Reviewed Preprint
v1 • October 2, 2025
Not revised

Reviewed Preprint
v2 • June 10, 2026
Revised by authors

For correspondence:
bihutchins@wisc.edu

Competing interests: BIH contributed to the development of the NIH iCite database, which was used for data collection and analysis.

Funding: See page 20

Reviewing editor: Peter Rodgers, eLife, United Kingdom

Comparing the outputs of intramural and extramural grants funded by National Institutes of Health

Xiang Zheng, Qiyao Yang, Jai Potnuri, Chaoqun Ni, B Ian Hutchins
Information School, University of Wisconsin-Madison, Madison, United States

eLife Assessment

This **important** study used five metrics to compare the cost-effectiveness of intramural and extramural research funded by the National Institutes of Health in the United States between 2009 and 2019. They found that each type of research had its own set of strengths: extramural research was more cost-effective in terms of publications, whereas intramural research was more cost-effective in terms of influencing clinical work. The evidence supporting these findings is **solid**.

https://doi.org/10.7554/eLife.108929.2.sa4

Abstract

Funding agencies use a variety of mechanisms to fund research. The National Institutes of Health in the United States, for example, employs scientists to perform research at its own laboratories (intramural research), and it also awards grants to pay for research at external institutions such as universities (extramural research). Here, using data from 1594 intramural grants and 97054 extramural grants funded between 2009 and 2019, we compare the scholarly outputs from these two funding mechanisms in terms of number of publications, relative citation ratio and clinical metrics. We find that extramural awards are more cost-effective for producing outputs commonly used for academic evaluation, such as publications and citations (per dollar), while intramural awards are more cost-effective for generating research that influences future clinical work, more closely in line with the agency's health goals. These findings provide evidence that institutional incentives associated with different funding mechanisms drive their comparative strengths.

Introduction

National funding agencies have a responsibility to ensure that the research projects they fund are successful, as measured by various metrics (such as number of scientific papers, number of early-career researchers trained, and societal and/or economic impact). Many of these agencies run their own research laboratories (intramural research) and also fund research at universities and other institutions (extramural research). Certain aspects of the grant funding system have been the focus of research, such as publication of highly influential papers (Azoulay, Zivin, & Manso, 2009) risk management (Goldstein & Kearney, 2020), funding disparities (T. A. Hoppe et al., 2019) and diminishing marginal returns to funding directed to single labs (Lauer, Roychowdhury, Patel, Walsh, & Pearson, 2017; Wahls, 2018a, 2018b) However, the relative merits of intramural and extramural funding have received little attention to date.

The United States National Institutes of Health is one of the largest funders of research in the world. It comprises 27 institutes and centers (24 of which fund grants), each with its own research agenda often related to a specific disease (such as the National Cancer Institute and the National Institute of Allergy and Infectious Diseases). In Fiscal Year 2022, the NIH spent approximately $5 billion on intramural research at its own laboratories, and $39 billion on extramural research at universities, medical schools and research institutes across the US ("NIH Budget," 2025). Most applications for extramural grants are peer reviewed and assigned a percentile ranking of overall impact merit score by a Study Section ("What happens to your application during and after review?," 2025), with the relevant NIH institute or center making a final decision on which applications are funded based on those scores. Intramural research is conducted in government laboratories run by Senior Investigators (Sampat, 2012), supported by a combination of staff scientists and postdoctoral fellows. Senior investigators do not have to apply for grants, but external Boards of Scientific Counselors review their performance on a regular basis (usually every four years). Questions about the most effective portfolio management approaches have been an ongoing source of contention (Supplemental Text).

A potential advantage of the extramural approach is that cost-sharing at universities may increase the scientific return on the NIH's investment. An advantage of the intramural approach are that NIH has the direct ability to hire scientists whose research closely aligns with agency goals, and researchers do not need to devote time and effort on preparing and submitting grant applications. Here, using data from 1594 intramural grants and 97054 extramural grants, we compare the distribution of research topics funded by the two mechanisms (Figure 1), and the values of the grants awarded under each mechanism (Figure 2). The large difference in the number of intramural vs. extramural awards is reflective of the larger size of intramural awards compared to extramural awards, combined with the much smaller proportion of the funding portfolio dedicated to intramural research. We also use five metrics to analyze the 621,138 papers that acknowledge at least one of these projects: number of papers; relative citation ratio, which is a field- and time-normalized measure of scientific influence; approximate potential to translate, which is a machine-learning prediction that a given paper will be cited by a clinical article; total clinical citation counts; and a binary measure of the number of papers that received at least one clinical citation (Figure 3). We also compare the cost effectiveness of the two approaches by, for example, comparing the average cost of each paper published (Figure 4), and also explore the influence of various factors, such as the high costs associated with human-focused research (Figure 5).

Results

Comparison of research topics for intramural and extramural projects

Our analyses reveal differences in the research topics investigated by researchers funded by the two mechanisms. By clustering projects based on their titles and abstracts, we find that intramural projects yield a higher-than-average number of projects on viral infection, cancer and genes, and a lower-than-average number of projects on adolescents, brain studies and maternal health (Figure 1 and Supplemental Figures 1 and 2). The overrepresentation of viral research is likely because of the outsize investment toward the intramural Vaccine Research Center, and the cancer/genetics overrepresentation due in part because National Cancer Institute intramural investigators conduct research at that institute as well as at the NIH Clinical Center and Center for Cancer Research for their human genetics work ("National Cancer Institute (NCI) Center for Cancer Research," 2025). Intramural projects begin in our dataset with a lower proportion of the total projects considered (Supplemental Figure 3). This is because projects that were matched in 2008 were excluded as being possible continuations of existing intramural projects. Because these new projects are more typically associated with hiring of new principal investigators than in the extramural program, where established investigators can apply for a new grant at any time, the ramp-up of these intramural projects occurred more slowly.

Comparison of funding for intramural and extramural projects

Next we compared funding for intramural and extramural research projects, and found that annual funding for extramural research projects was consistently lower than that for intramural research projects (Figure 2). For all comparisons, we use inflation-adjusted total costs, which for extramural grants includes indirect as well as direct costs. Indeed, the average funding for

Figure 1. **Research topics for intramural and extramural projects.**

The topics listed were identified by clustering projects based on their titles and abstracts via Word2Vec (see Methods). The relative ratio of intramural projects for each topic was calculated by taking a ratio of the proportions of total grants a topic represented in the intramural vs. extramural portfolios. A relative ratio >1 signifies a higher share of intramural projects on that topic relative to their share across all topics. For example, if a topic comprised 10% of grants in the intramural portfolio but only 5% of grants in the extramural portfolio, this would represent a 2:1 intramural:extramural relative ratio, or 2.0.

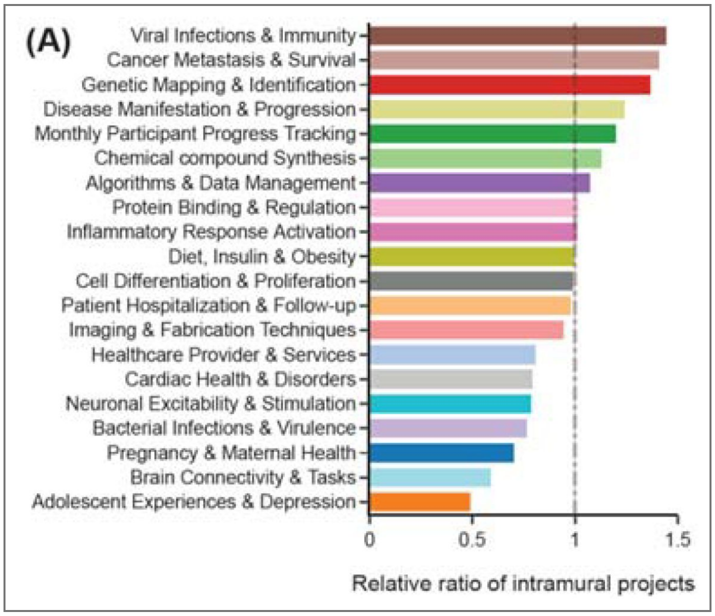
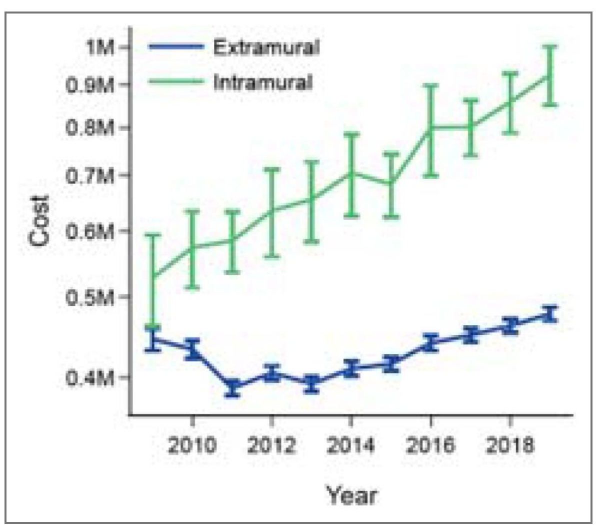


Figure 2. **Project funding for intramural and extramural projects.**

Mean project cost (on a log scale) versus year for intramural grants (green) and extramural grants (blue) between 2009 and 2019. Error bars denote 95% confidence intervals. Total costs were used rather than only direct costs in order to fully account for the degree of government investment. Error bars are larger for intramural data because of the smaller total number of awards (98,648 extramural and 1594 intramural).



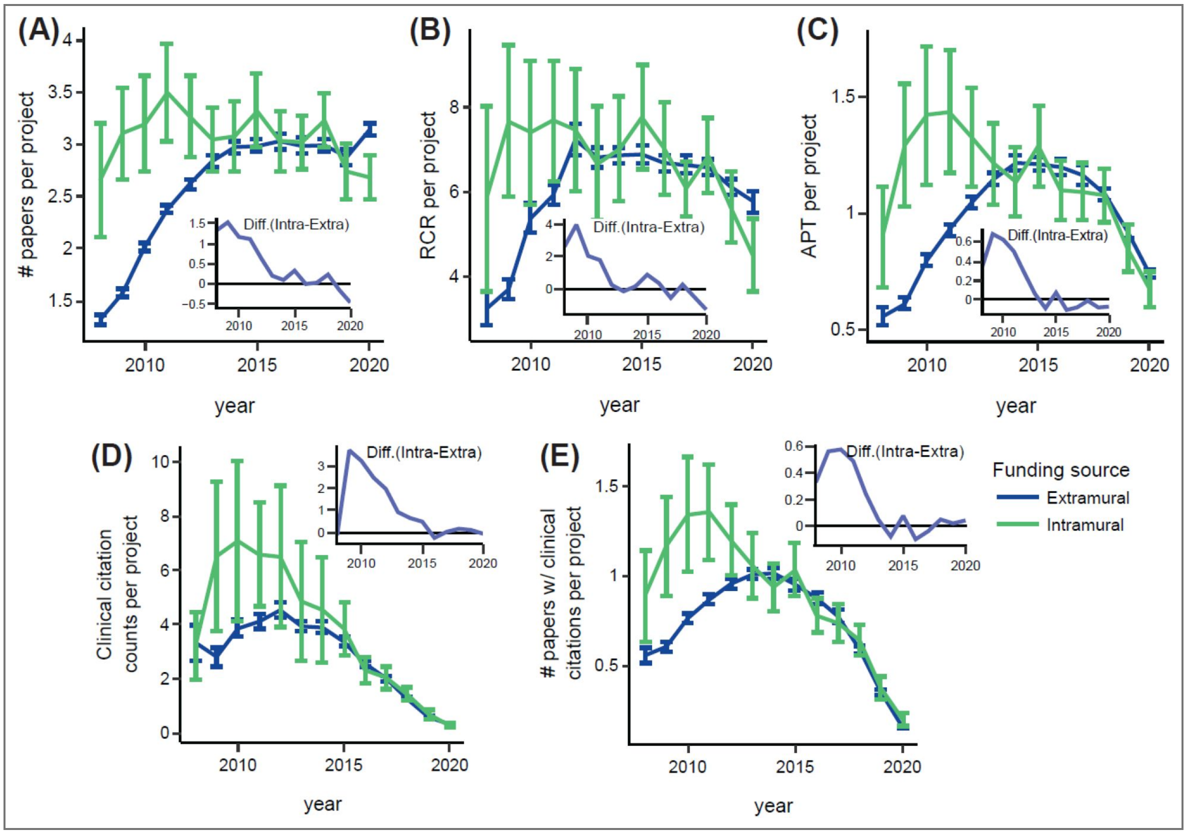


Figure 3. **Annual outputs from intramural and extramural projects.**

(A) Mean number of papers per project for intramural projects (green) and extramural projects (blue) between 2008 and 2020. The difference (inset) was close to 1.5 papers per project in 2008, but this gap closed over time. **(B)** Relative citation ratio per project. **(C)** Approximate potential to translate (ATP) per project. **(D)** Clinical citation counts per project. **(E)** Number of papers with at least one clinical citation per project. Error bars denote 95% confidence intervals. Error bars are larger for intramural data because of the smaller total number of awards (98,648 extramural and 1594 intramural).

