Experimentation at NIH

Kyle R. Myers,[1] Harvard Business School

May 2023

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

Summary

The U.S. National Institutes of Health (NIH) has funded some of the most important biomedical experiments in the past century (National Institutes of Health, 2016). But despite decades of progress, the NIH has yet to embrace the power of conducting its own experiments. Amidst the backdrop of increased international scientific competition, and in the wake of the Covid-19 pandemic, which highlighted the social value of scientific progress, the time is right for the NIH to turn the scientific method on itself.

Calls for NIH reforms and evaluations are not new (Anderson, 1991; U.S. Government Accountability Office, 1994). What is new is that we have much more experience conducting rigorous social-science experiments designed to inform policies as well as a growing community of scholars studying the “science of science” (e.g., how scientists choose projects or evaluate ideas; J-PAL, 2023; Azoulay et al., 2018). What is needed is a way of injecting these people and skillsets, along with a culture that values experimentation, into the NIH.

A low-cost way of initiating more experimentation would be for the NIH to establish a Chief Economist, who could work as a bridge between academic social scientists, NIH staff, and stakeholders to generate the “gold standard” – experimental evidence – to help usher the NIH into a new era. To be clear, this new era might involve a combination of both new and existing policies. But each decision to change, or not, would have the advantage of being based on evidence held to the same standards as the research the NIH itself funds.

This initiative should be led by social scientists with the expertise necessary to properly evaluate policy alternatives within the complicated environment of the NIH. The mission would not be academic. Rather, the mission would be to inform NIH policies with a broader goal to increase

---

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

Thanks are due to: Heidi Williams, Caleb Watney, Stuart Buck, and Matt Clancy for aiding in the development of this paper; and Azoulay (2012) for inspiration.

the productivity of the U.S. scientific enterprise. And there is a tremendous amount of value at stake – even a 1 percent reduction in mortality from cancer, just one of the major causes of death in the U.S., could be worth more than $500 billion to the U.S. economy (Murphy and Topel, 2006).

Challenge and Opportunity

The NIH of today looks remarkably similar to how it did 50 years ago. For example, the vaunted R01 grant – NIH's oldest funding mechanism, which awards approximately $500,000 in total costs per year for 4 to 5 years – remains mostly unchanged and continues to be used for most research project grants (National Institutes of Health, 2017). Likewise, the structure of NIH's peer-review system is still based on the same basic principles and rules from decades ago: a handful of scientists (often sourced from current NIH award recipients) report a score for each application, those scores are averaged, and that average score largely determines the fate of the application. Are there efficient grant structures not currently being offered (e.g., grants that provide smaller amounts of funding for longer periods)? Are there better ways of recruiting peer-reviewers or aggregating their evaluations to find the "best" ideas? We don't know. But we could if we ran experiments.

To be fair, the NIH has considered and made several changes to their operations in recent years. There have been: new grant structures introduced with the aims of attracting ideas that are more "high-risk high-reward" (National Institutes of Health, 2013a); additional reviews for scientists seeking large amounts of funding (National Institutes of Health, c); tweaks to the peer-review scoring system (National Institutes of Health, b); and an attempt to impose limits on the amount of resources individual scientists can obtain from the NIH (National Institutes of Health, 2017) which was ultimately abandoned at the behest of NIH stakeholders (Kaiser, 2017).

