Supplementary Online Material

Appendix A: Linking drug approvals to patents and publications

We find it is quite common for important new drugs to cite NIH-funded research in the bottom 40% of the rank ordering the NIH uses to prioritize which research to fund. To the extent citations reflect meaningful influences on the citing inventions, this would suggest that cutting the NIH budget by 40% would have jeopardized many of the medical advances we have enjoyed over the last several decades.

This is consistent with a variety of other studies [5,8]. Similarly, Mansfield [16] finds approximately 30% of drugs and medical devices developed between 1975 and 1994 could not have been developed without substantial delay, in the absence of recent academic research. Sampat and Lichtenberg [17] find 41 percent of drugs approved from 1988-2005 had a patent citing a government-funded publication.[a] Roach and Cohen [18] provide some validation of this approach: comparing citations to survey data, they find labs whose patents make more citations to public research also report in surveys a greater use of public research.

Our study departs from those above in several ways. First, this study focuses specifically on grants and publications that would be affected by a counterfactual 40% cut, and their linked patents and drugs. Second, we use citations in the text of the patent, rather than front-page prior art citations to scientific literature (which are used in [5,8]). Bryan et al. [19] show that these "in-text" citations reference a different body of knowledge and argue that (given their legal purpose) these citations are more likely to reflect intellectual influences than front-page citations. They provide evidence showing they are more closely related to survey-based measures of the use of public science than front-page citations.

One reason we may overstate linkages to potentially cut research relates to choice of threshold for connected drugs. We define a drug as linked to at-risk research if it has at least one citation to research funded by at-risk NIH grants, but the average drug linked to at-risk research cites 35 research publications. In many cases, only one of these citations goes to at-risk research (see supplemental figure S1, Panel B). If that research had not occurred, it may well be that the drug's development would have been largely unaffected. In other words, it is likely that some drugs linked to at-risk research would be discovered anyway, even though they cite research funded by an at-risk grant. More importantly, though in-text citations are potentially better measures of influence than front-page citations, it is likely that many of these citations too were not crucial for the development of the drugs.

However, there are also factors that cut in the opposite direction, false negatives. Drugs whose development does depend on at-risk research will not be linked to it if their associated patents do not cite the research. This can happen for three reasons. First, not all relevant enabling research needs to be cited in patents. Second, it might simply be that important research is not cited in the text of a patent, but is cited instead *only* in the front-page, or the research was patented and the patent is cited instead.

Third, it might be that a research publication has an impact on drug discovery that is mediated through follow-on research. For example, research funded by at-risk grants might not directly lead to a drug discovery, but may

[a] Other work not using citations also suggests strong linkages between NIH funding and drug development [20,21].

inspire follow-on research that does. If the follow-on research is cited, but not funded by at-risk grants, then our approach would not identify the drug discovery as connected to at-risk research. Ahmadpoor and Jones [22] provide evidence that these kind of second-order connections are very common in biomedical research. Their study maps out the citation network connecting papers to each other, and eventually to patents. For each paper, they identify the shortest number of citation links that lead to a patent. In some cases, papers are directly cited by patents (as is the case in our drugs linked to at-risk research). But for every life science subfield they find these direct links are less common than second-order (or longer links).[b]