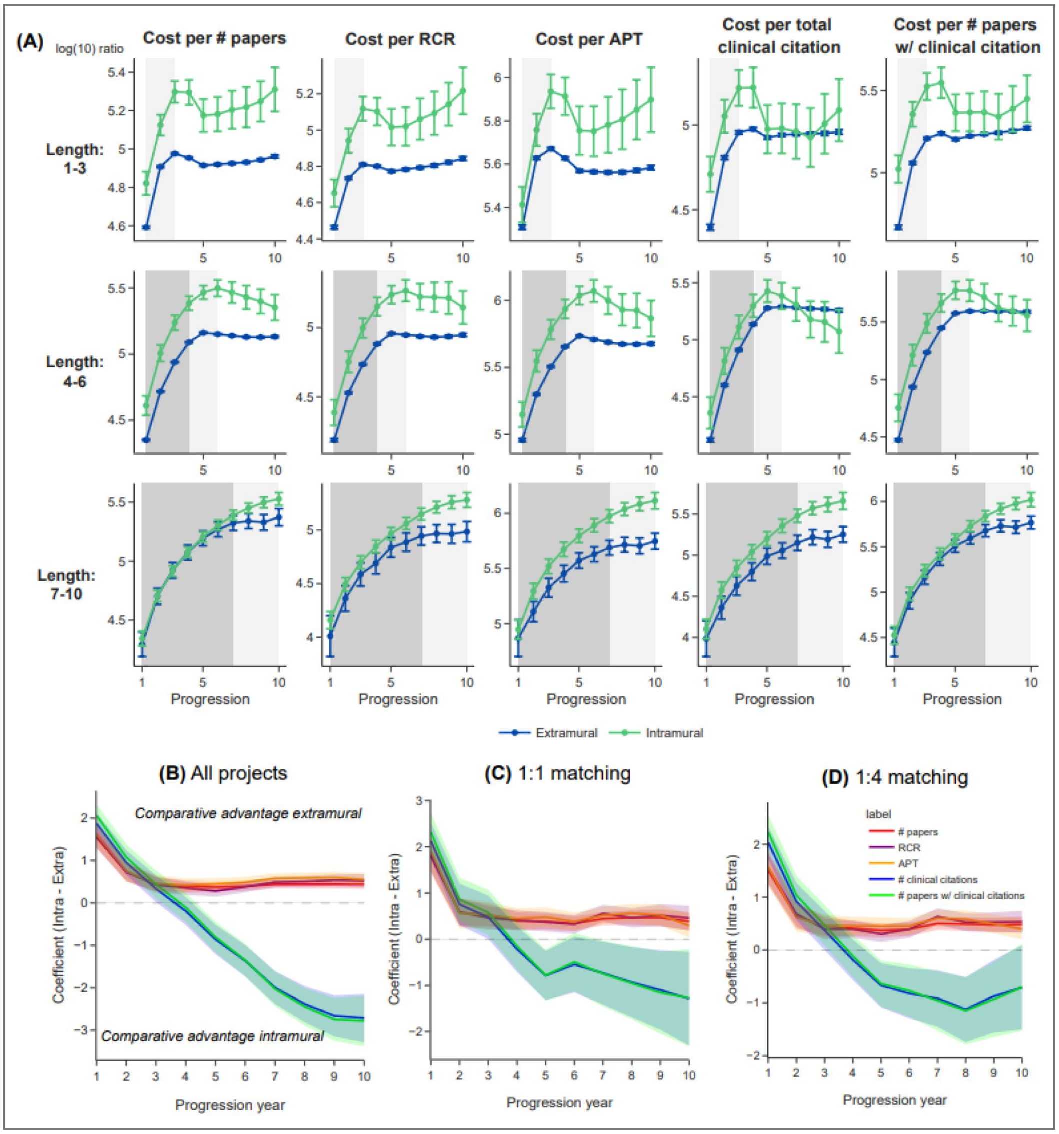


Figure 4. **Cost effectiveness of intramural and extramural projects.**

(A) A measure of cost effectiveness versus progression (i.e., year of grant) for intramural research (green) and extramural research (blue), for projects of different durations: 1–3 years (top row), 4–6 years (middle), and 7–10 years (bottom). These regressions do not control for other characteristics, but rather represent the raw ratios. For the first column, the Y-axis displays $\log_{10}$(ratio) +1, where ratio is the cumulative total costs to the cumulative total research output for each metric (cost:output, for the first column output = #papers); error bars denote the 95% confidence intervals. The remaining columns show measures of cost effectiveness for relative citation ratio, approximate potential to translate, total clinical citation counts, and a binary measure of clinical citations. To account for the fact that many papers are published after funding for the relevant grant has ended, grant amounts were multiplied by a deflator – this represents the proportion of papers published to date against the anticipated number of future publications, as determined by empirical measurements (Supplemental Table 1). In most cases, according to this analysis, extramural research is more cost effective than intramural research when observing uncontrolled regressions. **(B–D)** Linear regression results of the cost efficiency of research output measures against project types (intramural vs. extramural). The regression model was fitted for each year of the project's progression. Unlike panel (A), this regression model controls for grant, investigator, and collaboration characteristics in order to obtain a more accurate estimate of the relative cost efficiency of intramural vs. extramural projects. The Y-axis coefficient indicates the mean disparity in research output between intramural and extramural projects, controlling for these other variables (see Methods). Because there might be covariates that could confound the data, separate regressions were conducted for all projects (B, the default), and for balanced projects using 1:1 propensity score matching (1 extramural grant for every 1 intramural grant) in order to compare grants that were the most similar to reduce the influence of unobserved covariates (C) and (D) similarly to (C) 1:4 propensity matching as a robustness check.

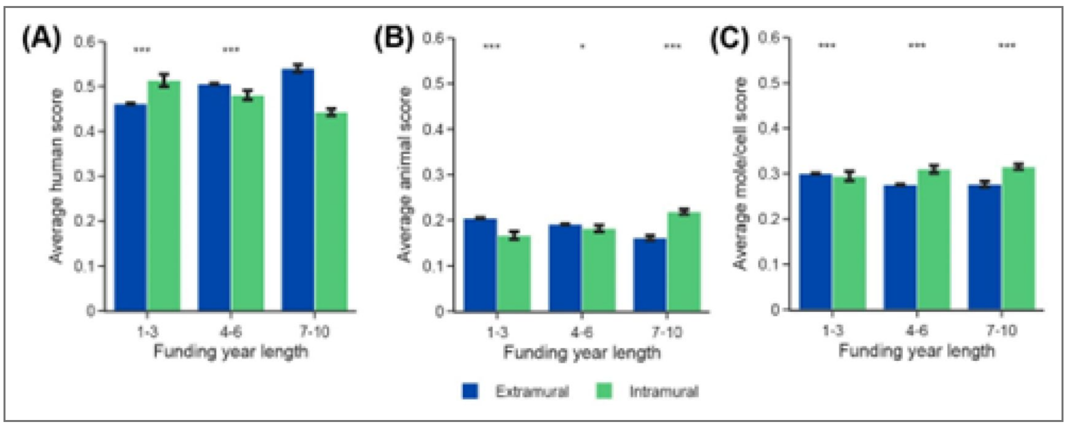


Figure 5. Comparison of scores for human-focused research, animal research, and molecular/cellular research for intramural and extramural projects.

(A-C) These represent the average Human, Animal and Molecular/Cellular scores for publications funded by extramural vs. intramural grants, respectively, which were downloaded from iCite (B. I. Hutchins, Davis, et al., 2019; iCite et al., 2019). **(A)** Average scores for human-focused research for intramural research (green) and extramural research (blue) for projects of different durations: 1–3 years (left), 4–6 years (middle), and 7–10 years (right).. **(B)** Average scores for animal research. **(C)** Average scores for molecular/cellular research. Mann-Whiteney U tests were conducted to test the difference between the scores for intramural and extramural projects. *** $p<0.001$, ** $p < 0.01$, * $p < 0.05$. No asterisk indicates the difference is not statistically significant.

extramural projects remained roughly constant, at below $500k per year between 2009 and 2019, whereas the average funding for intramural projects increased from about $0.42 million to about $0.45 million over the same period. Given that NIH funding for intramural research has remained relatively constant as a percent of total funding over the years, this indicates larger single awards for intramural research while extramural investigators may increasingly require multiple concurrent grants to sustain their labs.

This finding is consistent with the intramural research requiring higher financing due to the absence of cost-sharing with universities. It also reflects observations that extramural researchers need multiple federally funded grants to sustain a lab. While such institutions do receive audited indirect costs to cover overhead associated with research (Culliton, 1992), recent research indicates that significant cost-sharing greatly influences scientific productivity. Specifically, the extent of universities subsidies for student labor costs is a direct factor in university productivity (Zhang, Wapman, Larremore, & Clauset, 2022). In contrast to extramural awards, intramural funding as reviewed by the Board of Scientific Counselors and each Institute's scientific Director, can in principle fund an entire lab through a single, larger award. This frees the intramural investigators from the time commitment of securing grants. However, running in-house labs does entail that there is no cost sharing with external institutions, potentially raising the cost of such research (Culliton, 1992; Korn, 2015; Macilwain, 1999; Zhang et al., 2022).

Comparison of outputs from intramural and extramural projects

Next we used five metrics related to publications and citations – number of papers; relative citation ratio (Arabi, Ni, & Hutchins, 2025; B. I. Hutchins, Hoppe, Meseroll, Anderson, & Santangelo, 2017; B. I. Hutchins, Yuan, Anderson, & Santangelo, 2016); approximate potential to translate (B. I. Hutchins, Davis, Meseroll, & Santangelo, 2019); total clinical citation counts (B. I. Hutchins, Baker, et al., 2019; B. I. Hutchins, Davis, et al., 2019; iCite, Hutchins, & Santangelo, 2019); and a binary measure of clinical citations – to compare the outputs of intramural and extramural projects on a year-by-year basis (Figure 3). For all of the metrics apart from total clinical citation counts, intramural research scored highest before 2010, but the gap between intramural and extramural research had closed by 2020. This increased early productivity for intramural projects may reflect the extra time intramural investigators save because they do not have teaching and grant writing responsibilities. Because this dataset is truncated (e.g. only papers and their citations through 2020 are considered), we observe a decrease in most outcome measures in Figure 3 as the end of this window approaches. In our previous study (B. I. Hutchins, Davis, et al., 2019), we observed a lag of approximately seven years for clinical citations to accrue. We see the same trend in this dataset (Figure 3 D-E), approaching zero clinical citations as our time window approaches. The Approximate Potential to Translate metric, a prediction of future clinical citations, also decreases but to a lesser extent. This is because the forward citation network is used for making these predictions.