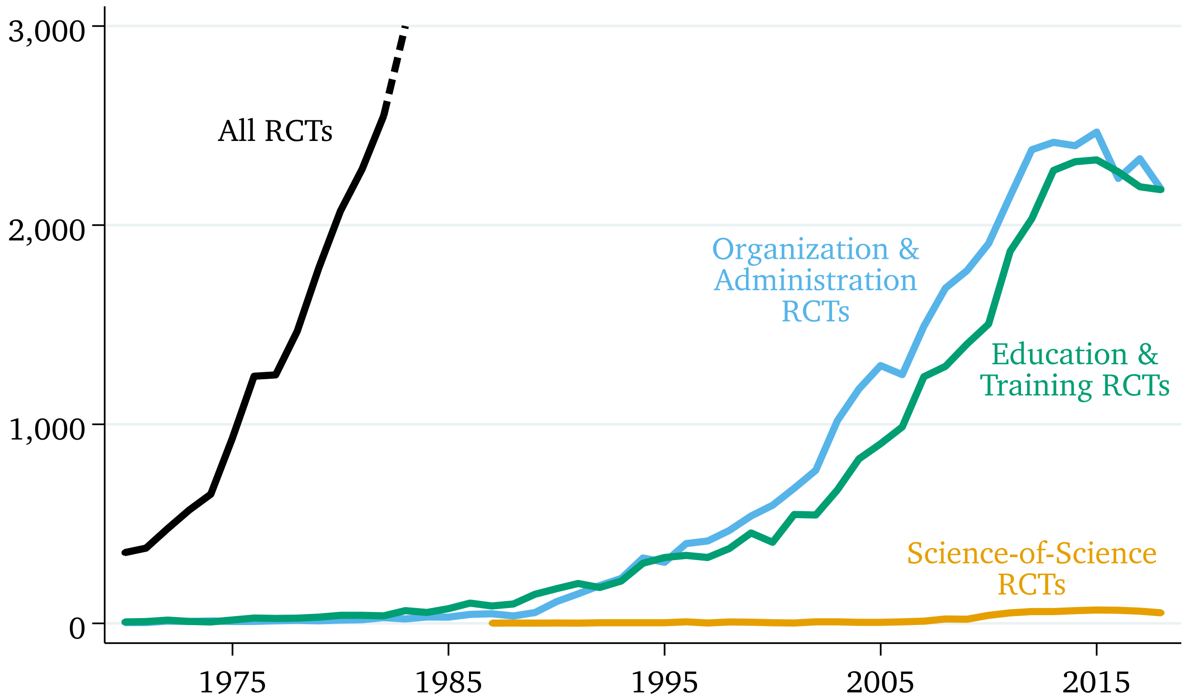
However, none of these or any other NIH policies have ever been motivated by experimental evidence, nor were they implemented in the form of an experiment to test if the policies had their intended effect. For example, the NIH Director's New Innovator Award (announced in 2007) has clearly funded impressive research projects (National Institutes of Health, 2013b). But it is extremely difficult, in the absence of an experiment, to make a convincing claim as to whether this program funded research *that would not have otherwise been funded*.[i]

Why are there so few science-of-science experiments?

Experiments, or Randomized Controlled Trials (RCTs), have always been a bedrock of biomedical research. In the past thirty years, biomedical scientists have expanded their use of RCTs beyond the bench to study broad policy questions of organization, administration, education, and training (e.g., policies such as medical resident duty-hour rules and workplace wellness programs [Silber et al., 2019; Song and Baiker, 2019]; see Figure 1). Likewise, the share of social scientists using RCTs continues to grow steadily, in no small part motivated by groundbreaking work that used experimental methods to evaluate policies in developing countries (Kleven, 2018; Burtless, 2019).

Why, then, has the RCT approach been used so sparingly to study peer-review, grant funding, and other "science-of-science" questions relevant to NIH policies (see Figure 1)?

Figure 1. Very Few Experiments on the "Science-of-Science" To-date



Count of RCT publications in the PubMed database, by select topics, 1970-2018.[ii]

There are two important constraints on performing science-of-science experiments: (1) scale and financial costs, and (2) initiative.

Individual academic researchers have been limited by the first. Given the uncertainties inherent to science, answering questions important to the NIH can easily require thousands of participants. And with grant sizes in the hundreds of thousands of dollars, costs become prohibitive to individuals very quickly.[iii]

The NIH is only limited by initiative. Financial costs are manageable. The NIH continues to be the largest single funder of biomedical research in the world – investments which have been very productive – so, relatively small percentages of NIH budgets can easily suffice to fund large-scale experiments with enormous potential payoffs (Lakdawalla et. al, 2010; Sampat and Lichtenberg, 2011; Lichtenberg, 2018; Azoulay et. al, 2019). And these payoffs could accrue immediately (e.g., by quickly observing the results of a trial and making operational changes) as well as over the long-term.[iv]

