NBER WORKING PAPER SERIES

RESTING ON THEIR LAUREATES?
RESEARCH PRODUCTIVITY AMONG WINNERS OF THE NOBEL PRIZE
IN PHYSIOLOGY OR MEDICINE

Jay Bhattacharya
Paul Bollyky
Jeremy D. Goldhaber-Fiebert
Geir H. Holom
Mikko Packalen
David M. Studdert

Working Paper 31352
http://www.nber.org/papers/w31352

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023

We thank Bruce Weinberg, Joel Blit, Neeraj Sood, Vetla Torvik, Neil Smallheiser, Gerald Marschke, and Partha Bhattacharyya for helpful discussions. We also thank seminar participants at the Institute for Fiscal Studies (London), the University of Illinois at Chicago Institute of Government and Public Affairs, Stanford Medical School, Ca'Foscari University of Venice, Johns Hopkins University, American Economic Association Annual Conference, Asia Pacific Innovation Conference (Beijing), Ohio State University, the University of Southern California, the Latin American, Caribbean Economics Association - Latin American Meeting of the Econometric Society (LACEA-LAMES) Annual Meeting, India Conference on Innovation, Intellectual Property, and Competition at the Indian School of Business (Hyderabad), and the National Bureau of Economic Research working group on Invention in an Aging Society for helpful feedback. Finally, Drs. Bhattacharya and Packalen thank the National Institute of Aging for funding this research through grant P01-AG039347. None of the other authors have any relevant funding to disclose. None of the authors have any conflicts of interest to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Resting on Their Laureates? Research Productivity Among Winners of the Nobel Prize in Physiology or Medicine
Jay Bhattacharya, Paul Bollyky, Jeremy D. Goldhaber-Fiebert, Geir H. Holom, Mikko Packalen, and David M. Studdert
NBER Working Paper No. 31352
June 2023
JEL No. I1,I23,O3

ABSTRACT

The Nobel Prize in Physiology or Medicine is the most prestigious and coveted award in medical research. Anecdotal evidence and related research suggest that receiving it may adversely affect research productivity. We compared the post-Nobel research output of laureates (prize years: 1950-2010) with their pre-Nobel output and with the output of a matched control group consisting of winners of the Lasker Award, another highly prestigious medical research prize. Pre-Nobel, laureates' publications were more voluminous, highly cited, and novel than those of (future) Lasker winners. Post-Nobel, laureates' productivity decreased sharply, eventually falling below that of Lasker winners on all three measures. These declines may reflect diversionary effects of the Prize, changed incentives, or intrinsically different career arcs for medical researchers who win the Nobel Prize.

Jay Bhattacharya
117 Encina Commons
CHP/PCOR
Stanford University
Stanford, CA 94305-6019
and NBER
jay@stanford.edu

Paul Bollyky
Stanford University
School of Medicine
241 Beckman Building
300 Pasteur Dr.
Stanford, CA 94305
pbollyky@stanford.edu

Jeremy D. Goldhaber-Fiebert
117 Encina Commons
CHP/PCOR
Stanford University
Stanford, CA 94305-6019
jeremygf@stanford.edu

Geir H. Holom
Department of Health Economics
and Health Management
Institute of Health and Society
University of Oslo
Oslo, Norway
and Stanford University
g.h.holom@gmail.com

Mikko Packalen
University of Waterloo
Department of Economics
200 University Avenue West
Waterloo, ON N2L 3G1
Canada
packalen@uwaterloo.ca

David M. Studdert
117 Encina Commons
CHP/PCOR
Stanford University
Stanford, CA 94305-6019
studdert@stanford.edu

INTRODUCTION

Upon his death in 1896, Alfred Nobel, the Swedish inventor and businessman, bequeathed most of his fortune to establish annual prizes in chemistry, physics, literature, physiology or medicine, and peace. He stipulated that the prizes should go to those who "conferred the greatest benefit on mankind." (Nobelprize.org, 1911) Between 1901 and 2016, 579 Nobel Prizes were awarded to 911 laureates; in Medicine or Physiology, 107 prizes were awarded to 211 laureates (Nobelprize.org, 2023).

In science and medicine, a Nobel Prize has unrivaled prestige. Media coverage of the announcements and ceremony is intense, especially in laureates' home countries (Baram-Tsabari and Segev, 2015; Bucchi, 2012; Ganetz, 2016; and Zuckerman, 1977). The accolade can be life-altering. Almost overnight, laureates experience a dramatic boost in notoriety and influence. They mix with heads of state and find newly-dedicated parking spaces at work. Many laureates also experience an onslaught of speaking invitations and travel, pressure to lead professional societies, and greater involvement in public affairs and discourse. Paul Samuelson, the 1970 Laureate in Economics, wryly suggested that laureates become "pontificating windbags" who "wither away into vainglorious sterility." True to his field, he urged estimates of "before-and-after age-corrected productivities on an age-corrected basis." (Samuelson, 2001)

This paper analyzes how researchers' productivity changes after winning the Nobel Prize in Physiology or Medicine. We focus on the volume, impact, and novelty of their publications and the extent of their collaboration with coauthors. We compare laureates' performance on these measures with that of two groups: (1) laureates themselves before the Prize; and (2) winners of the Lasker Award, another prestigious but less famous medical research prize. Given the increased demands on their time outside science, we hypothesize that winning a Nobel Prize is associated with substantially lower productivity.

Our study sample consists of all winners of three prizes between 1950 and 2009: the Nobel Prize in Physiology or Medicine, the Albert Lasker Basic Medical Research Award, and the Lasker-

DeBakey Clinical Medical Research Award.[1] Eligibility for all three prizes is limited to people alive when the Prize is announced. The Lasker Awards have been called “America’s Nobels,” and the two award categories we focus on have been awarded annually since 1946. Using official, publicly-available sources of information, we gather the name, date of birth, date of death (if applicable), and prize year of each winner (Lasker Foundation, 2023; Nobelprize.org, 2023). In addition, drawing on biographical information and prize citations, we classify each winner into one of 15 mutually-exclusive research fields. We base this classification on the nature of their training and research and the body of work for which they received the Prize.

Many studies have addressed the productivity effects of tournaments and awards, although relatively few have focused on “ultra-elite” scholars and highly prestigious prizes like the Nobel. Zuckerman (1977), studying 41 Nobel laureates, found short- and long-term decreases in publication volume after winning the Prize; the reductions were particularly large among older scientists who had been less eminent before their win. Borjas and Doran (2015) studied winners of the Fields Medal, the most prestigious Prize in mathematics, and found post-medal decreases in their papers, citations, and mentoring activity, relative to “contenders” who did not win; medalists were also more likely to shift their research into areas outside those in which they had made their name. By contrast, Chan et al. (2013) studied winners of the John Bates Clark Medal, awarded by the American Economic Association to outstanding scholars under 40 years; they went on to have more and more highly-cited publications than a comparable group of high-performing economists who did not win the medal.

[1] We focus on this time period to provide adequate time to observe outcomes for all Nobel winners in the years after they win the prize.

METHODS

We employ an accurate search strategy to identify each winner's publications. Because prize winner names are not generally unique and prize winners have sometimes been listed in publications with variants of their names, searching MEDLINE by author name can result in type I and type II errors. Instead, we identify relevant publications using author disambiguation. Disambiguation involves identifying clusters of publications likely to include the searched-for author based on name, common MeSH terms, and other predictors like common coauthors, institutions, etc. Torvik & Smallheiser (2009) show this approach produces low type I and type II error rates. We check the automated disambiguation process by validating the results against a database that contains the publication information of many US Nobel winners for Physiology and Medicine and the Lasker prizes.

We use meta-data available in MEDLINE to determine publication volume and to distinguish research articles from non-research articles (e.g., letters, editorials, reviews). For each researcher in each calendar year, we determine career age based on the years since an author's first publication appeared in MEDLINE. Our impact measures rely on the Expanded Science Citation Index (ESCI), compiled by the Web of Science from the Web of Science. These data provided us with the number of forward citations (that is, the number of papers that cite a published paper) for each paper published by the scientists in our sample in each year after the publication of the initial paper through 2014 (because our data access to the ESCI runs through 2014).

In addition to total lifetime citations, we analyze the number of forward citations received in the first five years after publication and between six and ten years after publication. These statistics enable us to distinguish papers' short-, medium, and long-run impacts as measured by citations to the paper.

Our novelty measure employs a comprehensive natural-language analysis of the text of the abstracts and titles of every English-language publication indexed in MEDLINE (Packalen and Bhattacharya, 2017). We determine the earliest appearance date in a peer-reviewed publication for each entry in the US National Library of Medicine's Unified Medical Language System

(UMLS) thesaurus. We hereafter refer to these UMLS terms as “ideas.” We then classify every paper based on the date of the youngest idea in the article. The resulting novelty measure is denominated in years. For instance, if the youngest idea in a paper published in 1982 is polymerase chain reaction (PCR), we assign a novelty score of zero years; by contrast, we would assign an article published in 2012 with the same youngest ideas a novelty score of 30 years.

For our statistical analysis, we match each Nobel prize winner to the complete set of Lasker winners who published in the same field as the Nobel winner and who was born within a decade of each other. We compare outcomes each year for Nobel winners against a matched set of Lasker controls. We align the career age for each matched scientist relative to the career age at which the Nobel winner in each matched set won the Nobel. So, for example, if a Nobel winner was awarded the Prize in the 20^{th} year after publishing his or her first paper, we designate career ages in the 20^{th} year or before as “before” the Prize and all career ages after the 20^{th} year as “after” for all members of the matched cohort.

For all of our statistical comparisons, we calculate the residual values of our outcome variables based on regressions that remove variation due to the fields of study and career and calendar age of the researchers, as well as the calendar year of publication of each research publication. Detailed information about these calculations and full regression results are available in the Appendix to this paper. We report bootstrapped confidence intervals for all our results (also reported in full in the Appendix), where we report the results of a wide variety of sensitivity analyses in which we vary our assumptions in our statistical procedure. The main results we present here are qualitatively and quantitatively robust to all these checks.

RESULTS

Our sample consists of the universe of Nobel Prize and Lasker Award winners between 1950 and 2010 – a total of 140 Nobel Prize winners and 176 Lasker Award winners. Table 1 shows the distribution over the fields of study of these two groups. Nobel winners and Lasker winners are drawn from a broad array of biomedical fields. However, Nobel winners tend to come from the basic sciences at higher rates and clinical fields at lower rates than Lasker winners. Nobel winners are more likely than Lasker winners to study cell biology and genetics, while they are less likely to study pharmacology, cardiovascular medicine, or surgery.

Table 2 compares the Nobel winners and Lasker winners in terms of their history of peer-reviewed publications, forward citation rate (which measures the rate at which other peer-reviewed papers cite papers published by the Nobel and Lasker winners), coauthorship history, and novelty score (that is, the age of the newest ideas mentioned or used within each paper). Both Nobel and Lasker winners are incredibly prolific scientists, but Lasker winners publish at a slightly higher rate than Nobel winners, with little difference in non-research publications. This higher rate adds up over a career; Nobel winners have 173 lifetime publications, while Lasker winners have 191. Nobel winners also have fewer first-authored papers per year, slightly fewer last-authored papers, and their papers have the same number of coauthors on average.

Nobel and Lasker winners differ in their papers' forward citation rates. Both publish widely influential papers, but the papers published by Nobel winners are substantially more influential, with ~52 more citations per paper on average. Nobel winners' papers receive an average of ~15 more citations in the first five years after publication than Lasker winners and ~12 more citations between six and ten years after publication.

Finally, Table 2 shows novelty scores, which we define as the year the youngest idea in the paper was first introduced into the biomedical literature minus the year that the paper was published. By definition, the novelty scores will always be less than or equal to zero, with a higher value indicating a more novel paper. Since these scientists publish multiple papers yearly, Table 2 reports the median and 25th percentiles of the novelty scores over their peer-reviewed papers.

The mean values of these novelty score measures do not reveal a large difference in novelty. At the median, the youngest ideas referenced by both are about 13 years old at the time of publication, while the 25th percentile paper in a given year references ideas that are about ten years old.

The results in Table 2 collapse changes that occur over time in the careers of these scientists, and in particular, what happens after the Nobel prize is awarded. Figure 1 displays publication outcomes, including the number of papers (Panel A), the number of first-authored papers (Panel B), the number of last-authored papers (Panel C), and the number of coauthors (Panel D). Each panel (like all the figures in this paper) plots residual values of these outcomes from regressions that adjust for the fields of study and career and calendar age of the researchers, as well as the calendar year of publication of each research publication. Each panel's vertical line (at t = 0) corresponds to the career age when the Nobel winner received the Prize. The x-axis of each panel represents the career age of the prizewinners normalized around this line. In each figure, the black dots are data from the Nobel winners, while the white dots are data from the matched Lasker winners.

Until about ten years before the Nobel prize, Nobel and Lasker winners have roughly the same number of publications. In the ten years before the Prize, Nobel winners publish about one paper more per year than matched Lasker winners. Then, after winning the Nobel, the productivity of Nobel winners drops substantially, so that by ten years after winning the Prize, Nobel winners publish about a paper per year fewer than Lasker winners. Panels B and C show a similar post-Nobel productivity drop for first-authored and last-authored papers. Finally, Panel D shows a sharp rise in the number of coauthors per paper, starting the year before the Nobel Prize and continuing for about five years after winning the Nobel.

Figure 2 shows forward citation outcomes for Nobel winners and matched Lasker winners. Panel A shows the average number of forward citations from the paper's publication date up to 2014 (the last year of our access to citation information). Panels B and C show average forward citations for the first five years and in years six to ten after publication, respectively. It is instructive to consider papers published by Nobel winners 20 years before the Prize. Those papers garner an average of 150 more citations than papers published at the same career stage by matched Lasker winners (Panel B). However, in the first five years after publication, those same papers earn about seven more citations than the papers by matched Lasker winners and about 20 more citations in years six to ten after publication. Most of the additional citations to those older papers occur after the Nobel Prize. By contrast, papers published by Nobel winners after winning the Prize garner about the same number of citations as matched Lasker at the same career stage.

Figure 3 shows our novelty outcomes, with Panel A reporting median novelty scores and Panel B reporting 25th percentile novelty scores. Qualitatively, both tell a similar story: the papers that Nobel winners publish before they win the Prize typically employ newer ideas than papers published by Lasker winners at the same career stage. During those years, Nobel winners employed ideas between one and three years younger than the matched Lasker winners. Our statistical Appendix shows that this result is statistically significant at the $p<0.01$ level. By contrast, after winning the Prize, Nobel winners tend to publish papers that employ older ideas than the matched Lasker winners.

DISCUSSION

Lasker and Nobel Prize winners in biomedicine are unique scientists whose work has a profound scientific and practical impact. The Nobel Prize in biomedicine serves as tangible recognition of this impact, which in turn fundamentally alters the lives of Nobel laureates. The prestige of the Nobel prize sends a signal to budding young scientists of society's importance on scientific knowledge. Given the long and arduous training required to conduct research in biomedical sciences at the highest level, this social sanction can help motivate students to enter science and stay the course through the many years of necessary and often thankless training. Furthermore, winning the Nobel prize can focus the scientific community's attention on the work of the prizewinners, expanding the use of the prizewinner's ideas to new areas and fields. At the same time, winning the Nobel prize can bring fame and outsized attention from non-scientists, students, journalists, and others that can deprive Nobel winners of time in their laboratories.

Our analysis documents the costs and some of the benefits of the Nobel Prize discuss in the paragraph above. We compare the scientific output of Nobel winners and matched Lasker winners (working in the same field, born within a decade of each other, and at the same career stage). We find that Nobel and Lakser winners publish at roughly the same rate until about ten years before the Nobel prize. At that point, Nobel winners publish about one or two papers more per year than Lasker winners.

The papers that Nobel winners publish before the Nobel earn more citations than those written by Lasker winners even before the Nobel. Over that period, Nobel winners publish papers with ideas about one to three years younger than Lakser winners. However, after winning the Nobel, the productivity of Nobel winners drops substantially below matched Lasker winners. They publish fewer papers per year with ideas that are no longer newer than those of Lasker winners and which receive roughly the same number of citations as Lasker winners' papers.

There are three plausible explanations for our results. The first and the most compelling is that the Nobel Prize reduces productivity by drawing its recipients away from research. Producing innovative and influential research output demands considerable time, effort, and focus. The

Nobel Prize in Medicine or Physiology provides a platform to serve as ambassadors for science. Laureates often step onto this platform, replacing time in the laboratory with time leading committees and institutions, serving on government and professional bodies, and writing books and delivering talks for general audiences. Zuckerman's interviews with laureates found abundant evidence of such "diverting consequences" (***Error! Bookmark not defined.***).

A second explanation is that laureates' decline in productivity stems more from changed incentives than reallocated effort. In the pre-Nobel phase of their career, eventual Nobel winners may (correctly) perceive their idea to have exceptional promise though many properties of the ideas they explored remain undeveloped. Convincing the wider scientific community requires considerable effort to develop, test, and validate the discovery. The Prize is a highly visible acknowledgment that the intellectual battle concerning the idea has been won. Post-prize, the incentive to pursue further studies to convince others of the idea is thus greatly diminished.

Furthermore, the cycle of grant writing, experimentation, data analysis, manuscript preparation, and supervision of students and staff wears on even the most talented scientists. With the most prestigious Prize in medical research behind them, laureates' appetite for this grind may diminish, especially when other enticing opportunities are knocking. One physicist analogized his flagging vigor after winning the Nobel Prize to "the lady from Boston who said, 'Why should I travel when I'm already here?'" (Bernstein, 1975). Lasker winners, on the other hand, may press on with their research with fewer distractions. A continuing desire to win the Nobel Prize may also be a factor in explaining their sustained productivity, particularly for winners of the Basic Medical Research Award, more than half of whom subsequently win the Nobel Prize.

Finally, scientists who win the Nobel Prize may tend to have intrinsically different productivity trajectories than other high-performing scientists. The spate of transformative ideas evident in the early stages of their career may be unsustainable over the longer run, regardless of the Prize. Such regression toward the mean is well documented in many other aspects of human performance (Kahneman, 2011).

Our study cannot disentangle the extent to which these three explanations or others account for the decline in post-Prize productivity we observed among laureates. The third explanation does not rely on any inference that the Nobel Prize itself reduces laureates' productivity, whereas the first and second explanations hinge on that causal claim.

Our study has several important limitations. First, our study design lends itself to an analysis of the scientific output of productive and accomplished scientists; it cannot be used to study the extent to which the prestige of the Nobel induced those scientists (or others) to devote themselves to science in the first place. Second, our study focuses on biomedical Nobel prize winners and may not generalize to Nobel winners from other disciplines like physics or economics. We analyze biomedical researchers because of the ready availability of comprehensive biomedical publication data; our methods could be readily applied to other scientific publication data, though we do not conduct this analysis here. Third, our statistical methods are necessarily retrospective and non-randomized, so we abstain from using causal language in interpreting our results. It is difficult, though, to imagine that a randomized intervention could be designed to study the consequences of winning a Nobel prize on productivity. Fourth, our list of productivity outputs (publication volume, coauthorship, citations, and novelty scores) is not exhaustive. There are certainly other measures of scientific productivity, such as the training of students, that we do not measure. Finally, readers should not interpret our results to determine whether a Nobel or Lasker winner's prize was incorrectly awarded or whether a scientist who received neither Prize should have won one or the other. The Nobel and Lasker Award committees appropriately consider many factors beyond publication-related outcomes in deciding who should win in any given year, and it is beyond our scope to evaluate how well they do so.

CONCLUSION

The possibility that the results we report may reflect a causal effect of winning the Nobel Prize on highly productive scientists' productivity raises questions about the social utility of the Nobel Prize in Medicine or Physiology. Could the Prize's diffuse social benefits, including promoting science and incentivizing discovery, justify a substantial body of "lost" research output from the world's best medical researchers? This cost-benefit question would be challenging to answer empirically because it requires valuing goods that are both intangible and incommensurable—specifically, weighing important discoveries that were not made (or were made later by someone else) against a diffuse array of prize-related benefits, such as promoting science to legislators and the public and inspiring bright young minds to enter the field and strive for excellence. Nonetheless, questions about productivity effects join other criticisms leveled at Nobel Prizes, including misattribution, gender and region bias, and incompatibility with modern methods of scientific discovery (Nature, 2019; Keating, 2018).