Comparison of cost effectiveness for intramural and extramural projects

So far we have seen that the average funding for intramural projects is higher than that for extramural projects (Figure 2), and that intramural projects also score higher than extramural projects on the five metrics for the outputs from a project that we computed (Figure 3). Next, therefore, we compare cost effectiveness by calculating the average cost of each published paper, and likewise for the other four publication/citation metrics (Figure 4A). We do this for projects of different durations: 1–3 years, 4–6 years, and 7–10 years.

Our analysis suggests that extramural research is more cost-effective than intramural research when considering number of papers, relative citation ratio and approximate potential to translate. However, when considering metrics based on clinical citations, the gap was smaller, and for some project durations, intramural research was as cost-effective as extramural research.

Some types of research are more expensive rather others (for example, animal research is more expensive than cell biology research, and human-focused research has higher regulatory overhead costs due to increased ethical concerns), so we decided to explore if such factors could explain some of the differences that we had observed between intramural and extramural projects. To do this we conducted a regression analysis that controlled for project topic and various factors related to the principal investigator who had received the grant.

Figure 4B shows a plot of the number of years that elapsed since the start of the grant (Progression year, x-axis) and the relative cost effectiveness of intramural vs. extramural grants at funding each metric measured (#papers, red; RCR, purple; APT, yellow; #clinical citations, blue; and #papers with at least one clinical citation, green). Each curve was a subset of the data frame focusing on each measure individually. These were generated by comparing the difference in regression coefficients of extramural vs. intramural research once controlling for grant, investigator and collaboration variables (for regression model details, see Methods). Curves above 1.0 indicate a comparative cost effectiveness advantage for extramural grants on that measure, while those below 1.0 indicate a comparative cost effectiveness advantage for intramural grants on that measure. We observe that over time, extramural and intramural grants show opposite patterns. Extramural grants excel in terms of cost advantage at measures commonly used in academic performance assessment (e.g. number of papers and citations). Intramural grants in comparison excel in cost advantage in measures more closely aligned with the NIH agency mission (i.e. knowledge that informs work to improve human health via knowledge flow to clinical trials).

We also performed propensity score matching between intramural and extramural projects by pairing each intramural project with its closest extramural counterpart (Figure 4C), and with its four closest extramural counterparts (Figure 4D). This allows us to sample intramural and extramural grants that are similar to one another at a 1:1 (e.g. 1594 intramural grants and 1594 matched extramural grants) based on their similarity in terms of topic area, prior publication record, and prior collaboration history. This propensity score matching approach accounts for potentially hidden covariates and reduces their influence. We observe the same pattern using 1:1 propensity score matching (Figure 4C) as we did with the full dataset (Figure 4B). Likewise, using propensity score matching of 1:4 (i.e. 4 matched extramural grants for every 1 intramural grant) to increase the sample size of extramural awards yielded the same result (Figure 4D). Taken together, these results indicate that regardless of sampling strategy, cost effectiveness aligns with the primary missions of the institutions at which investigators are housed (maximum knowledge generation and flow for academic institutions, and clinically relevant knowledge generation for NIH intramural research).

It is estimated that universities do not fully recover expenditures through indirect costs. The magnitude of this effect is estimated to be approximately 30% (Droegemeier, 2017). We therefore asked whether these trends hold when indirect costs are inflated by the same amount, to more completely reflect research investment expenditures. Supplemental Fig 4 shows that inflating indirect costs by 30% narrows the gap in estimated costs between intramural and extramural awards as shown in Figure 4A. Likewise, our regression analysis shows a larger cost effectiveness advantage for intramural awards in clinical citations, and a smaller but still significant advantage for extramural awards in number of papers and RCR (Supplemental Fig 4).

Comparison of scores for human-focused research, animal research, and molecular/cellular research

Next, we assign scores to papers and projects based on the extent to which they can be classified as human research, animal research or molecular/cellular research. Again, we do this for projects of different durations: 1–3 years, 4–6 years, and 7–10 years. Both intramural and extramural projects have the highest average scores for human research (Figure 5 and Supplemental Figure 5). However, long-duration intramural projects tend to have lower human scores, while long-duration extramural projects tend to have higher human scores. For animal and molecular/cellular scores, the trend is reversed: long-duration intramural projects tend to have higher animal and molecular-cellular scores, while long-duration extramural projects tend to have lower scores. The possibility that longer-term intramural projects are more human-focused, which might explain the clinical citation comparative advantage with respect to the extramural program, is therefore inconsistent with the data.

Discussion

Taken together, these results demonstrate comparative advantages for extramural and intramural funding mechanisms. In particular, extramural funding seems to excel at generating raw knowledge and facilitating its downstream flow. In contrast, intramural funding mechanisms seem to have a comparative advantage at generating research, basic or human-focused, that successfully informs downstream clinical research, aligning with the agency's mission. This could potentially be attributed to the selection process for directly hiring scientists whose research agendas closely match the agency's objectives, though this aspect wasn't directly assessed in our study. However, we do rule out an obvious explanation: that more human-focused work in the intramural program is more likely to be conceptually closer to clinical trials and, therefore, have a lower barrier to entry into clinical studies (Kim, Levine, Nehl, & Walsh, 2020;

Weber, 2013). Intramural research is characterized by long-duration projects in contrast to the extramural portfolio, and these appear less human-focused than the extramural portfolio. Even when accounting for hidden costs borne by universities (Supplemental Figure 4), the fundamental pattern holds: intramural and extramural mechanisms have distinct comparative advantages that align with their institutional incentives. This suggests that these differences may be primarily driven by structural features, like mission alignment, of these funding mechanisms.

This study is not without limitations. First, data about the intramural portfolio is only available from post-2008, which constrains the time frame for this study. Second, collaboration between intramurally and extramurally funded scientists introduces complexity to the comparative analysis, leading to the exclusion of jointly funded publications. Third, the primary analysis uses an agency perspective based on NIH expenditures recorded in RePORTER, which does not capture the full economic cost of extramural research borne by universities. To address this, we conducted a sensitivity analysis (Supplemental Figure 4) inflating extramural indirect costs to account for unrecovered expenditures. Finally, recent changes at the NIH and the Department of Health and Human Services (which oversees the NIH), as well as changes proposed by the Senate, could lead to significant changes in the budget and organization of the NIH, and also in the processes it uses to make funding decisions.

Critiques of NIH's grant review process often cite its conservatism, with a strong emphasis on preliminary data to mitigate project failure risks (Packalen & Bhattacharya, 2020). The recent creation of the Advanced Research Projects Agency for Health (ARPA-H) was in part for this reason (Collins, Schwetz, Tabak, & Lander, 2021). Although the high-risk, high-reward NIH portfolio seems to be largely effective at identifying and funding such projects (Tabak et al., 2019), its overall proportion of the total portfolio remains relatively small in favor of more traditional investigator-initiated research project grants. Because intramural researchers face retrospective rather than prospective review, this conservatism might be expected to manifest in a comparative advantage across a variety of measures for intramurally funded research. Competing theories suggest that extramural research may hold advantages on an investment-adjusted basis because of cost-sharing at universities, particularly for student labor (Zhang et al., 2022). Notably, intramural research focused on human/molecular or animal research seems to be particularly effective at generating clinically relevant research outputs (Supplemental Figure 5). Our findings reveal a nuanced reality: extramural institutions hold an edge in publication and citation rates aligned with their internal review procedures, while intramural research excels at stimulating bench-to-bedside translation on an investment-adjusted basis.

This analysis uses an agency-centered perspective to estimate comparative cost effectiveness, but there are other aspects of these two funding models that should be acknowledged (Drummond, Sculpher, Torrance, O'Brien, & Stoddart, 2005). First, while we estimated a more complete cost estimate through our modeling of a 30% indirect cost increase, this likely does not cover the complete degree of university investment in research. For example, startup costs are not modeled, nor are university investments that come from alternative funding streams like philanthropy, endowments, or state budgetary contributions. Extramural investigators also have hidden costs in the form of time spent preparing and managing grant applications, a cost that is not shared by intramural researchers. This represents an unreimbursed cost to the extramural funding system (Ioannidis, Gross, & Bergstrom, 2019), although this cost should be reflected in our productivity measures. Thus, our estimates of the comparative cost effectiveness of intramural research may be conservative. However, these alternative contributions to extramural research, which take the form of graduate student training (Zhang et al., 2022), faculty retention and infrastructure development (*Research Universities and the Future of America*, 2012), have real societal impacts. While the intramural program does train a limited number of graduate students, its postdoctoral workforce is its primary focus. Thus the extramural program's contributions to a more highly educated United States workforce pipeline should be taken into consideration as a strength alongside its research outputs.

The distinction between the NIH intramural and extramural programs is sometimes controversial. A 2024 Senate proposal from the Senate Health, Education, Labor, and Pensions Committee suggested reforms to both the extramural and intramural programs (Cassidy, 2024). Feedback from a Request for Information suggested that respondents believe that these two programs are in many cases functionally equivalent. One salient consideration was how to best differentiate the extramural and intramural portfolios. Recommended changes to the extramural program focused on stimulating innovation and encouraging high-risk, high-reward projects. Proposals for the intramural program focused on increasing collaboration among funding institutes, promoting replication studies, and introducing temporary rotating intramural investigators focused on addressing long term, unmet needs. Both programs were recommended to increase focus on basic research. NIH was also encouraged to improve transparency by providing the public with more data to conduct metascience studies, in order to learn from research failures and build on scientific successes.

The present work suggests two avenues to build on the comparative strengths of intramural and extramural funding mechanisms. First, although the Senate proposal seems to indicate that the NIH as a whole refocus on basic research, it is the extramural program that seems comparatively well positioned for this task. Federal expenditures per article and downstream knowledge flow are lower in the extramural portfolio. By contrast, the intramural program is well-positioned to build on its strengths informing clinical research during structural reforms, if the most effective parts of the research portfolio are targeted. Supplemental Figure 5 suggests that animal research or research that combines aspects of human, animal, and molecular/cellular biology simultaneously within this portfolio are particularly cost effective at this goal. This aligns well with a focus from the Senate proposal on interdisciplinary research, and these kinds of projects can be readily identified with the data that are already available. However, longer-term intramural projects show a decreasing correlation with human research activities (Figure 5). This trend may need to be reversed if the strengths in the intramural program are to be effectively aligned with long-term research horizons.

We find that the NIH extramural program is comparatively more cost effective at generating research products that are aligned with traditional markers of academic achievement, publications and citations. The intramural program is comparatively more cost effective at generating research outputs that align with the agency's stated mission, in this case, research that informs downstream clinical research. If this result generalizes to other agencies, it has implications for their portfolio management. The Department of Defense has a sizeable research portfolio in the biomedical space, most notably funded directly through the department itself or through the Congressionally Mandated Research Priorities program ("Congressionally Directed Medical Research Programs," 2025). These results imply that refocusing that research portfolio with a focus on basic research in the extramural portfolio and health-informing research in the intramural portfolio may build on the programs' respective incentive structures. However, the Department of Defense as well as other agencies like the National Science Foundation's Federally

Funded Research and Development Centers (FFRDCs) ("Federally Funded R&D Centers," 2026) or the Department of Energy's combination of national labs and FFRDCs may benefit from a refocus of their intramural programs toward research focused on advancing applied outcomes pertaining to their portfolios. If institutional incentive alignment is a significant contributor to the findings here, then these results are likely to be reflected in agencies whether or not their applied research goals are health-oriented.