But despite its many Institutes, Centers, and Divisions, the NIH has never had any positions or groups with an explicit mandate to experiment. There have been Congressional requests for evaluations, but these often go unanswered or unimplemented (Buck, 2022). This is certainly not for a lack of interest on the side of the NIH staff who continue to communicate a desire for progress and improvements. Instead, it is likely that the tremendous size and complexity of the NIH leaves little time or resources for the prospect of conducting formal experiments. Thus, a new initiative focused explicitly on developing the processes that would allow the NIH to test new ideas is essential. Stakeholders agree that experimentation is necessary.[v]

Intra-agency groups with mandates to evaluate policies are not unheard of (See Figure 2). Across the US government, several agencies have been recruiting social scientists trained in policy evaluation and providing them with the freedom, if not a mandate, to experiment. Looking beyond government agencies, businesses have rapidly embraced the power of experiments in their operations. Millions of dollars continue to be invested in RCTs and other evaluations as businesses continue to appreciate the value of employing skilled social scientists (The Economist, 2022).[vi]

Figure 2. Ongoing Efforts to Incorporate Policy Experiments at U.S. Agencies

Initiative	Notes
Center for Medicare & Medicaid Innovation	Develops and tests new healthcare payment and service delivery models to improve patient care, lower costs, and better align payment systems (see here for recent reports and program evaluations).
USPTO Office of the Chief Economist	Advises on the economic implications of intellectual property policies and has begun to perform experimental tests of programs designed to help U.S. inventors (see here for an example of a USPTO-led RCT).
Research at the US Census Bureau	The Census Program for Evaluations and Experiments has been conducting experiments and evaluations since 1950 and continues to test and improve their policies and procedures.
USDA Economic Research Service	Has recently conducted experiments to study farmers' risk preferences and their demand for crop insurance as well as the marketing and uptake of micro-loans.
Department of Education announces first Chief Economist	The mission of this new Chief Economist will be to "…build a culture of experimentation, including partnerships with leading social science researchers to pilot-test new ways to serve students and borrowers."

The need for social scientists

Conducting experiments within the context of the NIH will be no small feat. Unlike the cells under their microscopes, researchers are free to make their own decisions: whether or not to apply to the NIH; which Institute or which call for applications to submit their application to; what idea to submit and how much funding to request, etc. In short, while it will no doubt continue to attract researchers, the NIH cannot force anyone to engage with it. Handling this freedom when developing experiments will be challenging. However, economists have made good progress on integrating theoretical and empirical methods to account for the fact that NIH-funded scientists get to choose, for instance, their own pursuits, and their opinions of others' work (Myers, 2020; Li, 2017).

One particularly illustrative example of the value of combining experiments with economic theory comes from the world of education. There, researchers have been able to make accurate forecasts about the impact of both hypothetical and actual policy changes (Calsamiglia, 2020; Pathak and Shi, 2021). The Center proposed here could do the same (See Figure 3 for some example policy questions that could be solved with experiments).

Figure 3. Example Questions that can be Solved with Experiments

How can the NIH increase its support of younger and under-represented scientists?
What is the best way of aggregating peer review scores to identify socially valuable research ideas?
Do longer grant award periods incentivize scientists to alter their research plans in beneficial ways?
Are there portions of the application process that can be eliminated to reduce administrative burden without sacrificing scientific integrity?
Should investigators be allowed to request no-cost extensions of their grants more than once to increase the long-term flexibility of grants?
Do solicited calls for proposals (e.g., RFAs) fund research that would not have been funded otherwise?
Are there more effective ways of recruiting scientists to participate in study sections?

Acknowledging risks and defining success

Would experiments introduce unnecessary risk into the NIH's operations? No. There is already a large amount of randomness in the system. But through well-designed trials, the Center could leverage and focus that randomness to identify cost-effective ways of improving the NIH's operations.

Would all experiments be successful? Yes and no. Yes, because, regardless of what the experiments reveal, they will teach us how to fund science better. There will be unforeseen issues that may halt some experiments or render results inconclusive. But, if staffed and funded properly, even these "failures" would provide learning opportunities for the Center to improve their own operations.