In recent years, the biomedical sciences have experienced an unprecedented aging of its workforce. Young researchers who start research careers in biomedicine often leave the field before they develop into fully independent researchers. The age of earning a first large research grant has risen sharply in the US and elsewhere. The prospect of late-career recognition in the form of a Nobel or Lasker prize has not sufficed to stem this trend. Our results support the idea that funding agencies should increase investments in productive early career scientists who have a taste for newer ideas to encourage the work of budding future Nobel winners who tend to share that taste. Our results also support the development of improved systems of recognition for accomplished early career scientists to encourage them to continue.

We find no productivity decline related to winning the Lasker, after all. Anecdotally, winning such a prize does seem to inhibit productivity. For example, economics Nobelist Paul Samuelson who went on to a long, innovative, and productive career as an economist, was the first recipient of the prestigious John Bates Clark medal, given to "that American economist under the age of forty who is adjusted to have made a significant contribution to economic thought and knowledge." Future work should more explicitly study the productivity effects of winning an early career research award.

References

Baram-Tsabari A and Segev E. (2015) The half-life of a “teachable moment”: The case of Nobel laureates PUBLIC UNDERSTANDING OF SCIENCE 24(3): 326-337/

Bernstein J. Profiles—Physicist~1. The New Yorker, October 17, 1975.

Borjas GJ, Doran KB. (2015) Prizes and productivity: how winning the Fields Medal affects scientific output. J Human Res 2015;50:728-758

Bucchi M. (2012) Visible Scientists, Media Coverage and National Identity: Nobel Laureates in the Italian Daily Press in Scheiel B, Claessens M, Shi S (eds) Science Communication in the World: Practices, Theories and Trends. Springer 2012. pp 259-268

Chan HF, Frey BS, Gallus J, Torgler B.(2013) Does the John Bates Clark Medal boost subsequent productivity and citation success? Working Paper No. 111. February 2013. ISSN 1664-7041.

Ganetz H. (2016) The Nobel celebrity-scientist: genius and personality CELEBRITY STUDIES 7(2):234-248

Kahneman D. Thinking, fast and slow. New York, NY: Farrar, Straus and Giroux, 2011.

Lasker Foundation. (2023) The Lasker Awards. http://www.laskerfoundation.org/awards/

Keating B. (2018) Losing the Nobel Prize: a story of cosmology, ambition, and the perils of science’s highest honor. New York, NY: WW Norton and Co., 2018.

Nature (2019) Boosting inclusivity in the Nobels. Nature Editorial Staff. Nature 2019 Oct;574(7778):295. doi.org/10.1038/d41586-019-03115-0

Nobelprize.org (1911) Excerpt from the Will of Alfred Nobel. http://nobelprize.org/alfred_nobel/will/short_testamente.html

Nobelprize.org. (2023) All Nobel Laureates in Physiology or Medicine. https://www.nobelprize.org/nobel_prizes/medicine/laureates/

Packalen M and Bhattacharya J (2017) “Neophilia Ranking of Scientific Journals” Scientometrics 110(1):43-64 PMID: 28713181 PMCID: PMC5506293 doi: 10.1007/s11192-016-2157-1

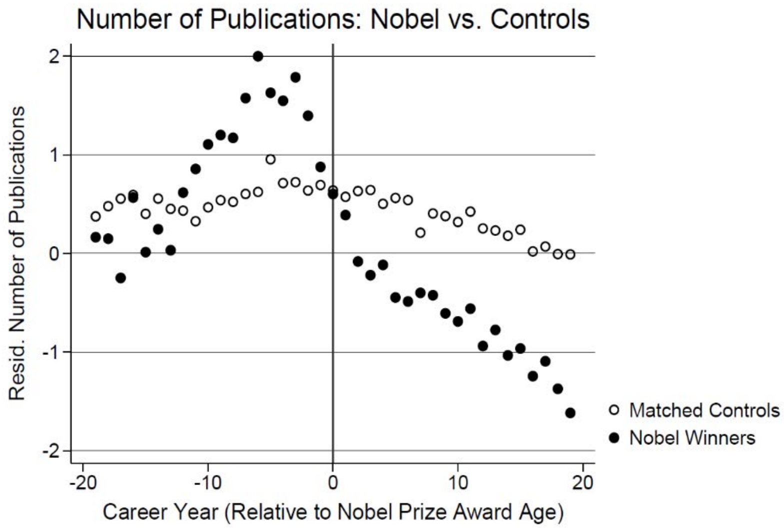
Samuelson P (2001) “Is There Life After Nobel Coronation?”, http://nobelprize.org/economics/articles/samuelson/index.html

Torvik CI & Smallheiser NR (2009) “Author Name Disambiguation in MEDLINE,” ACM Trans Knowl Discov Data 3(3): pii: 11. PMID: 20072710 PMCID: PMC2805000

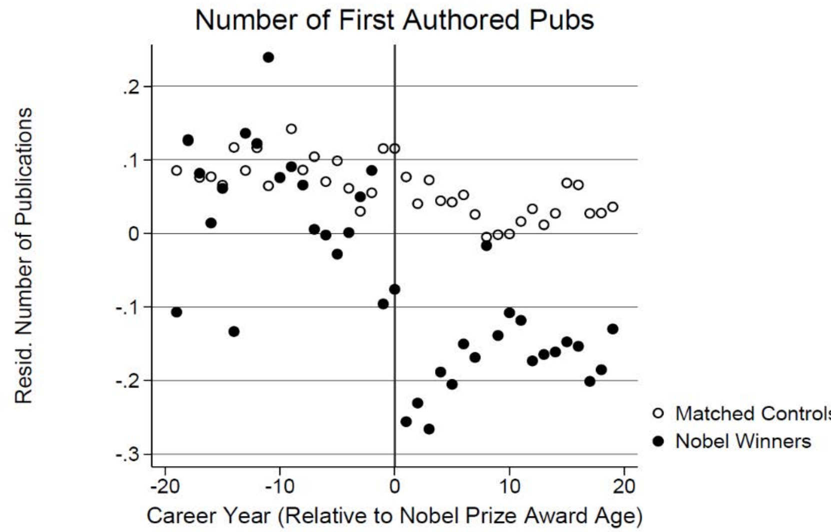
Zuckerman, H. (1977). The scientific elite: Nobel laureates in the United States. New York: Free Press. New edition, 1996, New Brunswick/London: Transactions.

Figure 1: Productivity Outcomes for Nobel Winners vs. Matched Controls

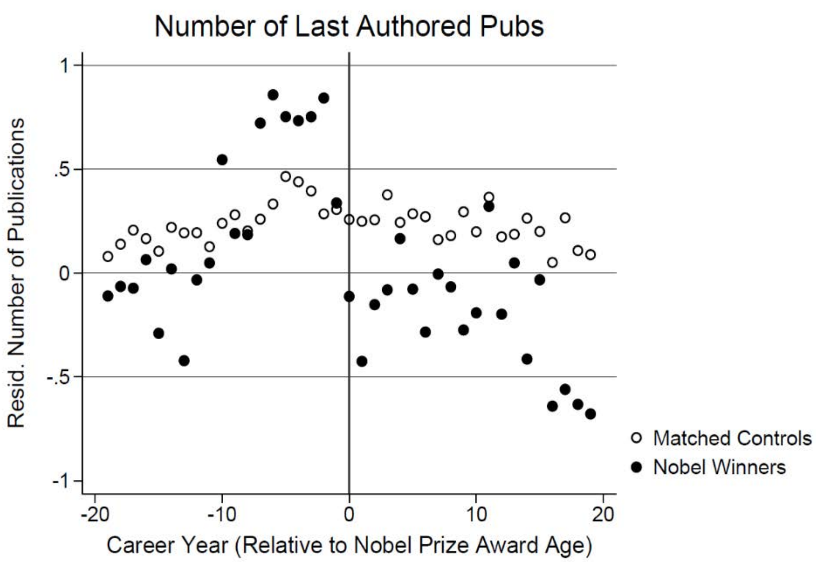
Panel A: Number of Publications



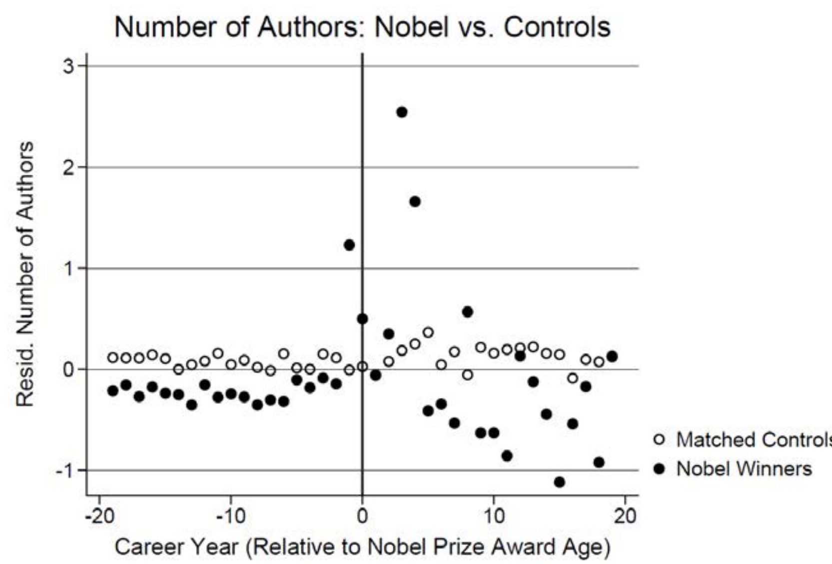
Panel B: Number of First-Authored Publications



Panel C: Number of Last-Authored Publications



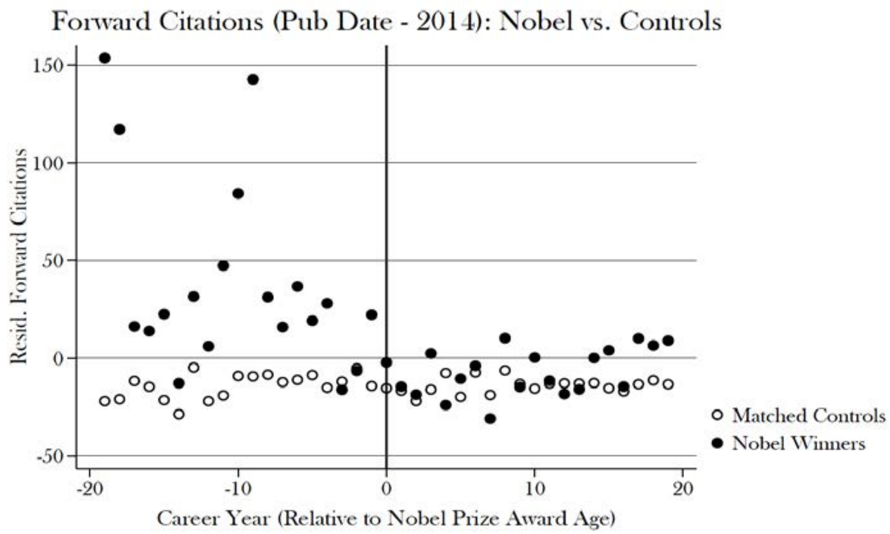
Panel D: Number of Coauthors



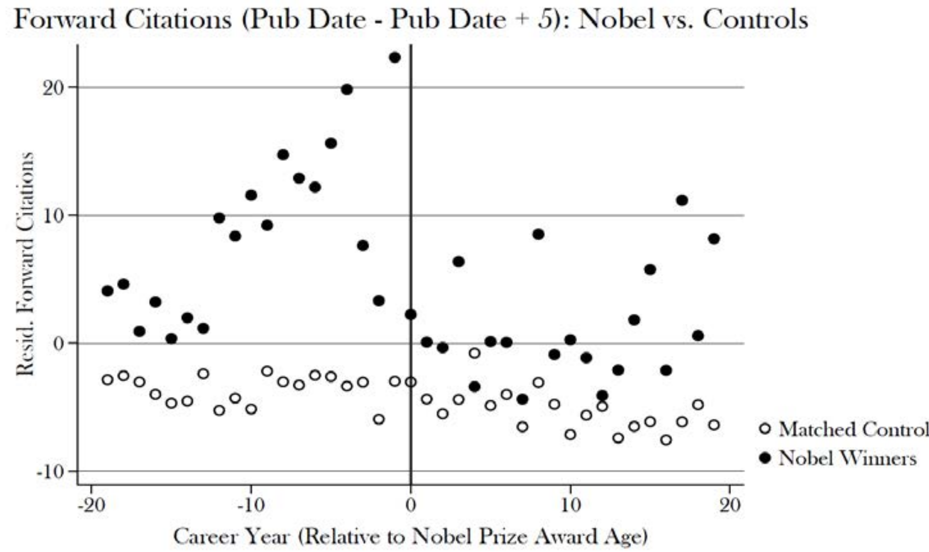
Note: The outcomes reported in this figure are residual values from regression analyses that adjust for publication year, career age, and scientist field of work. Full details about the statistical analysis underlying this figure (including the full set of regression results) are reported in the appendix accompanying this paper.

Figure 2: Citation Outcomes for Nobel Winners vs. Matched Controls

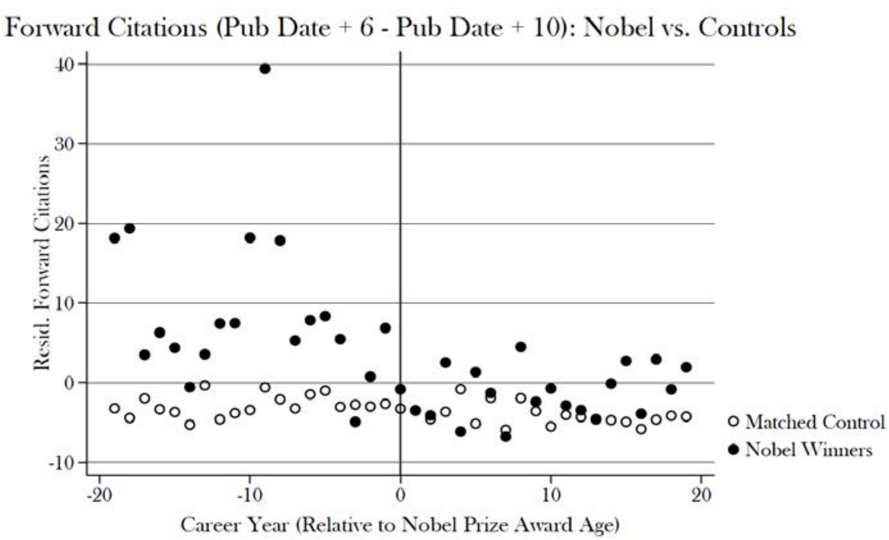
Panel A: Forward Citations Up to 2014 for Papers Published in Each Career Year



Panel B: Forward Citations in First Five years for Papers Published in Each Career Year



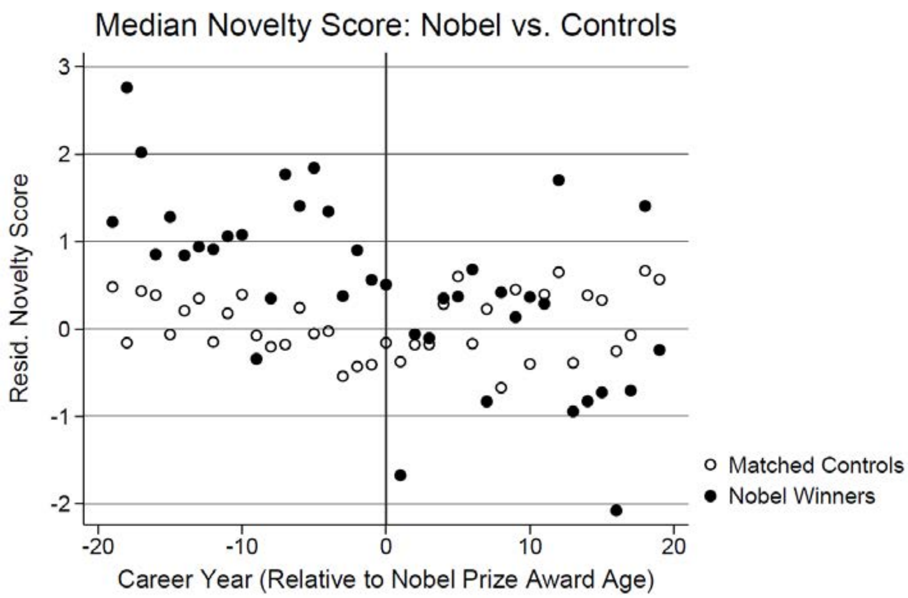
Panel C: Forward Citations between Years Six and Ten for Papers Published in Each Career Year



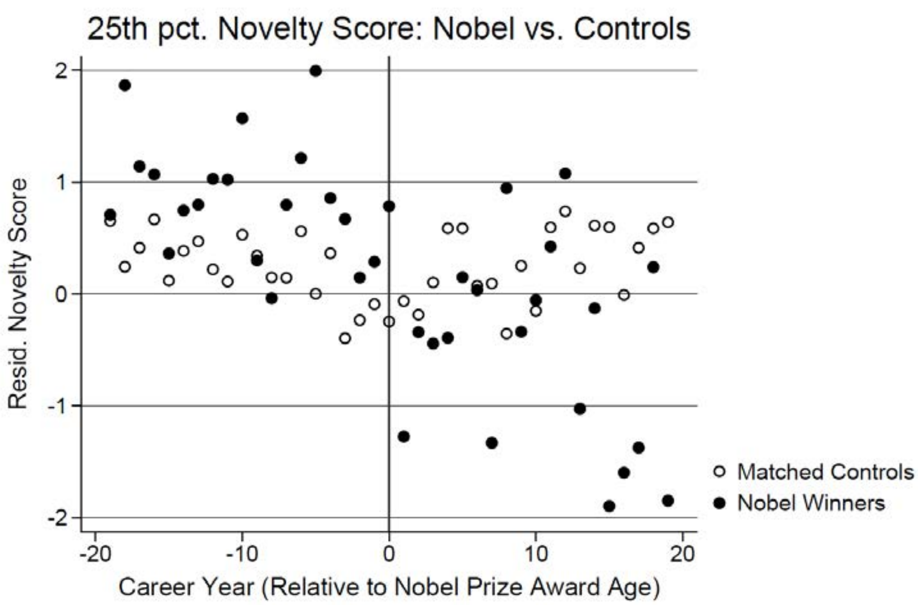
Note: The outcomes reported in this figure are residual values from regression analyses that adjust for publication year, career age, and scientist field of work. Full details about the statistical analysis underlying this figure (including the full set of regression results) are reported in the appendix accompanying this paper.

Figure 3: Novelty Outcomes for Nobel Winners vs. Matched Controls

Panel A: Median Novelty Score (Years) for Papers Published in Each Career Year



Panel B: 25th Percentile Novelty Score (Years) for Papers Published in Each Career Year



Note: The outcomes reported in this figure are residual values from regression analyses that adjust for publication year, career age, and scientist field of work. Full details about the statistical analysis underlying this figure (including the full set of regression results) are reported in the appendix accompanying this paper.

Table 1: Distribution of Field of Work – Nobel Winners vs. Lasker Winners

	All	Nobel	Lasker
N	316	140	176
biochemistry	9%	10%	7%
cell biology	14%	18%	10%
developmental biology	3%	5%	1%
genetics	25%	33%	18%
metabolism & endocrine	7%	9%	5%
pharmacology	14%	3%	23%
physiology	23%	20%	25%
audiology	4%	4%	3%
cardiovascular medicine	11%	4%	16%
hematology & oncology	13%	6%	18%
infectious disease & immunology	25%	25%	24%
neurology & cognition	11%	12%	9%
pediatrics & maternal-fetal medicine	4%	1%	6%
surgery	5%	2%	7%
molecular biology	26%	21%	30%

Note: Percentages sum to more than one hundred in each column because each scientist may have published in more than one field.