In order to illustrate the second-order chains missed by our analysis, we can focus on the example of Semaglutide (brand name Ozempic), a long-acting glucagon-like peptide-1 (GLP-1) receptor agonist that has reshaped therapy for type 2 diabetes and, more recently, obesity. Its approval in 2017 (for diabetes, with a higher-dose version for obesity in 2021) marked the culmination of decades of research [23]. The GLP-1 hormone itself was discovered in the mid-1980s by Habener and Mojsov, who identified GLP-1(7–37) as a potent insulinotropic peptide from proglucagon gene processing [24]. That foundational work—for which Habener, Mojsov, and Novo Nordisk scientist Lotte Knudsen received the 2024 Lasker-DeBakey Clinical Medical Research Award—launched a therapeutic hunt for stable GLP-1 mimetics. Bibliometrically, Novo Nordisk's lead patent US 8536122B2 "Acylated GLP-1 Compounds" cites only five technical publications (e.g., on peptide-analog optimization), none with NIH acknowledgments because the late-stage work was done largely inside the company. Yet the patent *does* reference Massachusetts General Hospital's earlier US5545618 A "GLP-1 analogs useful for diabetes treatment," which in turn cites the seminal 1986 *Journal of Biological Chemistry* paper by Mojsov et al. describing the insulinotropic fragment GLP-1 (7-37). That article in turn acknowledges NIDDK grant R01AM30834 ("Glucagon Biosynthesis and Metabolism"), awarded to Joel F. Habener in 1982 and renewed through 2008. This indirect citation chain illustrates how semaglutide's ostensibly industry-centric patent lineage still traces back to federally funded basic science—a linkage captured neither by straight publication counts nor by the narrow lens of patent-embedded literature, but documented in recent GLP-1 histories that credit Habener, Mojsov, and Knudsen for the key biological insight—GLP-1's insulinotropic activity.

Figure S2 plots the cumulative distribution of drug approvals by their citations to "at risk" NIH-acknowledged science. The horizontal axis shows, for each new molecular entity (N = 557), the fraction of scientific articles cited in its pre-approval U.S. patents that would be counterfactually cut under our hypothetical 40% reduction in the size of the NIH budget; the vertical axis plots the running count of approvals when NMEs are ordered from the lowest to the highest fraction, thus tracing a step-wise cumulative density curve. The dashed arrow emphasizes the gap between the zero-fraction cluster and the remainder of the distribution: $557 - 286 = 271$ approvals sit at the origin, while the remaining 286 approvals stretch across the 0-100 % range. Interpreting the curve left-to-right therefore illustrates how many drug approvals would lose successively larger proportions of their patent-cited literature if NIH-acknowledged science were "counterfactually removed."

Finally, a 40% long-run cut to the NIH would likely not play out exactly as we model over the course of several decades. It is possible that specific types of grants, or specific Institutes or Centers, would disproportionately be cut. Over time, the set of grants submitted to the NIH for funding would certainly diverge from our actual history, as studies that were counterfactually cut are re-submitted in subsequent years, and proposals that build

[b] It is also possible that that NIH research matters for drug discovery but the resulting publication does not need to be referenced given legal reasons for citation, or has impact through non-publication channels (e.g., labor markets).

on the result of counterfactually cut grants become infeasible. The set of grants evaluated by the NIH for funding would not look like it did in our own history, and so the funded set would probably not be the same as the top 60% of grants that were actually funded. It is also possible that other funders (e.g. the private sector) may fund this research instead, or the research could be undertaken without NIH funding. We do not consider this possibility here.[c]

For all these reasons, we take our main finding mostly as illustrating the importance of research in the bottom 40% of NIH grant priority ranking, not as generating a literal list of lost drugs from such a cut.

Appendix B: Data

Grant priority scores. We obtained NIH grant data for the 1980-2007 period, including priority review scores, from the NIH's IMPAC II system. To mimic the NIH priority system, we proceed in three steps. First, we rank each grant by order of priority score within each Initial review Group (IRG). Second, routing each grant to the component Institute within NIH that eventually funded the grant, we re-rank each grant within each Institute, thus creating a rank of rank. Ties are broken at random. Third, we search for the critical grant such that, within each Institute-fiscal year combination, the cumulative budget of all grants of lower ranks up to that critical grant equals 60% of the total extramural funding for that Institute-fiscal year combination. The grants of higher rank than this critical grant are then categorized as "at risk" grants. This methodology ignores many details that lead actual funding policies to deviate from the typical prioritization scheme. For instance, some Institute staff will occasionally use their discretion to fund grants "out of order," promoting some applications (and demoting others) reflecting Institute-level priorities. The prioritization scheme we constructed keeps the spirit of the actual procedure, while ignoring some details that would require access to more granular data. More details can be found in Appendix A of Azoulay et al. [5].