Methods

Data

We collected the original NIH project data from NIH RePORTER ("NIH ExPORTER," 2025), which contains 433,930 projects with funding information spanning from 1985 to 2019. We identified projects' activity categories by looking up their first three letters in the project number (activity code). We classified the projects into intramural and extramural projects by the initial letter of their activity codes. Specifically, projects with an activity code starting with 'Z' were intramural projects, and other projects were extramural projects. Using this strategy, we identified 9,225 intramural projects and 424,705 extramural projects in the raw dataset. We retrieved the publication records for these projects by PMID indexed by PubMed (Medicine, 2020). These include mostly journal articles, although a small number of NIH-funded preprints are included as well (Funk, Zayas-Caban, & Beck, 2024; Hong, Hutchins, & Ni, 2026; Lindsay Nelson et al., 2022; L. Nelson et al., 2022).

The data cleaning process is as follows. Since the scientific focus of a study may drift over time, we dropped 70,297 projects with renewal records in our data. Second, considering intramural projects might change their activity categories and the three initial project number letters (e.g., ZIA changed to ZIH), we normalized 3,105 intramural project numbers by matching the rest of the project numbers to avoid inconsistency. Third, to focus on activity categories intended as research-oriented and exclude practice-oriented activity categories, at the activity level, we selected activity categories where at least 75% of projects had produced at least one paper. This step kept 106 activity codes (including six intramural activity codes). Although this process was designed to rule out intramural projects that had start dates prior to the beginning of data collection at 2009, it is possible that some intramural projects had earlier start dates than the records indicate. For example, if there was a gap year of funding in 2008, our ability to detect these might be more limited, or for example if the grant serial number changed. This is a technical limitation to our approach.

In this study, we focus on projects initially funded after 2008 and select the ten years from 2009 to 2019 as our analysis period. A total of 122,815 projects fell in this period. To remove the rest of potential non-research-oriented projects, at the individual project level, we selected 1% of projects with the highest cumulative ratio of funding and publication number and 1% of projects with the lowest. We then trained a random forest model to predict the projects most likely to be non-research-oriented based on their title and abstracts. Based on the predicted probability, we excluded 5% of intramural (84) and extramural (5,347) projects. The final analytical sample consists of 98,648 projects, including 97,054 extramural projects and 1,594 intramural projects, which produced 621,138 papers during our time window. Papers with both intramural and extramural funding were excluded. However, papers with multiple intramural funding links or multiple extramural funding links were included; most papers acknowledged only one project (Supplemental Figure 7).

Topic level ratios were calculated as a ratio of the proportions of total grants a topic represented in the intramural vs. extramural portfolios:

$$R = \frac{I_{\text{topic}}}{N_{\text{topic}}} / \frac{I}{N}$$

Where I is the number of intramural awards and N is the number of intramural plus extramural awards.

Cost effectiveness

We used the primary project costs (direct costs, indirect costs for extramural institutions, and total costs, which combines these two) listed on NIH RePORTER as the cost data source. Subproject costs were not calculated. This approach takes an agency perspective, measuring cost effectiveness in terms of NIH expenditures. This does not capture the full investment in research expenditures at extramural institutions, so we also conduct a robustness check of our analyses where indirect costs are inflated by 30% (Supplemental Figure 4), estimated to be close to the unrecovered costs that universities invest (Droegemeier, 2017). Given the influence of price changes and inflations over years, we converted total costs (direct costs plus indirect costs) at the 2015 price level using NIH's Biomedical Research and Development Price Index.

We used five paper-level metrics to measure the research output: number of papers; relative citation ratio (Arabi et al., 2025; B. I. Hutchins et al., 2017; B. I. Hutchins et al., 2016); approximate potential to translate (B. I. Hutchins, Davis, et al., 2019; Santangelo, 2017; Weber, 2013); total clinical citation counts (Travis A. Hoppe, Arabi, & Hutchins, 2023; B. Ian Hutchins, 2021; B. I. Hutchins, Baker, et al., 2019; "iCite," 2015); and number of papers once received clinical citations. A project's total research performance regarding a certain metric in one year is approximated as the sum of that metric for every paper published in that year.

Based on the project costs and research outputs, we calculated the cost per output as follows.

$$R_{ij} = \frac{C_{ij}}{O_{ij}} * \frac{P_{ij}}{TP_i}$$

Where R_{ij} is the cost per output for project i in year j, C_{ij} the cumulative sum of funding costs for project i up to year j, O_{ij} is the cumulative sum of a certain research performance metric for project i up to year j, P_{ij} is the cumulative number of papers for project i up to year j, and TP_i is the total number of papers for project i until 2020.

Regression analysis

We run the following regression model at the project level to estimate the differences between extra- and intramural projects for every year after the projects started.

$$\log(y_{it} + 1) = \beta_0 + \beta_1 I_i + \beta_2 T_t + \sum \beta p_i + \alpha r_t + \delta s_i$$

where y_{it} is project i's deflated cost efficiency regarding a certain research performance t years after the funding start year; I_i is whether project i is an intramural project (1 if yes); T_t is the length of funding years until that time point, equal to the minimum of t and the total funding years; p_i stands for the PI-related variables; r_t and s_i are the year's and project topic's fixed effects. We transformed the y_{it} into $\log(y_{it} + 1)$ to mitigate the impact of uneven distribution.

PI-related variables include the number of past publications, number of past projects, number of PIs, share of clinical papers, past publications' average relative citation ratio, publication experience, project experience, number of collaborators. We downloaded the PubMed Knowledge Graph datasets (Xu et al., 2020) to help extract the PI level variables. The dataset has disambiguated the authors of PubMed indexed publications and assigned unique identifiers to the authors. We matched both project numbers and paper author names with the datasets to find the PI's assigned unique identifiers. We successfully retrieved the PI information for 98803 (94.9%) projects. The PI-related variables before the project funding started were extracted for every project, which played the proxy role of the input for the projects.

Another control variable, project topic, is calculated by performing a K-means clustering based on the NIH spending categories for all the projects in the sample. Project topic is defined by unsupervised clustering of semantic embeddings (Afshar, Yang, Thebault-Spieker, & Hutchins, 2026; Travis A. Hoppe, Arabi, & Hutchins, 2022; T. A. Hoppe et al., 2019) rather than graph clustering (Davis et al., 2025; Ni & Hutchins, 2025). To restrict the dimensionality, the 100 most frequent NIH spending categories were used in clustering, which cover 96.4% of all projects. We tried k=3, 4...10 and finally selected k=5 which generated the highest silhouette score.

As a robustness check, we used propensity score matching to reduce the potential confounding biases that may affect the outputs of interest and increase the comparability between intramural and extramural projects. For every year after the projects started, we used all project-level variables, including the length of funding years until that time point, PI-related variables, and project topics, to predict the propensity score of each project by fitting a logistic regression model. The propensity score shows the probability of a project to be an intramural project, based on the observable variables. For each intramural project, we selected one and four extramural projects, respectively, with the nearest propensity scores from a pool of extramural projects with the same funding start year. The regression model was run on the PSM sample again to check the robustness of previous results.

The PSM steps are as follows.

1. Estimate propensity scores. The study fits a propensity model (logistic) using covariates like productivity, collaboration, PI history, project duration, and project topic dummies. Multiple models are fit with balance checking for robust scoring.
2. Match within the same year: Treated and control records are paired only within the same year. For each treated project, the four closest controls are retained, and only very tight matches (diff < 0.001) are kept to enforce near-identical propensity scores.
3. Construct the matched cohort: Treated and control project IDs from the matched pairs are combined into a single dataset, forming the final matched sample.
4. Run outcome regressions on matched data: For each progression year, this study uses Stata reghdfe to run on log-transformed outcomes, with the project type as the key regressor and controls for productivity, collaboration, PI history, project duration, and project topic. The regressions contain fixed effects of project fiscal years.
5. Collect and visualize results: Coefficients, p-values, confidence bounds, and sample sizes are extracted per outcome and progression years, then plotted across progression years with shaded confidence bands and a zero reference line.

Paper features

We used the concatenated documents of a paper's title and abstract to train a word2vec model (Analysis, 2018; Analysis, Intelligence, & Institute, 2019) to classify the papers into clusters (Afshar et al., 2026; T. A. Hoppe et al., 2019). We removed common stop words, punctuation, and content lacking semantic information before training. During clustering, each paper's document is represented as a 300-dimension vector by summing its each unique word's vector weighted by its IDF. Principal component analysis (PCA) dimensionality reduction is applied to these 300-dimensional vectors to identify the 25 most influential components. We finally performed spectral clustering method using the document vectors and extract highly-frequent words to determine the cluster property and labels. Word2Vec nearest neighbor terms were uploaded to ChatGPT 3.5 to develop more human-readable labels.

Supplemental Materials

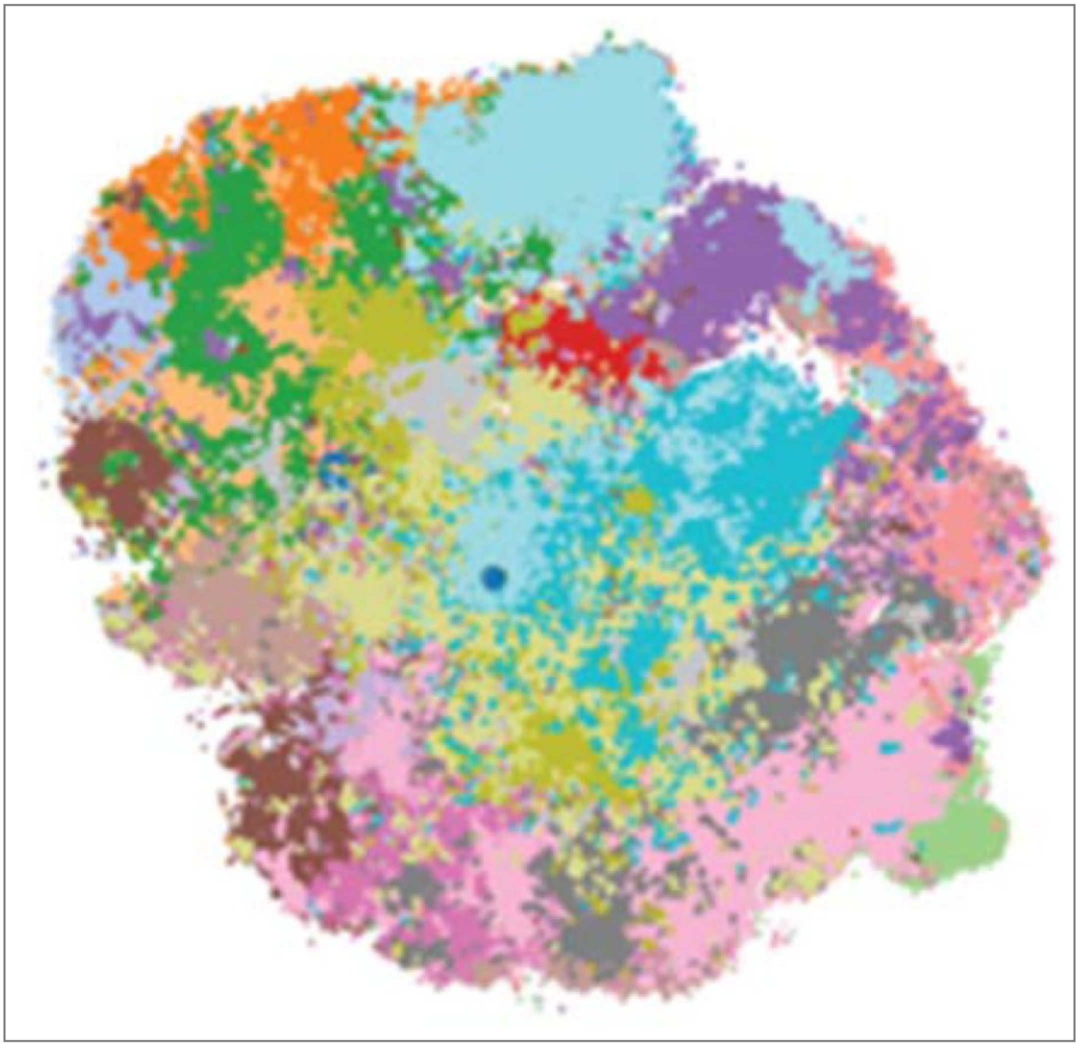
Supplemental Text

Questions about the most effective approaches to structure portfolio management for science funders have been a source of contention. This is primarily due to the conflicting priorities among government officials, the mission of funding agencies, and the perspectives of scientific researchers (Goldstein & Kearney, 2020). While the 2018 Evidence Act (Abraham & Haskins, 2017; Young, 2021) mandates that all science funders incorporate data-driven decision-making, the U.S. Congress played a significant role in catalyzing such efforts, particularly at the National Institutes of Health (NIH), through the establishment of divisions such as the Office of Portfolio Analysis in 2011 (Department, 2011). The division was created to advance these data-driven initiatives even before they were broadly implemented across other federal agencies. Consequently, NIH serves as an excellent case study for policy examination, given its more extensive and robust data infrastructure compared to other agencies.