Would it be difficult to scale? No. While many field experiments fail to scale, the NIH is in a unique position where the universe of scientists can be identified (List, 2022); even small, pilot studies can be performed with representative samples. This way, when policies are adopted for larger groups of scientists, there should be less of a chance for unintended consequences due to differences between those who participated in pilot studies and those who did not. Still, there are important differences across the NIH Institutes and Centers that would need to be accounted for when any trial would be designed because what works for one, may not work for the other.

Possible Next Steps

One simple next step to begin down this path would be for the NIH to make more use of the Intergovernmental Personnel Act (IPA) Mobility Program and bring in academics to work with individual Institutes and Centers on the development of experiments tailored to their specific needs and interests. This approach is flexible; however it also would require more initiative on the part of NIH staff to instigate and manage these relationships.

A more committed next step would be to more formally pursue the creation of a Chief Economist position at the NIH (possible also via the IPA program). With the right support, this position would provide more centralization, which could lessen the demands on NIH staff from participating in these important experiments. Furthermore, it would more easily allow for the coordination of experimental efforts across the NIH Institutes and Centers to avoid duplication and leverage the full scale of the NIH.

If this avenue is pursued, the Chief Economist should be someone capable of developing rigorous social-science experiments and communicating the results of these efforts to the NIH community. They could report directly to the NIH Director and would also serve as a connection to the social-science community to incorporate best-practices and attract new talent to the NIH. With the right leadership, autonomy, and authority, a more experimental approach could ensure that the NIH maintains is status as the most productive and preeminent scientific funding agency in the world.

References

Anderson, C. (1991). Storm clouds gather over research costs. *Nature*, *350*(6319), Article 6319. https://doi.org/10.1038/350543a0

Azoulay, P., Graff Zivin, J. S., Li, D., & Sampat, B. N. (2019a). Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules. *The Review of Economic Studies*, *86*(1), 117–152. https://doi.org/10.1093/restud/rdy034

Azoulay, P., Graff Zivin, J. S., Li, D., & Sampat, B. N. (2019b). Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules. *The Review of Economic Studies*, *86*(1), 117–152. https://doi.org/10.1093/restud/rdy034

Azoulay, P., Graff-Zivin, J., Uzzi, B., Wang, D., Williams, H., Evans, J. A., Jin, G. Z., Lu, S. F., Jones, B. F., Börner, K., Lakhani, K. R., Boudreau, K. J., & Guinan, E. C. (2018). Toward a more scientific science. *Science*, *361*(6408), 1194–1197. https://doi.org/10.1126/science.aav2484

Bowers, J., Higgins, N., Karlan, D., Tulman, S., & Zinman, J. (2017). Challenges to Replication and Iteration in Field Experiments: Evidence from Two Direct Mail Shots. *American Economic Review*, *107*(5), 462–465. https://doi.org/10.1257/aer.p20171060

Buck, S. (2022, June 28). *The Efforts to Reform Research Bureaucracy To Date – Good Science Project*. https://goodscienceproject.org/articles/the-efforts-to-reform-research-bureaucracy-to-date/

Burtless, G. (2019, October 23). Experimental economists win Nobel Prize (and deserved to win). *Brookings*. https://www.brookings.edu/opinions/experimental-economists-win-nobel-prize-and-deserved-to-win/

Calsamiglia, C., Fu, C., & Güell, M. (2020). Structural Estimation of a Model of School Choices: The Boston Mechanism versus Its Alternatives. *Journal of Political Economy*, *128*(2), 642–680. https://doi.org/10.1086/704573

Centers for Medicare & Medicaid Services. *Innovation Center*. Retrieved February 12, 2023, from https://innovation.cms.gov/

deGrazia, C., Pairolero, N., Pappas, P.-A., Teodorescu, M., & Toole, A. (2022, July 21). *Increasing the Representation of Women in the Patent System: Evidence from a Randomized Control Trial*. Innovation Growth Lab. https://www.innovationgrowthlab.org/content/increasing-representation-women-patent-system-evidence-randomized-control-trial

Framingham Heart Study. *About FHS*. Retrieved February 13, 2023, from https://www.framinghamheartstudy.org/fhs-about/

