The Effect of R&D Investment on Graduate Student Productivity: Evidence From the Life Sciences

*Alexandra Graddy-Reed**
*Lauren Lanahan**
Nicole M. V. Ross

Abstract

This study examines the role of graduate training and R&D investments on research productivity by focusing on the effect of federal funding for early-career graduate students. We employ a difference-in-differences research design drawing upon a sample of high-quality life science graduate students who either are award recipients or honorable mentions of the prestigious U.S. National Science Foundation's Graduate Research Fellowship Program. We find that a $91,000 grant over three years has a limited, yet positive impact on the awardee's productivity. These effects are driven by the sample of graduate students without publications prior to applying for the fellowship. © 2018 by the Association for Public Policy Analysis and Management.

INTRODUCTION

Academic research and development (R&D) is a key contributor to the creation of scientific knowledge. With an emphasis on basic science research and training, academic R&D provides foundational support for the economy and national competitiveness. In 2016, U.S. academic institutions spent $72 billion on R&D, which accounts for 16 percent of total U.S. R&D activity.[1] The federal government is the leading funder for academic R&D, providing 54 percent of this investment. While researchers have broadly examined the impacts of this vital funding source (e.g., Azoulay, Graff Zivin, & Manso, 2011; Blume-Kohout, Kumar, & Sood, 2015; Jacob & Lefgren, 2011a, 2011b; Stephan, 2012), the discussion has largely ignored a key population in the research enterprise: graduate students.

The development and support of students during graduate training has important implications for the advancement of science broadly. Decisions about the training and support of graduate students today shape future scientific innovations and lines of inquiry that we eventually depend on. Graduate students fill a unique role by providing key labor for senior scientists while they acquire knowledge and experience via apprenticeship-like training that will enable them to design future independent research projects. The expectation, as with any apprenticeship model, is that these students graduate to become advisors themselves, eventually leading their own laboratories and contributing to basic and applied knowledge and research production.

[1] National Science Board, Science & Engineering Indicators (2018). See https://www.nsf.gov/statistics/2018/nsb20181/report/sections/academic-research-and-development/highlights.

Journal of Policy Analysis and Management, Vol. 00, No. 0, 1–26 (2018)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.22083

Life science research has consistently dominated academic and industry R&D, generating innovations that span the biomedical, environmental, pharmaceutical, and biotechnology fields. In 2016, over 56 percent of federal academic R&D funding was directed to this academic division.[2] In addition, approximately 85,000 students enroll annually in life science Master's or doctorate programs in the United States to provide the labor that generates these public goods; this division accounts for 13 percent of the total graduate student workforce.[3]

During training, these graduate students are primarily supported through their advisors' laboratories (Stephan, 2012) with limited opportunities for independent funding. While external, competitive R&D funds are most commonly available for more senior researchers, there are a few options available for graduate students at the pre-candidacy stage to support the establishment of their own research projects.[4] The impact of such external funding for student-led research at this early-career stage, however, is unclear.

By designating the student as the principal investigator (PI) and providing financial support, these external R&D funds have the potential to disrupt the apprenticeship model of graduate training. Free from traditional research assistance or laboratory obligations, graduate recipients may pursue their own research inquiries. This independence may cultivate skills essential in research, such as leadership and self-directedness, while allowing students to produce original research outputs at an earlier career stage. On the other hand, graduates "freed" from the traditional training model may miss out on learning tangible and intangible research skills that may be best imparted while working under the tutelage of experienced senior researchers and faculty. Further, the collaborative nature of research in the life science fields (Buffington et al., 2016; Conti & Lui, 2015; Jones, Wutchy, & Uzzi, 2008) raises the question of whether funding independence is a net benefit to students. The effects of competitive, comprehensive graduate funding appear ambiguous. Thus, we offer the first study that examines the impact of external R&D for life science graduate students on measures of research productivity.

The federal government offers the largest source of external funding for science and engineering (S&E) graduate students through the U.S. National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) (2018). The GRFP provides three years of support and service buy-out for early-stage graduate students with less than 12 months of training. As of 2017, the GRFP annual budget was $276 million. We draw upon the GRFP to assess the impact of federal funding on research productivity for life science graduate students, as this division is the largest recipient of the program. In addition, controlling for the academic field allows for comparison to prior research, which has prominently focused on the life sciences. While these studies have found a nominal positive impact of R&D investment on productivity, they have examined the impacts of R&D for more senior researchers

[2] National Science Board, Science & Engineering Indicators (2018), Table 5(1). See https://www.nsf.gov/statistics/2018/nsb20181/data/tables.

[3] National Science Board, Science & Engineering Indicators (2018), Appendix Tables 2–23. See https://www.nsf.gov/statistics/2018/nsb20181/report/sections/higher-education-in-science-and-engineering/graduate-education-enrollment-and-degrees-in-the-united-states.

[4] National Science Foundation (NSF) GRFP and National Institutes of Health (NIH) F30 provide funds directly to pre-candidacy graduate students. Students are eligible to explore innovative research without service obligations to the program. More detail on the F30 program is available at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-084.html#. NIH offers additional options (R36 and DP5) for the student to serve as the principal investigator once they reach the candidacy stage. More commonly, however, graduate students are supported indirectly through their advisor's laboratory grants or advisor's training grants (e.g., NIH T32, T35, and T90/R90). The NSF Doctoral Dissertation Research Improvement Grant awards the advisor on behalf of the student.

(e.g., Azoulay, Graff Zivin, & Manso, 2011; Blume-Kohout, Kumar, & Sood, 2015; Jacob & Lefgren, 2011a, 2011b).

Jacob and Lefgren (2011b) analyze the youngest cohort of researchers to date: post-doctorate fellows and applicants for the National Institute of Health (NIH, 2018) F32 training program.[5] They find evidence that the fellowship leads to a 20 percent increase in research productivity five years following the program. We extend beyond their study by examining the impact of R&D approximately five years earlier, at the *beginning* of the researcher's graduate training. While both samples include early-career scientists with a commitment to research, our sample draws upon a larger labor pool beyond those that eventually place in post-doctorate fellowships.[6] Our study further enhances prior work most focused on measuring the impact of R&D on research production by also considering training implications and professional placement.

For the primary model, we employ a difference-in-differences (DD) research design with a sample of high-quality life science graduate students who are either GRFP award recipients or honorable mentions between 1995 and 2005. The average annual budget for the program during this timeframe was $182 million with an average individual award allocation of $91,000, disbursed over a three-year period. To complement this design, we estimate a series of extensions to the primary model including sample stratifications and an augmented DD that accounts for time trends. In addition, as robustness checks to further address endogeneity of the student researcher quality due to non-random assignment of the GRFP award, we include coarsened exact matching (CEM) procedures and fixed effects estimations. Our primary outcome measure is research productivity as measured by total peer-reviewed publications. This is a common metric for scientific contribution, an outcome of importance to society at large. Publications proxy for innovative output and are frequently used in professional evaluation for placement and advancement. This is especially true for prospective employers evaluating early-stage researchers with shorter professional histories. Publication activity serves as a positive indicator of research ability and innovative potential.

We also consider other professional metrics. These include additional bibliometric measures of research quality and leadership, measures of graduate school completion, and professional placement following graduate training. These additional measures also inform what factors moderate the impact on research productivity.

The primary results indicate that federal funding is associated with a small improvement to the research record in the 10 years following funding receipt. Specifically, receiving a GRFP award, an average $91,000 investment over a three-year period, is associated with less than one additional peer-reviewed publication (0.636). This corresponds to an 11.8 percent increase in productivity 10 years following the award. These effects are driven by the sample of graduate students *without* publications prior to applying for the GRFP. Results from the augmented DD with time trends find a small but significant increase in the relative slopes between awardees and honorable mentions before and after the proposal; this effect is also moderated by the sample of students without publications prior to applying for the GRFP. The effect attenuates toward zero and is not significant for the sample of students with a prior record of publications.

While the GRFP is designed to provide early-stage funding, the impact of the funding is limited. The small impact on research productivity raises important questions

[5] See https://researchtraining.nih.gov/programs/fellowships/F32.

[6] Approximately half of the sample in this study secures research post-doctorate positions following their graduate training. This mirrors trends from other studies (e.g., Sauermann & Roach, 2012).

about the use of federal dollars to fund graduate students for the purpose of increased productivity. Moreover, these findings prompt an important discussion of the allocation of scarce federal dollars in regard to the trade-offs between scientific innovations and graduate training.

THEORETICAL MOTIVATION: FUNDING ACQUISITION FOR GRADUATE STUDENTS ON RESEARCH PRODUCTION

To motivate this study, we consider the mechanisms by which external resource acquisition can impact graduate training and subsequent research production. While competitive external funding confers a signal of research quality and potential, it also provides research and financial autonomy for the graduate student (Campbell & O'Meara, 2013; O'Meara et al., 2014). Namely, receipt of external funding may allow the graduate student to pursue a research agenda without the obligation to serve as a research or teaching assistant. Independence of this sort offers different training and research opportunities for the student. We consider the potential benefits and possible trade-offs made available by external funding.