Table 2: Publication Output – Nobel Winners vs. Lasker Winners

	All	Nobel	Lasker
Lifetime Publications (total)	183 (177)	173 (157)	191 (191)
Research Publications per Year (mean)	3.3 (2.9)	3.1 (2.5)	3.5 (3.2)
Non-Research Publications per Year (mean)	0.3 (0.4)	0.4 (0.4)	0.3 (0.4)
First-Authored Papers per Year (mean)	0.5 (0.5)	0.4 (0.4)	0.6 (0.5)
Last-Authored Papers per Year (mean)	1.9 (1.9)	1.8 (1.8)	1.9 (2.0)
Coauthors (mean)	3.5 (2.1)	3.5 (2.8)	3.5 (1.3)
Citations per paper (mean)	100.3 (255.3)	128.6 (319.7)	76.6 (181.7)
Citations per paper in first five years after publication (mean)	34.4 (56.3)	42.4 (71.3)	27.7 (38.3)
Citations per paper, 6 – 10 years after publication (mean)	23.8 (58.8)	30.2 (74.4)	18.3 (40.7)
Citation rate per paper per year (mean)	3.2 (3.7)	4.1 (4.0)	2.5 (3.3)
Median Novelty Score (years)	-13.2 (4.7)	-13.1 (5.3)	-13.2 (4.2)
25th Percentile Novelty Score (years)	-10.0 (4.9)	-10.2 (5.6)	-9.9 (4.2)
N	316	140	176

Note: Variance are below each statistic in parentheses. Novelty scores are measured in years based on the age of the ideas contained in each paper published by the prize winners at the time of publication. Novelty scores negative because we calculate them as minus one times idea age. We orient novelty scores in this way so that higher novelty scores indicate a reliance on newer ideas, just as higher citation scores indicate more impactful papers.

Methods Appendix

In this appendix, we provide a detailed description of our data sources and methods used to calculate the statistics that we report in the main paper. In addition, we describe sensitivity analyses that we have performed that assess the robustness of the findings we report to alternate statistical assumptions.

A. Data Sources

Our data sources comprise four types of information about prize winners. The first is which prize or prizes each won and in which years. The second is biographical and career information. The third is information about each winner's corpus of publications, and the fourth is information about the citations that corpus has received over time by other peer-reviewed scientific papers. We describe the data sources and methods for collecting them in this section, providing details about the measures we construct and analytic approaches we take in subsequent sections of the Appendix.

We construct a complete list of the names and nicknames of all individuals who won one or more of the following prizes between 1950 and 2010 (inclusive) and the year(s) in which they won: 1) the Nobel Prize in Physiology or Medicine; 2) the Albert Lasker Basic Medical Research Award; and 3) the Lasker-DeBakey Clinical Medical Research Award. To do so, we visited the websites of the prize granting organizations.[1, 2] We categorize scientists as Nobel winners if they have won a Nobel even if they have also won one or more Lasker awards; otherwise, we categorize them as Lasker winners.

For each winner, we collect biographical and scientific career information. We obtain their birth date (and death date if they died before 2018) from the prize websites and their own websites and Wikipedia. We also categorize their field of scientific research based on 15 broad categories listed in Table A1. Winners are permitted to belong to multiple categories depending on their research focus. Prof. Paul Bollyky, a microbiologist and immunologist and a coauthor on this paper, assigned each scientist to their fields of work. We designed the categories with the following goals in mind: (1) to follow general divisions of scientific fields commonly used and which may be characterized by different patterns, frequency, or volume of publishing; and (2) to help in matching Nobel winners with one or more similar Lasker winners (see section on matching below).

We identify the corpus of biomedical published work that winners generated during each year of their careers. To do so, we search PubMed MEDLINE for all publications through the calendar year 2014.[3] The goal is to identify the PubMed identifier and year of publication for each published work for each winner along with meta-data like the names and order of co-author(s), the type of publication (e.g., research publication, review, letter to the editor, etc.), and research topics of the paper as characterized by Medical Subject Heading (MeSH) terms.[4]

We employ an accurate search strategy to identify each winner's publications. Because prize winner names are not in general unique and because prize winners sometimes identify themselves in publications with variants of their names (e.g., pre/post-marriage, with/without middle initial), searching PubMed by author name can result in both type I and type II errors. Instead, we identify appropriate publications using an author disambiguation algorithm developed by Smalheiser and Torvik in 2009, who show that the algorithm produces low rates of misclassification errors.[5] The algorithm identifies clusters of publications that are most likely written by a scientist based on the scientist's name, as well as common MeSH terms and other predictors like common co-authors, institutions, etc. The disambiguation algorithm identifies groups of publications that are more or less likely to be produced by the author of interest. Hence additional steps are required to determine who published works, especially where attribution is most uncertain. Therefore, we augmented the disambiguation process and validated the automated disambiguation process using a database provided by Pierre Azoulay which contains the publication information of many U.S. winners of the Nobel for Physiology and Medicine and of the Lasker prizes. Additionally, for all publications that are difficult to disambiguate with the algorithm as well as a large random subsample of other publications, we manually rechecked them by looking them up in PubMed as well as comparing them against lists of the winners' publications when available. As multiple processes are used to generate publication lists, we eliminate duplicate records so that each publication for each winner is listed only once.

Once we have collected the corpus of biomedical publications and their meta-data for each winner, we code each winner's corpus in the following ways.

First, based on the calendar year of each winner's earliest paper, we compute the career age of each of his/her publications by subtracting this earliest calendar year from the calendar years of subsequent publications. We term this value the "career age" of a publication for a given winner, which is zero for the earliest publication and a positive integer for publications in subsequent years. The purpose of career age is to be able to track quantities like the number of publications per year since a winner first started publishing in biomedicine.

Second, we categorize publications as research or non-research publications using an algorithm applied to the PubMed metadata field Publication Type, which can have multiple values.[6] For example, it can be a "journal article" as well as an "interview." The purpose of categorizing publications as research or not is because we wish to look at measures like the number of research publications per year separate from other types of publications that PubMed holds. Our specific rule for classifying publications as research is:

1) Define a publication as a research publication if its publication type includes "journal article" *unless* it also includes any from the following list – addresses; autobiography; bibliography; biography; clinical conference; collected works; comment; congresses; consensus development conference; consensus development conference (NIH); dataset; dictionary; directory; duplicate publication; editorial; expression of concern; guideline; interactive tutorial; interview; introductory journal article; lectures; legal cases; legislation; letter; news; newspaper article; patient education handout; periodical index; portraits; practice guideline; publication components; publication formats; publication type category; published erratum; review; scientific integrity review; study characteristics; video audio media; and webcasts.
2) Define a publication as a non-research publication if its publication type does not include "journal article" *and* its publication type is not in the following list – adaptive clinical trial; case reports; clinical study; clinical trial; clinical trial phase i; clinical trial phase ii; clinical trial phase iii; clinical trial phase iv; comparative study; controlled

clinical trial; equivalence trial; evaluation studies; meta-analysis; multicenter study; observational study; twin study; and validation studies.

Finally, from the Expanded Science Citation Index, compiled by the Web of Science, we obtain information about the forward citations that each publication received from other peer-reviewed publications in the years after its publication. [7]

B. Notation Summary and Construction of Scientist-Year Level Panel Dataset

In this section of the appendix, we introduce some notational conventions to facilitate the discussion about a dataset with so much bookkeeping and describe the construction of the main outcome variables of our paper.

- Let $n_1 \ldots n_S$ enumerate the S Nobel winners in our sample and $N = \{n_1 \ldots n_S\}$ denote the set of Nobel winners;
- Let $\ell_1 \ldots \ell_M$ enumerate the M Lasker winners in our sample and $\mathcal{L} = \{\ell_1 \ldots \ell_M\}$ denote the set of Lasker winners; [1]
- Let i denote a winner of either type, so $i \in N \cup \mathcal{L}$;
- Let t denote calendar year (either a publication of a paper or the awarding of a prize);
- Let $byear(i)$ be the birth year of scientist i;
- Let $cyear(i, t)$ be the career age of scientist i in calendar year t, defined as the number of years since scientist i's first peer-reviewed published paper;
- Let $nyear(n_i)$ be the career age of scientist n_i in the calendar year the Nobel prize was awarded;
- Let $K = \{f_1, f_2, \ldots, f_{15}\}$ enumerate the complete set of fields of study of our scientists (listed in Table A1); and
- Let $f_k(i)$ be an indicator variable that equals one if scientist i works in field k, and zero otherwise, for each scientist $i \in N \cup \mathcal{L}$ and field $k \in K$.

[1] We set the values of $\ell_1 \ldots \ell_M$ and $n_1 \ldots n_S$ such that there is no overlap in these numbers. Substantively, this means that each prizewinner is assigned to either N or $\mathcal{L}$, but not both. This modeling choice is discussed in Section D of this appendix below.

From the MEDLINE publication data and the biographical data we have collected on each scientist, we construct a scientist-year level panel data set. This panel dataset is an annual, longitudinal dataset which tracks publication information about the prize winners over their entire careers, augmented with biographical information about each prize winner. From the MEDLINE publication data collected from the list of winners and the Web of Science, Science Citation Index, we calculate variables: (1) the total number of publications by winner i in calendar year t $(pubs_{it})$; (2) the median novelty score (*nov*) among the papers published by winner i in calendar year t $(nov50_{it})$; (3) the 25th percentile of the novelty score (*nov*) among the papers published by winner i in calendar year t $(nov25_{it})$; (4) the mean number of coauthors $(auth_{it})$ among winner i's papers in calendar year t; (5) the total number of first authored papers by winner i in year t $(fpubs_{it})$; (6) the total number of last authored papers by winner i in year t $(lpubs_{it})$; (7) the total number of forward citations on papers published by winner i in year t $(cites_{it})$; (8) the total number of forward citations on papers published by winner i in year t in the first five years after publication (that is, between t and $t+5$, inclusive) $(cites5_{it})$; (9) the total number of forward citations on papers published winner i in year t between six and ten years after publication (that is, between $t+6$ and $t+10$, inclusive) $(cites10_{it})$; and (10) the average citation rate (defined as the number of forward citations a paper receives per year over the lifetime of its publication) of papers published by winner i in year t $(cite\_rate_{it})$.

The construction of our citation outcomes and novelty scores deserves some additional explanation. The latter is described in more detail in Appendix Section C immediately below. We derive our citation statistics from the Web of Science data. For each paper (denoted j) published in calendar year t, we observe the total number of forward citations $\left(cites_{jt\tau}\right)$ received in each subsequent year, $\tau \geq t$. This object is defined at the article-year level, but we need a scientist-year level measure of citation frequency $(cites_{it})$. We calculate this by counting the forward citations that each paper receives over time, and attribute this number to the scientist and year in which the paper was published. Formally, let P_{it} denote the set of papers published by scientist i in calendar year t. We define:

$$cites_{it} = \sum_{j \in P_{it}} \sum_{\tau=t}^{2014} cites_{jt\tau}.$$

The inner sum ends at 2014 because this is the last calendar year we have obtained permission to use from Web of Science. This induces a slight underestimate on the citation frequency of more recent prize winners, which applies in equal measure to both Nobel winners and Lasker winners.

While $cites_{it}$ provides a measure of the long run influence of the work that scientist i performed in year t, it is interesting to observe the extent to which a scientist's work was influential in the short and medium term, a few years after publication. To this end, we define $cites5_{it}$ and $cites10_{it}$ analogously to $cites_{it}$, except we limit the number of years after publication that citations are counted to zero to five years after publication in the case of $cites5_{it}$ and six to ten years after publication in the case of $cites10_{it}$. So,

$$cites5_{it} = \sum_{j \in P_{it}} \sum_{\tau=t}^{max(t+5,\ 2014)} cites_{jt\tau},$$

and

$$cites10_{it} = \sum_{j \in P_{it}} \sum_{\tau=t+6}^{max(t+10,\ 2014)} cites_{jt\tau}.$$

Finally, to address the problem caused by the fact that papers published longer ago have a longer time to collect citations, we calculate a citation rate statistic, $cite\_rate_{it}$, that accounts for this fact by counting the rate at which publications in year t receive forward publications in subsequent years divided by the number of years at risk. We define the citation rate as follows:

$$cite\_rate_{it} = \sum_{j \in P_{it}} \sum_{\tau=t}^{2014} \frac{cites_{jt\tau}}{(2014 - t + 1)}.$$

We organize our data with observations defined only in every year in which a winner has at least one publication, and winners represented once for every year that they publish any papers; this constitutes our *unmatched* panel dataset. In years where a prize winner has no publications, we treat each of these variables as missing, except $pubs_{it}$, $fpubs_{it}$, and $lpubs_{it}$, which we set to zero. We use this unmatched panel dataset for creating the final versions of our outcome variables described in Appendix Section E below.

C. Novelty measure construction

For our corpus of medical research papers, we must first determine which published papers rely upon on new ideas and which rely upon on older ideas. Our strategy, which builds on a previously published paper, is to analyze the entire corpus of over 23 million research papers in the MEDLINE database (including those published by non-prize winners).[8] MEDLINE represents a nearly comprehensive index of peer-reviewed journal articles in life sciences, with a concentration on biomedicine. For each research paper, especially those published from 1960 onwards, MEDLINE provides a title and abstract which we analyze for their textual content.

To analyze the textual content of the title and abstract of all of the papers indexed MEDLINE, we take advantage of the availability of a large and well-accepted thesaurus, the United Medical Language System ("UMLS"), which is maintained by the U.S. National Library of Medicine.[9] We allow each term in this thesaurus to represent an idea, broadly interpreted. To determine which ideas each paper builds upon, we first search the title and abstract of each paper for the presence of each of the more than 5 million terms that appear in the UMLS thesaurus. This search provides us with a mapping from each UMLS term to the year in which the idea represented by the term was first introduced into the biomedical literature. One major advantage of the UMLS thesaurus is that it reveals which terms are synonyms, allowing us to treat synonyms as representing the same idea. Thus, the year an idea entered the corpus is the earliest year that any of its synonymous terms were used in a title or abstract of an indexed paper.

With this mapping in hand, we revisit each paper published by the prize winners in our sample. For each paper in this sample, we determine the vintage of each term that appears in its title and abstract, based on the paper's publication year and the year in which the term first appeared in the published biomedical literature. This calculation permits us to determine the age of the newest term that appears in each paper – which we term the novelty score. For example, for papers that mention a term introduced in the same year that the paper was published, we assign a novelty score of zero years. We thus define the novelty score (*nov*) as the minus one times number of years since the newest idea in each paper was initially introduced into the biomedical peer-reviewed corpus. We orient *nov* such that a higher novelty score indicates a reliance on newer ideas, in order to ease interpretation (in most of the other outcomes we analyze such as our citation and publication outcomes, a higher score is typically better or more desirable). We use *nov* to calculate percentiles of the distribution of novelty scores among the set of papers that each prize winner published each year. In particular, we calculate the 25th percentile ($nov25_{it}$) and median novelty ($nov50_{it}$), scores from this distribution for each winner in each year.

D. Matching Nobel Winners and Lasker Winners

In this section, we describe our construction of a matched group of Lasker winner for each Nobel winner. To start, we must address one difficulty in creating such a match -- scientists who won both a Nobel and a Lasker during our observation period. Our approach is to assign each such joint Nobel/Lasker laureate to the set N only. That is, each scientist prize winner in our sample is assigned either set N or set $\mathcal{L}$, never both, so $N \cap \mathcal{L} = \emptyset$, and a joint winner is assigned to N. The reason for this exclusive assignment is two-fold. First, the Nobel is the more exclusive prize, with a smaller set of winners who are the primary focus of this paper. A dual prize winner is, for our purposes, primarily a Nobel winner. Second, our empirical strategy requires creating matched sets of Nobel and Lasker prize winners. Assigning dual prize winners to both sets N and $\mathcal{L}$ would lead to the dual winners being matched to themselves, which we avoid with our exclusive assignment rule.

To match each Nobel winner in N to Lasker winners in $\mathcal{L}$, we select the set of (exclusive) Lasker winners who share at least one field as the Nobel winner, and who were born within a decade of the Nobel winner. That is, for each $n_s \in N$, we select a matched subset of the Lasker winners, $\mathcal{L}(n_s) \subset \mathcal{L}$, such that:

$$\mathcal{L}(n_s) = \{\ell_m \in \mathcal{L} \mid |byear(n_s) - byear(\ell_m)| \leq 10 \text{ and } \exists k \text{ such that } f_k(n_s) = f_k(\ell_m) = 1\}.$$

There is no guarantee, given this definition, that $\mathcal{L}(n_s)$ contains any matched Lasker winners at all, or that there will only be one matched winner. Figure A1 shows the distribution over all the Nobel winners of the number of matched Lasker winners. In fact, only two Nobel winners (Albert Claude and Edward Calvin Kendall) have no matched Lasker winners, and we drop them from our main analysis. In a sensitivity analysis (reported in Section G below), we modify the definition of $\mathcal{L}(n_s)$ to permit matches based on fields alone (and no birth date restriction). Using this modified definition, every Nobel winner matches at least one Lasker winner. The results from this analysis are both qualitatively and quantitatively nearly identical to the results we present in the main paper. Additionally, the same Lasker winners may meet the match criteria for multiple Nobel winners, which is a reasonable outcome given our objectives, but pose some problems in statistical inference that we address in Section F of this appendix.

Since our goal is to create a panel dataset with annual observations on each scientist, we need to define our time variable appropriately. The natural choice – calendar year – is not appropriate since a Nobel winner and matched Lasker winners are likely to be at different career stages when the Nobel winner won the Nobel prize. Instead, we center our time measurement around the career age of the Nobel winner n_s when the Nobel was awarded – $nyear(n_s)$. For each matched Lasker winner in $m \in \mathcal{L}(n_s)$, we define an affine transformation of $cyear(m, t)$:

$$\overline{cyear}_{\ell_m, n_s}(t) \equiv cyear(\ell_m, t) - nyear(n_s).$$

The first subscript on $\overline{cyear}_{\ell_m, n_s}(t)$ represents the scientist whose career age is being measured at time t, while the second subscript represents the match group, $\{\mathcal{L}(n_s) \cup \{n_s\}\}$, to which this scientist belongs. With this notational convention, we can apply this transformation for each Nobel winner as well to track career age centered around the year the Nobel prize was awarded:

$$\overline{cyear}_{n_s, n_s}(t) \equiv cyear(n_s, t) - nyear(n_s).$$

Our matched dataset takes each Nobel/Lasker matched group, $\{\mathcal{L}(n_s) \cup \{n_s\}\}$, and merges in all the scientist-year level publication information from the unmatched dataset that we describe in Appendix Section B. As we note above, Lasker winners who match to multiple Nobel winners have their publication information included multiple times in this final matched dataset.

E. Measuring Career Output as Function of Career Age

The simplest possible statistical analysis would compare the mean values of our ten outcome variables ($pubs$, $nov50$, $nov25$, $auth$, $fpubs$, $lpubs$, $cites$, $cites5$, $cites10$, and $cite\_rate$) for the set of Nobel winners against the mean values of those outcomes for at the same career age of their matched controls . Recall from Appendix D, that we measure career age relative to the calendar year when the Nobel winner in each matched set won. This construction guarantees that w*ithin* each Nobel/matched Lasker group, unadjusted means by career age would compare the outcomes of the Nobel winner against the outcomes of the matched Lasker winners *at the same career age*, both before and after awarding of the Nobel. However, *across* Nobel winners and their matched Lasker groups, $\mathcal{L}(n_s)$, a simple comparison of means by career age would not hold career age constant, since different Nobel winners won their prize at different career ages. Additionally, a simple comparison of mean outcomes by career age would be confounded by the fact that we analyze data from more than 60 years of publications and that different winners worked and published in different fields. Secular trends in publishing over that 60-year period and different publishing norms in different scientific fields would thus confound any simple comparison of means.