Kaiser, J. (2017, June 8). *Updated: NIH abandons controversial plan to cap grants to big labs, creates new fund for younger scientists*. https://www.science.org/content/article/updated-nih-abandons-controversial-plan-cap-grants-big-labs-creates-new-fund-younger

Kleven, H. J. (2018, July). *Language Trends in Public Economics*.
Lakdawalla, D. N., Sun, E. C., Jena, A. B., Reyes, C. M., Goldman, D. P., & Philipson, T. J. (2010). An economic evaluation of the war on cancer. *Journal of Health Economics*, *29*(3), 333–346. https://doi.org/10.1016/j.jhealeco.2010.02.006

Li, D. (2017). Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics*, *9*(2), 60–92. https://doi.org/10.1257/app.20150421

Lichtenberg, F. R. (2019). The long-run impact of new medical ideas on cancer survival and mortality. *Economics of Innovation and New Technology*, *28*(7), 722–740. https://doi.org/10.1080/10438599.2018.1557421

List, J. (2022). *The Voltage Effect*. https://www.thevoltageeffect.com/

Luca, M., & Bazerman, M. (2021). *The Power of Experiments*. The MIT Press. https://mitpress.mit.edu/9780262542272/the-power-of-experiments/

Murphy, K. M., & Topel, R. H. (2006). The value of health and longevity. Journal of Political Economy, 114(5), 871–904. https://doi.org/10.1086/508033 Myers, K. (2020). The Elasticity of Science. *American Economic Journal: Applied Economics*, *12*(4), 103–134. https://doi.org/10.1257/app.20180518

National Institutes of Health. (a). *Director's New Innovator Award Outcomes Evaluation*.

National Institutes of Health. (b). *NOT-OD-09-025: Enhancing Peer Review: The NIH Announces Enhanced Review Criteria for Evaluation of Research Applications Received for Potential FY2010 Funding*. Retrieved February 12, 2023, from https://grants.nih.gov/grants/guide/notice-files/not-od-09-025.html

National Institutes of Health. (c). *NOT-OD-22-049: Notice of Change to NIHs Policy on Special Council Review of Research Applications*. Retrieved February 12, 2023, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-22-049.html

National Institutes of Health. (2013a, June 26). *High-Risk, High-Reward Research Program*. https://commonfund.nih.gov/highrisk

National Institutes of Health. (2013b, June 26). *New Innovator Award Program*. https://commonfund.nih.gov/newinnovator/programhighlights

National Institutes of Health. (2016, March 7). *Our Stories*. https://www.nih.gov/about-nih/what-we-do/impact-nih-research/our-stories

National Institutes of Health. (2017, May 2). *New NIH Approach to Grant Funding Aimed at Optimizing Stewardship of Taxpayer Dollars*. https://www.nih.gov/about-nih/who-we-are/nih-director/statements/new-nih-approach-grant-funding-aimed-optimizing-stewardship-taxpayer-dollars

Pathak, P. A., & Shi, P. (2021). How well do structural demand models work? Counterfactual predictions in school choice. *Journal of Econometrics*, *222*(1, Part A), 161–195. https://doi.org/10.1016/j.jeconom.2020.07.031

Rosch, S. D., Cooper, J. C., Holt, C., Sproul, T. W., & Tulman, S. (2019). Risk Preferences and Demand for Crop Insurance. *2019 Annual Meeting, July 21-23, Atlanta, Georgia*, Article 291273. https://ideas.repec.org//p/ags/aaea19/291273.html

Sampat, B. N., & Lichtenberg, F. R. (2011). What Are The Respective Roles Of The Public And Private Sectors In Pharmaceutical Innovation? *Health Affairs*, *30*(2), 332–339. https://doi.org/10.1377/hlthaff.2009.0917

Silber, J. H., Bellini, L. M., Shea, J. A., Desai, S. V., Dinges, D. F., Basner, M., Even-Shoshan, O., Hill, A. S., Hochman, L. L., Katz, J. T., Ross, R. N., Shade, D. M., Small, D. S., Sternberg, A. L., Tonascia, J., Volpp, K. G., Asch, D. A., & iCOMPARE Research Group. (2019). Patient Safety Outcomes under Flexible and Standard Resident Duty-Hour Rules. *The New England Journal of Medicine*, *380*(10), 905–914. https://doi.org/10.1056/NEJMoa1810642