For the counterfactual analyses we assume the "at risk" publications (and the associated research advances) would disappear under the proposed budget cuts. It is also possible that other funders (e.g. the private sector) may fund this research instead, or the research could be undertaken without NIH funding. We do not consider this possibility here.

Publications. We use data from NIH RePORTER to link the NIH grants to the publications that acknowledge their support.[d] We drop publications that appear more than seven years after the end of a grant's final cycle. More details can be found in Appendix D1 of Azoulay et al. [3].

Drugs and Drug Patents. We started with an FDA compilation of new molecular entity and new biologic drug approvals between 2000 and 2023. We focus on drugs approved starting in 2000, rather than 1980, because there is a lag between when a grant is made and when its research is completed. This restriction allows us to focus on drugs that could possibly have been influenced by the grants in our sample. We limit our analysis to small molecule drugs—new drug applications (NDAs) associated with new molecular entities (NMEs). We excluded biologics since the FDA's patent listing provisions we rely on only apply to NDAs, not biologic license

[c] See Azoulay et al. [5] for a careful treatment of crowd-out and related issues.
[d] These data are available beginning in 1980, and can be obtained from the NIH Exporter web site at https://reporter.nih.gov/exporter/linktables.

applications (BLAs).[e] For each drug, we identified relevant patents using data from current and archival versions of the FDA's Orange Book [25] using the dataset described in Hemphill and Sampat [7].

Overall there are 611 NMEs approved between 2000 and 2023, of which 566 have at least one Orange Book listed patent (by October 2023). We focus the analysis on pre-approval patents, dropping all patents filed after drug approval, using filing date information from USPTO PatentsView database and approval date information from the FDA list above.[f] Our final set of drugs is the 557 NMEs approved 2000 to 2023 with at least one pre-approval patent.

Direct Linkages to NIH grants. To determine direct linkages to NIH grants, we merged the resulting list of pre-approval patents for all NMEs approved 2000 to 2023 to the Ouellette and Sampat [6] dataset on drugs where the government has Bayh-Dole rights. For each such drug, we used information from USPTO government-interest statements and NIH RePORTER to determine the specific grant number(s) that supported the drug's patents. In a few cases a specific grant number was not listed, but the patent government interest statements acknowledged NIH or Public Health Service support. In these instances we used information on the patent inventor, filing date, and patent text, together with grant and PI information in NIH RePORTER, to determine the most likely grant number.

In-text patent citations to scientific publications. We improve on previous research by using in-text citations to NIH-funded research [19]. We extract the in-text citations from Marx and Fuegi [10] to link patents to cited publications.[g] We focus on patent-publication linkages with a confidence score of 4 or higher, to limit false positives. While in-text citations may be a better proxy for intellectual influence than front-page citations for reasons discussed in Appendix A, we do not interpret our results as literally identifying which specific drugs would not exist if we lived in an alternative history where the NIH budget was much smaller. This is because in-text citations, too, may include false positives and false negatives when it comes to identifying intellectual contributions to patents. Just as is true of citations in academic papers, not all citations in the text of patents represent papers that were crucial for the development of the citing patent.[h] Although strategic citation is less common in in-text references than on a patent's front page [19], even these citations may be included for legal coverage rather than to reflect true intellectual influence. For that reason—and for the additional caveats noted in [19]—in-text citations remain an imperfect proxy for a publication's actual impact on the citing patent.

Patent value estimates. Kogan et al. [14] estimate patent values by analyzing stock market reactions of publicly traded firms to patent approval announcements. Specifically, they measure the abnormal returns observed in a narrow window around patent grant dates, isolating the market's immediate valuation of each patent. This approach relies on the efficient market hypothesis, assuming stock prices promptly reflect the incremental future profits attributable to newly granted patents. By aggregating these abnormal returns, KPSS generate an estimate of patent-specific economic value, enabling comparative analysis across patents and firms. We leverage these estimates by matching the KPSS dataset to the set of pre-approval patents for each of the drug approvals

[e] This arrangement has its roots in the FDA processes for approving generic versions of small-molecule drugs, typically once they are off patent.