To address this problem, we regress each of our ten outcomes on career age $\left(cyear(i,t)\right)$, career age squared, a complete set of field indicators $\left(f_1(i) \ldots f_{15}(i)\right)$, and calendar year (t) using the unmatched version of the panel dataset that we describe in Appendix Section B. For instance, we estimate the following regression for $pubs_{it}$:

$$pubs_{it} = \beta_0^{pubs} + \beta_1^{pubs} cyear(i,t) + \beta_2^{pubs} cyear(i,t)^2 + \beta_3^{pubs} t + \sum_{k=1}^{15} \gamma_k^{pubs} f_k(i) + \varepsilon_{it}.$$

In this equation, ε_{it} represents the regression error. The coefficient estimate vectors ($\hat{\beta}$, $\hat{\gamma}$) and other statistics from these regressions are presented in Tables A2 (productivity outcomes), A3 (citation outcomes), and A4 (novelty outcomes).[2] Using these coefficient estimates, we calculate *residual values* of the outcomes (r_pubs_{it}, r_nov50_{it}, r_nov25_{it}, r_auth_{it}, r_fpubs_{it}, r_lpubs_{it}, r_cites_{it}, r_cites5_{it}, $r\_cites10_{it}$, and $r\_cite\_rate_{it}$) implied by the regressions. For instance, we calculate r_pubs_{it} as follows:

$$r\_pubs_{it} \equiv pubs_{it} - \left(\hat{\beta}_0^{pubs} + \hat{\beta}_1^{pubs} cyear(i,t) + \hat{\beta}_2^{pubs} cyear(i,t)^2 + \hat{\beta}_3^{pubs} t + \sum_{j=1}^{15} \hat{\gamma}_k^{pubs} f_k(i) \right)$$

By construction, these residual values are purged of variation due to differences between scientists as a result of the included control variables, including career age, calendar year, and the field(s) of study of each scientist. Since we do not include a control variable indicating whether the scientist belongs to N (rather than $\mathcal{L}$), the residuals are appropriately not purged of variation due to a scientist winning a Nobel prize.

Our primary results, Figures 1-3 of the main paper, are non-parametric plots of how the mean residual values of our outcomes vary with career age for Nobel winners and matched Lasker winners separately. Let $\theta_1(a|\text{Nobel})$, $\theta_2(a|\text{Nobel})$, … $\theta_{10}(a|\text{Nobel})$ be the plots corresponding to r_pubs_{it}, r_nov50_{it}, r_nov25_{it}, r_auth_{it}, r_fpubs_{it}, r_lpubs_{it}, r_cites_{it}, r_cites5_{it}, $r\_cites10_{it}$, and $r\_cite\_rate_{it}$ respectively for the Nobel winners, with $\theta_j(a|\text{Lasker})$ for $j = 1 \dots 10$ defined analogously for the matched Lasker group. Each of these plots are functions of career age, denoted by a.

To construct the $\theta_j(a|\text{Nobel})$ plots for the Nobel winners, we calculate mean values of our outcomes over all Nobel winners at the appropriate time points in their career. For a given Nobel winner, i, we calculate the calendar year when the Nobel winner had a career age of a:

$$\overline{cyear}_{n_s,n_s}(t) = a.$$

We denote the value of t that solves this equation as $t'_{n_s}(a)$ to emphasize that it will take different values for different Nobel winners (since they won the Nobel at different career ages).

[2] In Table A2, we obtain estimates for all of the field dummies $f_1 \dots f_{15}$. There is no excluded field because some prize winners work in multiple fields over their career.

We then calculate $\theta_1(a|\text{Nobel})$ as the following conditional mean, taken over all Nobel winners:

$$\theta_1(a|\text{Nobel}) \equiv E\left[r\_pubs_{st} \middle| s \in N, t = t'_{n_s}(a)\right].$$

We define $\theta_2(a|\text{Nobel}) \ldots \theta_{10}(a|\text{Nobel})$ analogously.

The calculation for the Lasker winners is a bit more complicated since there may be multiple Lasker winners matched to each Nobel winner, and a Lasker winner might match to multiple Nobel winners. We handle this complication by first calculating a mean value of each outcome variable for the Lasker winner within each matched group, $\mathcal{L}(n_s)$ for $i \in N$, and then taking the mean across all the groups (as we did for the Nobel winners). To do this, for each outcome and for each matched group, we first calculate the calendar year when each Lasker winner in a match group $\ell_m \in \mathcal{L}(n_s)$ had a career age of a:

$$\overline{cyear}_{\ell_m,n_s}(t) = a.$$

We denote the value of t that solves this equation as $t'_{\ell_m,n_s}(a)$ to emphasize that it will take different values for different Lasker winners (denoted by the first subscript, ℓ_m) in different match groups (denoted by the second subscript, n_i). We then take the mean over the set of matched Lasker winners in $\mathcal{L}(n_s)$:

$$\theta_1\left(a \middle| \mathcal{L}(n_s)\right) \equiv E\left[r\_pubs_{mt} \middle| m \in \mathcal{L}(n_s), t = t'_{\ell_m,n_s}(a)\right].$$

Finally, we average these values across all the match groups to obtain:

$$\theta_1(a|\text{Lasker}) \equiv \frac{1}{S} \sum_{n_s \in N} \theta_1\left(a \middle| \mathcal{L}(n_s)\right).$$

Recall that there are S Nobel winners, and hence S match groups. As with Nobel winners, we calculate $\theta_2(a|\text{Lasker}) \ldots \theta_{10}(a|\text{Lasker})$ analogously.

With any matching methodology, it is helpful to consider the weight that each observation receives in the estimators of interest. In the case of Nobel winners, our matching scheme yields a simple outcome – all Nobel winners who published at least one paper at career age a are equally weighted in the $\theta_j(a|\text{Nobel})$. The weight of each Nobel winner at career age a is thus proportional to 1.[3]

In the case of Lasker winners, the weighting scheme is more complicated since (1) Lasker winners can be represented in multiple match sets, $\mathcal{L}(n_s)$; and (2) the weight received in each set depends inversely on the total number of other Lasker winners in that set. Of course, each matched set is equally represented in our calculation, since there is exactly one set per Nobel winner.

Given these considerations, the weight of Lasker winner ℓ_m in the $\theta_j(a|\text{Lasker})$ is proportional to

$$\sum_{n_s \in N} \frac{1\big(\ell_m \in \mathcal{L}(n_s)\big)}{|\mathcal{L}(n_s)|},$$

where $|\mathcal{L}(n_s)|$ is the number of matched Lasker winners in $\mathcal{L}(n_s)$. This expression reduces to 1 in a setting (not ours) where each Nobel winner matches to exactly one Lasker winner and each Lasker winner matches to exactly one Nobel winner, since by definition, for each n_i $|\mathcal{L}(n_s)| = \sum_{\ell_m \in \mathcal{L}} 1\big(\ell_m \in \mathcal{L}(n_s)\big)$. Therefore, our matching estimator weights Lasker winners who match to multiple Nobel winners with few other matches relatively more than other Lasker winners publishing in the same career age.

In Appendix Section H, we conduct a sensitivity analysis in which we adopt a statistical method that assigns all Nobel and Lasker winners equal weight in the years they publish at least one paper.

F. Statistical Inference

Statistical inference is complicated by at least two considerations. First, the main objects on which we want to conduct hypothesis tests, $\theta_j(a|.)$ are functions of career age, rather than scalars. Second, our matching methodology necessarily reuses Lasker winners in multiple

[3] The weighting scheme is a bit more complicated than discussed here, since Nobel winners' careers span differing numbers of years. Since accounting for this would add substantial bookkeeping without clarifying our key point, we abstract away from this fact in our discussion.

Nobel/Lasker match groups, so no independence assumption is possible which might simplify calculation of sample statistics necessary for inference. We solve the first problem by defining a summary statistic that characterizes the outcomes for Nobel winners and their matched Lasker winners before and after the career age at which the Nobel winner was awarded the prize. We solve the second problem by implementing a block bootstrap, in which we separately resample the Nobel winners and the Lasker winners. An alternative bootstrapping approach would involve resampling the matched groups, $\{\mathcal{L}(n_s) \cup \{n_s\}\}$, but this approach would not account for the presence of Lasker winners in multiple matched groups.

Recall that within each matched group, $\{\mathcal{L}(n_s) \cup \{n_s\}\}$, we define career age relative to the career age at which Nobel winner, n_s, was awarded the Nobel prize. One major advantage of this approach is that for every matched group, the career ages where $a \leq 0$ represent the years before the Nobel prize was awarded to n_i, while the career ages where $a > 0$ represent the years after. For the Lasker winners in $\mathcal{L}(n_s)$, $a = 0$ represents the career age at which the Nobel winner who defines the group won the Nobel. With only a slight abuse of notation, for matched group $\mathcal{L}(n_s)$, let $t'_{n_s}(a \leq 0)$ and $t'_{n_s}(a > 0)$ be the set of calendar years before and after Nobel winner n_i won the Nobel. Analogously, let $t'_{\ell_m,n_s}(a \leq 0)$ and $t'_{\ell_m,n_s}(a > 0)$ be the set of calendar years before and after Lasker winner ℓ_m in matched set $\mathcal{L}(n_s)$ attains a career age equal to year in which Nobel winner n_i won the Nobel.

With those (admittedly tedious) definitions in hand, let $\mu_1(\text{before, Nobel})$ and $\mu_1(\text{after, Nobel})$ be the mean value for r_pubs_{it} among all Nobel winners before and after the Nobel was awarded; and let $\mu_1(\text{before, Lasker})$ and $\mu_1(\text{after, Lasker})$ be the mean value of r_pubs_{it} among Lasker winners in the career ages before and after their matched Nobel winner was awarded the Nobel. Let $\mu_2(.,.), \ldots, \mu_{10}(.,.)$ be analogously defined for our outcomes r_nov50_{it}, r_nov25_{it}, r_auth_{it}, r_fpubs_{it}, r_lpubs_{it}, r_cites_{it}, r_cites5_{it}, $r\_cites10_{it}$, and $r\_cite\_rate_{it}$ respectively.

We calculate $\mu_1(.,.)$ as follows (with the others calculated analogously):

$$\mu_1(\text{before, Nobel}) = E\left[r\_pubs_{st} \middle| s \in N, t = t'_{n_s}(a \leq 0)\right]$$

$$\mu_1(\text{after, Nobel}) = E\big[r\_pubs_{st} \big| s \in N, t = t'_{n_s}(a > 0)\big],$$

$$\mu_1(\text{before, Lasker}) = \frac{1}{S}\sum_{n_s \in N} E\big[r\_pubs_{mt} \big| m \in \mathcal{L}(n_s), t = t'_{\ell_m, n_s}(a \leq 0)\big], \text{and}$$

$$\mu_1(\text{after, Lasker}) = \frac{1}{S}\sum_{n_s \in N} E\big[r\_pubs_{mt} \big| m \in \mathcal{L}(n_s), t = t'_{\ell_m, n_s}(a > 0)\big].$$

For our block bootstrap exercise, we separately and randomly resample with replacement from the list of Nobel winners and Lasker winners. Let $N^{(b)}$ and $\mathcal{L}^{(b)}$ be the b^{th} bootstrap sample. We draw $B = 200$ samples from each population in total, where the size of each bootstrap sample matches the size of the original samples (M Lasker winners and S Nobel winners). For each bootstrap sample, we perform each of the steps in the calculation that we conducted with the actual samples, including matching Lasker winners to Nobel winners, merging in publication information by year, aligning career ages within match groups to the career age in which a Nobel winner won, regression adjusting to calculate residuals, and calculating our statistics of interest, $\mu_1^{(b)}(.,.)$, … $\mu_{10}^{(b)}(.,.)$, which are now indexed by the bootstrap sample from which they are derived. The p-values that we report in the main paper are based upon percentiles of the bootstrap distributions over these statistics.

Figures A2 – A11 show the mean values of $\mu_1^{(b)}(.,.)$, … $\mu_{10}^{(b)}(.,.)$, as well as 95% confidence intervals around these means. Figure A2 shows a sharp and statistically significant ($p < 0.01$) decline of about 1.25 papers per year (*r_pubs*) in the years after winning the Nobel prize (relative to before), but no change before and after the same career age for matched Lasker winners. Since this statistic is constructed on the residual number of publications, *r_pubs*, it is purged of variation due to differences in career age, birth year, the calendar year of measurement, and field of study. Figures A3 and A4 shows a similar, statistically significant drops in first-authored (r_fpubs, $p < 0.05$) and last-authored papers (r_lpubs, $p < 0.01$) – about a paper a year each, in both cases – after winning the Nobel Prize, but no similar drop for Lasker winners. Figure A5 shows no statistically significant difference in the change in the residual number of coauthors (*r_auth*) for either Nobel or Lasker winners.

Figure A6, which plots the sum of forward citations (r_cites) for papers published by Nobel winners and Lasker winners, shows two interesting facts. First, the papers that Nobel winners published before winning the Nobel prize are incredibly well cited – on average, their papers receive about 60 more citations than papers published by matched Lasker winners at the same field, career age, and calendar year (p < 0.01). Second, the papers the Nobel winners publish after winning the Nobel receive, on average, the same number of citations as matched Lasker winners (again holding field, career age, and calendar year fixed). Though all of these scientists – Nobel and Lasker winners alike – are publishing well cited papers, the papers that Nobel winners publish before are an order of magnitude better cited than both the papers published by Lasker winners and their own papers in the years after winning the Prize.

Figures A7 and A8 plot the sum of short run and medium run forward citations respectively – that is, r_cites5 and $r\_cites10$ – again separately for Nobel and matched Lasker winners. Qualitatively, the results are similar to those shown in Figure A6: a high rate of citations relative to matched Lasker winners in the short and medium run for Nobel prize winners for papers they published before winning the Nobel, as well as a decline in short run forward citations for papers published after the Nobel prize is awarded relative to themselves before.

Quantitatively, though, the results show a *smaller* difference in the short run and medium run citation counts between Nobel winners and Lasker winners for the papers published before the Nobel was awarded than we observed in long run citation counts. Despite being smaller in magnitude, these differences remain statistically significant, with p < 0.01. Papers published in any given year by Nobel winners prior to the Nobel earned about 10 to 12 more short run and medium run citations than those published by matched Lasker winners at a similar career age (after adjustment for calendar year and field). By contrast, there was an analogous difference of about 60 more citations in long run citations as noted above. This pair of results suggests that at least part of the difference in long run citations between Lasker winners and Nobel winners arises after the awarding of the Nobel prize; perhaps scientific authors discover and cite the past work of the Nobel winner as a result of the Nobel prize being awarded.

Figure A9 plots the adjusted citation rate for Nobel winners and Lasker controls, before vs. after winning the Nobel. The figure shows a rescaled version of our outcomes (r_cite_rate) that adjusts for the number of years that a paper is "at risk" to be cited by other papers in the scientific literature. The results confirm the results we observed in Figure A6 – higher citation rate for papers published by Nobel winners before winning the Nobel (about one extra citation per year at risk, $p< 0.01$), and a drop in the citation rate of papers published by the Nobel winner in the years after winning.

The last two figures present the results from our analysis of the novelty of work of Nobel winners relative to the matched controls. Figure A10 shows an increase in the (residual of) the median age of ideas (r_*nov50*) in the published papers of Nobel winners by about 2.5 years in the years after winning the Nobel ($p < 0.01$) – they were working on older ideas after winning the prize than before (after adjusting for career age, calendar year, and field of study), but no change before and after the same career age for matched Lasker winners. Finally, Figure A11 shows a similar result for the 25th percentile of the novelty distribution (*r_nov25*) ($p < 0.01$).

G. Sensitivity Analysis: Nobel/Lasker Match Regardless of Birth Year

In this section, we describe the results of a sensitivity analysis in which we tweak our match algorithm to permit Lasker winners to match to Nobel winners, even if they are born beyond the ten-year window that we enforce in our preferred match algorithm. We conduct this sensitivity analysis because our preferred algorithm requires us to drop two Nobel winners from our analysis because they do not match any Lasker winners. In this sensitivity analysis, by relaxing the match requirement, every Nobel winner has at least one match. We continue to require that the Lasker winner and Nobel winner share at least one field of study in common for a match. In particular, we use the following matched set of Lasker winners for each Nobel winner:

$$\mathcal{L}'(n_s) = \{\ell_m \in \mathcal{L} \mid \exists k \text{ such that } f_k(n_s) = f_k(\ell_m) = 1\}$$

Figure A12 shows the distribution over the number of matched Lasker winners per Nobel winner. Without the restriction to Lasker winners born within a decade of the Nobel winner, the number of matched Lasker winners per Nobel winner is substantially higher – nearly double that of the distribution shown in Figure A1, with the whole distribution shifted to the right as expected. In particular, there are no Nobel winners who match no Lasker winners in this analysis.

Figures A13-A15 (in ten panels) replicate the output of Figures A2-A11 (corresponding to the mean values of $\mu_1^{(b)}(.,.)$, … $\mu_{10}^{(b)}(.,.)$) with our new match algorithm. Recall that these are the means of our ten outcome variables, r_pubs_{it}, r_nov50_{it}, r_nov25_{it}, r_auth_{it}, r_fpubs_{it}, r_lpubs_{it}, r_cites_{it}, r_cites5_{it}, $r\_cites10_{it}$, and $r\_cite\_rate_{it}$. Even a cursory glance at these figures shows that the result from this sensitivity analysis are qualitatively identical to the corresponding results in Figures A2-A11. Though these results confirm our main results, it is nevertheless appropriate to limit the match to winners born within a decade of one another since scientists born in very different time periods likely faced very different scientific opportunities and environments from one another, so we present the results with the birth year proximity restriction as our primary analysis in the main paper.

H. Sensitivity Analysis: Equally Weighted Prize Winners Regressions

In this section, we report the results of a regression-based sensitivity analysis of our ten outcome variables in which we adopt a statistical method that assigns all Nobel and Lasker winners equal weight in the years they publish at least one paper. This method stands in contrast to our primary methodology, which relies on matching Nobel winners to Lasker winners based on birth year proximity and field of work, but which consequently places greater weight on Lasker winners who match to more Nobel winners or who match to Nobel winners with fewer (other) Lasker matches.

We adopt a method – difference-in-difference analysis – drawn from the economics literature. The main idea is to compare the outcome variables Nobel winners before versus after they win the Nobel prize against Lasker winners before versus after they win the Nobel. Unfortunately, this method poses an immediate difficulty – we do not know when or if a Lasker winner will ever win a Nobel prize, so we cannot define which career ages come before or after the prize for Lasker winners. However, under the null hypothesis that Lasker winners possess the same distribution of measured outcomes as Nobel winners, we can randomly assign Nobel Prize-winning career years to Lasker winners from the empirical distribution of these career ages that we observe among the Nobel winners. Our strategy is to repeatedly create an array of datasets using this random assignment methodology, perform our difference-in-difference analysis on all the datasets in the set, and calculate the mean treatment effect of interest over these analyses. In each analysis, each Nobel winner and each Lasker winner is equally weighted. As in our main analysis, we apply a bootstrap analysis to this methodology to obtain the statistical parameters we need for statistical inference.

Given this plan, we need some bookkeeping. The sample for our regression analysis consists of all Nobel winners and all Lasker winners, $i \in N \cup \mathcal{L}$, who we enumerate with the subscript i. Each winner, k, contributes an observation for every calendar year after the first year they publish a paper, until the last year in which we observe any papers published by them.

Recall that for Nobel winner n, $nyear(n)$ represents the career age in which the Nobel prize was awarded. Let $G(x) = P[nyear(n) < x]$ represent the empirical cumulative distribution function of $nyear(n)$ over all the Nobel winners in our sample. We assign each Lasker winner, ℓ_m, 100 random draws from $G(x)$ to assign a set of random career ages at which they "win" a Nobel prize. Let $nyear_d(\ell_m)$ be the d^{th} draw from $G(x)$ assigned to Lasker winner ℓ_m. This calculation is analogous to our calculation in our main analysis in which we find the career age a in which the Nobel winner (n_i) matched to a Lasker winner (ℓ_m), won the Nobel prize, and assign that career age, $t'_{\ell,n}(a)$, to the Lasker winners in that matched set as the year in which they would have "won" a Nobel prize if they had followed the career trajectory of the matched Nobel winner.

For each Lasker winner, ℓ, and each $G(x)$ draw d, we define an indicator variable, $after_t^d(\ell)$ which is equal to zero for calendar years before $nyear_d(\ell)$, and one otherwise. We define $after_t(n)$ analogously for each Nobel winner. We write $after_{it}^d$ to refer to these variables for an arbitrary individual i in our sample. We also specify a dummy variable, $nobel_i$ which equals one if an individual $i \in N$, and zero otherwise. Finally, let X_{it} represent a set of regression covariates that we will specify shortly.

Our difference-in-difference regression specification to analyze the annual number of publications, $pubs_{it}$, is as follows:

$$pubs_{it} = \beta_0^{pubs,d} + \beta_1^{pubs,d} nobel_i + \beta_2^{pubs,d} after_{it}^d +$$
$$\beta_3^{pubs,d} nobel_i \; after_{it}^d + \beta_4^{pubs,d} X_{it} + \varepsilon_{it}.$$

The regression coefficients, $\beta_{\cdot}^{pubs,d}$ are indexed by the left hand side variable because we also run similar regressions for our other outcome variables, $nov50_{it}$, $ov25_{it}$, and $auth_{it}$. They are also indexed by d because they will take on different values for different draws from the $G(x)$ distribution. Our regression error is denoted by ε_{it}, for which we adopt the standard assumption that it has a zero mean and is orthogonal to all the regression covariates.