External R&D grant fellowships are designed to provide resources for research advancement. This can offer advantages for the student as the autonomy may provide an opportunity to carve their own research niche and develop leadership and independence. According to interviews with graduate student recipients of NIH-funded traineeships and fellowships, preferences among the students are clear—the ability of students to focus exclusively on their own dissertation research was the most valued aspect of the fellowship (National Research Council [NRC], 2005). The resources granted to the student and the legitimacy conferred by securing a competitive grant can act as a positive feedback mechanism, help socialize the student deeper into the field, and perhaps spark further research productivity. From an organizational standpoint, financial autonomy for graduate students can also result in added benefits for their institution or department, as internal funds may be made available to fund other students or projects (Freeman, Chang, & Chiang, 2009).

Although external grants are intended to improve research capacity and professional options,[7] additional resources do not necessarily ensure increased productivity. Given the student's early-career stage, there may be unanticipated trade-offs that diverge from the primary intentions of the federal grant and graduate training program. From one angle, the grant may provide resources to increase research independence at the expense of collaborating with the members in the department or laboratory. However, life science research production is an increasingly collaborative activity (e.g., Conti & Lui, 2015). Certain tangible and intangible skills may be best learned while working for more experienced senior researchers. External fellowship programs may act as a substitute for training opportunities that are available through more standard assistantship positions. As a result, students may miss out on important benefits unique to these opportunities.

From another angle, grant recipients may change their research behavior by expanding their interests. Or, the resources may give them the flexibility to engage in riskier projects.[8] Although potentially this could yield significant contributions, interdisciplinary or high-risk projects are harder to publish. Often, they take additional time to validate given their heightened level of innovation (Mansfield, 1995).

[7] The NSF GRFP explicitly states in its mission: "NSF Fellows are anticipated to become knowledge experts who can contribute significantly to research, teaching, and innovations in science and engineering." (See https://www.nsfgrfp.org/general_resources/about.)
[8] See https://www.nsf.gov/about/transformative_research/.

This could lead to a reduction in the pace of research production, which may have negative implications for the student as they advance through the training regimen. Given the number of professional transitions that take place at this early-stage, changes in the pace of research production can account for career differences in the long run.

Taken together, the implications of external funding allocation for graduate students are unclear. We are centrally interested in examining the impact of external R&D for life science graduate students on measures of research productivity. This analysis not only provides an assessment of the return from public R&D investment on scientific knowledge production, it also offers insight into an alternative graduate training regimen—procurement of independent research support —that differs from the more prevalent apprenticeship-training model.

EMPIRICAL CONTEXT, SAMPLE, AND DATA

GRFP Program Overview

Funding students since 1952, the NSF GRFP has a demonstrated history of supporting promising graduate students across S&E disciplines. Notable recipients of the award include Sergey Brin, founder of Google, Steven Chu, former Secretary of Energy, 42 Nobel Laureates, and over 450 members of the U.S. National Academy of Sciences. Over the program's tenure, roughly 50,000 of more than 500,000 applicants have received awards.[9] As of 2017, NSF offers $138,000 for the full GRFP award, apportioned as a $34,000 annual stipend and $12,000 annual educational allowance for three years. For comparison, NSF offered $69,000 in 1995 and $121,500 in 2005 for the full award, with $14,400 (1995) and $30,000 (2005) for the annual stipend and $8,600 (1995) and $10,500 (2005) for the annual educational allowance.[10] The educational allowance is intended to cover the student's cost of tuition and fees.

This paper uses the GRFP to assess the impact of funding on future research productivity and on training and placement. The GRFP is optimal to study such an effect due to its unique attributes. In contrast to other NSF funding mechanisms, eligible GRFP applicants are at an earlier point in their research training.[11] Students are eligible to apply only through their first 12 months of graduate training.[12] The GRFP proposal consists of four components: (i) three-page personal statement; (ii) two-page research statement; (iii) three to five reference letters; and (iv) academic transcripts. Moreover, the graduate student serves as the PI. For the large majority of other NSF funding opportunities, PIs are more senior researchers. This even holds for NSF's Doctoral Dissertation Research Improvement Grants (DDRIG) program, which formally funds the doctoral candidate's advisor on behalf of the student.

Though NSF has firm guidelines about restricting application data with the public, the GRFP program provides the exception: competitive applicants just shy of receiving the award are publicized as esteemed honorable mentions. NSF explicitly groups these two together in their solicitation as they seek to "determine the successful applicants from these recommendations, with Fellowships [awardees] and Honorable Mentions offered based on the GRFP portfolio within the context of NSF's mission."[13] While honorable mentions do not receive the financial

[9] See https://www.nsfgrfp.org/general_resources/about.
[10] GRFP Program Solicitation, NSF 05-601.
[11] NSF GRFP recipients are required to be U.S. citizens, have national or permanent resident status, and be enrolled at U.S. institutions. See https://www.nsf.gov/pubs/2016/nsf16588/nsf16588.htm.
[12] During this study's timeframe, students are eligible to apply the year prior to applying to graduate school through their first 12 months of graduate training.
[13] GRFP Program Solicitation, NSF 15-597. See http://www.nsf.gov/pubs/2015/nsf15597/nsf15597.htm.

benefits of the GRFP award, this acknowledgment is revered as a signal of intellectual merit and research promise. During our timeframe of interest, on average, there were 870 awardees and 1,380 honorable mentions acknowledged annually across the entire S&E program. This accounts for approximately the top 20 percent of the total applicant pool each year. We draw upon awardees and honorable mentions to identify a sample of graduate students with comparable research quality around the theoretical award cutoff. It is with this award cutoff that we address non-random assignment of the R&D investment on graduate student productivity.

To substantiate identification, the unique structure of the GRFP proposal alleviates some concerns of plausible variation between the two groups. In contrast to other NSF funding opportunities, the length of the two-page research statement is shorter than NSF's standard 15-page project description. The succinct application focuses most on the applicant's background and research goals rather than presenting a definitive research design and plan for implementation; this stands in contrast to NSF's general proposal requirements.[14] Moreover, applicants may submit their proposals concurrently with applications to graduate programs. This limits the student's ability to specify the advisor, laboratory, program, and even institution where they will execute the research. Even for applicants that apply during their first year of graduate training, the GRFP proposal is due the fall of their first semester before many students have finalized their laboratory assignment. These program elements—format, eligibility criteria, and submission timeline—provide few definitive items for reviewers to evaluate proposal viability.

From this limited information, reviewers must explicitly distinguish the top 20 percent that are deemed meritorious from the rest. That is to say, the external reviewers note which students meet the quality threshold to receive an award or an honorable mention. From this curated sample of applicants, a separate NSF panel determines which applicants are awardees and which are honorable mentions (Freeman, Chang, & Chiang, 2009). We use the distinction in the GRFP recognition—namely, awardee versus honorable mention—as a baseline for identification. We also use additional individual-level research metrics to further address issues of endogeneity related to researcher quality. We elaborate on this below.

GRFP Sampling and Data Construction

To examine the effect of GRFP funding on research productivity, we built an individual-level panel data set of GRFP graduate students with details on their graduate training and research output. To construct this extensive data set, we first selected a sample of GRFP awardees and honorable mentions and then gathered annual metrics on the following: graduate training activity at the individual and organizational levels; professional placement following the completion of graduate training; and research production outputs in terms of peer-reviewed publication activity.

We construct the data set by drawing upon a series of third-party data sources. In doing so, we are able to triangulate the information across various sources to ensure accuracy of the data (Feldman & Lowe, 2015). This approach yields complete data over an extended panel for a larger sample of individuals—effectively producing a higher "response rate" than comparable surveys. This follows prior work (e.g., Jacob & Lefgren, 2011a, 2011b), yet it offers useful additions by including a more extensive set of metrics related to training and professional outcomes.

[14] Proposals, awards, policies, and procedures guide; see https://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp.

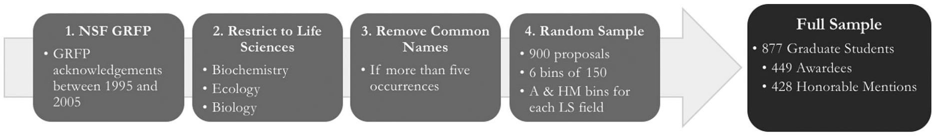
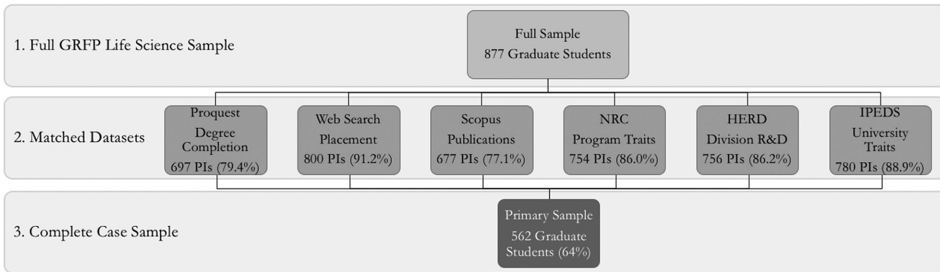


Figure 1. Sampling Flow Chart.



Notes: Sample restricted to primary sample—complete case of those who ever publish (same sample as reported in Table 1); stratified by those with and without publications prior to GRFP year. Panels A, B, and C include activity for 562, 142, and 384 students, respectively.

Figure 2. Data Construction Flow Chart.