[f] https://patentsview.org/download/data-download-tables

[g] Available at https://relianceonscience.org/

[h] Teplitskiy et al. [26] is suggestive here; it studies the citations made by research publications to other research publication, rather than patents, but they find the large majority of such citations do not exert a major or very major influence on research.

in our data.[i] We sum the values for each pre-approval patent for a given drug approval. Since these patents are not typically granted in a single year, we deflate the amounts using the consumer price index (CPI). Not all of the drug patent-holders are public firms. We can compute a KPSS patent value for only 329 (59%) of the 557 drug approvals in the data.

Appendix C: Alternative budget cut analysis

There is a natural reason to use the 40% cutoff for the counterfactual budget cut, based on the President's proposal. However, the ultimate NIH budget for FY2026 will depend on Congressional action. We examine the impact of funding cuts of different sizes, ranging from 0 to 100 percent, and summarize the results in Figure S3. On the x-axis, we plot the size of the budget cut and on the y-axis, the number of drug approvals affected using the indirect (i.e., patent to publication to grant) linkages. The curve's slope is initially very steep and subsequently tapers off as the share of grants cut increases. This indicates that initial funding cuts impacts more grants on the extensive margin (that is, many drugs see at least one of their cited publications impacted), while further funding cuts impact grants on the intensive margin (that is, the share of impacted publications increase).

We could, in principle, also consider different ways of selecting which grants to cut, under budget cut scenarios. There is a natural reason to illustrate the potential impact of budget cuts using contemporaneous priority scores, since priority scores are actually used to allocate scarce funding to grant proposals. However, it is also worth briefly discussing the extent to which priority scores allocate funding to research with social impact.

Li and Agha [27] examine the correlation between NIH priority score data (the same data used in our analysis) and various metrics of subsequent research output, such as number of publications, citations, and patents. Across all these proxies for impact, stronger priority scores are robustly correlated with more impact, including after controlling for other observable characteristics of grant applicants. Park, Lee, and Kim [28] perform a complementary analysis, examining similar metrics for subsequent research impact for NIH grants whose priority scores put them below the normal funding cut-off, but which were unexpectedly funded as part of the 2008 American Recovery and Reinvestment Act. Using this approach, they also find that priority scores are correlated with the number of publications and citations associated with a grant.

These papers suggest selecting grants with lower priority scores is one way to reduce the impact of budget cuts, since on average lower priority scores are associated with lower research impact. Nonetheless, our paper indicates that even when the funder uses this approach to select which grants to cut, a relatively large share of drugs are nonetheless linked to the cut research.

References

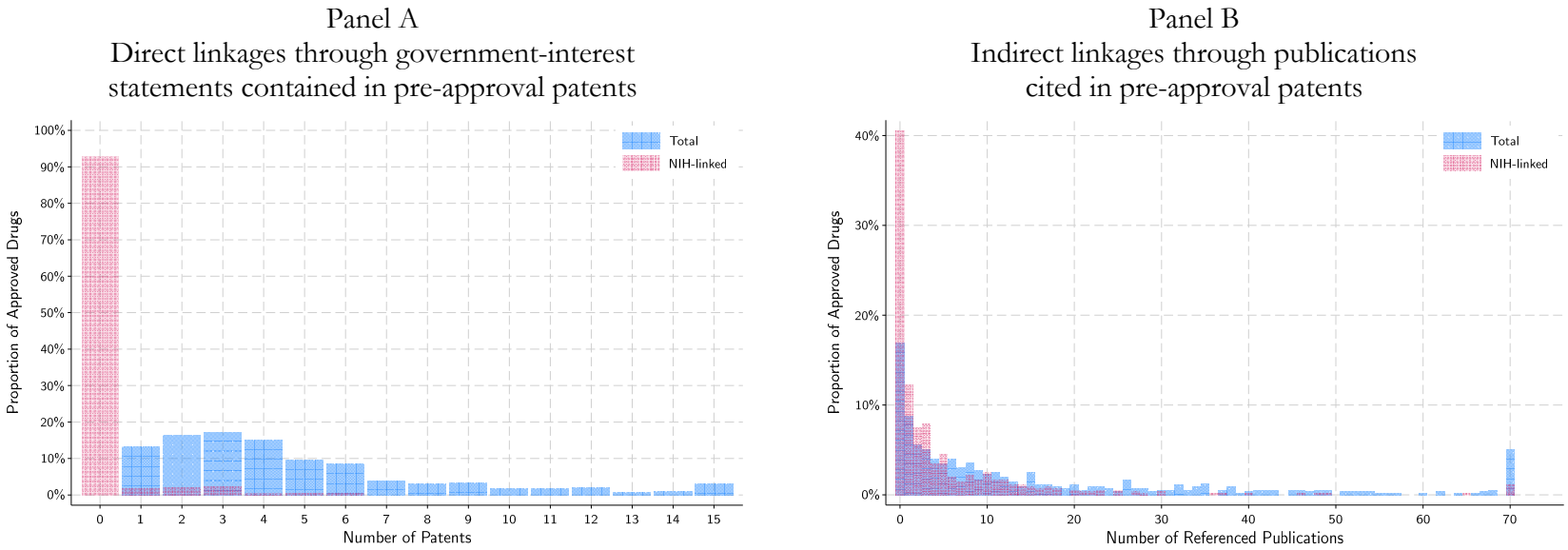
1. B. N. Sampat, Mission-oriented biomedical research at the NIH. *Research Policy* **41**, 1729-1741 (2012).