Our primary regression coefficient of interest is $\beta_3^{pubs,d}$; it is not hard to show that this coefficient reflects the difference-in-difference in publications between before and after winning the Nobel prize for the Nobel winner relative to the Lasker winner:

$$\begin{aligned}\beta_3^{pubs,d} = & \left\{E\left[pubs_{it} | nobel_i = 1, after_{it}^d = 1, X_{it}\right] - E\left[pubs_{it} | nobel_i = 1, after_{it}^d = 0, X_{it}\right]\right\} \\ & - \left\{E\left[pubs_{it} | nobel_i = 0, after_{it}^d = 1, X_{it}\right]\right. \\ & \left. - E\left[pubs_{it} | nobel_i = 0, after_{it}^d = 0, X_{it}\right]\right\}\end{aligned}$$

We estimate this regression $d = 1 \ldots 100$ times, once for each time we draw from the $G(x)$ distribution to assign Nobel "win" ages to the Lasker winners. With all these regressions in hand, we calculate:

$$\beta_3^{pubs} = \frac{1}{100} \sum_{d=1}^{100} \beta_3^{pubs,d}.$$

This coefficient reflects how differently Nobel winners change in the number of publications and other outcomes before vs. after winning the Nobel prize, using the change in the number of publications for the Lasker winners before versus after a similar point in their careers to measure the expected change not due to winning the Nobel.

We conduct five different versions of this regression analysis, distinguished by an increasingly larger set of variables included among the X_{it} control variables. In the minimal specification, there are no additional controls, and X_{it} is a null vector. In a second specification, we add variables for the calendar year of the observation (t), and the physical age of the scientist in year t, so $X_{it} = (t, age_{it})$. In a third specification, we add controls for the career age of the scientist in year t, $cyear(i,t)$, as well as career age squared $cyear(i,t)^2$, to permit non-linear changes in the outputs over the course of a career:

$$X_{it} = (t, age_{it}, cyear(i,t), cyear(i,t)^2).$$

In the fourth specification, we add a complete set of field indicator variables, so:

$$X_{it} = \left(t, age_{it}, cyear(i,t), cyear(i,t)^2, f_k(i)\right).$$

Finally, for the fifth specification, we add a control variable for the total number of publications scientist *k* has had over his or her entire career, $totpubs_i = \sum_t pubs_{it}$.

$$X_{it} = (t, age_{it}, cyear(i,t), cyear(i,t)^2, f_k(i), totpubs_i).$$

We exclude this fifth specification for our analyses of total publications since it is inappropriate to include control variables that are direct functions of the dependent variable in a regression analysis.

For statistical inference, we conduct a block bootstrap analysis, using the same 200 bootstrap datasets that we describe in Section F of this Appendix, where the Nobel winners and Lasker winners are sampled with replacement, independently from one another.

Tables A5 through A14 report the complete set of β coefficients from these analyses, along with the standard errors that we calculate from the block bootstrap analyses reported below each coefficient. In these tables, a single asterisk indicates that the coefficient is statistically significant at the $p < 0.05$ level for a two-tailed t-test, while two asterisks indicate that $p < 0.01$. No asterisk means that the coefficient of interest is not statistically significant at the $p < 0.05$ level. Table A5 reports the results for the number of publications outcome ($pubs_{it}$); Table A6 reports the results for the first authored papers outcome ($fpubs_{it}$); Table A7 reports the results for the last authored papers outcome ($lpubs_{it}$); Table A8 reports the results for the number of coauthors outcome ($auth_{it}$); Table A9 reports the results for the forward citations analysis ($cites_{it}$); Table A10 reports the results of the analysis of forward citations limited to zero to five years after publication of each paper ($cites5_{it}$); Table A11 reports the results of the analysis of forward citations limited to six to ten years after publication of each paper ($cites10_{it}$); Table A12 reports the results for the forward citation rate analysis, in which the citation count is adjusted by the number of years each paper is at risk to garner forward citations ($cite\_rate_{it}$); Table A13 reports the results for the analysis of 50th percentile of the novelty score ($nov50_{it}$); and finally, Table A14 reports the results for the analysis of 25th percentile of the novelty score ($nov25_{it}$).

The difference-in-difference results largely confirm those that we report for our main analyses, both in the main paper and in this Appendix. We focus our verbal summary of these results to a discussion of the coefficient on "Nobel * After" in these regressions, which is the difference-in-difference estimate of the effect of winning the Nobel prize that we develop above. In Table A5, the results show that Nobel winners publish between 0.76 and 1.07 papers fewer after winning the Nobel relative to before the prize and relative to the Lasker control difference. In the specification with the greatest number of control variables, we find a difference in difference effect of 0.93 papers per year fewer by Nobel winners after winning the Nobel. Table A6 shows that that Nobel winners publish between 0.057 and 0.17 fewer first authored papers after winning the Nobel prize, relative to the Lasker controls and to before winning the Nobel prize. In the specifications with more control variables, this result is not statistically significant. Table A7 shows a drop in number of last authored papers for Nobel winners after winning the prize, relative to controls, but like the result for first authored papers, this result is not statistically significant. Table A8 shows no statistically measurable difference between the average number of coauthors in published work by Nobel winners before versus after winning the prize, relative to the analogous difference for Lasker winners.

Table A9 shows that papers published by Nobel winners after they win the prize garner between 40.8 and 53.1 citations fewer compared to the papers they published before winning, relative to the Lasker controls. These results are all statistically significant ($p < 0.01$). Table A10 shows that this result is robust to limiting the outcome to the first five years after publication of the paper. The papers published by Nobel winners after winning the prize garner between 2.5

and 6.0 fewer citations in the first five years after publication than papers they published before winning, relative to controls. The specifications the largest number of control variables are statistically significant at $p < 0.05$. Table A11 shows the analogous results when limiting to citations earned between six and ten years after publication of a paper. By this measure, papers published after winning the Nobel garner between 7.9 and 10.7 fewer citations relative to papers published before and relative to Lasker controls. All of the difference-in-difference estimate specifications in this table are statistically significant ($p < 0.01$). Table A12 shows that the citation rate of papers by Nobel winners after winning the Nobel drops by between 0.71 and 1.1 citations per year after publication, with all specifications statistically significant ($p < 0.05$). This confirms that the result shown in Table A9 is robust to adjusting the citation variable for years at risk to be cited.

Finally, Table A13 shows that after winning the Nobel prize, the newest ideas referenced in papers published by Nobel winners are between 0.68 and 1.79 years *older* relative the newest ideas referenced in papers before the prize (again relative to the analogous difference for Lasker winners, with the range given over the five different regression specifications, and the dependent variable measured at the median among papers published each year). In the specification with the largest set of controls, the measured difference is 1.52 *older* and statistically significant at the $p < 0.01$ level. We thus confirm that Nobel prize winners publish older ideas after winning the Nobel, relative to what one might expect from observing changes in the age of ideas worked on by Lasker winners. Table A14 shows similar findings for the newest ideas in published papers, measured at the 25th percentile of novelty rather than the median.

Table A1. Fields of Research

1	**Biochemistry**
2	Cell Biology
3	Developmental Biology
4	Genetics
5	Metabolism & Endocrine
6	Pharmacology
7	Physiology
8	Audiology & Cardiology
9	Cardiovascular
10	Hematology & Oncology
11	Infectious Disease & Immunology
12	Neurology & Cognition
13	Pediatrics & MFM
14	Surgery
15	Molecular Biology

Table A2: Regression Estimates for Residual Outcome Calculation (Productivity Outcomes)

	Number of Publications per Year	Number of First Authored Pubs per Year	Number of Last Authored Pubs per Year	Number of Coauthors
Career Age	0.274	0.315	0.315	0.009
	(28.11)**	(24.07)**	(24.07)**	(0.77)
Career Age Squared	-0.003	-0.004	-0.004	0.000
	(30.45)**	(25.21)**	(25.21)**	(0.11)
Age	-0.118	-0.112	-0.112	-0.021
	(20.32)**	(14.03)**	(14.03)**	(3.04)**
Calendar Year	0.029	0.034	0.034	0.081
	(10.14)**	(9.77)**	(9.77)**	(26.14)**
Field: biochemistry	0.223	-0.115	-0.115	0.379
	(1.42)	(0.62)	(0.62)	(2.34)*
Field: cell biology	-0.640	-0.769	-0.769	-0.586
	(4.85)**	(4.91)**	(4.91)**	(4.28)**
Field: developmental biology	-0.688	-0.785	-0.785	1.365
	(2.76)**	(2.65)**	(2.65)**	(5.33)**
Field: genetics	-0.478	-0.095	-0.095	0.209
	(3.57)**	(0.60)	(0.60)	(1.51)
Field: metabolism & endocrine	2.268	2.730	2.730	0.225
	(12.74)**	(13.06)**	(13.06)**	(1.23)
Field: pharmacology	-0.057	0.117	0.117	0.430
	(0.38)	(0.62)	(0.62)	(2.61)**
Field: physiology	0.157	0.240	0.240	-0.008
	(1.29)	(1.61)	(1.61)	(0.06)
Field: audiology & cardiology	-1.352	-0.901	-0.901	-0.492
	(6.04)**	(3.20)**	(3.20)**	(1.95)
Field: cardiovascular	-0.197	0.298	0.298	0.386
	(1.14)	(1.34)	(1.34)	(2.02)*
Field: hematology & oncology	0.592	0.233	0.233	0.129
	(4.45)**	(1.49)	(1.49)	(0.94)
Field: infectious disease & immunology	1.149	1.570	1.570	0.237
	(8.33)**	(9.38)**	(9.38)**	(1.62)
Field: neurology & cognition	0.723	0.906	0.906	-0.178
	(4.95)**	(5.15)**	(5.15)**	(1.16)
Field: pediatrics & maternal-fetal medicine	-1.519	-1.616	-1.616	-0.591
	(6.45)**	(5.43)**	(5.43)**	(2.33)*
Field: surgery	1.754	2.327	2.327	0.606
	(9.30)**	(10.24)**	(10.24)**	(3.03)**
Field: molecular biology	0.079	0.177	0.177	-0.370
	(0.62)	(1.20)	(1.20)	(2.85)**
Constant	-50.432	-61.836	-61.836	-156.313
	(9.20)**	(9.05)**	(9.05)**	(25.68)**
R^2	0.11	0.10	0.10	0.09

* $p<0.05$; ** $p<0.01$

Table A3: Regression Estimates for Residual Outcome Calculation (Citation Outcomes)

	Citations per Year	Citations per Year (0-5 years)	Citations per Year (6-10 years)	Citations per Paper per Year
Career Age	0.767	0.682	0.452	0.040
	(1.03)	(4.18)**	(2.63)**	(1.58)
Career Age Squared	-0.015	-0.015	-0.007	-0.001
	(1.54)	(6.99)**	(3.22)**	(3.24)**
Age	-2.889	-0.748	-0.768	-0.094
	(6.52)**	(7.70)**	(7.51)**	(6.24)**
Calendar Year	1.364	1.016	0.619	0.166
	(6.67)**	(22.66)**	(13.11)**	(23.88)**
Field: biochemistry	-12.009	1.724	-2.705	-0.209
	(1.25)	(0.82)	(1.22)	(0.64)
Field: cell biology	36.680	7.370	6.914	0.836
	(4.53)**	(4.15)**	(3.70)**	(3.03)**
Field: developmental biology	19.211	3.973	-2.381	1.179
	(1.31)	(1.23)	(0.70)	(2.36)*
Field: genetics	27.269	9.092	5.719	1.266
	(3.32)**	(5.05)**	(3.02)**	(4.53)**
Field: metabolism & endocrine	9.472	8.334	2.602	0.610
	(0.87)	(3.48)**	(1.03)	(1.64)
Field: pharmacology	-34.305	-8.310	-7.026	-0.727
	(3.42)**	(3.77)**	(3.03)**	(2.12)*
Field: physiology	-5.838	-0.973	-0.629	0.138
	(0.74)	(0.56)	(0.35)	(0.52)
Field: audiology & cardiology	56.452	-0.534	5.806	1.412
	(3.69)**	(0.16)	(1.64)	(2.71)**
Field: cardiovascular	7.370	0.886	-0.264	0.300
	(0.63)	(0.35)	(0.10)	(0.75)
Field: hematology & oncology	-0.826	3.337	-0.755	0.035
	(0.10)	(1.90)	(0.41)	(0.13)
Field: infectious disease & immunology	-6.284	1.710	-2.194	-0.150
	(0.72)	(0.90)	(1.09)	(0.51)
Field: neurology & cognition	30.253	0.442	0.500	0.711
	(3.29)**	(0.22)	(0.24)	(2.27)*
Field: pediatrics & maternal-fetal medicine	-61.627	-14.107	-12.140	-1.817
	(4.02)**	(4.19)**	(3.43)**	(3.48)**
Field: surgery	-27.461	-7.676	-6.153	-0.657
	(2.31)*	(2.94)**	(2.24)*	(1.62)
Field: molecular biology	-14.665	1.799	-3.670	-0.454
	(1.94)	(1.08)	(2.10)*	(1.76)
Constant	-2,454.349	-1,944.744	-1,163.355	-321.009
	(6.11)**	(22.08)**	(12.55)**	(23.46)**
R^2	0.04	0.10	0.04	0.08

* $p<0.05$; ** $p<0.01$

Table A4: Regression Estimates for Residual Outcome Calculation (Novelty Outcomes)

	Median Novelty Score	25th percentile Novelty Score
Career Age	0.318	0.397
	(11.94)**	(14.97)**
Career Age Squared	-0.005	-0.006
	(13.76)**	(17.41)**
Age	-0.223	-0.229
	(14.28)**	(14.74)**
Calendar Year	-0.016	0.000
	(2.10)*	(0.06)
Field: biochemistry	1.664	1.838
	(4.86)**	(5.38)**
Field: cell biology	1.110	0.598
	(3.92)**	(2.12)*
Field: developmental biology	-3.632	-2.266
	(7.36)**	(4.60)**
Field: genetics	0.788	0.533
	(2.77)**	(1.88)
Field: metabolism & endocrine	2.586	2.427
	(6.68)**	(6.28)**
Field: pharmacology	-0.874	-0.604
	(2.45)*	(1.70)
Field: physiology	-1.198	-0.890
	(4.38)**	(3.27)**
Field: audiology & cardiology	-3.709	-4.459
	(7.10)**	(8.55)**
Field: cardiovascular	0.803	0.100
	(1.94)	(0.24)
Field: hematology & oncology	0.995	0.836
	(3.61)**	(3.04)**
Field: infectious disease & immunology	0.553	0.484
	(1.82)	(1.60)
Field: neurology & cognition	-2.074	-1.827
	(6.51)**	(5.75)**
Field: pediatrics & maternal-fetal medicine	-2.895	-2.665
	(5.45)**	(5.03)**
Field: surgery	-2.869	-1.433
	(6.96)**	(3.49)**
Field: molecular biology	0.109	-0.034
	(0.42)	(0.13)
Constant	28.295	-2.311
	(1.83)	(0.15)
R^2	0.17	0.16

* $p<0.05$; ** $p<0.01$

Table A5: Equally Weighted Regressions -- Number of Publications Per Year

	(1)	(2)	(3)	(4)
Nobel winner	0.143	-0.150	-0.060	-0.010
	(0.314)	(0.326)	(0.336)	(0.402)
After	-0.393	0.096	-0.228	-0.307
	(0.163)*	(0.193)	(0.157)	(0.162)
Nobel * After	-1.070	-0.758	-0.803	-0.934
	(0.277)**	(0.302)*	(0.321)*	(0.301)**
Year	.	0.041	0.022	0.031
		(0.007)**	(0.007)**	(0.008)**
Age	.	-0.064	-0.094	-0.112
		(0.008)**	(0.019)**	(0.020)**
Career Age	.	.	0.288	0.300
			(0.034)**	(0.035)**
Career Age Squared	.	.	-0.004	-0.004
			(0.000)**	(0.000)**
Field: biochemistry	.	.	.	0.367
				(0.486)
Field: cell biology	.	.	.	-0.596
				(0.388)
Field: developmental biology	.	.	.	-1.023
				(0.578)
Field: genetics	.	.	.	-0.280
				(0.509)
Field: metabolism & endocrine	.	.	.	2.493
				(0.899)**
Field: pharmacology	.	.	.	-0.188
				(0.579)
Field: physiology	.	.	.	0.011
				(0.418)
Field: audiology & cardiology	.	.	.	-1.464
				(0.672)*
Field: cardiovascular	.	.	.	-0.408
				(0.636)
Field: hematology & oncology	.	.	.	0.603
				(0.537)
Field: infectious disease & immunology	.	.	.	1.031
				(0.499)*
Field: neurology & cognition	.	.	.	0.744
				(0.591)
Field: pediatrics & maternal-fetal medicine	.	.	.	-1.833
				(0.843)*
Field: surgery	.	.	.	1.764
				(1.540)
Field: molecular biology	.	.	.	-0.267
				(0.464)
Constant	3.909	-73.177	-39.006	-56.112
	(0.215)**	(12.546)**	(14.375)**	(16.115)**
R^2	0.012	0.036	0.090	0.134

* $p<0.05$; ** $p<0.01$

Table A6: Equally Weighted Regressions – First Authored Papers Per Year

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	-0.090	-0.149	-0.149	-0.139	-0.112
	(0.064)	(0.064)*	(0.063)*	(0.062)*	(0.066)
After	-0.312	-0.068	-0.064	-0.061	-0.059
	(0.033)**	(0.027)*	(0.023)**	(0.022)**	(0.023)*
Nobel * After	-0.174	-0.104	-0.107	-0.076	-0.057
	(0.055)**	(0.056)	(0.055)	(0.059)	(0.064)
Year	.	-0.011	-0.011	-0.009	-0.011
		(0.002)**	(0.002)**	(0.002)**	(0.002)**
Age	.	-0.005	-0.004	-0.009	-0.001
		(0.002)*	(0.004)	(0.004)*	(0.003)
Career Age	.	.	-0.002	0.002	-0.005
			(0.006)	(0.006)	(0.005)
Career Age Squared	.	.	0.000	-0.000	-0.000
			(0.000)**	(0.000)**	(0.000)**
Field: biochemistry	.	.	.	-0.175	-0.197
				(0.091)	(0.075)**
Field: cell biology	.	.	.	-0.020	0.027
				(0.076)	(0.069)
Field: developmental biology	.	.	.	-0.073	-0.019
				(0.126)	(0.115)
Field: genetics	.	.	.	-0.009	-0.001
				(0.078)	(0.067)
Field: metabolism & endocrine	.	.	.	0.178	0.003
				(0.118)	(0.119)
Field: pharmacology	.	.	.	0.149	0.134
				(0.156)	(0.145)
Field: physiology	.	.	.	-0.006	-0.026
				(0.085)	(0.084)
Field: audiology & cardiology	.	.	.	-0.017	0.076
				(0.130)	(0.116)
Field: cardiovascular	.	.	.	-0.099	-0.048
				(0.127)	(0.104)
Field: hematology & oncology	.	.	.	0.003	-0.019
				(0.106)	(0.101)
Field: infectious disease & immunology	.	.	.	0.052	-0.037
				(0.099)	(0.081)
Field: neurology & cognition	.	.	.	0.384	0.339
				(0.126)**	(0.116)**
Field: pediatrics & maternal-fetal medicine	.	.	.	-0.139	-0.012
				(0.155)	(0.120)
Field: surgery	.	.	.	0.531	0.333
				(0.264)*	(0.187)
Field: molecular biology	.	.	.	-0.045	-0.029
				(0.072)	(0.071)
Lifetime Pubs	.	.	.	.	0.001
					(0.000)**
Constant	0.766	22.472	21.922	19.494	22.555
	(0.046)**	(3.173)**	(3.510)**	(3.768)**	(3.619)**
R^2	0.037	0.086	0.087	0.122	0.158