A pressing question that often surfaces, particularly when facing inflation-adjusted budgetary declines, concerns the comparative efficiency of externally funded grants, usually awarded to universities, medical institutions, and research centers, in contrast to intramurally funded projects where scientists are employed directly as government personnel and conduct research within federal facilities. The 2013 sequester (Fox, 2013), a budget reduction mechanism that abruptly removed a sizeable fraction of government funding for scientific research, revealed significant contention in the scientific community about the extent to which extramural versus intramural funding should shoulder the burden of budgetary declines (Scientopia, 2014).

Various theories exist to highlight the respective merits of these two funding models. Extramural institutions are thought to engage in extensive cost-sharing that might reduce the degree of government investment necessary to stimulate scientific advancement (Culliton, 1992; Korn, 2015; Macilwain, 1999). In effect, despite the negotiated indirect costs that are paid to offset institutional overhead that supports scientific research at extramural institutions, these institutions often contribute additional resources that foster science advance. Research indicates that institutional contributions, particularly in terms of trainee labor (Zhang, Wapman, Larremore, & Clauset, 2022), are important for stimulating scientific productivity, supporting this theory. Moreover, it is essential to recognize the substantial portion of extramural funding typically dedicated to training students and early career researchers. This investment not only aids in producing the next generation of researchers in the field but also contributes to the long-term sustainability of the research workforce (Harris, 2014). On the other hand, the direct hiring of scientists by the government under the intramural funding model allows for the selection of researchers whose research agendas more closely align with the agency's mission. Furthermore, despite shouldering the entire cost of intramural research, intramural PIs are freed from the time and resource burdens associated with grant applications. This freedom allows them to focus entirely on advancing scientific knowledge in their respective fields. Nonetheless, intramural grants may encounter constraints on autonomy due to their affiliation with a larger government institution, potentially restricting their freedom through manuscript clearance processes. This affiliation also implies that intramural researchers may not be entirely shielded from potential bureaucratic hurdles and unwarranted administrative burdens that can impede the progression of scientific endeavors. Therefore, each funding approach has unique strengths and considerations, making it essential to carefully weigh the advantages and disadvantages of extramural and intramural funding when allocating resources for scientific research.

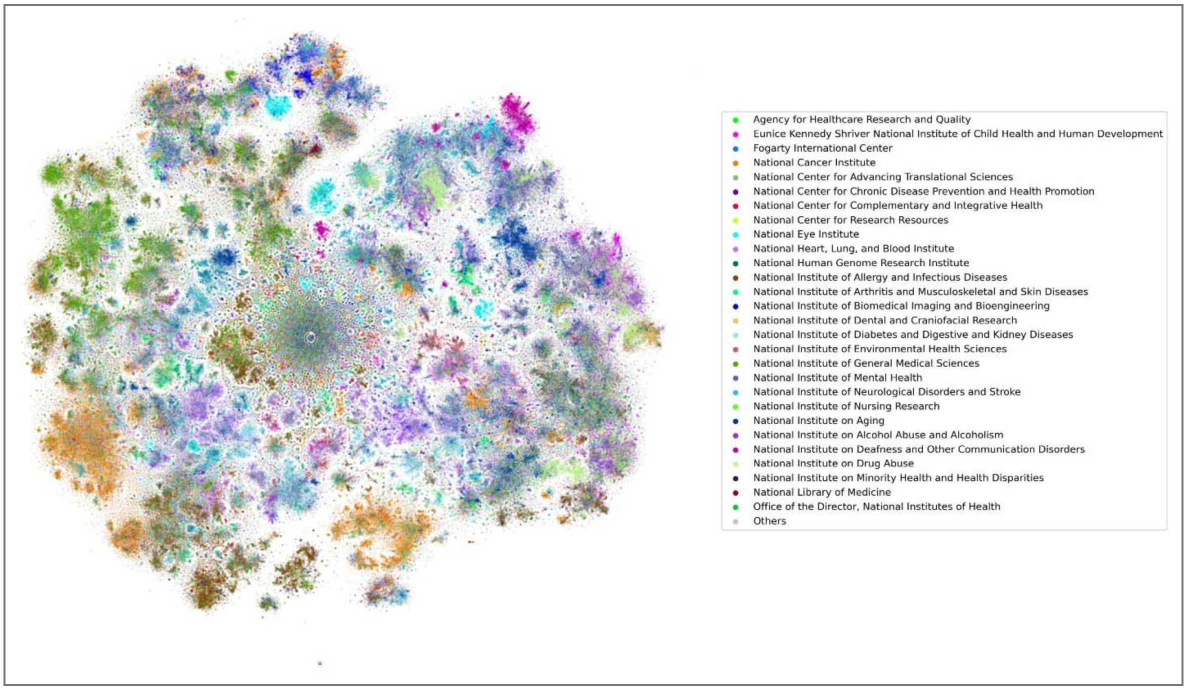
Supplemental Figures



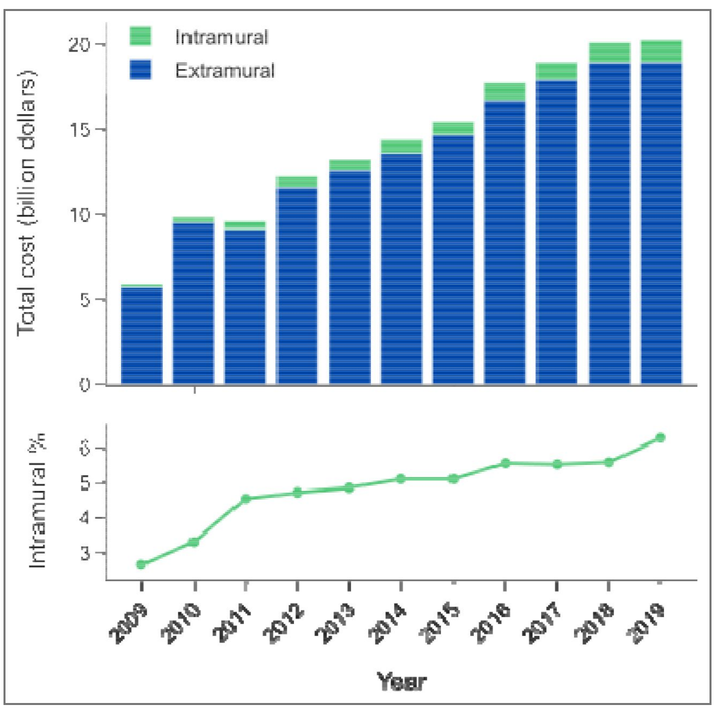
Supplemental Figure 1. T-SNE plot illustrating the distribution of topic clusters, with colors consistent with those in Figure 1. This shows a dimensionality reduction of the locations of these grants in Word2Vec space, indicating their relative proximity to one another. Grants with similar semantic meaning appear closer to one another, while those with less semantic similarity will appear farther away. Similar grants tend to cluster together in T-SNE visualizations.

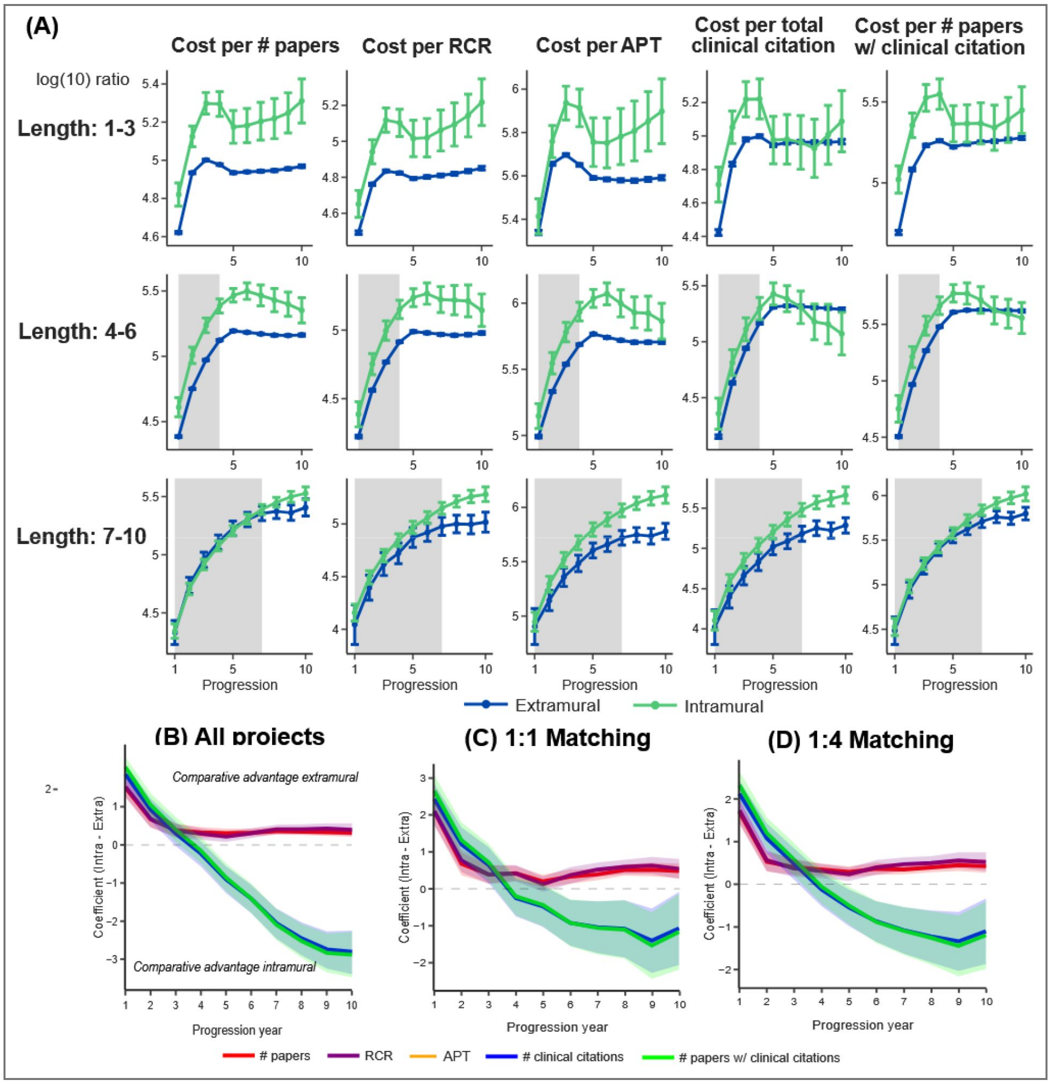
Supplemental Figure 2. **T-SNE plot illustrating the distribution of grants in NIH institutes, using the same methodology as Supplemental Figure 1.**

This shows a dimensionality reduction of the locations of these grants in Word2Vec space, indicating their relative proximity to one another. Grants with similar semantic meaning appear closer to one another, while those with less semantic similarity will appear farther away. Similar grants tend to cluster together in T-SNE visualizations.