[i] Available at https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data.

2. J. Berg, *NIH Changes Since January 20, 2025* (2025); https://jeremymberg.github.io/jeremyberg.github.io/. Accessed 29 July 2025.
3. A. Zimmermann, *FY 2026 R&D Appropriations Dashboard* (American Association for the Advancement of Science, 2025); https://www.aaas.org/news/fy-2026-rd-appropriations-dashboard. Accessed 29 July 2025.
4. F. R. Lichtenberg, K. Krstovski, The impact of biomedical innovation on US mortality, 1999–2019: evidence partly based on 286 million descriptors of 27 million PubMed articles. *J Demogr. Econ.*, https://doi.org/10.1017/dem.2024.27 (2025).
5. P. Azoulay, J. S. Graff Zivin, D. Li, B. N. Sampat, Public R&D investments and private-sector patenting: Evidence from NIH funding rules. *Rev. Econ. Stud.* **86**, 117-152 (2019).
6. L. L. Ouellette, B. N. Sampat, The Feasibility of Using Bayh-Dole Act March-In Rights to Lower US Drug Prices *JAMA Health Forum* **5**, e243775 (2024).
7. C. S. Hemphill, B. N. Sampat, Patents, innovation, and competition in pharmaceuticals: The Hatch-Waxman Act after 40 years. *J. Econ. Perspect.* **39**, 27-52 (2025).
8. D. Li, P. Azoulay, B. N. Sampat, The applied value of public investments in biomedical research. *Science* **356**, 78-81 (2017).
9. W. M. Cohen, R. R. Nelson, J. P. Walsh, Links and impacts: The influence of public research on industrial R&D. *Manag. Sci.* **48**, 1–23 (2002).
10. M. Marx, A. Fuegi, Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *J. Econ. Manag. Strategy* **31**, 369-392 (2022).
11. T. Hunter, Treatment for chronic myelogenous leukemia: the long road to imatinib. *J. Clin. Invest.* **117**, 2036-2043 (2007); doi:10.1172/JCI31691.
12. F. A. Shepherd, J. R. Pereira, T. Ciuleanu, E. H. Tan, V. Hirsh, S. Thongprasert, D. Campos, S. Maoleekoonpiroj, M. Smylie, R. Martins, M. van Kooten, M. Dediu, B. Findlay, D. Tu, D. Johnston, A. Bezjak, G. Clark, P. Santabárbara, L. Seymour, Erlotinib in previously treated non-small-cell lung cancer. *N. Engl. J. Med.* **353**, 123-132 (2005).
13. U.S. Food and Drug Administration, *Priority Review* (2025); https://www.fda.gov/patients/fast-track-breakthrough-therapy-accelerated-approval-priority-review/priority-review. Accessed 29 July 2025.
14. L. Kogan, D. Papanikolaou, A. Seru, N. Stoffman, Technological innovation, resource allocation, and growth. *Q. J. Econ.* **132**, 66-712 (2017).
15. P. Azoulay, M. Clancy, D. Li, B. Sampat. Data and code for "Alternative history: What if the NIH had been 40% smaller? An alternative history." Available at https://doi.org/10.5281/zenodo.17065831
16. E. Mansfield, Academic research and industrial innovation: An update of empirical findings. *Res. Policy* **26**, 773-776 (1998).
17. B. N. Sampat, F. R. Lichtenberg, What are the respective roles of the public and private sectors in pharmaceutical innovation? *Health Aff.* **30**, 332-339 (2011).
18. M. Roach, W. M. Cohen, Lens or prism? Patent citations as a measure of knowledge flows from public research. *Manag. Sci.* **59**, 504-525 (2013).
19. K. A. Bryan, Y. Ozcan, B. N. Sampat, In-text patent citations: A user's guide. *Res. Policy* **49**, 103946 (2020).
20. I. M. Cockburn, R. M. Henderson, Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *J. Ind. Econ.* **46**, 157-182 (1998); doi:10.1111/1467-6451.00067.
21. E. G. Cleary **et al.**, Contribution of NIH funding to new drug approvals 2010-2016. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2329-2334 (2018); doi:10.1073/pnas.1715368115.
22. M. Ahmadpoor, B. F. Jones, The dual frontier: Patented inventions and prior scientific advance. *Science* **357**, 583-587 (2017).
23. D. J. Drucker, Discovery of GLP-1–based drugs for the treatment of obesity. *N. Engl. J. Med.* **392**, 612–615 (2025); doi:10.1056/NEJMcibr2409089.
24. S. Mojsov, G. Heinrich, I. B. Wilson, M. Ravazzola, L. Orci, J. F. Habener, Preproglucagon gene expression in pancreas and intestine diversifies at the level of post-translational processing. *J. Biol. Chem.* **261**, 11880-11889 (1986).