* $p<0.05$; ** $p<0.01$

Table A7: Equally Weighted Regressions – Last Authored Papers Per Year

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	-0.103	-0.199	-0.192	-0.200	0.013
	(0.206)	(0.212)	(0.215)	(0.260)	(0.209)
After	0.080	0.177	-0.037	-0.080	-0.085
	(0.126)	(0.124)	(0.100)	(0.107)	(0.075)
Nobel * After	-0.243	-0.170	-0.202	-0.317	-0.154
	(0.222)	(0.235)	(0.223)	(0.229)	(0.199)
Year	.	0.028	0.016	0.018	0.004
		(0.005)**	(0.005)**	(0.006)**	(0.004)
Age	.	-0.032	-0.069	-0.066	-0.005
		(0.006)**	(0.016)**	(0.017)**	(0.008)
Career Age	.	.	0.249	0.247	0.198
			(0.027)**	(0.028)**	(0.019)**
Career Age Squared	.	.	-0.003	-0.003	-0.004
			(0.000)**	(0.000)**	(0.000)**
Field: biochemistry	.	.	.	0.296	0.152
				(0.408)	(0.231)
Field: cell biology	.	.	.	-0.352	0.034
				(0.315)	(0.188)
Field: developmental biology	.	.	.	-0.507	-0.083
				(0.481)	(0.354)
Field: genetics	.	.	.	-0.157	-0.080
				(0.376)	(0.193)
Field: metabolism & endocrine	.	.	.	1.141	-0.238
				(0.567)*	(0.327)
Field: pharmacology	.	.	.	-0.649	-0.745
				(0.315)*	(0.344)*
Field: physiology	.	.	.	-0.185	-0.340
				(0.337)	(0.177)
Field: audiology & cardiology	.	.	.	-0.756	-0.004
				(0.425)	(0.193)
Field: cardiovascular	.	.	.	0.107	0.501
				(0.460)	(0.223)*
Field: hematology & oncology	.	.	.	-0.024	-0.196
				(0.421)	(0.278)
Field: infectious disease & immunology	.	.	.	0.584	-0.103
				(0.372)	(0.246)
Field: neurology & cognition	.	.	.	0.155	-0.195
				(0.430)	(0.244)
Field: pediatrics & maternal-fetal medicine	.	.	.	-1.497	-0.518
				(0.541)**	(0.223)*
Field: surgery	.	.	.	1.152	-0.304
				(1.110)	(0.234)
Field: molecular biology	.	.	.	0.172	0.292
				(0.404)	(0.207)
Lifetime Pubs	.	.	.	.	0.009
					(0.001)**
Constant	2.356	-51.831	-29.364	-33.749	-9.847
	(0.164)**	(8.867)**	(9.778)**	(12.056)**	(7.765)
R^2	0.003	0.018	0.083	0.117	0.301

* $p<0.05$; ** $p<0.01$

Table A8: Equally Weighted Regressions – Mean Number of Couthors

	(1)	(2)	(3)	(4)	(5)
Nobel winner	-0.285	-0.248	-0.249	-0.306	-0.273
	(0.116)*	(0.091)**	(0.091)**	(0.135)*	(0.140)
After	0.803	-0.025	-0.039	-0.050	-0.044
	(0.083)**	(0.062)	(0.064)	(0.066)	(0.064)
Nobel * After	0.399	0.249	0.263	0.351	0.368
	(0.303)	(0.294)	(0.299)	(0.368)	(0.362)
Year	.	0.077	0.076	0.081	0.079
		(0.006)**	(0.006)**	(0.005)**	(0.005)**
Age	.	-0.007	-0.011	-0.022	-0.015
		(0.004)	(0.010)	(0.009)*	(0.010)
Career Age	.	.	0.001	0.007	0.002
			(0.014)	(0.014)	(0.015)
Career Age Squared	.	.	0.000	0.000	0.000
			(0.000)**	(0.000)**	(0.000)**
Field: biochemistry	.	.	.	0.328	0.308
				(0.153)*	(0.150)*
Field: cell biology	.	.	.	-0.577	-0.517
				(0.191)**	(0.193)**
Field: developmental biology	.	.	.	1.549	1.618
				(1.944)	(1.939)
Field: genetics	.	.	.	0.248	0.260
				(0.253)	(0.252)
Field: metabolism & endocrine	.	.	.	0.202	0.035
				(0.148)	(0.138)
Field: pharmacology	.	.	.	0.334	0.322
				(0.169)*	(0.167)
Field: physiology	.	.	.	-0.021	-0.038
				(0.120)	(0.115)
Field: audiology & cardiology	.	.	.	-0.524	-0.426
				(0.286)	(0.276)
Field: cardiovascular	.	.	.	0.338	0.394
				(0.204)	(0.192)*
Field: hematology & oncology	.	.	.	0.035	0.023
				(0.208)	(0.198)
Field: infectious disease & immunology	.	.	.	0.206	0.136
				(0.183)	(0.177)
Field: neurology & cognition	.	.	.	-0.166	-0.215
				(0.151)	(0.147)
Field: pediatrics & maternal-fetal medicine	.	.	.	-0.688	-0.561
				(0.557)	(0.586)
Field: surgery	.	.	.	0.551	0.354
				(0.269)*	(0.247)
Field: molecular biology	.	.	.	-0.417	-0.397
				(0.200)*	(0.203)
Total Lifetime Publications	.	.	.	.	0.001
					(0.000)**
Constant	3.563	-147.218	-145.994	-155.847	-151.710
	(0.076)**	(12.096)**	(11.119)**	(9.696)**	(10.293)**
R^2	0.013	0.108	0.108	0.126	0.130

* $p<0.05$; ** $p<0.01$

Table A9: Equally Weighted Regressions – Forward Citations All Time to Papers Published in Each Year Per Year

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	74.504	62.664	62.558	54.597	53.366
	(9.019)**	(8.672)**	(8.693)**	(9.016)**	(8.650)**
After	-19.929	3.137	1.594	1.671	1.806
	(2.904)**	(2.709)	(2.710)	(3.062)	(2.962)
Nobel * After	-53.108	-41.641	-40.880	-44.985	-46.524
	(9.005)**	(8.691)**	(8.715)**	(8.704)**	(8.632)**
Year	.	1.516	1.411	1.159	1.249
		(0.186)**	(0.204)**	(0.235)**	(0.238)**
Age	.	-2.773	-3.272	-2.431	-2.823
		(0.223)**	(0.442)**	(0.468)**	(0.488)**
Career Age	.	.	1.379	0.870	1.206
			(0.819)	(0.827)	(0.811)
Career Age Squared	.	.	-0.014	-0.015	-0.014
			(0.009)	(0.009)	(0.009)
Field: biochemistry	.	.	.	-5.566	-4.808
				(7.193)	(7.139)
Field: cell biology	.	.	.	35.371	32.289
				(10.447)**	(10.540)**
Field: developmental biology	.	.	.	22.611	18.938
				(21.447)	(21.217)
Field: genetics	.	.	.	20.277	19.833
				(8.983)*	(8.780)*
Field: metabolism & endocrine	.	.	.	15.259	24.495
				(9.845)	(9.889)*
Field: pharmacology	.	.	.	-19.957	-18.936
				(8.781)*	(8.212)*
Field: physiology	.	.	.	-0.140	0.960
				(11.516)	(11.251)
Field: audiology & cardiology	.	.	.	66.871	61.645
				(42.684)	(41.741)
Field: cardiovascular	.	.	.	15.116	11.904
				(16.483)	(15.882)
Field: hematology & oncology	.	.	.	4.805	5.807
				(7.032)	(7.099)
Field: infectious disease & immunology	.	.	.	3.334	7.216
				(9.097)	(9.285)
Field: neurology & cognition	.	.	.	29.842	32.605
				(14.867)*	(15.029)*
Field: pediatrics & maternal-fetal medicine	.	.	.	-49.706	-56.274
				(15.399)**	(14.816)**
Field: surgery	.	.	.	-15.466	-4.778
				(11.509)	(14.280)
Field: molecular biology	.	.	.	6.223	5.019
				(8.132)	(8.410)
Lifetime Pubs	.	.	.	.	-0.058
					(0.015)**
Constant	82.008	-2773.500	-2561.958	-2107.085	-2259.928
	(3.539)**	(359.512)**	(398.668)**	(456.416)**	(460.146)**
R^2	0.020	0.037	0.037	0.047	0.049

* $p<0.05$; ** $p<0.01$

Table A10: Equally Weighted Regressions – Forward Citations in First Five Years After Publication to Papers Published in Each Year Per Year

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	17.200	13.167	13.164	13.071	12.980
	(2.315)**	(1.841)**	(1.845)**	(1.888)**	(1.893)**
After	-2.321	-0.371	-0.699	-1.261	-1.256
	(0.954)*	(0.880)	(0.758)	(0.813)	(0.804)
Nobel * After	-5.987	-2.473	-2.745	-5.265	-5.433
	(3.090)	(2.824)	(2.713)	(2.682)*	(2.720)*
Year	.	1.080	1.054	0.926	0.935
		(0.064)**	(0.070)**	(0.070)**	(0.071)**
Age	.	-0.918	-1.001	-0.627	-0.664
		(0.052)**	(0.130)**	(0.141)**	(0.145)**
Career Age	.	.	0.885	0.685	0.717
			(0.197)**	(0.191)**	(0.193)**
Career Age Squared	.	.	-0.014	-0.014	-0.014
			(0.002)**	(0.002)**	(0.002)**
Field: biochemistry	.	.	.	3.019	3.053
				(2.267)	(2.268)
Field: cell biology	.	.	.	6.947	6.637
				(2.686)**	(2.766)*
Field: developmental biology	.	.	.	6.704	6.285
				(7.588)	(7.522)
Field: genetics	.	.	.	6.927	6.872
				(3.089)*	(3.107)*
Field: metabolism & endocrine	.	.	.	9.318	10.220
				(4.129)*	(4.216)*
Field: pharmacology	.	.	.	-5.186	-5.045
				(3.022)	(2.971)
Field: physiology	.	.	.	0.991	1.050
				(2.698)	(2.712)
Field: audiology & cardiology	.	.	.	2.023	1.518
				(5.964)	(5.961)
Field: cardiovascular	.	.	.	3.132	2.776
				(3.528)	(3.520)
Field: hematology & oncology	.	.	.	3.745	3.793
				(2.867)	(2.923)
Field: infectious disease & immunology	.	.	.	4.675	5.010
				(3.078)	(3.147)
Field: neurology & cognition	.	.	.	0.933	1.129
				(2.685)	(2.774)
Field: pediatrics & maternal-fetal medicine	.	.	.	-10.606	-11.250
				(4.965)*	(4.929)*
Field: surgery	.	.	.	-3.957	-2.895
				(3.427)	(3.824)
Field: molecular biology	.	.	.	8.657	8.499
				(2.923)**	(2.967)**
Lifetime Pubs	.	.	.	.	-0.006
					(0.004)
Constant	28.349	-2059.747	-2013.465	-1781.236	-1797.102
	(1.186)**	(124.857)**	(137.767)**	(138.125)**	(138.873)**
R^2	0.021	0.096	0.101	0.119	0.120

* $p<0.05$; ** $p<0.01$

Table A11: Equally Weighted Regressions – Forward Citations in Six to Ten Years After Publication to Papers Published in Each Year Per Year

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	16.341	13.355	13.338	12.265	12.029
	(2.166)**	(1.951)**	(1.943)**	(2.086)**	(2.022)**
After	-2.564	0.362	-0.073	-0.252	-0.179
	(0.759)**	(0.733)	(0.687)	(0.729)	(0.721)
Nobel * After	-10.754	-8.014	-7.916	-9.261	-9.624
	(2.360)**	(2.244)**	(2.155)**	(2.212)**	(2.211)**
Year	.	0.651	0.618	0.563	0.581
		(0.048)**	(0.053)**	(0.059)**	(0.059)**
Age	.	-0.684	-0.829	-0.654	-0.734
		(0.056)**	(0.104)**	(0.116)**	(0.123)**
Career Age	.	.	0.560	0.466	0.534
			(0.198)**	(0.193)*	(0.189)**
Career Age Squared	.	.	-0.007	-0.007	-0.007
			(0.002)**	(0.002)**	(0.002)**
Field: biochemistry	.	.	.	-1.271	-1.129
				(1.628)	(1.634)
Field: cell biology	.	.	.	6.665	6.038
				(2.260)**	(2.337)**
Field: developmental biology	.	.	.	-1.256	-2.021
				(4.812)	(4.777)
Field: genetics	.	.	.	4.231	4.138
				(2.417)	(2.384)
Field: metabolism & endocrine	.	.	.	3.973	5.846
				(2.801)	(2.788)*
Field: pharmacology	.	.	.	-4.015	-3.790
				(2.279)	(2.155)
Field: physiology	.	.	.	0.751	0.959
				(2.330)	(2.312)
Field: audiology & cardiology	.	.	.	8.355	7.306
				(7.289)	(7.178)
Field: cardiovascular	.	.	.	1.687	1.008
				(3.388)	(3.278)
Field: hematology & oncology	.	.	.	0.112	0.289
				(1.864)	(1.889)
Field: infectious disease & immunology	.	.	.	0.263	1.046
				(2.482)	(2.550)
Field: neurology & cognition	.	.	.	0.584	1.108
				(2.733)	(2.818)
Field: pediatrics & maternal-fetal medicine	.	.	.	-9.439	-10.785
				(3.462)**	(3.357)**
Field: surgery	.	.	.	-3.569	-1.371
				(2.991)	(3.469)
Field: molecular biology	.	.	.	1.380	1.118
				(2.252)	(2.329)
Lifetime Pubs	.	.	.	.	-0.012
					(0.004)**
Constant	19.053	-1232.196	-1167.886	-1067.274	-1098.651
	(0.881)**	(92.028)**	(102.544)**	(114.512)**	(114.657)**
R^2	0.016	0.045	0.046	0.053	0.055

* $p<0.05$; ** $p<0.01$

Table A12: Equally Weighted Regressions – Mean Citation Rate per Year for Papers Published in Year

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	2.212	1.740	1.739	1.474	1.425
	(0.343)**	(0.281)**	(0.280)**	(0.297)**	(0.282)**
After	0.118	-0.070	-0.085	-0.110	-0.098
	(0.133)	(0.109)	(0.099)	(0.101)	(0.102)
Nobel * After	-1.090	-0.706	-0.727	-0.823	-0.892
	(0.430)*	(0.379)	(0.368)*	(0.364)*	(0.356)*
Year	.	0.166	0.165	0.157	0.161
		(0.010)**	(0.010)**	(0.011)**	(0.010)**
Age	.	-0.106	-0.107	-0.079	-0.095
		(0.009)**	(0.016)**	(0.018)**	(0.018)**
Career Age	.	.	0.052	0.037	0.051
			(0.026)*	(0.026)	(0.025)*
Career Age Squared	.	.	-0.001	-0.001	-0.001
			(0.000)**	(0.000)**	(0.000)**
Field: biochemistry	.	.	.	-0.049	-0.018
				(0.253)	(0.245)
Field: cell biology	.	.	.	0.804	0.676
				(0.400)*	(0.404)
Field: developmental biology	.	.	.	1.516	1.361
				(1.138)	(1.121)
Field: genetics	.	.	.	1.073	1.056
				(0.415)**	(0.403)**
Field: metabolism & endocrine	.	.	.	0.753	1.135
				(0.452)	(0.442)*
Field: pharmacology	.	.	.	-0.397	-0.353
				(0.374)	(0.341)
Field: physiology	.	.	.	0.350	0.397
				(0.378)	(0.374)
Field: audiology & cardiology	.	.	.	1.762	1.545
				(1.206)	(1.174)
Field: cardiovascular	.	.	.	0.555	0.421
				(0.568)	(0.541)
Field: hematology & oncology	.	.	.	0.027	0.069
				(0.318)	(0.317)
Field: infectious disease & immunology	.	.	.	0.204	0.367
				(0.352)	(0.358)
Field: neurology & cognition	.	.	.	0.774	0.889
				(0.377)*	(0.393)*
Field: pediatrics & maternal-fetal medicine	.	.	.	-1.448	-1.722
				(0.701)*	(0.668)**
Field: surgery	.	.	.	-0.259	0.190
				(0.516)	(0.593)
Field: molecular biology	.	.	.	0.288	0.238
				(0.294)	(0.306)
Lifetime Pubs	.	.	.	.	-0.002
					(0.001)*
Constant	2.742	-320.454	-319.142	-303.800	-310.198
	(0.157)**	(18.781)**	(20.368)**	(20.594)**	(20.378)**
R^2	0.013	0.081	0.082	0.091	0.094

* $p<0.05$; ** $p<0.01$

Table A13: Equally Weighted Regressions – Median Novelty Score

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	1.585	0.710	0.735	1.077	1.171
	(0.444)**	(0.401)	(0.409)	(0.407)**	(0.397)**
After	-2.178	0.197	-0.018	-0.089	-0.134
	(0.257)**	(0.209)	(0.207)	(0.203)	(0.193)
Nobel * After	-1.594	-0.676	-0.733	-1.785	-1.520
	(0.524)**	(0.522)	(0.522)	(0.500)**	(0.471)**
Year	.	0.000	-0.013	-0.018	-0.025
		(0.015)	(0.017)	(0.018)	(0.018)
Age	.	-0.188	-0.261	-0.206	-0.156
		(0.016)**	(0.036)**	(0.037)**	(0.033)**
Career Age	.	.	0.368	0.327	0.272
			(0.050)**	(0.048)**	(0.044)**
Career Age Squared	.	.	-0.005	-0.005	-0.005
			(0.001)**	(0.001)**	(0.001)**
Field: biochemistry	.	.	.	1.891	1.823
				(0.672)**	(0.633)**
Field: cell biology	.	.	.	1.201	1.562
				(0.697)	(0.625)*
Field: developmental biology	.	.	.	-3.825	-3.393
				(1.298)**	(1.258)**
Field: genetics	.	.	.	0.906	0.951
				(0.654)	(0.587)
Field: metabolism & endocrine	.	.	.	2.802	1.572
				(0.883)**	(0.825)
Field: pharmacology	.	.	.	-0.715	-0.912
				(0.760)	(0.746)
Field: physiology	.	.	.	-1.161	-1.327
				(0.663)	(0.604)*
Field: audiology & cardiology	.	.	.	-3.259	-2.625
				(1.679)	(1.661)
Field: cardiovascular	.	.	.	0.839	1.249
				(1.009)	(1.059)
Field: hematology & oncology	.	.	.	1.047	0.904
				(0.717)	(0.639)
Field: infectious disease & immunology	.	.	.	0.735	0.210
				(0.779)	(0.706)
Field: neurology & cognition	.	.	.	-1.978	-2.293
				(1.000)*	(0.876)**
Field: pediatrics & maternal-fetal medicine	.	.	.	-3.063	-2.272
				(1.821)	(1.840)
Field: surgery	.	.	.	-2.636	-3.942
				(0.948)**	(0.978)**
Field: molecular biology	.	.	.	0.235	0.344
				(0.821)	(0.712)
Lifetime Pubs	.	.	.	.	0.007
					(0.001)**
Constant	-12.168	-2.659	21.991	29.423	40.662
	(0.275)**	(28.859)	(32.906)	(34.418)	(34.229)
R^2	0.029	0.101	0.121	0.182	0.206