The rest of this sub-section details the methodology for sampling and data construction. We present this as two steps. In Step One we overview the sampling approach for selecting GRFP awardees and honorable mentions. Figure 1 illustrates this process. In Step Two we detail the methodology for augmenting the full GRFP sample with additional research-related metrics drawn from a series of third-party data sources. Figure 2 overviews this process and provides detail on the match rate across the various sources. Refer to Appendix A for additional detail on data collection, sampling, and the matching process.[15]

Step One: GRFP Sampling

We draw a sample of NSF GRFP awardees and honorable mentions who applied for the award between 1995 and 2005. We use 1995 to define the lower bound of our sample selection, as it is the first year that the complete list of honorable mentions is available. We selected 2005 for the upper bound to allow for a 10-year post period.[16] For each individual, we construct a 16-year timeframe; this includes five years prior to the GRFP proposal submission ($-5 \leq t \leq -1$), the GRFP year—the sixth year ($t = 0$), and the following 10-year time period ($1 \leq t \leq 10$).[17] This provides a sufficient timeframe to study the follow-on research productivity, graduate completion, and initial professional placement for these students.

Second, while the NSF GRFP program is a foundation-wide program—thus supporting graduate students across all six S&E divisions[18] —we restrict the sample to

[15] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[16] Data construction for this project took place from fall 2015 to spring 2016.
[17] To illustrate, for students who applied for the GRFP in 1995, their 16-year timeframe extends from 1990 to 2005, while someone who applied in 2005 would have a timeframe from 2000 to 2015.
[18] NSF directorates include Biological Sciences (life science division); Computer & Information Science & Engineering; Education & Human Resources; Engineering; Directorate for Geosciences; Mathematical & Physical Sciences; and Social, Behavioral & Economic Sciences.

GRFP activity with the largest funding allocation, the life science division. Moreover, time and resource constraints prohibit an exhaustive data collection process for students across all divisions. We restrict the sample to the three most prominent fields within the life science division as defined by GRFP performance. These are Ecology, Biochemistry, and Biology, which comprise 21.32, 20.54, and 12.36 percent of the life science GRFP activity, respectively.

Third, we restrict the sample by removing observations with common last names. This approach follows prior studies reliant on third-party data sources to help ease the process of verifying whether the correct individual is being tracked over time across various data sources (Jacob & Lefgren, 2011a). Importantly, this approach reduces the presence of false positives in the analysis. We do not expect that individuals with more common names would behave or be treated differently in their research, so this restriction should not bias the results. Pragmatically, among the students listed in the three life science fields, we remove those with a last name that occurs more than five times.[19] One trade-off with this approach is that individuals with common Asian-American surnames are removed at a disproportionately greater rate. This limits the generalizability to this demographic at the expense of improving the accuracy of the data.

Fourth, we divide our sample by GRFP proposal status (awardee or honorable mention) and by the three life science fields. This yields a total of six bins. Within each bin, we randomly sample without replacement 150 unique proposals yielding a total of 900 student-proposal-year observations. We defined this number to ensure sufficient statistical power[20] while also limiting the size given the notable search costs with the triangulation of data across numerous sources (Step Two). Of the 900 proposals, 23 students appeared in the data set twice. This occurred as a result of an individual initially receiving an honorable mention and then an award or honorable mention in the following year. To define the full sample, we identify the 877 unique individuals. For those with more than one record we document the most recent observation.

Step Two: Data Construction

To trace research-related activity for the set of 877 unique individuals, we relied on the following data sources: ProQuest, online web searches, Scopus, NRC's decennial survey of U.S. doctoral programs, NSF Higher Education Research and Development (HERD) survey, and National Center for Education Statistics (NCES) Integrated Postsecondary Education Data System (IPEDS). To elaborate on each, ProQuest serves as the largest central repository of dissertations and theses, housing over three million entries. We rely on this source not only to track degree completion activity, but also to obtain data on each student's dissertation, advisor, institution, and subject area of research. Using the student name and institutional affiliation from the GRFP awards database, we were able to retrieve data for 697 graduate students (79.4 percent of the GRFP sample).

Second, we systematically searched online sources—including LinkedIn, institutional webpages, personal webpages, and news outlets—to identify initial placement

[19] We remove 10 percent of the sample (389 observations) that have five or more last name duplicates.
[20] We ran power tests to ensure the sample size was sufficiently large to estimate significance tests. The minimum size for difference in means test with data in wide form is 303 for 80 percent power and 405 for 90 percent. For regression analysis with data in long form, it is 5,152 for 80 percent power and 6,944 for 90 percent. Our primary sample yielded 562 student observations in wide form and 8,430 student-year observations in long form. We also have sufficient power when we stratify by those without prior publications. When we stratify by having prior publications, the long form sample of students with prior publications contains 2,280 observations; thus, we have insufficient statistical power with this subset.

data immediately following graduate training. This online search also allowed us to corroborate ProQuest matching. Moreover, we vetted degree completion activity for the set of observations not found in ProQuest. We retrieved professional placement data for 800 graduate students (91.2 percent).

Third, we pulled publication activity from Scopus. This data source, a subsidiary of Elsevier, houses the largest collection of bibliometric data with over 60 million records. Using names,[21] keywords available via ProQuest, and institutional affiliation (further confirmed by the ProQuest and online searches), we identified 677 graduate students (77.1 percent) that published 5,900 peer-reviewed[22] academic publications over their respective standard 16-year timeframe.[23] Given the breadth of Scopus' database, for those 200 individuals with no record, we conclude they have no peer-reviewed publications over their standardized 16-year window.

Fourth, to account for differences arising between students based on institutional affiliation, we gathered organizational-level measures from three U.S.-based higher education data sources. The NRC provides university-field-level data on faculty publication activity, program size, and rank indicators. The HERD data provide university-division-level data on federal R&D financing. Lastly, IPEDS data provide university-level data on Carnegie classifications and institutional governance. We draw upon the applicant's field, broader division, institution, and year of proposal acknowledgment to match with each respective data source. Of note, 70 students from the full GRFP sample ended up graduating from non-U.S. institutions so we are unable to gather data for this set. With that said, we retrieved data from these three sources for the sample of students as follows: NRC (754), HERD (756), and IPEDS (780).

For the primary analysis, we focus on the sample of research-active graduate students—specifically, students with *any* publication activity over their standardized 16-year timeframe. This research-active sample includes 677 students. Among this set, we have data from the complete set of additional external sources for 562; 296 are awardees and 266 are honorable mentions. We refer to this research-active sample that allows for complete case analysis as the *primary sample*. From the full sample (877 students), we gathered complete data for 66 percent of awardees and 62 percent of honorable mentions. Based on results from power tests, this combined sample size is sufficiently large to detect statistical significance.[24] We also estimate the primary model for the extended sample; this includes all applicants—including those who never publish over their standardized 16-year timeframe—with complete data for the full list of controls (697 students). The results are robust and are discussed later.

Descriptive Statistics

Table 1 presents summary statistics of the data set for the parameters of interest. Proportions or means for the primary sample are presented in column 1 and stratified samples by awardee and honorable mention are presented in columns 2 and 3, respectively. Column 4 denotes the statistical significance of the difference in means between the two subsamples. Panel A details the statistics for variables of

[21] By pulling data from multiple sources, we were able to trace name changes. This applied most prominently to women who married over the timeframe.
[22] We only include peer-reviewed journal articles—a standard bibliometric measure for research production (e.g., Adams & Griliches, 1998; Stephan, 2012). Moreover, we follow Jacob and Lefgren's (2011a) methodology—removing reports, books, conference proceedings, trade journals, and other publications.
[23] This timeframe includes five years prior to the GRFP proposal submission ($-5 \le t \le -1$), the GRFP year—the sixth year ($t = 0$), and the following 10-year time period ($1 \le t \le 10$).
[24] Refer to Footnote 20.

Table 1. Descriptive statistics by award type.

	(1) Primary sample	(2) Awardees	(3) Honorable mentions	(4)
Panel A: Publication activity				
Publications				
Total number of journal articles (1 to 86)	13.49	14.50	12.37	**
	(12.87)	(13.24)	(12.37)	
Years from GRFP to first publication (−4 to 11)	2.32	2.11	2.57	**
	(2.98)	(2.86)	(3.09)	
Any prior publications 5 years before GRFP	0.27	0.29	0.24	*
If prior publications, total number (1 to 4), $N = 152$	1.54	1.56	1.51	
	(0.06)	(0.08)	(0.10)	
Publication leadership				
Any first author publications	0.91	0.93	0.88	**
Years from GRFP to first author position (−4 to 11), $N = 509$	3.93	3.80	4.09	
	(3.00)	(2.99)	(3.00)	
If first author publications, total number (1 to 21), $N = 509$	4.13	4.40	3.82	**
	(3.06)	(3.32)	(2.70)	
Of total, share first author publications	0.35	0.35	0.34	
Publication quality				
Number of publications in top 1% of Journals (0 to 10)	0.34	0.42	0.26	**
	(0.88)	(1.04)	(0.66)	
Number of publications in top 5% of journals (0 to 13)	0.89	1.02	0.73	**
	(1.58)	(1.76)	(1.34)	
Number of publications in top 10% of journals (0 to 13)	1.29	1.45	1.10	**
	(1.87)	(2.03)	(1.67)	
Panel B: PI, professional placement, and program characteristics				
Characteristics				
Female	0.50	0.52	0.47	*
Field				
Biochemistry	0.34	0.33	0.35	
Ecology	0.34	0.35	0.32	
Biology	0.33	0.32	0.33	
Degree completion and placement				
Completed a PhD, $N = 561$	0.93	0.95	0.91	**
Years from proposal to degree (0 to 12), $N = 552$	5.30	5.36	5.24	
	(1.62)	(1.47)	(1.76)	
Academic position, $N = 542$	0.66	0.68	0.64	
Research position, $N = 542$	0.83	0.86	0.81	*
Post-doctoral research position	0.56	0.62	0.49	***
Program traits				
University-field publications per faculty (0.46 to 3.6)	2.18	2.21	2.14	
	(0.60)	(0.57)	(0.63)	
Federally financed HE R&D (5 to 186,284)	53,295	55,205	51,170	*
	(34,808)	(34,310)	(35,297)	
University-field GRFP 2-year capacity (0 to 24.8)	5.27	5.65	4.86	**
	(4.66)	(4.66)	(4.63)	
Program ranked in top tercile	0.84	0.87	0.79	***

Table 1. Continued.