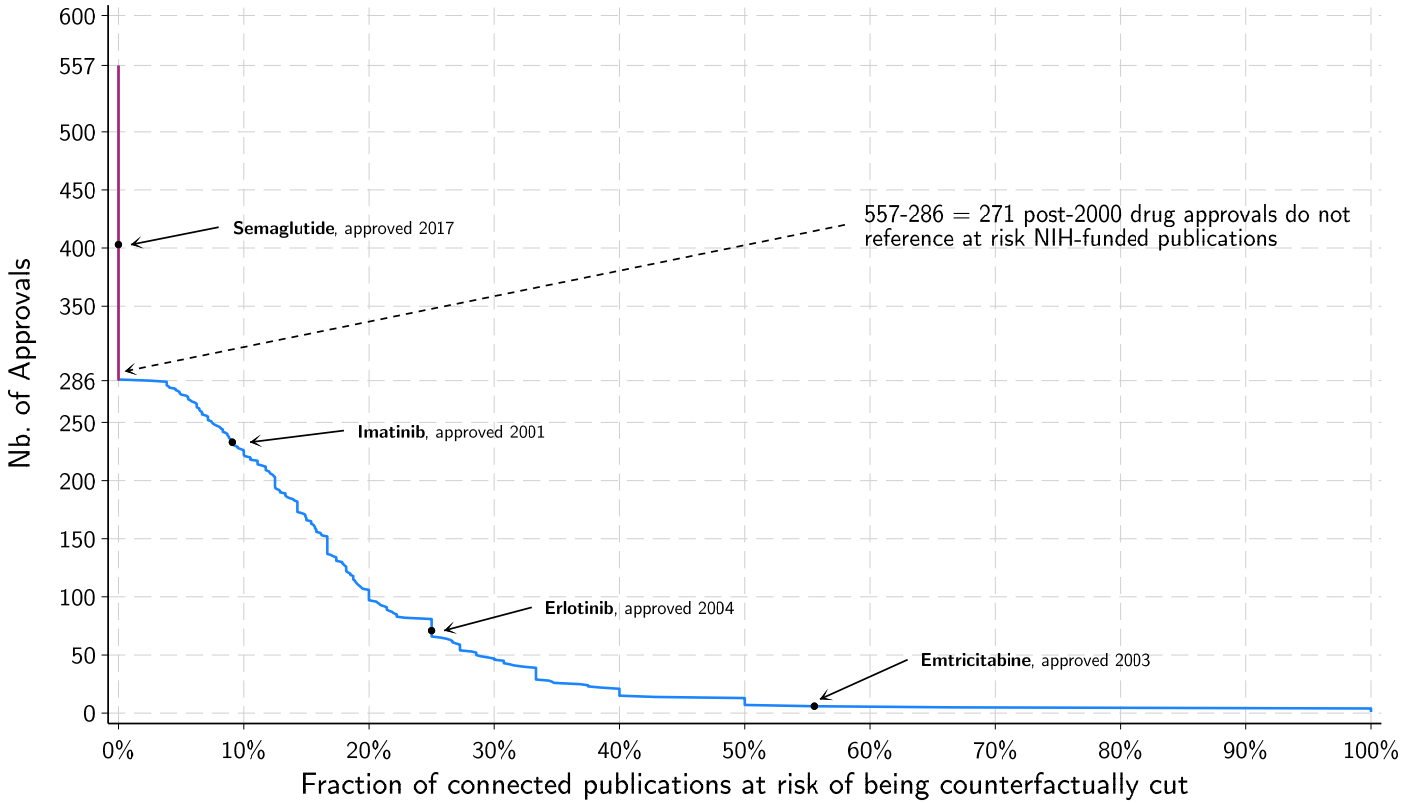
25. M. Durvasula, C. S. Hemphill, L. L. Ouellette, B. N. Sampat, H. L. Williams, The NBER Orange Book Dataset: A user's guide. *Res. Policy* **52**, 104686 (2023); doi:10.1016/j.respol.2023.104686.
26. M. Teplitskiy **et al.**, How status of research papers affects the way they are read and cited. *Res. Policy* **51**, 104480 (2022).
27. D. Li, L. Agha, Big names or big ideas: Do peer-review panels select the best science proposals? Science 348, 434–438 (2015).
28. H. Park, J. Lee, B.-C. Kim, Project selection in NIH: A natural experiment from ARRA. Res. Policy 44, 1145–1159 (2015).