* $p<0.05$; ** $p<0.01$

Table A14: Equally Weighted Regressions – 25th Percentile Novelty Score

	(1)	(2)	(3)	(4)	(5)
Nobel Winner	1.230	0.387	0.412	0.513	0.637
	(0.377)**	(0.346)	(0.350)	(0.394)	(0.350)
After	-1.987	0.188	-0.083	-0.158	-0.176
	(0.238)**	(0.207)	(0.200)	(0.186)	(0.196)
Nobel * After	-1.728	-0.852	-0.923	-1.737	-1.439
	(0.505)**	(0.487)	(0.494)	(0.489)**	(0.469)**
Year	.	0.015	0.002	0.001	-0.008
		(0.013)	(0.015)	(0.016)	(0.015)
Age	.	-0.180	-0.256	-0.213	-0.151
		(0.016)**	(0.038)**	(0.040)**	(0.033)**
Career Age	.	.	0.444	0.408	0.339
			(0.052)**	(0.052)**	(0.047)**
Career Age Squared	.	.	-0.006	-0.006	-0.006
			(0.001)**	(0.001)**	(0.001)**
Field: biochemistry	.	.	.	1.986	1.902
				(0.616)**	(0.564)**
Field: cell biology	.	.	.	0.678	1.123
				(0.684)	(0.588)
Field: developmental biology	.	.	.	-2.604	-2.064
				(1.222)*	(1.138)
Field: genetics	.	.	.	0.751	0.804
				(0.624)	(0.517)
Field: metabolism & endocrine	.	.	.	2.654	1.136
				(0.742)**	(0.726)
Field: pharmacology	.	.	.	-0.725	-0.976
				(0.764)	(0.682)
Field: physiology	.	.	.	-0.991	-1.194
				(0.619)	(0.535)*
Field: audiology & cardiology	.	.	.	-4.185	-3.402
				(1.611)**	(1.548)*
Field: cardiovascular	.	.	.	-0.065	0.442
				(0.903)	(0.890)
Field: hematology & oncology	.	.	.	0.771	0.597
				(0.714)	(0.613)
Field: infectious disease & immunology	.	.	.	0.514	-0.133
				(0.768)	(0.664)
Field: neurology & cognition	.	.	.	-1.749	-2.126
				(0.981)	(0.816)**
Field: pediatrics & maternal-fetal medicine	.	.	.	-3.123	-2.140
				(1.581)*	(1.530)
Field: surgery	.	.	.	-1.358	-2.979
				(1.102)	(1.000)**
Field: molecular biology	.	.	.	-0.196	-0.052
				(0.760)	(0.618)
Lifetime Pubs	.	.	.	.	0.009
					(0.001)**
Constant	-8.417	-28.563	-4.063	-3.419	10.671
	(0.246)**	(25.961)	(30.373)	(31.267)	(30.303)
R^2	0.027	0.090	0.121	0.170	0.207

* $p<0.05$; ** $p<0.01$

Figure A1: Distribution of Number of Matched Lasker Winners

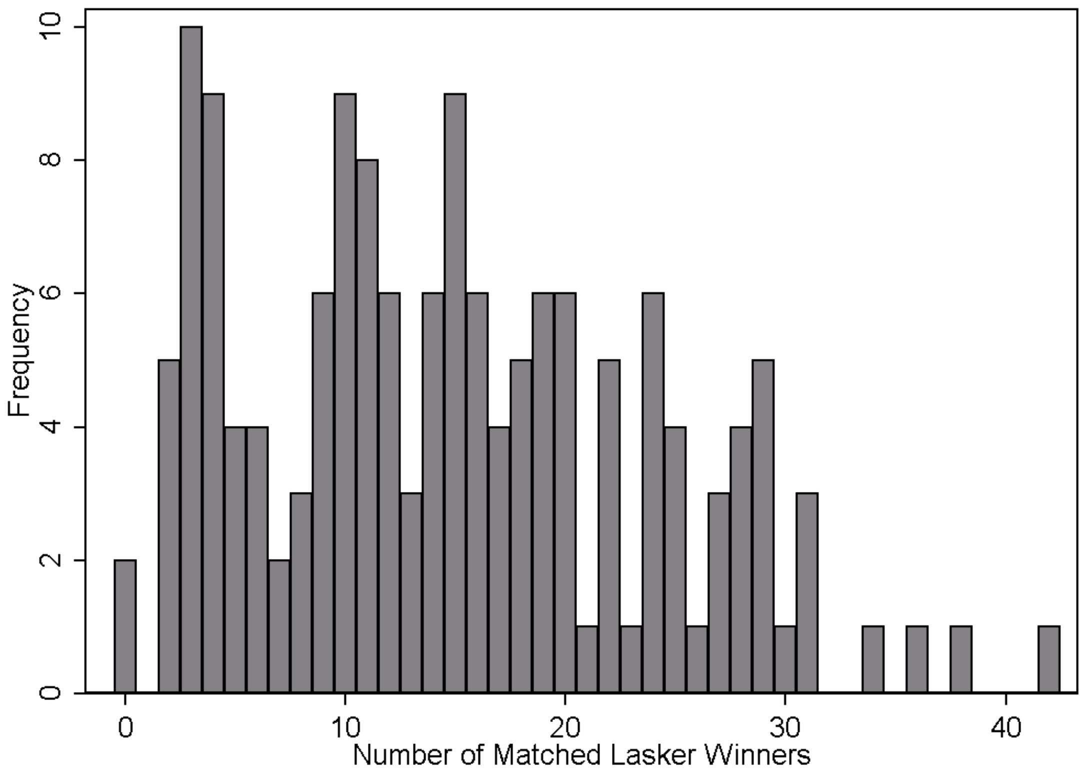


Figure A2: Residual Number of Publications per Year (Nobel Winners vs. Matched Lasker Controls)

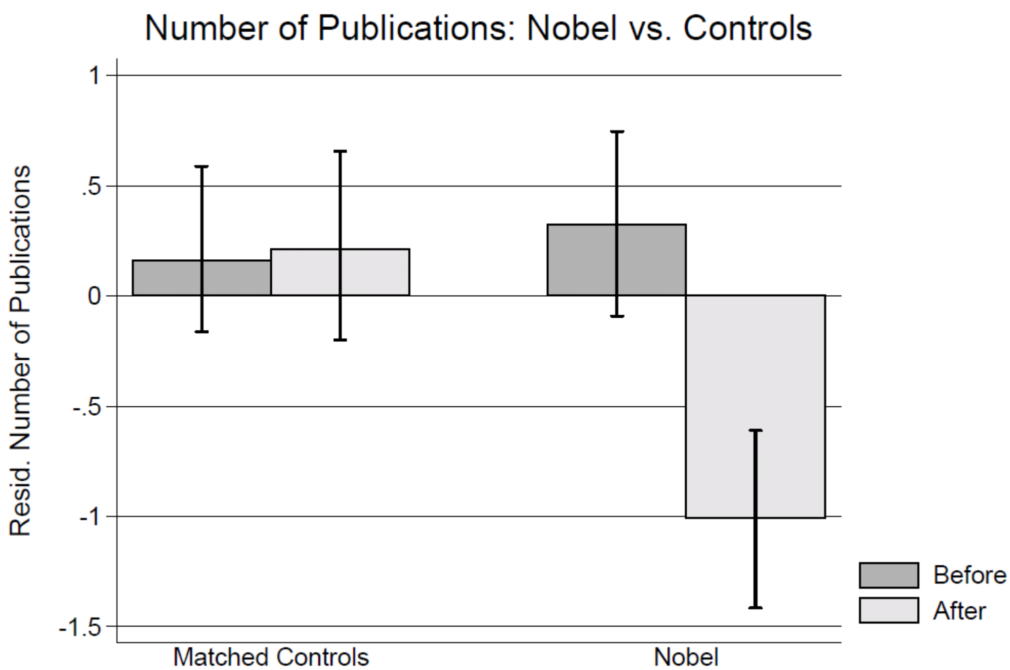
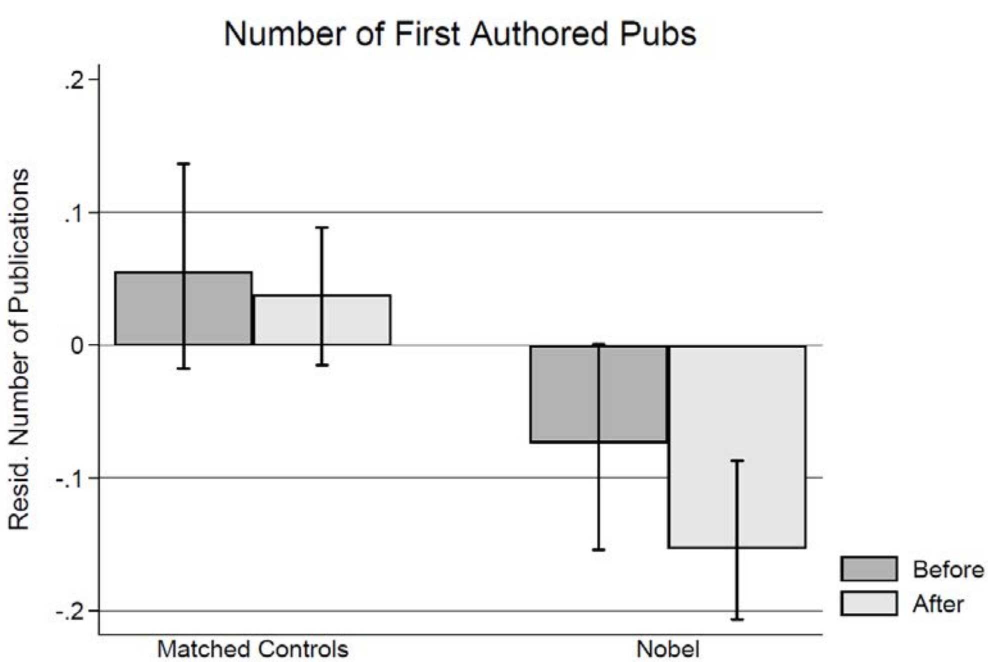


Figure A3: Residual Number of First Authored Publications per Year (Nobel Winners vs.



Matched Lasker Controls)

Figure A4: Residual Number of Last Authored Publications per Year (Nobel Winners vs. Matched Lasker Controls)

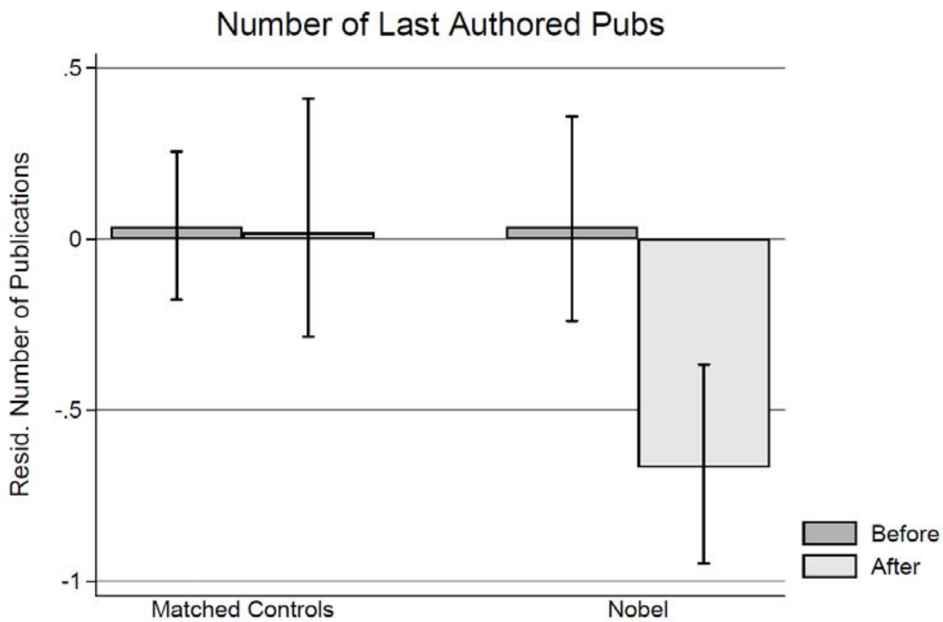


Figure A5: Residual Number of Co-Authors (Nobel Winners vs. Matched Lasker Controls)

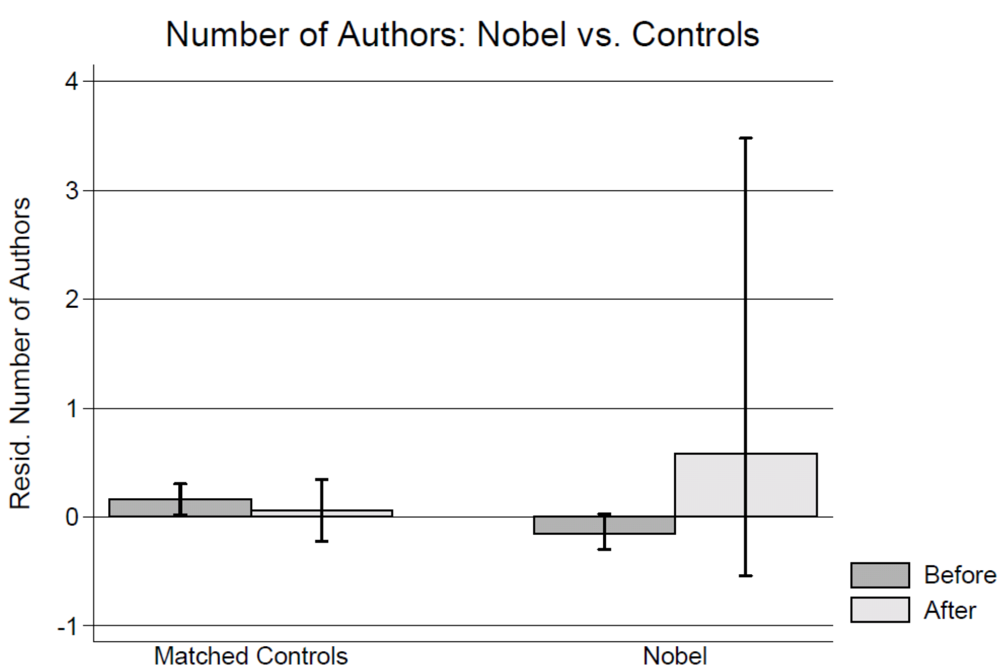


Figure A6: Residual Forward Citations All Time to Papers Published in Each Year Per Year (Nobel Winners vs. Matched Lasker Controls)

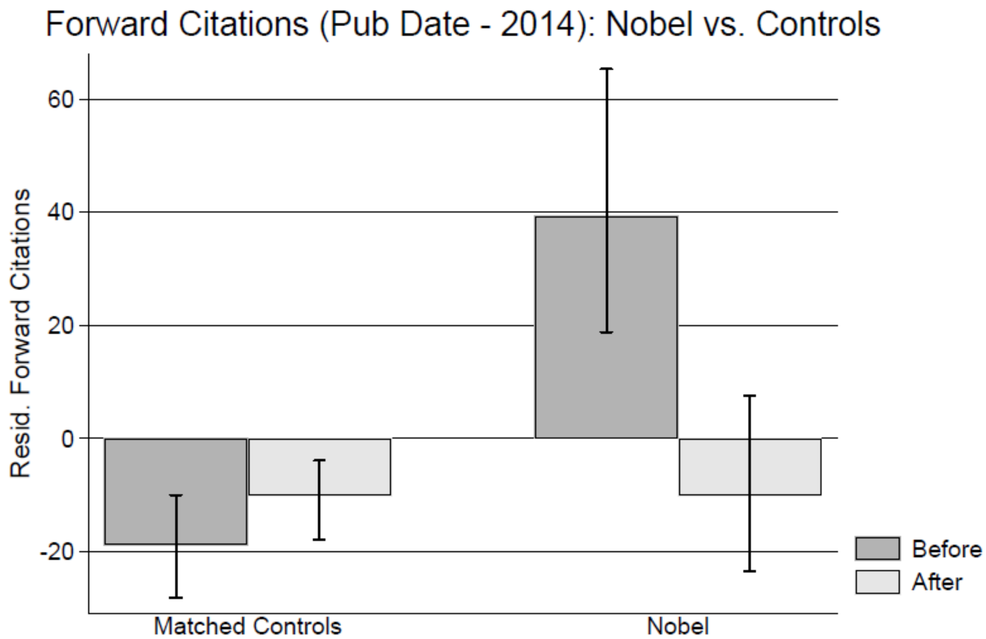


Figure A7: Residual Forward Citations in First Five Years After Publication to Papers Published in Each Year Per Year (Nobel Winners vs. Matched Lasker Controls)

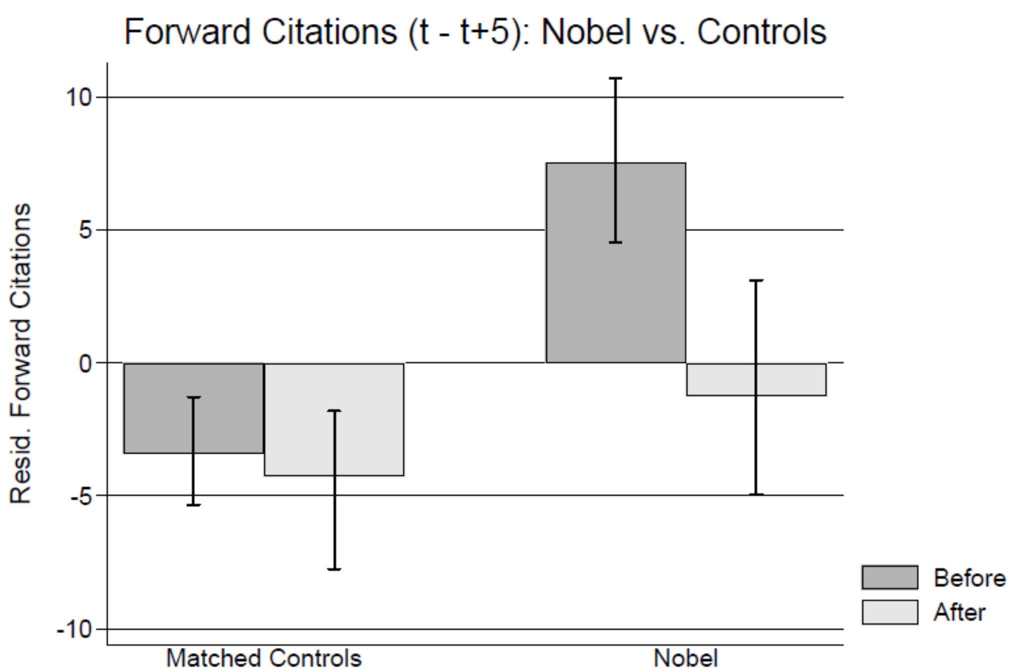


Figure A8: Residual Forward Citations in Six to Ten Years After Publication to Papers Published in Each Year Per Year (Nobel Winners vs. Matched Lasker Controls)

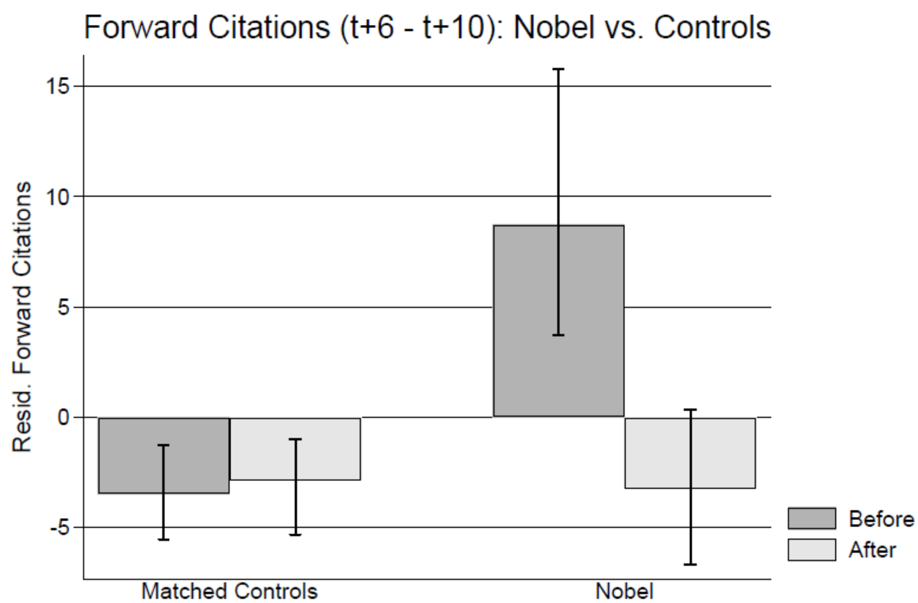


Figure A9: Forward Citation Rate (Nobel Winners vs. Matched Lasker Controls)