	(1) Primary sample	(2) Awardees	(3) Honorable mentions	(4)
Program size by quartiles				
Quartile 1 (lowest)	0.02	0.01	0.03	
Quartile 2	0.07	0.07	0.06	
Quartile 3	0.55	0.56	0.54	
Quartile 4 (highest)	0.36	0.36	0.37	
University traits				
Carnegie very high research classification	0.92	0.91	0.93	
Public institution	0.52	0.49	0.55	*
Observations	562	296	266	

Notes: Descriptive statistics for primary sample reflect the complete case sample of individuals who publish at least once during the 16-year timeframe. Means (standard deviations) or proportions are presented in columns (1), (2), and (3). Statistically significant differences in t tests between Awardee and Honorable Mention means indicated in (4) by ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. The variable *prior publications* omits outliers above the 99th percentile ($n > 4$). Federally financed University R&D of the life sciences division deflated (base year 2009) and scaled to thousands of dollars.

publication activity, while panel B provides statistics on variables of professional placement and program characteristics.

On average, the total publications level is 13.49. It took on average 2.32 years from the GRFP proposal submission event to first publish, with the award group demonstrating a faster pace of 2.11 years verses 2.57 years for the honorable mention group (p-value = 0.03). One should note that these means are pulled towards zero by the student observations whose first publication took place *before* the GRFP proposal submission.

In an effort to examine baseline differences of research productivity between the two groups, we examined publication trends in the five years preceding the GRFP proposal submission. Among the primary sample, 27 percent published *before* applying for the GRFP. Twenty-nine percent of awardees have prior publication activity, while 24 percent of honorable mentions do (p-value = 0.09). However, for those with prior publication activity in each sample, the mean prior publication level between awardees (1.56) and honorable mentions (1.51) is not statistically different (p-value = 0.33). These numbers suggest that awardees and honorable mentions with prior publication activity have comparable averages along this measure.

When we examine additional detail on publication trends, we find 91 percent had a first-author publication over the 16-year window. The position of first-author within the field of life sciences indicates the individual who led the research inquiry. Other authors and laboratory members are listed after, with the laboratory director listed in the last position (Dance, 2012; Tscharntke et al., 2007). These first-authored publications accounted for 35 percent of the student's total publication activity over the 16-year timeframe. Awardees are statistically more likely to have a first-author publication over the full timeframe (93 percent compared to 88 percent for honorable mentions; p-value = 0.04) and to have a higher level of first-author publications (4.40 for awardees versus 3.82 for honorable mentions; p-value = 0.02). Not surprisingly, it took roughly 19 more months to publish as first author in contrast to any publication for the primary sample (3.93 years from the GRFP event compared to 2.32 years). It takes slightly more time for honorable mentions to achieve a first-author publication (4.09 years from the GRFP event) compared to awardees (3.80 years), although this difference is not statistically significant (p-value = 0.14). Turning to publication quality, as measured by the rank of the journal outlet, awardees exhibit slightly higher levels of publications than honorable mentions. Across the various journal ranking bibliometrics, these differences are statistically significant.

In panel B, we examine post-graduate training metrics that include degree completion and professional placement. Awardees exhibit a slightly higher graduation rate (95 percent) compared to honorable mentions (91 percent); this difference is statistically significant (p-value = 0.04). However, the time to completion is slightly over five years and does not vary by group (p-value = 0.18). Regarding placement, the results indicate the two groups are comparable in terms of attaining an academic-research position following graduation; this includes post-doctoral, tenure-track, and research appointments. However, looking at this more narrowly, we do find differences in post-doctorate appointments specifically (62 percent for awardees in contrast to 49 percent honorable mentions; p-value = 0.00), which may be a function of the GRFP.

METHODS

Primary Model

We estimate the effect of R&D funding on graduate student research production. The central design threat with a study of this nature lies in accounting for unobservable measures of individual research quality. Without proper model correction, we run the risk of estimating biased results by (most likely) overestimating the treatment effect. To address this, scholars have most often relied on quasi-experimental designs drawing upon panel scoring data around the award cutoff (Arora & Gambardella, 2005; Azoulay, Graff Zivin, & Manso, 2011; Jacob & Lefgren, 2011a, 2011b). However, NSF does not provide public information on the panel's merit review rankings; thus, we are unable to estimate a regression discontinuity design. Even so, given the unique features of this program that lend a heightened ambiguity to the review process, including an applicant pool at a nascent research stage and an abbreviated proposal, panel scores would offer less value than scores from programs that review more detailed projects from senior researchers with demonstrated records.

To estimate the effect of early-stage R&D investment on graduate student productivity, we employ a DD research design. We use the two prestigious distinctions from the GRFP program—awardee and honorable mention—to serve as treatment and control groups, respectively. This provides a baseline for identifying graduate students with comparable research quality, as designated by the merit review process. Moreover, we trace their research productivity over a standardized 16-year timeframe both prior to ($-5 \leq t \leq -1$) and following the GRFP event ($1 \leq t \leq 10$). For the empirical estimations of the DD model, we drop the sixth year ($t = 0$), the year of the GRFP acknowledgment, to define a clear cutoff. This offers a standard window for each individual.

Equation (1) is an individual-year-level model estimating the effect of the GRFP award receipt on graduate student productivity, where i denotes the individual graduate student and t denotes the standardized 16-year time period.

$$Y_{it} = \beta_0 + \delta_1 Post_t + \delta_2 Award_i + \delta_3 (Award_i * Post_t) + \beta_z X + \varepsilon_{it}. \quad (1)$$

Y_{it} captures graduate student productivity, which we measure as the number of peer-reviewed, academic journal publications authored by the graduate student.[25]

[25] As a sensitivity check to the level form, we estimate the outcome in logarithmic functional form. We discuss these results in the section on Results.

The parameter of interest is the coefficient on the interaction term $Award_i * Post_t$, δ_3, which theoretically measures the treatment effect of external R&D funding on research productivity. The parameter δ_1 captures the mean change in all publication activity in the 16-year timeframe.[26] The coefficient for $Award_i$, δ_2, measures the differences between eventual awardees and honorable mentions prior to GRFP proposal submission. Theoretically, the set of awardees and honorable mentions should be nearly equivalent in this regard; however, this serves as an added control for baseline differences.[27]

Additionally, we include a series of individual- and institutional-level controls, X,[28] which have been shown to be associated with higher rates of GRFP attainment at the program level (Graddy-Reed, Lanahan, & Ross, 2017). We rely on these variables both as controls for the main model, but also rely on some of them to assess heterogeneity in subsequent models (refer to the next section). First, we include individual binary indicators of the applicant's gender and for whether the applicant has publication activity prior to the GRFP, respectively. Second, we include a series of institutional measures: a measure for the average number of a faculty member's annual publications;[29] a set of dummies for graduate program size and rank; a program-level measure for the number of GRFP awards and honorable mentions received in year t and year t-1 relative to the year the student applied for the GRFP; and a continuous measure of federal R&D funding at the division level. For program rank, we use the NRC's R ranking, a regression-based ranking derived from a faculty survey of peer programs (Ostriker et al., 2011). Specifically, we estimate two bins from the tercile rankings—rank 1 (high) and ranks 2 (mid) and 3 (low). We also include an indicator for the university research rank ("very high") and an indicator for university governance type (public or private).

Finally, we include two additional sets of controls. The first are binary indicators for each of the three fields sampled within the life sciences division (Biochemistry, Ecology, and Biology) to control for disciplinary discrepancies that may vary within division at the field-level and affect professional or publication norms. Second, we include a set of calendar year dummies (1995 to 2005) of the year the student submitted their GRFP proposal to control for annual macroeconomic trends and changes to the NSF GRFP conferment process.

Primary Model Extensions

To strengthen the DD design given that award assignment is not random, we estimate two specifications as extensions to the primary model. First, we rely on the binary indicator that measures whether an applicant had publication activity prior to the GRFP to stratify the primary sample before re-estimating equation (1). This more directly controls for baseline variation in research capacity among the sample of students. We estimate the primary model with two subsamples, respectively: students with peer-reviewed publication activity *prior to* applying for the GRFP and those without. Although this is built from the outcome variable, we focus on activity in the pre-period, which is a critical factor considered during the GRFP panel review (Freeman, Chang, & Chiang, 2009). Moreover, it is our strongest observable measure to reduce endogeneity between the two groups in the DD model. For the

[26] $Post_t$ is coded one for annual time periods after the GRFP proposal event ($t > 0$) and zero otherwise.
[27] $Award_i$ equals one for individuals that ever received the GRFP award and zero for honorable mentions.
[28] We do not include subscripts for the vector of X in equation (1) given that some are time varying while others are time-invariant. Additionally, the level of the measure varies from the individual to the institution.
[29] Derived from faculty activity in the life science division at the student's graduate institution.