Supplemental Figure 3. **Cost breakdown for our sample of projects (see Methods) by Extramural or Intramural origin.**



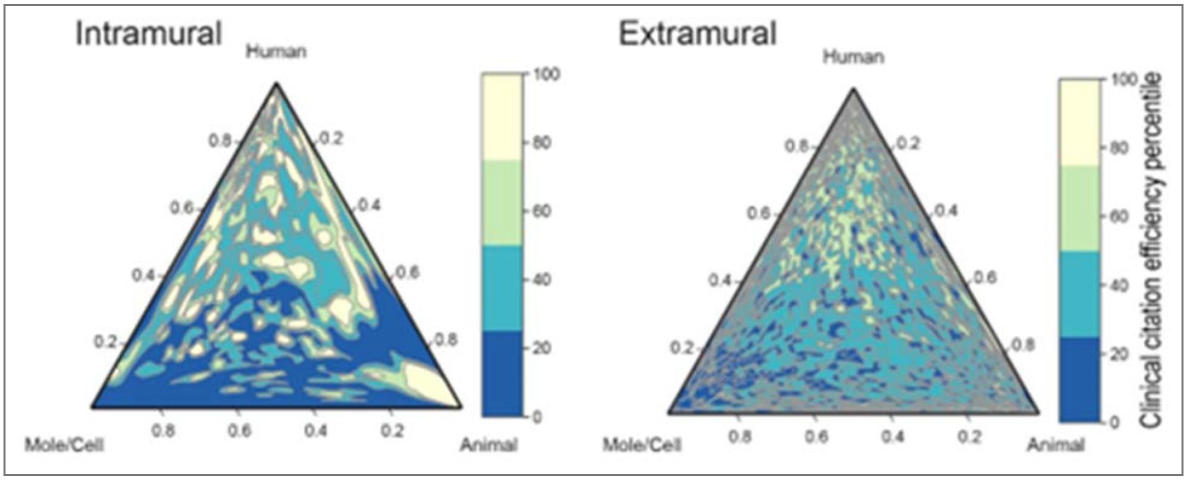


Supplemental Figure 4. Cost effectiveness of intramural and extramural projects, with Indirect costs for Extramural projects inflated by 30%.

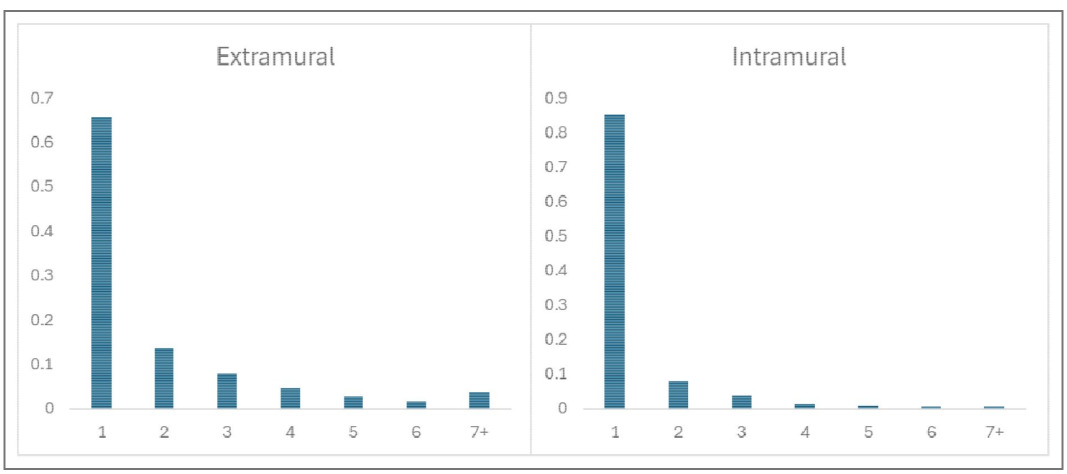
(A) A measure of cost effectiveness versus progression (i.e., year of grant) for intramural research (green) and extramural research (red), for projects of different durations: 1–3 years (top row), 4–6 years (middle), and 7–10 years (bottom). These regressions do not control for other characteristics, but rather represent the raw ratios. For the first column, the Y-axis displays log10(ratio) +1, where ratio is the cumulative total costs to the cumulative total research output for each metric (cost:output, for the first column output = #papers); error bars denote the 95% confidence intervals. The remaining columns show measures of cost effectiveness for relative citation ratio, approximate potential to translate, total clinical citation counts, and a binary measure of clinical citations. To account for the fact that many papers are published after funding for the relevant grant has ended, grant amounts were multiplied by a deflator – this represents the proportion of papers published to date against the anticipated number of future publications, as determined by empirical measurements (Supplemental Table 1). In most cases, according to this analysis, extramural research is more cost effective than intramural research when observing uncontrolled regressions. (B–D) Linear regression results of the cost efficiency of research output measures against project types (intramural vs. extramural). The regression model was fitted for each year of the project's progression. Unlike panel (A), this regression model controls for grant, investigator, and collaboration characteristics in order to obtain a more accurate estimate of the relative cost efficiency of intramural vs. extramural projects. The Y-axis coefficient indicates the mean disparity in research output between intramural and extramural projects, controlling for these other variables (see Methods). Because there might be covariates that could confound the data, separate regressions were conducted for all projects (B, the default), and for balanced projects using 1:1 propensity score matching (1 extramural grant for every 1 intramural grant) in order to compare grants that were the most similar to reduce the influence of unobserved covariates (C) and (D) similarly to (C) 1:4 propensity matching as a robustness check.

Supplemental Figure 5. Ternary contour plots representing the clinical citation efficiency in the human, animal, and molecular/cellular score system for intramural (D) and extramural projects (E) (Hutchins, Davis, Meseroll, & Santangelo, 2019; Weber, 2013).

Here, efficiency was the percentile of the cost per output in descending order. Each contour line denotes a constant efficiency percentile. Yellow/green are high-efficiency areas of the triangle, and blues are low-efficiency areas.



Supplemental Figure 6. Distribution of papers acknowledging more than one funding source, divided, Extramural and Intramural.



Activity code prefix	# Papers after funding ended	# Total papers	% Funded papers published after grant ended
D	9,213	13,039	70.66%
F	18,649	32,097	58.10%
G	283	1,095	25.84%
K	43,633	102,892	42.41%
P	20,766	50,650	41.00%
R	225,490	418,989	53.82%
S	1,061	2,176	48.76%
T	5,324	12,839	41.47%
U	49,222	112,459	43.77%
Z	1,794	19,871	9.03%

Supplemental Table 1. Number and proportion of papers published after grant ended by activity code prefix.

Data availability

Data used for this analysis are publicly available at: Article data: *https://doi.org/10.35092/yhjc.c.4586573* Grant data: *https://reporter.nih.gov/exporter* Author data: *http://er.tacc.utexas.edu/datasets/ped*

Additional information

Funding

Funder	Grant reference number	Author
University of Wisconsin-Madison (UW)		B Ian Hutchins
University of Wisconsin-Madison (UW)		Chaoqun Ni
Wisconsin Alumni Research Foundation (WARF)		B Ian Hutchins
Wisconsin Alumni Research Foundation (WARF)		Chaoqun Ni
U.S. Department of Defense (DOD)	W911NF2210294	B Ian Hutchins

Author ORCID iDs

Xiang Zheng: https://orcid.org/0000-0002-6619-5504
Chaoqun Ni: https://orcid.org/0000-0002-4130-7602

References

Afshar A. S., Yang Q., Thebault-Spieker J., Hutchins B. I. (2026) Quantifying (mis)alignment between reader focus and editor citation in scholarly biomedical topics in Wikipedia. *Quantitative Science Studies* **7**:326-348 https://doi.org/10.1162/QSS.a.457

Office of Portfolio Analysis at the National Institutes of Health (2018) National Institutes of Health word2vec pipeline. https://github.com/NIHOPA/word2vec_pipeline

Office of Portfolio Analysis at the National Institutes of Health (2019) Natural Language Preprocessing (NLPre) 2.0. GitHub. https://github.com/NIHOPA/NLPre

Arabi S., Ni C., Hutchins B. I. (2025) Most researchers would receive more recognition if assessed by article-level metrics than by journal-level metrics. *PLoS Biol* **23**:e3003532 https://doi.org/10.1371/journal.pbio.3003532 | PubMed

Azoulay P., Zivin J. S. G., Manso G. (2009) Incentives and creativity: evidence from the academic life sciences. *RAND J. Econ* **42**:527-554 https://doi.org/10.1111/j.1756-2171.2011.00140.x

Cassidy B. (2024) NIH in the 21st Century: Ensuring Transparency and American Biomedical Leadership. Senate Committe on Health, Education, Labor & Pensions. https://www.help.senate.gov/imo/media/doc/nih_modernization_5924pdf.pdf

Collins F. S., Schwetz T. A., Tabak L. A., Lander E. S. (2021) ARPA-H: Accelerating biomedical breakthroughs. *Science* **373**:165-167 https://doi.org/10.1126/science.abj8547 | PubMed

Congressionally Directed Medical Research Programs (2025) https://cdmrp.health.mil/

Culliton J. J. (1992) "Indirect" Costs Are Real Costs at Massachusetts Institute of Technology. *Science* **255**:778-779 https://doi.org/10.1126/science.1535998 | PubMed

Davis M. T., Busse B. L., Arabi S., Meyer P., Hoppe T. A., Meseroll R. A., Santangelo G. M. (2025) Prediction of transformative breakthroughs in biomedical research. *bioRxiv* https://doi.org/10.64898/2025.12.16.694385 | PubMed

Droegemeier K. (2017) Testimony before the House Appropriations Subcommittee on Labor, Health and Human Services, Education, and Related Agencies. https://docs.house.gov/meetings/AP/AP07/20171024/106525/HHRG-115-AP07-Wstate-DroegemeierK-20171024.pdf

Drummond M. E., Sculpher M. J., Torrance G. W., O'Brien B. J., Stoddart G. L. (2005) *Methods for the Economic Evaluation of Health Care Programmes* Oxford University Press. https://doi.org/10.1093/oso/9780198529446.001.0001

Federally Funded R&D Centers (2026) https://ncses.nsf.gov/resource/master-gov-lists-ffrdc

Funk K., Zayas-Caban T., Beck J. (2024) Phase 1 of the National Institutes of Health Preprint Pilot: Testing the viability of making preprints discoverable in PubMed Central and PubMed. *bioRxiv* https://doi.org/10.1101/2022.12.12.520156 | PubMed

Goldstein A. P., Kearney M. (2020) Know when to fold 'em: An empirical description of risk management in public research funding. *Research Policy* **49** https://doi.org/10.1016/j.respol.2019.103873

Hong X., Hutchins B. I., Ni C. (2026) Faster science, penalties in evaluation, and concerns on quality and impact: Researchers' use and perceptions of preprints. *bioRxiv* https://doi.org/10.64898/2026.03.02.709147