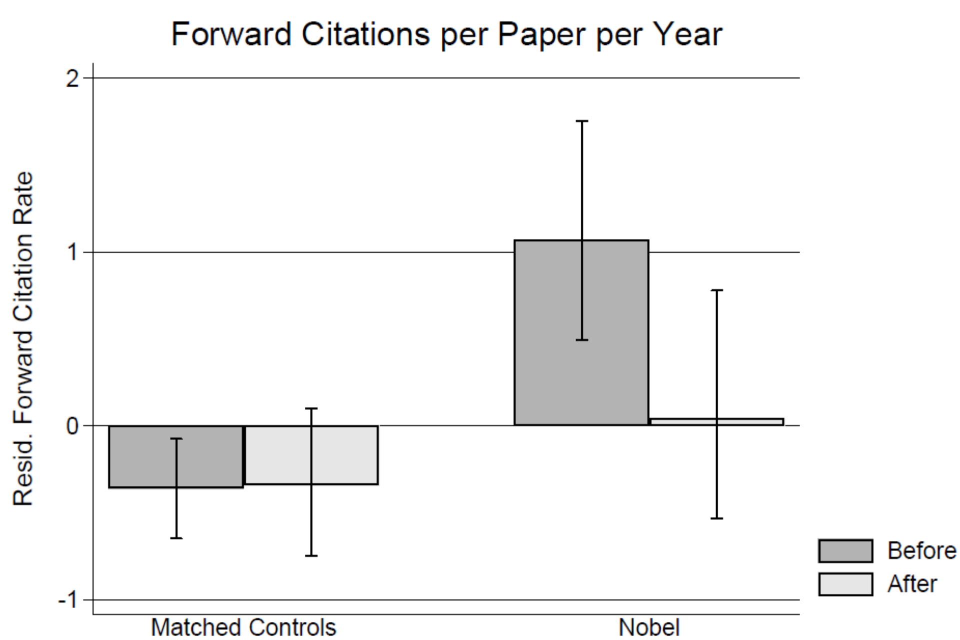


Figure A10: Residual Median Novelty Score (Nobel Winners vs. Matched Lasker Controls)

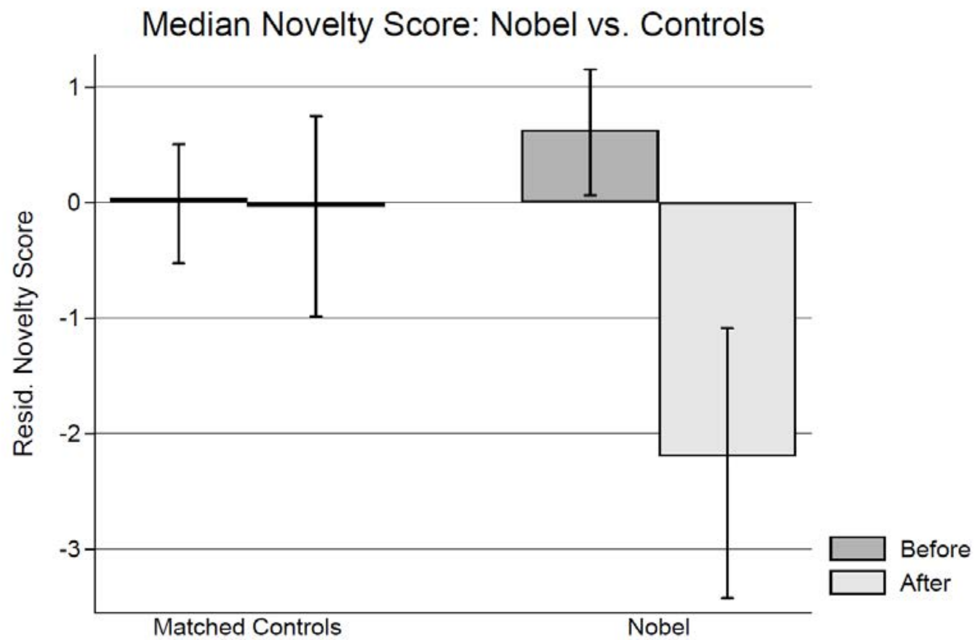


Figure A11: Residual 25th Pct. Novelty Score (Nobel Winners vs. Matched Lasker Controls)

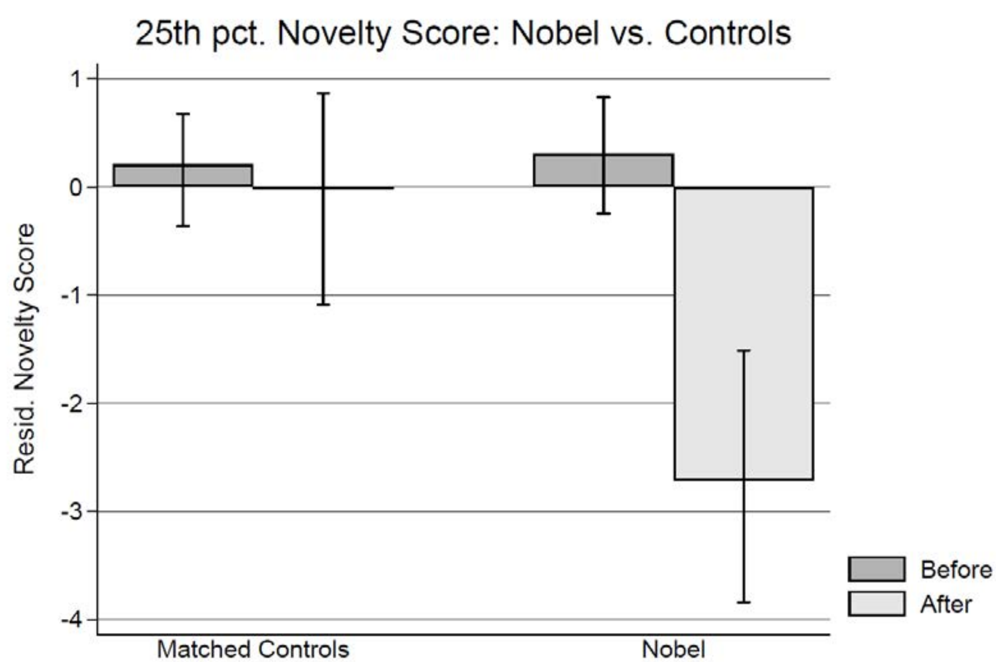


Figure A12: Distribution of Number of Matched Laskers – Match Not Limited by Birth Year

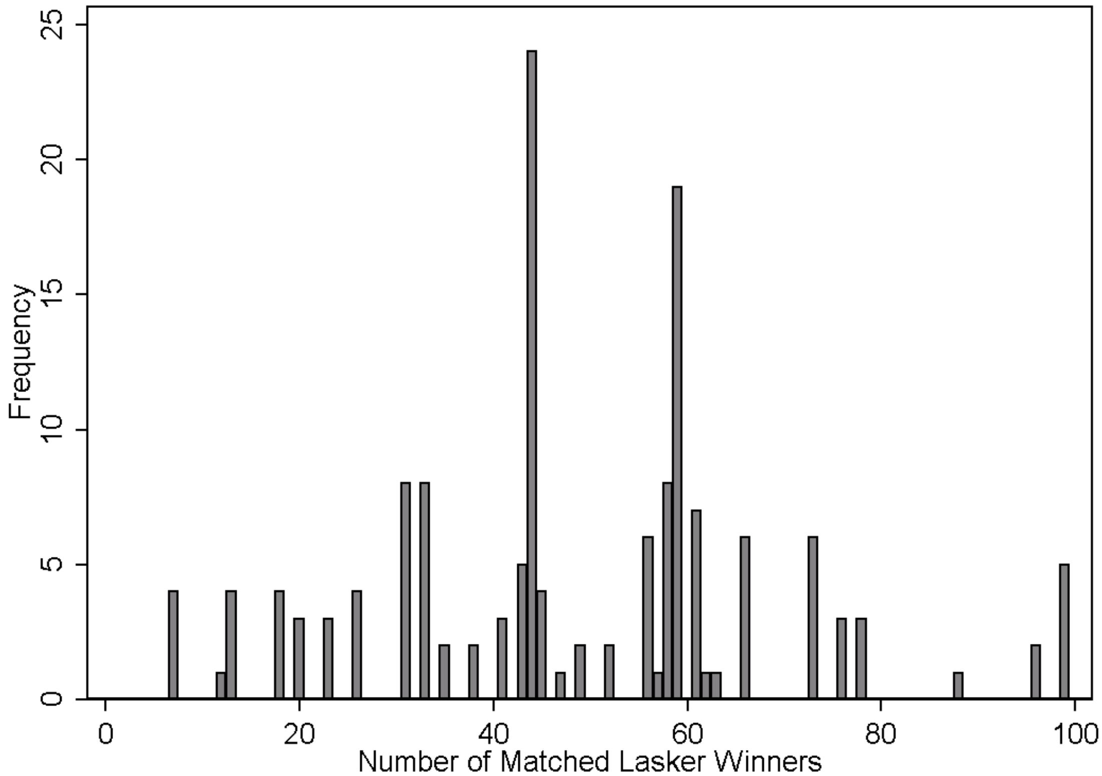


Figure A13: Sensitivity Analysis (Authorship) – Match Not Limited by Birth Year

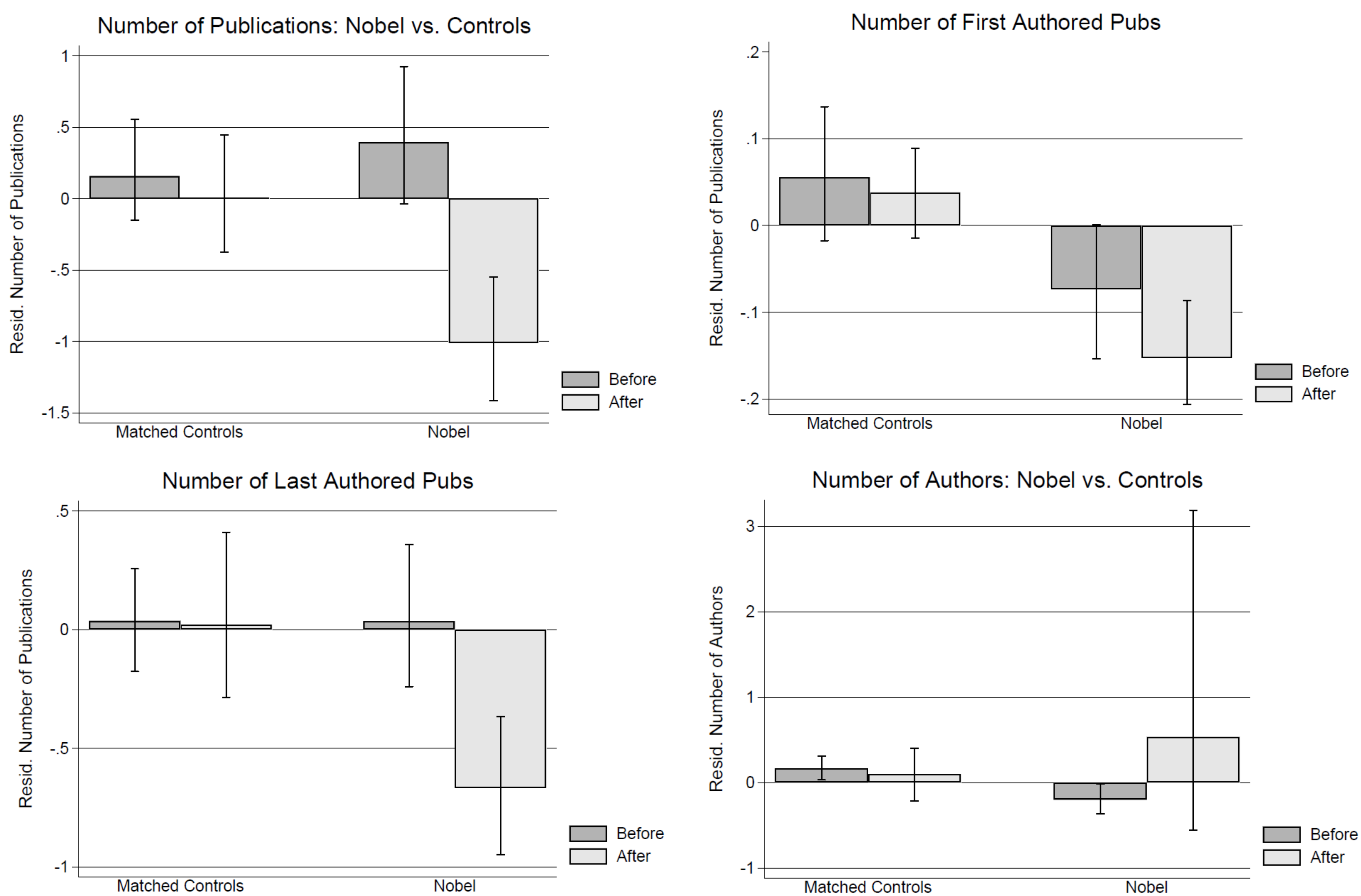


Figure A14: Sensitivity Analysis (Citations) – Match Not Limited by Birth Year

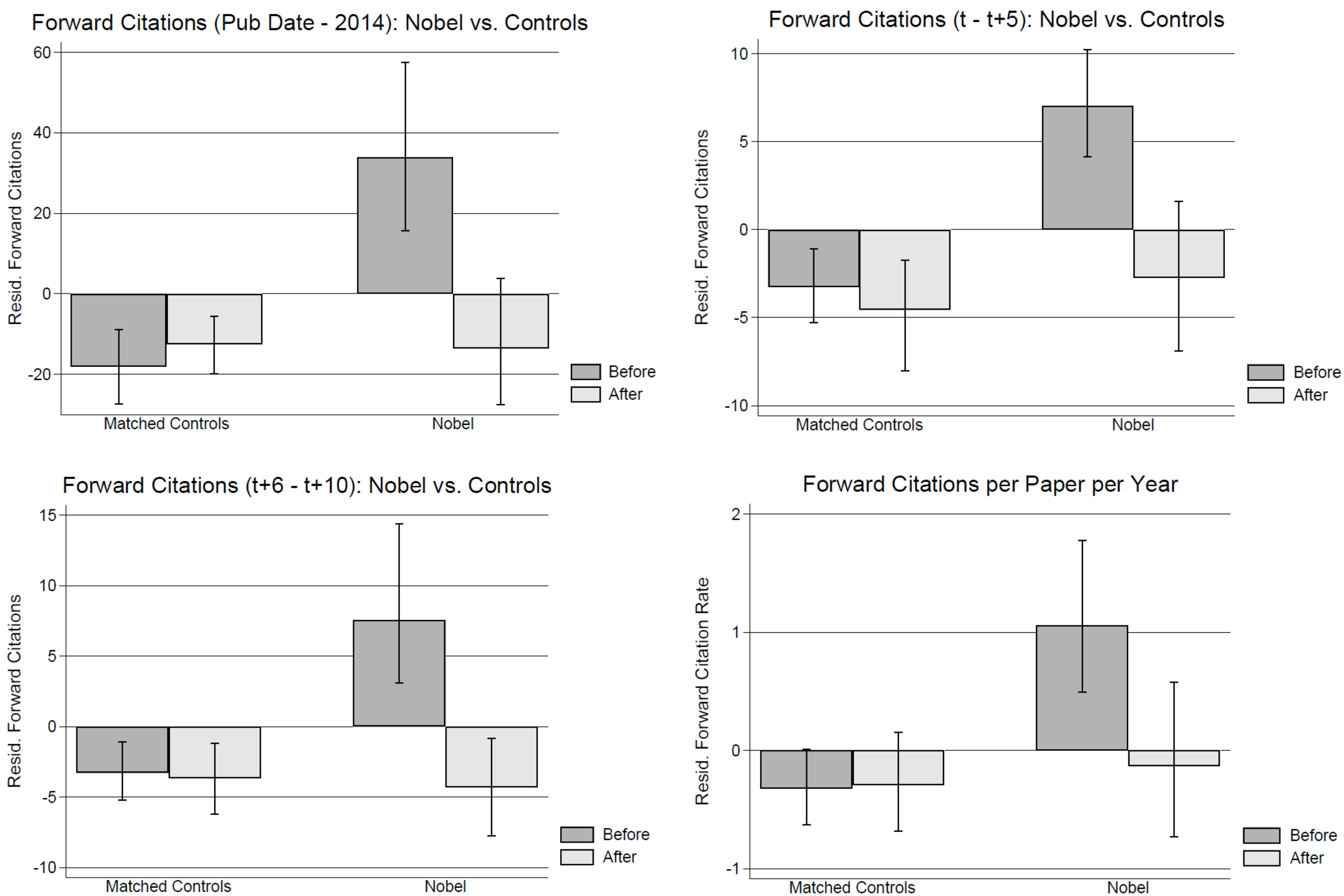
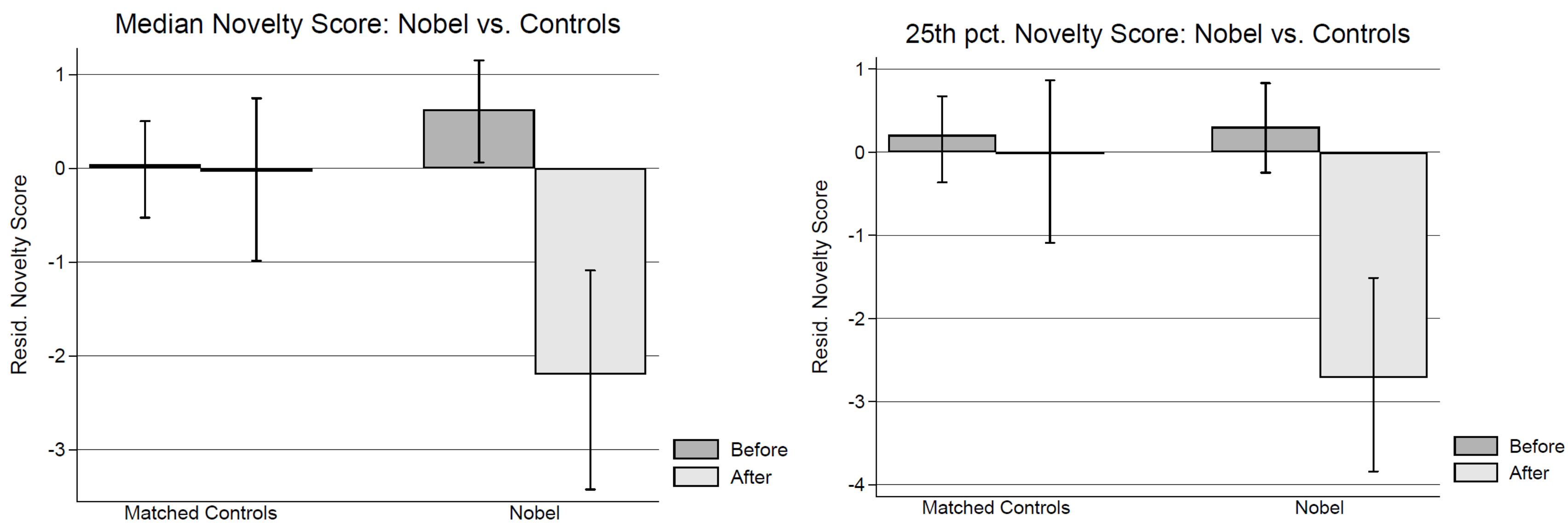


Figure A15: Sensitivity Analysis (Novelty) – Match Not Limited by Birth Year



Notes

1. Nobel Prize, "All Nobel Prizes in Physiology or Medicine," https://www.nobelprize.org/prizes/lists/all-nobel-laureates-in-physiology-or-medicine/ (last accessed February 25, 2019)
2. Albert and Mary Lasker Foundation, "The Lasker Awards," http://www.laskerfoundation.org/awards/ (last accessed February 25, 2019)
3. National Library of Medicine, "PubMed/MEDLINE," https://www.ncbi.nlm.nih.gov/pubmed/ (last accessed February 25, 2019)
4. National Library of Medicine, "MeSH," https://www.ncbi.nlm.nih.gov/mesh (last accessed February 25, 2019).
5. Torvik CI & Smallheiser NR (2009) "Author Name Disambiguation in MEDLINE," *ACM Trans Knowl Discov Data* 3(3): pii: 11. PMID: 20072710 PMCID: PMC2805000
6. https://www.nlm.nih.gov/bsd/mms/medlineelements.html
7. Clairvate Analytics, Web of Science, Science Citation Index, https://clarivate.com/products/web-of-science/ (last accessed on February 25, 2019)
8. Packalen M and Bhattacharya J (2017) "Neophilia Ranking of Scientific Journals" *Scientometrics* 110(1):43-64 PMID: 28713181 PMCID: PMC5506293 doi: 10.1007/s11192-016-2157-1
9. U.S. National Library of Medicine (2018) *Unified Medical Language System (UMLS)*. https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/ (last accessed 10/24/2018)