Figure S1: Approval-level Variation in Linkages



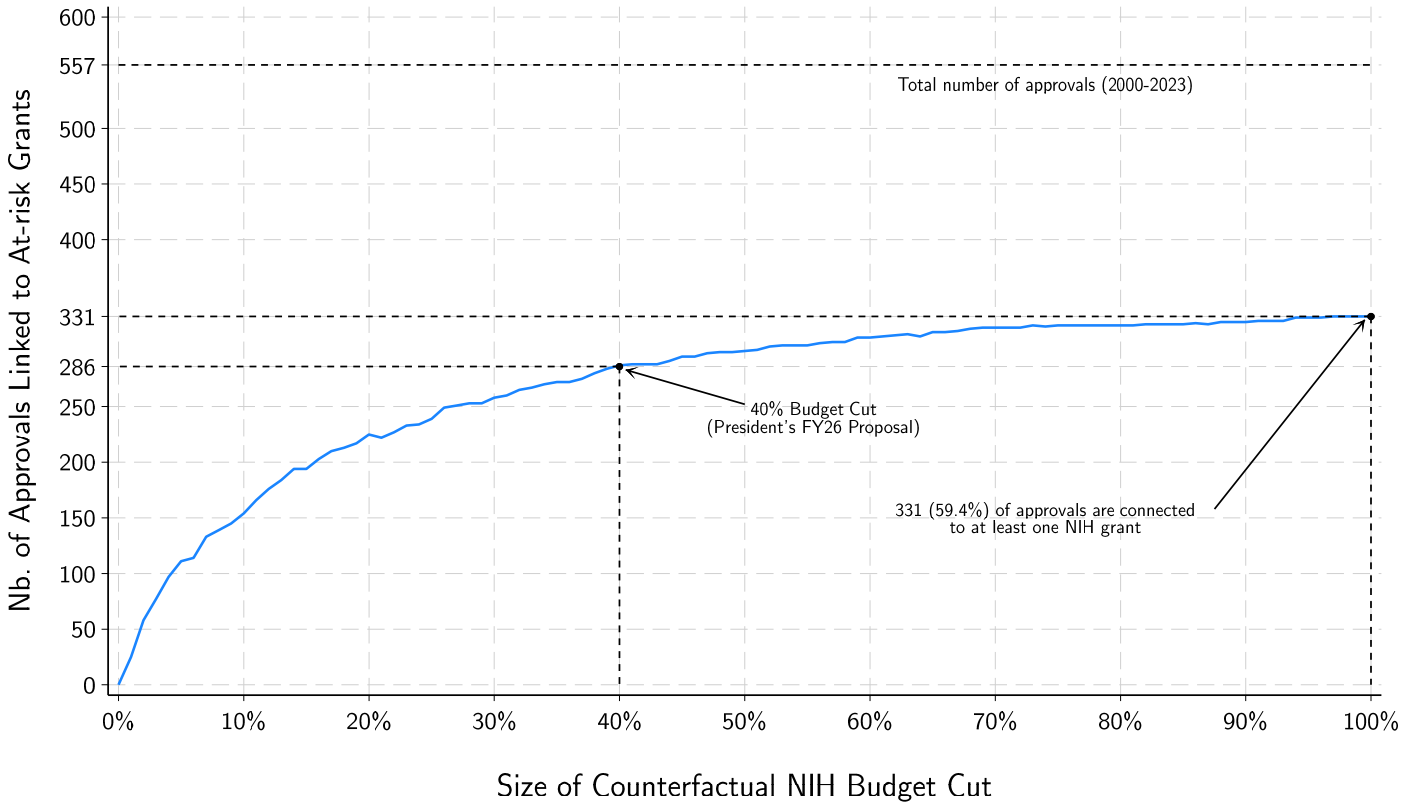
Note: **Panel A (direct links).** Distribution across approved drugs of the number of pre-approval patents: total (blue) and those directly linked to NIH grants (red), where "direct" means patents with government-interest statements and/or patents associated with the grant in NIH RePORTER. We top-code extreme values: 17 approvals have 15 or more pre-approval patents (max.=34). **Panel B (indirect links via citations).** Distribution across approved drugs of the number of publications cited in the text of pre-approval patents: total (blue) and publications acknowledging NIH funding (red). We treat citations to NIH-funded publications as an indirect link between the grant and the drug. 226 approvals (40.6%) do not cite any NIH-funded publications, either because they have no in-text publication citations or none with NIH funding. We top-code extreme values: 28 approvals reference 70 or more total publications (max.=390). In both panels, links include all NIH-funded research, regardless of "at-risk" status.

Figure S2: Cumulative distribution of drug approvals by their citations to "at risk" NIH publications (FDA-approved NMEs, 2000-2023)



Note: The horizontal axis shows, for each new molecular entity (N = 557), the fraction of scientific articles cited in its pre-approval U.S. patents that would be counterfactually cut under a 40% reduction in the size of the NIH budget; the vertical axis plots the running count of approvals when NMEs are ordered from the lowest to the highest fraction, thus tracing a step-wise cumulative density curve. The purple vertical segment at 0% reflects the 271 approvals (48 % of the total) whose patents contain no references to at-risk research publications.

Figure S3: Sensitivity Analysis—Varying the size of the proposed budget cut



Note: The figure depicts the number of NME drug approvals (2020-2023) that would be affected (in the sense of being linked to at-risk research referenced in a pre-approval patent) as the size of the proposed NIH budget cut increases from 0 to 100%. For a 100% budget cut, every approval linked to any one publication acknowledging NIH support would be at risk, or 331 approvals (59.4%).