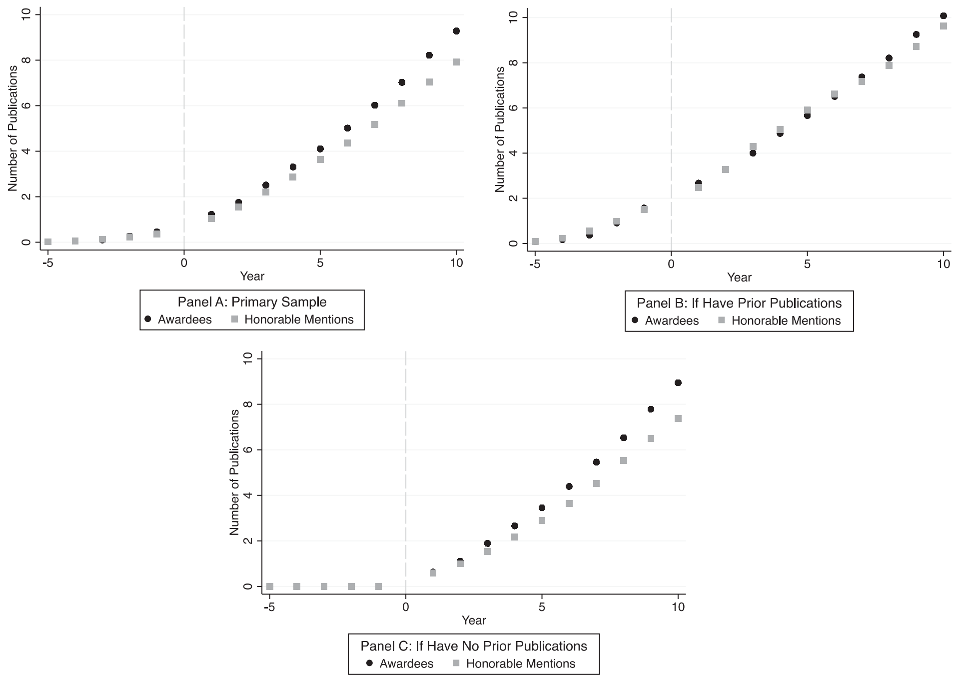


Figure 3. Annual Average Publication Level by Award Status.

sample of students with no prior publications, the two groups of awardees and honorable mentions have the same pre-GRFP trend of zero publications; thus, for this stratification, we technically estimate a single difference.

Second, while we rely on the interaction term—*Award * Post*—to estimate the treatment effect, we also estimate an augmented DD with time trends, which includes an additional interaction with *Year*. This model estimates the relative slope changes pre- and post-GRFP between the awardees and honorable mentions. This estimation helps to inform whether the treated group experiences a change in their publication activity *trend* over time as compared to the control group. At this early-career stage, this model extension offers useful additional insight to the primary model by allowing us to examine whether the award places students on different publication activity trajectories. Operationally, we interact the primary indicators of the DD model with $Year_t$, where t refers to years within each student's respective 16-year annual timeframe. We are primarily interested in the triple interaction term: $Award_i * Post_t * Year_t$. Equation (2) reports this model; the parameter of interest, δ_7, reports linear trend changes in publication activity for awardees pre- and post-award relative to honorable mentions.

$$Y_{it} = \beta_0 + \delta_1 Post_t + \delta_2 Award_i + \delta_3 Year_t + \delta_4 (Award_i * Post_t) + \delta_5 (Post_t * Year_t) + \delta_6 (Year_t * Award_i) + \delta_7 (Award_i * Post_t * Year_t) + \beta_z X + \varepsilon_{it}. \quad (2)$$

Primary Model Diagnostics

Before estimating the DD model of GRFP on total publication levels (equation 1), we map out the average cumulative activity by GRFP acknowledgment status over time (Figure 3). We label the 16-year timeframe on the x-axis to reflect the five years leading up to the GRFP event ($t = 0$) and the 10 years following. Panel A illustrates publication activity for the primary sample, those who ever publish within their standardized 16-year timeframe with complete data. The figure shows publication activity prior to the GRFP overlaps considerably for the primary sample—though there is a slight, albeit very small, increase with the awardees in the years immediately prior to the GRFP. The differences in the number of publications in the year immediately prior to the GRFP is 0.36, economically insignificant and weakly statistically significant. The difference in distributions (Kolmogorov–Smirnov test) is 0.05 and not statistically significant. When considering the post trends, the annual average publication level differs between awardees and honorable mentions with the difference growing over time, substantiating our investigation of trend differences with the augmented DD model (equation 2).

Turning to the stratified samples, the pre-trends are nearly identical for samples with prior publications in panel B. Moreover, the difference in means (t-test) and distributions (Kolmogorov–Smirnov test) are statistically insignificant. When considering the post trends, awardees and honorable mentions share a nearly identical set of averages along the timeframe. For the pre-trends for panel C, they are identical by construction for the sample without prior publication. However, when considering the post trend, awardees exhibit an increase in publication levels after GRFP receipt that exceeds the rate of change for honorable mentions.

RESULTS

Primary Model Results

The primary results estimating equation (1) are presented in Table 2.[30] We estimate the DD with the annual publication levels for the primary sample (562 students). The key coefficient is the interaction of award recipient and post-GRFP timeframe ($Award * Post$). The results indicate that awardee status, as compared to honorable mention status, is associated with approximately two-thirds of an additional publication (0.64) 10 years after the GRFP submission (95 percent confidence interval [CI]: 0.06 to 1.21), on average (column 1). When we stratify the sample by publication activity prior to the GRFP submission, we find it has a moderating effect: those *without* prior publications have a larger effect with an increase of 0.71 total publications (95 percent CI: 0.10 to 1.32), on average (column 3).[31] We find a smaller positive, but insignificant, effect for the sample with prior publications (column 2). As mentioned in Footnote 20, we lack statistical power in this subsample of 2,280 observations (152 students) to detect a smaller effect size; however, panel B of Figure 3 indicates that it is unlikely there is an economically significant difference.

The coefficients for *Award* are relatively small and statistically insignificant across the three samples, which suggests comparability in baseline values between the awardees and honorable mentions. Moreover, the coefficients for *Post* are positive and significant as we expect. The graduate student's gender has a consistent negative association on publication activity such that women publish less than their male counterparts over the full time period. This suggests that there are discrepancies in

[30] We tested for multicollinearity but found no concern with a calculated variance inflation factor (VIF) under 10 for each estimation.

[31] When we stratify the sample by prior publication status it does change the interpretation of the DD for those without prior publications. For this sample, both the treatment and control groups have the same pre-trend of zero publications, making it a single difference estimation; however, the identification is improved.

Table 2. DD model—total publications levels (primary and stratified samples).

	(1) Primary sample	(2) Prior publications	(3) No prior publications
Award * Post	0.636**	0.147	0.709**
	(0.291)	(0.666)	(0.310)
Award	0.031	0.243	0.034
	(0.096)	(0.299)	(0.096)
Post	4.029***	5.428***	3.577***
	(0.205)	(0.524)	(0.202)
Female	−1.265***	−1.846***	−0.937***
	(0.189)	(0.416)	(0.192)
Prior publications	1.654***		
	(0.230)		
Average faculty publications	0.510**	0.770	0.215
	(0.232)	(0.543)	(0.228)
Program ranked in top tercile	−0.322	−1.977*	−0.062
	(0.397)	(1.190)	(0.350)
Two-year GRFP capacity	−0.057**	−0.108*	−0.013
	(0.025)	(0.057)	(0.026)
Federal R&D funding (LN)	0.014	0.000	−0.015
	(0.116)	(0.197)	(0.110)
Carnegie very high research institution	−0.431	0.417	−0.458
	(0.390)	(0.675)	(0.408)
Public institution	0.003	0.648	−0.244
	(0.194)	(0.427)	(0.216)
Constant	−0.037	0.446	0.952
	(1.580)	(1.890)	(1.585)
Observations	8,430	2,280	6,150
Graduate students	562	152	410
Adjusted R^2	0.274	0.336	0.238
Clustered by student	Yes	Yes	Yes
Year applied FE	Yes	Yes	Yes
Field and program size controls	Yes	Yes	Yes

Notes: OLS long-form DD (equation 1); outcome: total publication; clustered standard errors in parentheses; $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Primary sample reflects those who ever publish in the 16-year timeframe with complete data along the full set of measures included in the model. Columns 2 and 3 report the stratified samples of students with prior and without prior publications, respectively. Column 3 reports a single difference model as this sample has no prior publications by design. Federal funding in deflated 2009 values. Additional controls of field and program size are not reported but are available upon request.

research productivity by gender. Further research is needed to examine the relationship between graduate training experiences and discrepancies in productivity between genders. The coefficient for faculty productivity is positive and significant for the primary sample.

Table 3 presents the results estimating the augmented DD with time trends (equation 2) for the primary and stratified samples. The triple interaction, $Post_t * Award_i * Year_t$, informs trend changes in publication activity for awardees pre- and post-GRFP relative to honorable mentions. The coefficient for triple interaction is weakly statistically significant for the primary sample (0.11, 95 percent CI: −0.013 to 0.236) and significant for the subsample *without* prior publications (0.16, 95 percent CI: 0.022 to 0.302). The positive and significant effect for the subsample without prior publications suggests that the relative publication trajectory is increasing for

Table 3. DD with time trends on total publication levels.