Song, Z., & Baicker, K. (2019). Effect of a Workplace Wellness Program on Employee Health and Economic Outcomes: A Randomized Clinical Trial. *JAMA*, *321*(15), 1491–1501. https://doi.org/10.1001/jama.2019.3307

The Abdul Latif Jameel Poverty Action Lab (J-PAL). *Handbook of Field Experiments*. Retrieved February 12, 2023, from https://www.povertyactionlab.org/page/handbook-field-experiments

Thomke, S. H. (2020). *Experimentation works: The surprising power of business experiments*. Harvard Business Review Press.

United States Patent and Trademark Office. *Office of the Chief Economist*. Retrieved February 12, 2023, from https://www.uspto.gov/about-us/organizational-offices/office-policy-and-international-affairs/office-chief-economist

US Census Bureau. (a). *Census Evaluations and Experiments*. Retrieved February 12, 2023, from https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/evaluate/eae.html

US Census Bureau. (b). *Research*. Census.Gov. Retrieved February 12, 2023, from https://www.census.gov/topics/research.html

US Census Bureau. (2019, July 31). *Randomized Experiments, Evaluations, Administrative Records and Focus Groups*. Census.Gov. https://www.census.gov/library/stories/2019/07/how-we-develop-improve-the-census.html

US Department of Education. (2022, June 8). *Announcing the First-Ever Chief Economist at ED*. https://blog.ed.gov/2022/06/announcing-the-first-ever-chief-economist-at-ed/

US Government Accountability Office. (1994, July 26). *Peer Review: Reforms Needed to Ensure Fairness in Federal Agency Grant Selection*. https://www.gao.gov/products/pemd-94-1

Why economists are flocking to Silicon Valley. (2022, September 7). *The Economist*. https://www.economist.com/business/2022/09/07/why-economists-are-flocking-to-silicon-valley

Notes

---

[i] To the NIH's credit, a program evaluation of New Innovator awards was conducted in 2016. However, the researchers who conducted that evaluation were extremely limited in their ability to make claims about the *causal effect* of these awards. Another new grant structure introduced in 2007 was the K99 grant, intended to help prepare young investigators for tenure-track faculty positions and funding independence. The effectiveness of these grants is still unknown.

[ii] Note: The figure plots the annual publication counts of articles in the PubMed database which are flagged as "Randomized Control Trials" and are also flagged with either (1) the "Organization and Administration" Medical Subject Heading (MeSH) term, (2) the "Education" MeSH term, or (3) any of the following MeSH terms related to science-of-science topics: Peer review; Research Support as Topic; Scholarly Communication; Knowledge Discovery; Community-Based Participatory Research. For reference, PubMed contains more than 27,000 randomized control trials published in 2018.

[iii] For example, consider a study to examine the effect of NIH investigators receiving an additional $250,000 beyond the traditional size of an R01 grant, which is approximately a 25% increase relative to the status quo, on the investigators' research output. Under standard assumptions, an experiment would need to recruit roughly 3,000 scientists to participate. Assuming half would randomly receive the treatment, the total cost of treatment would be $375 million. For even a large team of social scientists, this amount of funding is virtually impossible to consider. However, for the NIH this amounts to about 1% of *a single year's* extramural research budget. Compared to the social value such an experiment could generate – by helping us understand the return to public investments in science – these costs may be trivial.

[iv] The NIH has a long history of investing in ideas that take years to decades to pay off (with great rates of return). One famous example is the Framingham Heart Study, first funded by the NIH more than 70 years ago, which continues to provide important insights into cardiovascular health.

[v] As the Federation of American Societies for Experimental Biology stated in 2015, "…experimentation and analysis should be undertaken to assess the impact of efforts to broaden the distribution of research funding."

[vi] For more on businesses' recruitment of Ph.D. social-scientists and their growing adoption of large-scale high-stakes experiments, see "The Power of Experiments" (2021) by Michael Luca and Max Bazerman, or "Experimentation Works: The Surprising Power of Business Experiments" (2020) by Stefan Thomke.