh-Dole Act March-In Rights to Lower US Drug Prices. JAMA Health Forum. 2024 Nov 1;5(11):e243775. doi: 10.1001/jamahealthforum.2024.3775. PMID: 39485334; PMCID: PMC11530938.
[187] Azoulay, Pierre, Danielle Li, and Bhaven Sampat. 2017. "The Applied Value of Public Investments in Biomedical Research." Science 356 (Issue 6333), pp. 78-81, 2017. DOI: 10.1126/science.aal0010

[188] Marx, Matt and Aaron Fuegi, 2020. "Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles." Strategic Management Journal 41(9):1572 1594.https://onlinelibrary.wiley.com/doi/full/10.1002/smj.3145
[189] Marx, Matt and Aaron Fuegi, 2022. "Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations." Journal of Economics and Management Strategy 31 (Issue 2, Summer): 369-392
[190] Ouellette LL, Sampat BN. Using Bayh-Dole Act March-In Rights to Lower US Drug Prices. JAMA Health Forum. 2024 Nov 1;5(11):e243775. doi: 10.1001/jamahealthforum.2024.3775. PMID: 39485334; PMCID: PMC11530938.
[191] National Academies of Sciences, Engineering, and Medicine. 2022. Assessment of the SBIR and STTR Programs at the National Institutes of Health. Washington, DC: The National Academies Press. https://doi.org/10.17226/26376.
National Academies of Sciences, Engineering, and Medicine. 2022. Assessment of the SBIR and STTR Programs at the National Institutes of Health. Washington, DC: The National Academies Press. https://doi.org/10.17226/26376.
[192] Michael D. Frakes and Melissa F. Wasserman, "Deadlines Versus Continuous Incentives: Evidence from the Patent Office," NBER Working Paper 32066 (2024), https://doi.org/10.3386/w32066.
[193] https://www.accessdata.fda.gov/scripts/cder/ob/index.cfm
[194] https://www.nih.gov/about-nih/science-health-public-trust
[195] Pool, L.R., Wagner, R.M., Scott, L.L., RoyChowdhury, D., Berhane, R., Wu, C., Pearson, K., Sutton, J.A. and Schaffer, W.T. (2016), Size and characteristics of the biomedical research workforce associated with U.S. National Institutes of Health extramural grants. The FASEB Journal, 30: 1023-1036. https://doi.org/10.1096/fj.14-264358
[196] https://nigms.nih.gov/Research/DRCB/SEPA/Pages/default
[197] doi.org/10.1038/sdata.2016.18
[198] Santo Fortunato et al. ,Science of science.Science359,eaao0185(2018).DOI:10.1126/science.aao0185
[199] Alexander Krauss, Science of science: A multidisciplinary field studying science, Heliyon, Volume 10, Issue 17, 2024, e36066, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2024.e36066.