	(1) Primary sample	(2) Prior publications	(3) No prior publications
Award * Post * Year	0.111*	0.050	0.162**
	(0.063)	(0.127)	(0.071)
Award	−0.034	0.204	0.034
	(0.120)	(0.306)	(0.096)
Year	0.088***	0.358***	0.000
	(0.012)	(0.029)	(0.000)
Post period	−4.590***	−2.384***	−5.304***
	(0.318)	(0.638)	(0.354)
Award * Year	0.022	0.013	−0.000
	(0.017)	(0.037)	(0.000)
Award * Post	−0.827*	−0.535	−1.154**
	(0.487)	(0.877)	(0.565)
Year * Post	0.685***	0.414***	0.772***
	(0.042)	(0.096)	(0.044)
Constant	−0.300	−0.630	0.952
	(1.580)	(1.885)	(1.586)
Observations	8,430	2,280	6,150
Graduate students	562	152	410
Adjusted R^2	0.470	0.487	0.468
Clustered by student	Yes	Yes	Yes
Year applied FE	Yes	Yes	Yes
PI, program, and uni. controls	Yes	Yes	Yes

Notes: OLS long-form, DD interacted with time trends (equation 2); outcome: total publications; clustered standard errors in parentheses; ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Primary sample reflects those who ever publish in the 16-year timeframe with complete data along the full set of measures included in the model. Columns 2 and 3 report the stratified samples of students with prior and without prior publications, respectively. The same set of controls used in equation (1) are included with this estimation. Uni. refers to university.

awardees after the GRFP. The effect for the subsample with prior publications is smaller and not statistically significant.[32]

Sensitivity Checks

Heterogeneity Analyses

Table 4 presents the results from a variety of subsamples to assess the sensitivity of the primary findings. We first estimate the primary DD model on the extended sample, *inclusive* of those who both ever and never publish in our 16-year timeframe but who have complete data along the set of control variables (697 students). Column 1 of Table 4 presents these results. For the extended sample, the treatment effect is robust with an average increase of 0.66 in total publications 10 years after the GRFP submission for awardees as compared to honorable mentions (95 percent CI: 0.13 to 1.18).

[32] While we lack statistical power to detect an effect on the subsample with prior publications, panel B of Figure 3 indicates it is unlikely there is an economically significant difference between these two groups.

Table 4. Sensitivity analyses to primary model.

	(1) Extended sample	(2) Post-doc placement	(3) Rank 1	(4) Ranks 2 and 3
Award * Post	0.655**	0.958**	0.584*	1.230
	(0.268)	(0.404)	(0.302)	(0.952)
Award	−0.003	0.032	0.083	−0.847**
	(0.092)	(0.152)	(0.101)	(0.426)
Post period	3.180***	4.640***	3.947***	4.344***
	(0.185)	(0.282)	(0.221)	(0.515)
Constant	−1.071	2.084	−0.458	−0.005
	(1.292)	(2.684)	(1.859)	(2.307)
Observations	10,455	4,680	7,035	1,395
Graduate students	697	312	469	93
Adjusted R^2	0.244	0.299	0.286	0.251
Clustered by student	Yes	Yes	Yes	Yes
Year applied FE	Yes	Yes	Yes	Yes
PI, program, and uni. controls	Yes	Yes	Yes	Yes
DD time trend interaction	0.114**	NS	0.121*	NS

Notes: OLS long-form DD; outcome: total publications; clustered standard errors in parentheses; ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. DD time trend interaction denotes coefficient and significance level from triple interaction term estimated from the DD interacted with time trends (equation 2). NS refers to "not statistically significant." Column 1 reports for the extended sample that includes students who did not publish, but who have complete data on the set of controls. This allows for complete case analysis. Columns 2, 3, and 4 are subsamples drawn from the primary sample of 562 students who are either first-placed in a post-doctoral fellowship, graduated from a top ranked graduate program, or graduated from a lower ranked graduate program, respectively. Uni. refers to university.

We also estimate the model on a more restricted sample of those with an initial professional placement in a post-doctorate fellowship position. Given the statistically significant difference between awardees and honorable mentions initially placing in post-doctorate fellowships, there is concern that honorable mentions could be in professional positions with lower expectations to produce peer-reviewed publications. This would then be driving the effect rather than the GRFP award conferred years beforehand. Although this placement occurs *after* the GRFP submission event, we account for whether initial placement affects publishing by reducing the sample to only those with post-doctorate placements (the most prevalent initial job placement). When we reduce the sample, the treatment effect is significant and increases in magnitude to nearly an additional publication (0.96, 95 percent CI: 0.16 to 1.76—column 2).

Lastly, we examine heterogeneity by program rank, which serves as a proxy for access to research resources (Graddy-Reed, Lanahan, & Ross, 2017; Ostriker et al., 2011). Prior studies have found that the value of external R&D funding is inversely related to the resource capabilities of the recipient—particularly when funds are directed to those at an early-stage (Lanahan & Feldman, 2018).[33] This, suggests that value of the award may be larger for students with access to fewer resources. In our primary model, we include a control for the program rank by tercile; however, here we use this measure to stratify the sample and estimate the primary model

[33] While the context differs (early-stage startups, rather than graduate students), both are relatively early in their career stage with heightened levels of uncertainty.

for the sample of students in programs in the top tercile of their field (column 3) and by those in either the second or third terciles (column 4). We combined the lower two ranks due to smaller sample size. For students in the top tercile (Rank 1), the treatment effect is weakly significant with an average of 0.58 additional publications (95 percent CI: −0.01 to 1.18). The coefficient for the lower terciles is larger in magnitude but statistically insignificant. We lack the statistical power to detect the effect in this subsample as it includes only 93 unique graduate students.

Across each of these sensitivity analyses, we also estimated the augmented DD accounting for time trends (equation 2). The triple interaction term is positive and significant for the extended sample. It is weakly significant for the subsample with students in the top tercile (Rank 1).

Additional Sensitivity Analyses

Appendix B provides additional empirical extensions.[34] First, we vary the timeframe for the primary model with the level outcome to provide a linear approximation (Athey & Imbens, 2006). We draw upon varying combinations of two time points over the standardized 16-year window; one prior to applying for the GRFP and the other in the post period. The size of the effect increases as we increase the length of the timeframe.[35]

Second, we alter the functional form of the outcome variable to logarithmic form and re-estimate the primary DD model and main model extensions.[36] The results for equation (1) are comparable with a positive and significant treatment effect for the primary sample and for those without prior publications. The effect size for the primary sample is approximately an 18.6 percent increase in publications over the 10-year timeframe. The results for the augmented DD with time trends (equation (2)) with this functional form, however, are not robust across each stratification.

Third, we estimate the main model with a Poisson rather than ordinary least squares (OLS) distribution. While the latter assumes a normal distribution, the former carries greater methodological concern in the DD context. Namely, the coefficient for *Award * Post* cannot be interpreted when the outcome is nonlinear. Still, for comparison purposes, the estimated marginal effect for the primary interaction is 0.234 (p-value = 0.782). The estimated effect is roughly a third the size of the OLS estimation and statistically insignificant.

Last, we estimate the primary DD model with total publications with alternate clusters of standard errors. One might argue that since publications are increasing over time, standard errors will also increase over time, potentially biasing our standard errors. Thus, we estimate the model clustering first by year and then by using two-way clustering with year and graduate student. Both estimations produce robust results.[37]

Robustness Checks

We estimate two alternative research designs—coarsened exact matching (CEM) and fixed effects modeling—to assess the validity of the primary DD model. As discussed above, the central design threat for this analysis is endogeneity of researcher quality

[34] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[35] Columns 1 through 3 in Table B1 in Appendix B report these results.
[36] Columns 4 through 6 in Table B1 in Appendix B mimic Table 2 but with the outcome in logarithmic form. For all publication measures, prior to adjusting to the logarithm we increase the value by 0.1.
[37] We report the results in Table B5 in Appendix B.

due to non-random assignment of the award. This is illustrated along differences in means among individual, program, and university traits between the two groups as reported in Table 1.[38] These two designs offer unique empirical features to address this concern.

Coarsened Exact Matching

To enhance the primary DD model and further improve identification, we reduce imbalance between the awardees and honorable mentions at the baseline before the GRFP conferment using CEM (Blackwell et al., 2009; Iacus, King, & Porro, 2011). Operationally, we coarsen the sample by drawing upon the full set of pretreatment variables. This reduces the imbalance along this set of measures between the two groups of students such that they are statistically indistinguishable. While one trade-off with this approach is a loss in efficiency due to reduced sample size, the loss in this context is minimal. With the coarsened sample we estimate the primary model with 85.8 percent of the treated group and 83.8 percent of the control group. With the coarsened sample, we then estimate the impact of the award using a balanced subsample of the two groups—awardees and honorable mentions. This is a stronger matching technique than traditional propensity score matching procedures often used. The results are robust and comparable. Additional details on this design and estimation are available in Appendix B.[39]

Fixed Effects Model

Given that the GRFP award is meritocratic, theoretically, inherent research quality may vary between awardees and honorable mentions. While we argue above that the unique features of this program introduce noise into the ultimate funding decision between awardees and honorable mentions and thus help reduce this concern, unaccounted for differences in individual research quality between these two groups can lead to bias. Thus, we additionally estimate a student-level fixed effects model, which offers the benefit of controlling for unobservable time invariant measures, which includes inherent student quality. Equation (3) reports the model and includes the interaction term—*Award * Post*—to estimate the primary treatment effect. In contrast to the DD model (equation 1), the individual-level fixed effect, δ_i, captures time invariant regressors (including, but not limited to *Award* and the vector of controls). By design, this model is more parameterized than the primary DD model (equation 1); we estimate the model on multiple samples to maximize variation. These samples range from the primary sample of 562 students to the full sample of 877 students.[40]

$$Y_{it} = \alpha_1 + \beta_1(Award_i * Post_t) + \tau_t Year\ FE + \delta_i Individual\ FE + \varepsilon_{it}. \tag{3}$$

The results from equation (3) are robust to our main estimates. The differential effect of the award is associated with a statistically significant increase in total

[38] These include measures of activity prior to or during graduate training: female, graduate program rank, program-level measure for GRFP awards and honorable mentions received in years t and t-1 relative to the year the student applied, university governance indicator, and the measure for federally financed HE R&D.
[39] Results are presented in Tables B2 and B3. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[40] While we have complete data from the GRFP program for the full sample of students, we assume students without a publishing record in Scopus have no publications.

publications of 0.77 10 years following the program for the full sample and of 0.64 for the primary sample.[41] The comparability of the interaction term from the fixed effect model (equation 3) and DD model (equation 1) suggest that estimates in the DD model are not driven by unobservable student quality.