Hoppe T. A., Arabi S., Hutchins B. I. (2022) Predicting causal citations without full text. *bioRxiv* https://doi.org/10.1101/2022.07.05.498860

Hoppe T. A., Arabi S., Hutchins B. I. (2023) Predicting substantive biomedical citations without full text. *Proceedings of the National Academy of Sciences* **120** https://doi.org/10.1073/pnas.2213697120 | PubMed

Hoppe T. A., Litovitz A., Willis K. A., Meseroll R. A., Perkins M. J., Hutchins B. I., Santangelo G. M. (2019) Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Sci Adv* **5**:eaaw7238 https://doi.org/10.1126/sciadv.aaw7238 | PubMed

Hutchins B. I. (2021) A tipping point for open citation data. *Quantitative Science Studies* **2**:433-437 https://doi.org/10.1162/qss_c_00138 | PubMed

Hutchins B. I., Baker K. L., Davis M. T., Diwersy M. A., Haque E., Harriman R. M., Santangelo G. M. (2019) The NIH Open Citation Collection: A public access, broad coverage resource. *PLoS Biol* **17**:e3000385 https://doi.org/10.1371/journal.pbio.3000385 | PubMed

Hutchins B. I., Davis M. T., Meseroll R. A., Santangelo G. M. (2019) Predicting translational progress in biomedical research. *PLoS Biol* **17**:e3000416 https://doi.org/10.1371/journal.pbio.3000416 | PubMed

Hutchins B. I., Hoppe T. A., Meseroll R. A., Anderson J. M., Santangelo G. M. (2017) Additional support for RCR: A validated article-level measure of scientific influence. *PLoS Biol* **15**:e2003552 https://doi.org/10.1371/journal.pbio.2003552 | PubMed

Hutchins B. I., Yuan X., Anderson J. M., Santangelo G. M. (2016) Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level. *PLoS Biol* **14**:e1002541 https://doi.org/10.1371/journal.pbio.1002541 | PubMed

iCite (2015) https://icite.od.nih.gov/

iCite, Hutchins B. I., Santangelo G. M. (2019) iCite Database Snapshots (NIH Open Citation Collection). https://doi.org/10.35092/yhjc.c.4586573

Ioannidis J. P. A., Gross K., Bergstrom C. T. (2019) Contest models highlight inherent inefficiencies of scientific funding competitions. *PLOS Biology* **17** https://doi.org/10.1371/journal.pbio.3000065 | PubMed

Kim Y. H., Levine A. D., Nehl E. J., Walsh J. P. (2020) A Bibliometric Measure of Translational Science. *Scientometrics* **125**:2349-2382 https://doi.org/10.1007/s11192-020-03668-2 | PubMed

Korn D. (2015) Indirect costs: Cash is no gravy train. *Nature* **517**:438-438 https://doi.org/10.1038/517438b | PubMed

Lauer M., Roychowdhury D., Patel K., Walsh R., Pearson K. (2017) Marginal Returns and Levels of Research Grant Support among Scientists Supported by the National Institutes of Health. *bioRxiv* https://doi.org/10.1101/142554

Macilwain C. (1999) NSF told to ease up cost-share demands. *Nature* **399**:95-95 https://doi.org/10.1038/20039

National Library of Medicine (2020) Download MEDLINE/PubMed Data. https://www.nlm.nih.gov/databases/download/pubmed_medline.html

National Cancer Institute (NCI) Center for Cancer Research (2025) https://www.nih.gov/about-nih/nih-almanac/national-cancer-institute-nci

Nelson L., Ye H., Schwenn A., Lee C., Arabi S., Hutchins B. I. (2022) Robustness of evidence reported in preprints during peer review. *Research Square* https://doi.org/10.21203/rs.3.rs-1344293/v1

Nelson L., Ye H., Schwenn A., Lee S., Arabi S., Hutchins B. I. (2022) Robustness of evidence reported in preprints during peer review. *Lancet Glob Health* **10**:e1684-e1687 https://doi.org/10.1016/S2214-109X(22)00368-0 | PubMed

Ni C., Hutchins B. I. (2025) Framework for assessing the risk to a field from fraudulent researchers: A case study of Alzheimer's disease. *Journal of the Association for Information Science and Technology* **76**:1162-1173 https://doi.org/10.1002/asi.25009

NIH Budget (2025) https://www.nih.gov/about-nih/organization/budget

NIH ExPORTER (2025) https://reporter.nih.gov/exporter

Packalen M., Bhattacharya J. (2020) NIH funding and the pursuit of edge science. *Proceedings of the National Academy of Sciences* **117**:12011-12016 https://doi.org/10.1073/pnas.1910160117 | PubMed

Research Universities and the Future of America (2012)

Sampat B. N. (2012) Mission-oriented biomedical research at the NIH. *Research Policy* **41**:1729-1741 https://doi.org/10.1016/j.respol.2012.05.013

Santangelo G. M. (2017) Article-level assessment of influence and translation in biomedical research. *Mol Biol Cell* **28**:1401-1408 https://doi.org/10.1091/mbc.E16-01-0037 | PubMed

Tabak L., Lee B., Carpten J., Carnes M., Miller B., Palin A., Valantine H. (2019) Report of the ACD Working Group on High-Risk, High-Reward Research. https://acd.od.nih.gov/documents/presentations/06132019HRHR_B.pdf

Wahls W. P. (2018a) Diminishing marginal returns on NIH grant funding to institutions. *bioRxiv* https://doi.org/10.1101/367847

Wahls W. P. (2018b) The NIH must reduce disparities in funding to maximize its return on investments from taxpayers. *eLife* **7** https://doi.org/10.7554/eLife.34965 | PubMed

Weber G. M. (2013) Identifying translational science within the triangle of biomedicine. *J Transl Med* **11**:126 https://doi.org/10.1186/1479-5876-11-126 | PubMed

What happens to your application during and after review? (2025) https://public.csr.nih.gov/ForApplicants/InitialReviewResultsAndAppeals/applicationduringafterreview

Xu J., Kim S., Song M., Jeong M., Kim D., Kang J., Ding Y. (2020) Building a PubMed knowledge graph. *Scientific Data* **7** https://doi.org/10.1038/s41597-020-0543-2 | PubMed

Zhang S., Wapman K. H., Larremore D. B., Clauset A. (2022) Labor advantages drive the greater productivity of faculty at elite universities. *Science Advances* **8** https://doi.org/10.1126/sciadv.abq7056 | PubMed

Xu J, Kim S, Song M, Jeong M, Kim D, Kang J, Rousseau JF, Li X, Xu W, Torvik VI, *et al.* (2020) PubMed knowledge graph. PKG2020S4 (1781-Dec. 2020), Version 4. ID ped http://er.tacc.utexas.edu/datasets/ped

iCite, B.Ian Hutchins and George Santangelo (2021) iCite Database Snapshots (NIH Open Citation Collection). Figshare doi:10.35092/yhjc.c.4586573. https://doi.org/10.35092/yhjc.c.4586573

Abraham K. G., Haskins R. (2017) The Promise of Evidence-Based Policymaking. https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2019/03/Full-Report-The-Promise-of-Evidence-Based-Policymaking-Report-of-the-Comission-on-Evidence-based-Policymaking.pdf

Department H. a. H. S. (2011) National Institutes of Health Statement of Organization, Functions, and Delegations of Authority. Federal Register: Library of Congress. https://www.federalregister.gov/documents/2011/02/09/2011-2848/national-institutes-of-health-statement-of-organization-functions-and-delegations-of-authority

Fox J. L. (2013) Sequestration to slash research grants and delay biosimilars. *Nature Biotechnology* **31**:271-271 https://doi.org/10.1038/nbt0413-271 | PubMed

Harris A. (2014) Young, Brilliant, and Underfunded. New York Times. https://www.nytimes.com/2014/10/03/opinion/young-brilliant-and-underfunded.html

Scientopia (2014) Berg posts data on NIH Intramural funding. https://drugmonkey.scientopia.org/2014/02/05/berg-posts-data-on-nih-intramural-funding/

Young S. D. (2021) Evidence-Based Policymaking: Learning Agendas and Annual Evaluation Plans. (M-21-27). https://www.whitehouse.gov/wp-content/uploads/2021/06/M-21-27.pdf

Peer reviews

Reviewer #1 (Public review):

Summary:

This paper carefully compares intramural vs. extramural National Institutes of Health funded research during 2009-2019, according to a variety of bibliometric indices. They find that extramural awards more cost-effectively fund outputs commonly used for academic review such as number of publications and citations per dollar, while intramural awards are more cost-effective at generating work that influences future clinical work, more closely in line with agency health goals.

Strengths:

Great care was taken in selecting and cleaning the data, and in making sure that intramural vs. extramural projects were compared appropriately. The data has statistical validation. The trends are clear and convincing.

https://doi.org/10.7554/eLife.108929.2.sa3

Reviewer #2 (Public review):

This article reports a cost-effectiveness comparison of intramural and extramural that NIH funded between 2009 and 2019. Using data obtained from NIH RePORTER, they linked total project costs to publication output, using robust validated metrics including Relative Citation Ratio (RCR), Approximate Potential to Translate (APT), and clinical citations. They find that after adjusting for confounders in regression and propensity-score analyses, extramural projects were generally more cost-effective, though intramural projects were more cost effective for generating clinical citations. They also describe differences in the topics of intramural- and extramural-funded publications, with intramural projects more likely to generate papers on viral infections and immunity or cancer metastases and survival, but less likely to generate papers on pregnancy and maternal health, brain connectivity and tasks, and adolescent experiences and depression. The authors aptly describe the different natures of the intramural and extramural funding models, including that extramural researchers spend much time writing grant applications and that the work described in extramural publications often receives funding from sources other than NIH grants.

Strengths:

The authors leveraged publicly available data (including RePORTER and the iCite repository) and used robust validated metrics (RCR, APT, clinical citations). They carefully considered a large number of confounders, including those related to the PI, and performed several well-described regression analyses.

https://doi.org/10.7554/eLife.108929.2.sa2

Reviewer #3 (Public review):

This article demonstrates a comparative study on two funding mechanisms adopted by the National Institutes of Health (NIH). The authors adopted a quantitative approach and introduced five metrics to compare the output of intramural and extramural grants. These findings reveal the impacts of intramural and extramural grants on the scientific community, providing funders with insights into the future decisions of funding mechanisms they should take.

Strengths:

The authors clearly presented their methods for processing the NIH project data and classifying projects into either intramural or extramural categories. The limitations of the study are also well-addressed.

https://doi.org/10.7554/eLife.108929.2.sa1

Author response:

The following is the authors’ response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Strengths:*
>
> *Great care was taken in selecting and cleaning the data, and in making sure that intramural vs. extramural projects were compared appropriately. The data has statistical validation. The trends are clear and convincing.*

We thank the reviewer for highlighting the strengths of the manuscript.

> *Weaknesses:*
>
> *The Discussion is too short and descriptive, and needs more perspective - why are the findings important and what do they mean? Without recommending policy, at least these should discuss possible implications for policy.*

The Discussion has been substantially expanded. We added several new paragraphs discussing: the 2024 Senate HELP Committee proposal for NIH reform; implications for portfolio management (positioning extramural for basic research, intramural for clinical translation); generalizability to other agencies (DoD, NSF FFRDCs, DoE national labs); and the extramural program's role in workforce training as a societal benefit distinct from research outputs.

> *The biggest problem I have with this submission is Figure 3, which shows a big decrease in clinical-related parameters between 2014 and 2019 in both intramural and extramural research (panels C, D and E). There is no obvious explanation for this and I did not see any discussion of this trend, but it cries out for investigation. This might, for example, reflect global changes in funding policies which might also influence the observed closing gaps between intramural and extramural research.*

We added an explicit explanation in the Results: because the dataset is truncated at 2020, clinical citations naturally approach zero near the window's end, consistent with the ~7-year lag for clinical citations to accrue documented in prior work (Hutchins et al., 2019). The APT metric declines less steeply because it uses the forward citation network for predictions.