ADDITIONAL OUTCOMES

We also consider a range of additional outcome measures including: research leadership—as measured by total first author publications, research quality—which considers the rank of journal outlets, graduate training milestones—as measured by degree completion and time to degree, and professional placement. These additional outcomes help to more richly characterize the effect of external fellowship receipt on both intermediate graduate training experiences and other measures of early-career research production.

First, we estimate the impact of the award on additional bibliometric measures of research leadership and quality. We measure the former by the number of first-authored publications. As mentioned previously, within life sciences first authorship indicates that the graduate student led the research inquiry. Typically, all other authors or laboratory members are listed after with the laboratory director listed last (Dance, 2012; Tscharntke et al., 2007). While the GRFP offers the unique opportunity for the student to fund their own salary, they are still typically conducting their research within their advisor's laboratory and using their advisor's resources. Thus, we would expect the advisor to still be in the last position as the source of funding for the laboratory. The position of first author of the student signals leadership in conducting the research itself.

For the latter measure of research quality, we estimate the number of publications in high-ranking journals. We identify the level of publications in high ranked life science journals—specifically, the top 1, 5, and 10 percent ranks, respectively. Details on these metrics are provided in Appendix B.[42]

We use equation (1), the DD model, to estimate the impact of the award on these alternative bibliometric measures. Table 5 presents the key results for these four additional outcomes. We find positive and statistically significant effects for each outcome, but the coefficient is smaller in magnitude compared to the primary result. These range from 0.36 for first-authored publications to 0.09 for publications in the top 1 percent of journals. The triple interaction from the augmented DD with time trends (equation 2) is positive and significant for first-authored publications. However, we find no such difference in the relative rate between groups for the outcome measures of journal rank. Figures of these trends are presented in Appendix B.[43]

Finally, we examine the award's impact on non-publication-related outcomes such as degree completion and professional placement as a way of examining more intermediate graduate outcomes that also influence more distal publication production.[44] Because these outcomes are single instances, we use the wide form data rather than long form to estimate the relationship. Of the 877 graduate students from the full sample, 731 completed a PhD, 77 completed another type of degree,

[41] We report the results in Table B4 in Appendix B. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[42] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[43] Refer to Figures B1 and B2.
[44] Refer to Table B7 in Appendix B.

Table 5. Additional bibliometric outcomes.

	(1) First author pubs	(2) Top 1% journal pubs	(3) Top 5% journal pubs	(4) Top 10% journal pubs
Award * Post	0.363**	0.087**	0.170**	0.188**
	(0.146)	(0.041)	(0.074)	(0.086)
Award	0.036	−0.002	−0.015	−0.029
	(0.048)	(0.012)	(0.025)	(0.028)
Post period	1.793***	0.139***	0.397***	0.587***
	(0.099)	(0.021)	(0.045)	(0.057)
Constant	−1.077	−0.134	−0.430**	−0.271
	(0.677)	(0.123)	(0.211)	(0.262)
Observations	8,430	8,430	8,430	8,430
Graduate students	562	562	562	562
Adjusted R^2	0.241	0.0804	0.146	0.153
Clustered by student	Yes	Yes	Yes	Yes
Year applied FE	Yes	Yes	Yes	Yes
PI, program, and uni. controls	Yes	Yes	Yes	Yes
DD time trend interaction	0.073**	NS	NS	NS

Notes: OLS long-form DD; models are estimated with primary sample; outcome specified in column title. Clustered standard errors are in parentheses. Publications measures are in level form. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. DD time trend interaction denotes coefficient and significance level from triple interaction term estimated from the DD interacted with time trends (equation 2); NS refers to "not statistically significant." Uni. refers to university.

and 69 were undetermined or incomplete. We estimate that receiving the award is associated with a small increase in the probability of completing a degree; however, there is no statistically significant impact on time-to-degree after applying for the GRFP. As for professional placement, receiving the award is associated with an increased likelihood of placing initially in a post-doctorate fellowship. Based on prior studies finding evidence that early-career academic achievements exhibit a "Matthew Effect"[45] (Bol, de Vaan, & van de Rijt, 2018), it is plausible that the GRFP award offers important early-career advantages for future funding, research, and professional placement. Future work could examine longer-term professional implications. Regarding the last outcome, we do not find a statistically significant impact on placing in a tenure-track position.

RESEARCH AND POLICY IMPLICATIONS

Our study contributes to the debates over federal R&D investment and higher education training by examining the impacts of federal investment on graduate students. We find a consistent positive, yet small, association between being a GRFP awardee—relative to an honorable mention—on research productivity in the 10 years following GRFP application. Overall, awardee status, which comes with a $91,000 grant over three years, is associated with approximately two-thirds of a single additional publication 10 years following the program. This corresponds to an 11.8 percent increase in research productivity. In terms of total output volume, the

[45] The "Matthew Effect" refers to the phenomenon of accumulated advantage, where the "rich get richer and the poor get poorer."

effect sizes from GRFP funding are relatively small given the extended time period over which we observe students.

Moreover, we find evidence that these results are driven by students *without* a research record prior to applying for the funding. In contrast, among those with a prior research record, GRFP funding does not appear to differentiate awardees from honorable mentions in terms of research productivity. The results from our stratified subsamples suggest that the value of external R&D is larger for those without an established research record. This offers important implications for the GRFP and NSF more broadly when considering the broader mission aims of their programs. While the public program espouses research excellence, the federal mission agency has placed increasing precedence on broader impacts.[46] With respect to this aim, the larger, positive results for students without a prior research record provide evidence of some level of programmatic success for the GRFP by increasing productivity for a broader scientific research community.

Placing this study in context to prior scholarship, our findings are modest and consistent with previous research on R&D investments for life science researchers at later career stages. Jacob and Lefgren find that receiving an NIH R01 grant[47] leads to a publication increase of 7 percent (2011a) and a post-doctorate training grant yields an increase of 20 percent productivity (2011b); for both studies these results are within five years of the award. While these studies are focused on researchers later in their career, the authors (2011a) attribute the low effect size of an NIH grant to researchers seeking other external funding sources. Although this could also be the case with this study, it is less likely as there are few opportunities for graduate students to receive external funding so early in their careers. Moreover, there are very few grants at the level of funding equivalent to the GRFP. We turned to the NIH pre-doctoral training programs to assess whether the sample of students—either awardee or honorable mention—additionally secured one of the competitive NIH grants given that they also support life science research.[48] Among the primary sample, only one individual secured a pre-candidacy training grant. Thus, we are not concerned that this source of funding is confounding the results. The NSF DDRIG is an alternate, well-known award for candidates; however, this award provides only one-year of doctoral research funding with a grant amount in 2017 of approximately $20,000 for direct costs with no salary allowance for the student.[49] It seems unlikely this funding source would serve as a substitute for the GRFP.

Looking beyond the commonly used metric of total publications, there are other outcomes to assess with regard to scientific contribution and educational outcomes of graduate training and placement. We find support that the grant yields a modest increase in terms of research leadership and quality as measured by first author publications and journal rank, respectively. When considering some completion and placement outcomes, we find a small, positive effect of award status for completing the graduate degree and a larger, yet modest effect with securing a post-doctorate position as a first placement. While this is an initial effort to expand the breadth of outcome measures, more work remains.

[46] NSF added *broader impacts* as a second merit review criteria in 1997 (in addition to *intellectual merits*). See https://www.nsf.gov/pubs/2007/nsf07046/nsf07046.jsp. As a side, although this administrative change took place during the study's timeframe, we do not find empirical evidence that publication activity differs among recipients prior to or following this change.
[47] The Research Project Grant (R01) is intended for more senior personnel. The average age for this award is 42. (See https://www.nia.nih.gov/research/blog/2015/04/r01-teams-and-grantee-age-trends-grant-funding.)
[48] This includes the following NIH programs: F30, F31, F99, R36, R90, T32, and T90 programs.
[49] See https://nsf.gov/pubs/2017/nsf17506/nsf17506.htm.