> ***Reviewer #2 (Public review):***
>
> *Strengths:*
>
> *The authors leveraged publicly available data (including RePORTER and the iCite repository) and used robust validated metrics (RCR, APT, clinical citations). They carefully considered a large number of confounders, including those related to the PI, and performed several well-described regression analyses.*

We thank the reviewer for highlighting these strengths of the manuscript

> *Figure 3A shows intramural projects producing about 2.75 papers per year in 2009, whereas extramural projects are producing just over 1 paper per year. Extramural projects appear to catch up over the next five years. While the authors attempt to explain the difference in their figure legend, another explanation is that the intramural projects started well before 2009 but, as the authors state, intramural data only became available in 2009.*

We added a methodological note acknowledging that some intramural projects may have had start dates prior to 2009 that are not captured in the data, and that the ramp-up of new intramural projects is slower because they are more tied to new PI hiring. We also note the exclusion of projects matched in 2008 as possible continuations. However, the slow ramp-up of Intramural costs in Supplemental Figure 3 is consistent with hiring-associated lagged investment suggesting that our filtering of continuing projects was very successful. Nevertheless, because we cannot completely rule out some continuing projects made it through despite our efforts, we have made the caveats mentioned above in the “Comparison of research topics” section of the Results and the Data section of the Methods.

> *As the authors note, funding information is often complex and difficult to characterize for an analysis like this. How did the authors handle: i) publications linked to multiple extramural grants; ii) publications linked to intramural and extramural grants; iii) publications linked NIH grants and non-NIH grants?*
>
> *I would think it necessary to somehow apportion credit, as otherwise it would appear that extramural projects are more productive than they truly are.*

We have now explicitly stated that papers with both intramural and extramural funding links were excluded, while papers with multiple links within the same funding type were retained. A new Supplemental Figure 6 was added showing the distribution of papers by number of funding sources for both extramural and intramural grants, demonstrating that the vast majority acknowledged only one project. These changes are in the Methods, Data section and Supplemental Figure 6

Apportioning credit among a many-to-many graph like the ones used here is indeed a high value problem to solve, but one with many researcher-degrees-of-freedom about analytical design decisions that impact the results. We are working on a rigorous methodology for this, but the amount of time required to do this well is its own research project, and out of scope for manuscript revisions.

> *Also, it is not clear if the authors took account of the indirect costs paid by the NIH to universities that have received extramural grants.*

We added explicit language clarifying that all cost comparisons use inflation-adjusted total costs (direct + indirect) for extramural grants. We also added a new sensitivity analysis (Supplemental Figure 4) inflating extramural indirect costs by 30% to approximate unrecovered university expenditures, with the finding that the fundamental pattern holds even under this adjustment. These are found in the “Comparison of funding” and “Comparison of cost effectiveness” sections of the Results, as well as Supplemental Figure 4.

> ***Reviewer #3 (Public review):***
>
> *Strengths:*
>
> *The authors clearly presented their methods for processing the NIH project data and classifying projects into either intramural or extramural categories. The limitations of the study are also well-addressed.*

We thank the reviewer for highlighting these strengths of the manuscript

> *Weaknesses:*
>
> *The article would benefit from a more thorough discussion of the literature, a clearer presentation of the results (especially in the figure captions), and the inclusion of evidence to support some of the claims.*

The Introduction was updated with more specific framing of prior literature (e.g., explicit mention of risk management, funding disparities, and diminishing marginal returns as the focus of prior work). New references were added throughout, including Sampat (2012) on mission-oriented NIH research, Ioannidis et al. (2019) on grant competition inefficiencies, Drummond et al. (2005) on health economic evaluation methods, and the Cassidy (2024) Senate report, throughout the introduction and discussion.

> ***Recommendations for the authors:***
>
> ***Reviewer #2 (Recommendations for the authors):***

> *The article would benefit from a more detailed analysis/discussion about the recovery of indirect costs for extramural research.*
>
> *I note that the authors are from the University of Wisconsin, which is part of the IRIS network (https://iris.isr.umich.edu/iris-members-map/). They could work with IRIS (also called UMETRICS) to get a better sense as to the true costs of extramural research for each project (e.g., all labor costs, all equipment costs). The IRIS data are extraordinarily robust. Here's an example of an IRIS / UMETRICS paper: https://www.science.org/doi/10.1126/sciadv.abb7348.*
>
> *They could, for example, re-do the analyses assuming that the recorded indirect cost covers only 70% of the true indirect costs. Thus, if they get $700,000 indirect costs from RePORTER, they should assume that the true indirect costs were $1,000,000. Similarly, they can add the costs of the time the PI spent writing the grant proposal, using the Bergstrom paper as a guide.*
>
> *Another option would be to conduct sensitivity analyses taking into account ~30% incomplete indirect cost recovery (see https://docs.house.gov/meetings/AP/AP07/20171024/106525/HHRG-115-AP07-Wstate-DroegemeierK-20171024.pdf) and lost efficiency due to excess time writing grant proposals (see https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000065).*

We conducted a sensitivity analysis as requested inflating extramural indirect costs by 30%, citing the Droegemeier (2017) Congressional testimony as the basis for this estimate. The cost of grant-writing time is now acknowledged in the Discussion as an unreimbursed hidden cost of the extramural system, citing Ioannidis et al. (2019). This narrowed the gap between extramural research and intramural research, but did not close it completely. In addition, our updated regression (Supplemental Figure 4) showed similar trends as our main Figure 4, but with the Intramural advantage heightened and the Extramural advantage diminished. Both remained significant. We have also added to the discussion that there are additional costs and benefits that may not be fully captured in an analysis such as ours.

> *The authors appear to have used an agency-perspective for their cost-effectiveness analyses. Generally, it is preferable to use a wider societal perspective. While that may be difficult, the article would benefit from some discussion from the perspective of the government and universities.*

We added a new paragraph explicitly acknowledging the agency-centered perspective and its limitations, noting that it does not capture the full economic cost borne by universities (startup costs, philanthropy, endowments, state contributions, graduate student training, faculty retention, infrastructure). The extramural program's contribution to the US workforce pipeline is specifically highlighted as a societal benefit not captured by the cost-effectiveness metrics.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *Line 84-87: "The overrepresentation of viral research is likely because of the outsize investment toward the intramural Vaccine Research Center, and the cancer/genetics overrepresentation due in part because National Cancer Institute intramural investigators conduct research at that institute as well as at the NIH Clinical Center for their human genetics work." What evidence is there to support this claim?*

A citation to the NCI Center for Cancer Research website was added to support the claim about NCI intramural investigators working at the Clinical Center and Center for Cancer Research, where vaccine research is extensively discussed.

> *Lines 107-109. "Given that NIH funding for intramural research has remained relatively constant as a percent of total funding over the years, this indicates larger single awards for intramural research while extramural investigators may increasingly require multiple concurrent grants to sustain their labs." Authors may consider adding a panel to Figure 2 showing the percentage of total funding of intramural vs. extramural funding.*

Rather than adding a panel to Figure 2, we added a new Supplemental Figure 3 showing the cost breakdown and intramural percentage of total funding by year.

> *Discussion section: Are any of the findings of this study relevant to other funding agencies in the US (such as the National Science Foundation, the Department of Energy, and the Department of Defense)?*

A new paragraph to the Discussion was added discussing implications for the Department of Defense (including the Congressionally Directed Medical Research Programs), NSF FFRDCs, and the Department of Energy's national labs and FFRDCs, arguing that the incentive-alignment logic likely generalizes across agencies.

> *Methods section: Please add an explanation of the technique used for propensity score matching.*

A detailed step-by-step description of the PSM procedure was added, covering propensity score estimation, within-year matching, matched cohort construction, outcome regression on matched data, and visualization of results.

> *Figure 1: Please clarify if the relative ratio of intramural projects is calculated from the numbers of grants (as suggested in lines 95-96 and 98-100) or the numbers of publications (as suggested in lines 82-83 and 97-98).*
>
> *Also, this figure would be more intuitive if, for each topic, it showed the relevant intramural number (as it currently does) and also the relevant extramural number.*

The caption and Methods were updated to clarify that clustering and ratio calculation are based on projects/grants, not publications. A formula was added to the Methods to make the ratio calculation explicit. The figure itself was not modified to add extramural bars, though the ratio calculation already implicitly encodes both.

> *Figure 2: Please change "(red)" to "(blue)" in the caption, and remove the A as there is only one panel in this figure*
>
> *Figure 4: Please change "(red)" to "(blue)" in the caption.*

These changes have been made.

> *Lines 19-21: I suggest rewriting this sentence as follows:*
>
> *"We find that extramural awards are more cost-effective for producing outputs commonly used for academic evaluation, such as publications and citations per dollar, while intramural awards are more cost-effective for generating research that influences future clinical work, more closely in line with agency's health goals."*

The sentence was rewritten substantially in line with the reviewer's suggestion, now reading more clearly with "per dollar" removed as a parenthetical and the structure of the comparison clarified.

> *Lines 31-34: Please rewrite this sentence along the following lines to provide more context on previous research into the grant funding system:*

> *Certain aspects of the grant funding system have been the focus of research, such as AAAA (Azoulay et al., 2009), BBBB (Goldstein and Kearney, 2020), CCC (Hoppe et al., 2019), DDDD (Lauer et al., 2017), EEEE (Wahls, 2018a) and FFFF (Wahls, 2018b), but the relative merits of intramural and extramural funding have received little attention to date.*

The sentence was rewritten to name specific contributions of each cited paper (e.g., risk management, funding disparities, diminishing marginal returns), replacing the generic list of citations.

> *Lines 41-44: Please explain "merit score" and please add a reference to an article or website that explains the review process at the NIH.*

"Merit score" was revised to "percentile ranking of overall impact merit score" and a citation to the NIH CSR website ("What happens to your application during and after review?," 2025) was added.

> *Lines 53-54: Please change Intramural to intramural (two instances, and also in line 284), and Extramural to extramural.*

"Intramural" and "Extramural" were corrected to lowercase throughout.

> *Line 65-67: This sentence ("Potential advantages of the intramural approach are that researchers in the NIH's own laboratories allow the NIH to hire researchers whose research agendas more closely align with its mission.") reads awkwardly. Please clarify.*

The sentence was rewritten to read more clearly: "An advantage of the intramural approach are that NIH has the direct ability to hire scientists whose research closely aligns with agency goals, and researchers do not need to devote time and effort on preparing and submitting grant applications."

> *Line 95-97: Authors should consider including an equation to help explain the following sentence: "The relative ratio of intramural projects for each topic was calculated by taking a ratio of the proportions of total grants a topic represented in the intramural vs. extramural portfolios. A relative ratio >1 signifies a higher share of intramural project publications on that topic relative to their share across all topics."*

A formula was added to the Methods defining the topic-level ratio calculation explicitly.

> *Line 143: The phrase "may reflect the extra attention intramural investigators are afforded" reads awkwardly - please reword.*

Reworded to "may reflect the extra time intramural investigators save because they do not have teaching and grant writing responsibilities."

> *Lines 303-304: This sentence ("First, as the renewal of project contracts may alter the topic and arrangement of the projects, we dropped 70,297 projects with renewal records in our data.") reads awkwardly. Please clarify.*

Reworded to "Since the scientific focus of a study may drift over time, we dropped 70,297 projects with renewal records in our data."

> *Line 378-379: Please specify the model of ChatGPT used.*

Done.

https://doi.org/10.7554/eLife.108929.2.sa0