The small, positive effect across all of these outcomes has important economic implications given the size of the federal program and level of investment. The modest impacts from this study raise the question of whether such a large federal investment is warranted. With that said, the broader literature on academic R&D indicates that federal funding *crowds-in* other sources (Lanahan, Graddy-Reed, & Feldman, 2016). As the primary funder of academic R&D, if federal funding were cut, the academic research enterprise would likely suffer. Nonprofit and industry funders are likely unable to substitute for federal funding, especially for early-stage graduate students.

Moreover, adjustments to federal R&D support also likely impact universities. Given that universities are resource-constrained, students who obtain the GRFP funding allow *other* students access to internal funding. If GRFP funding decreases, universities may reallocate internal funding away from students on the margin. This is particularly salient given that there is growing evidence to suggest that this disproportionately impacts minority groups (e.g., Buffington et al., 2016; Fealing, Lai, & Myers, 2015; Moss-Racusin et al., 2012). The push to expand and diversify the STEM pipeline is thus reliant on such federal funding. Further investigation is needed to assess the institutional impacts of adjusting such a substantial source for graduate funding.

Federal support for academic R&D serves a critical role due to its competitive importance and the societal benefits of basic research.[50] The flexibility granted to students by this program may be viewed by students as an asset, yet the modest results indicate that there are trade-offs. Future work should examine how on-the-ground training experiences differ between awardees and honorable mentions to better understand what is driving the results. Within the larger debate over federal R&D funding, policymakers need to consider the implications for students and their training, along with the ramifications on the scientific research community at large.

ALEXANDRA GRADDY-REED is an Assistant Professor at the Price School of Public Policy, University of Southern California, 650 Childs Way, Los Angeles, CA 90089 (e-mail: graddyre@price.usc.edu).

LAUREN LANAHAN is an Assistant Professor at Lundquist College of Business, the University of Oregon, 1208 University Street, Eugene, OR 97403 (e-mail: llanahan@uoregon.edu).

NICOLE M. V. ROSS is a Doctoral Candidate at the Department of Public Policy, University of North Carolina at Chapel Hill, Abernethy Hall, C.B. 3435, Chapel Hill, NC 27599 (e-mail: nmvross@live.unc.edu).

ACKNOWLEDGMENTS

This research was funded in part by the National Science Foundation Science of Science and Innovation Policy Program (NSF 1548288) and the Lundquist College of Business Summer Research Grant. We thank Jonathan Eyer, Jeremy Moulton, Josh Graff Zivin, Rodney Ramcharan, Jade V. M. Jenkins, and Matt Ross for their comments on earlier versions of this analysis. This paper also benefited from feedback at the 2016 Association for Public Policy and Management annual meeting and the USC Sol Price junior faculty research seminar. We are greatly indebted to our research assistants at the University of Oregon for their invaluable

[50] Vannevar Bush's influential report, *Science, The Endless Frontier* (1945) served as the platform for substantiating the creation of federal mission agencies—including NSF. (See https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm.)

contributions with building the database; specifically, we would like to thank Georgiy Sichinava, Soro Soukpafolo, and Zachary Trudo. In addition, we would like to thank Emily Plews and Jung-Hyun Kwon for additional research assistance. We would also like to thank Austin McLean for his assistance with the Proquest data and John McDonald for his assistance with the Scopus data. All errors are our own. Alexandra Graddy-Reed and Lauren Lanahan contributed equally.

REFERENCES

Adams, J., & Griliches, Z. (1998). Measure science: An exploration. Proceedings of the National Academy of Sciences, 93, 12664–12670.

Arora, A., & Gambardella, A. (2005). The impact of NSF support for basic research in economics. Annales d'Economie et de Statistique, July/December, 91–117.

Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. Econometrica, 74, 431–497.

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: Evidence from the academic life sciences. RAND Journal of Economics, 42, 527–554.

Blackwell, M., Iacus, S. M., King, G., & Porro, G. (2009). CEM: Coarsened exact matching in Stata. Stata Journal, 9, 524–546.

Blume-Kohout, M. E., Kumar, K. B., & Sood, N. (2015). University R&D funding strategies in a changing federal funding environment. Science and Public Policy, 42, 355–368.

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. Proceedings of the National Academy of Sciences, 19, 4887–4890.

Buffington, C., Cerf, B., Jones, C., & Weinberg, B. A. (2016). STEM training and early career outcomes of female and male graduate students: Evidence from UMETRICS data linked to the 2010 Census. American Economic Review, 106, 333–338.

Bush, V. (1945). Science: The endless frontier. Transactions of the Kansas Academy of Science (1903), 48, 231–264.

Campbell, C. M., & O'Meara, K. (2013). Faculty agency: Departmental contexts that matter in faculty careers. Research in Higher Education, 55, 49–74.

Conti, A., & Liu, C. C. (2015). Bringing the lab back in: Personnel composition and scientific output at the MIT Department of Biology. Research Policy, 44, 1633–1644.

Dance, A. (2012). Authorship: Who's on first? Nature, 489, 591–593.

Fealing, K. H., Lai, Y., & Myers, S. L. Jr (2015). Pathways vs. pipelines to broadening participation in the STEM workforce. Journal of Women and Minorities in Science and Engineering, 21, 271–293.

Feldman, M., & Lowe, N. (2015). Triangulating regional economies: Realizing the promise of digital data. Research Policy, 44, 1785–1793.

Freeman, R. B., Chang, T., & Chiang, H. (2009). Supporting "the best and brightest" in science and engineering: NSF graduate research fellowships. In R. Freeman & D. Goroff (Eds.), Science and engineering careers in the United States: An analysis of markets and employment (pp. 19–57). Chicago, IL: University of Chicago Press.

Graddy-Reed, A., Lanahan, L., & Ross, N. (2017). Influences of academic institutional factors on R&D funding for graduate students. Science and Public Policy, 44, 834–854.

Iacus, S., King, G., & Porro, G. (2009). CEM: software for coarsened exact matching. Journal of statistical Software, 30(13), 1–27.

Iacus, S., King, G., & Porro, G. (2011). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20, 1–24.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political analysis, 20(1), 1–24.

Jacob, B. A., & Lefgren, L. (2011a). The impact of research grant funding on scientific productivity. Journal of Public Economics, 95, 1168–1177.

Jacob, B. A., & Lefgren, L. (2011b). The impact of NIH postdoctoral training grants on scientific productivity. Research Policy, 40, 864–874.

Jones, B. F., Wutchy, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. Science, 322, 1259–1262.

Lanahan, L., & Feldman, M. P. (2018). Approximating exogenous variation in R&D: Evidence from the Kentucky and North Carolina SBIR State Match programs. Review of Economics and Statistics. Retrieved from https://doi.org/10.1162/REST_a_00681. Last accessed July 23, 2018

Lanahan, L., Graddy-Reed, A., & Feldman, M. P. (2016). The domino effects of federal research funding. PLoS One, 11, e0157325.

Mansfield, E. (1995). Academic research underlying industrial innovations: Sources, characteristics, and financing. Review of Economics and Statistics, 70, 55–65.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. Proceedings of the National Academy of Sciences, 109, 16474–16479.

National Institutes of Health (NIH). (2018). *The F30 program*. Retrieved May 17, 208, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-084.html#

National Research Council (NRC). (2005). Graduate fellowships and traineeship programs, assessment of NIH minority research and training programs: Phase 3. Washington, DC: National Academies Press.

National Science Board, Science & Engineering Indicators. (2018). *Academic research and development*. Retrieved May 17, 2018, from https://www.nsf.gov/statistics/2018/nsb20181/report/sections/academic-research-and-development/highlights

National Science Foundation (NSF) GRFP. (2018). *About*. Retrieved from https://www.nsfgrfp.org/general_resources/about. The last access date is: July 23, 2018.

O'Meara, K., Jaegar, A., Eliason, J., Grantham, A., Cowdery, K., Mitchall, A., & Zhang, K. (2014). By design: How departments influence graduate student agency in career advancement. International Journal of Doctoral Studies, 9, 155–179.

Ostriker, J., Holland, P. W., Kuh, C. V., & Voytuk, J. A. (2011). A data-based assessment of research-doctorate programs in the United States. Washington, DC: National Academies Press. ISBN: 978-0-309-16030-8.

Sauermann, H., & Roach, M. (2012). Science PhD career preferences: Levels, changes, and advisor encouragement. PloS One, 7, e36307.

Stephan, P. E. (2012). How economics shapes science. Cambridge, MA: Harvard University Press.

Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., & Krauss, J. (2007). Author sequence and credit for contributions in multi-authored publications. PLoS Biology, 5, e18.

APPENDIX A: GRFP SAMPLING AND DATA CONSTRUCTION

In the third section of the main paper, we detail the GRFP Sampling (Step One) and Data Construction (Step Two). Here, we provide additional detail on both steps. Regarding the presentation of the data, we present the material in a manner that reflects the sequence of steps as depicted by Figures 1 and 2 in the paper. Figure 1 reflects Step One; Figure 2 reflects Step Two.

Step One: GRFP Sampling

We turn to the NSF GRFP program to define the sample of graduate students. Each unique proposal record[51] contains the following information: student name, baccalaureate institution, field of study, proposed graduate institution, current institution, year of proposal submission, and status as awardee or honorable mention. To be clear, only applications recognized as an honorable mention or granted an award are listed in the database.

Define GRFP Timeframe

First, we restricted the GRFP data by year of application—specifically 1995 to 2005. Though published lists of award winners are available starting in 1952, lists of honorable mentions are only available from 1995. A lower bound of 1995 ensures that we sample records for both honorable mentions and awardees, while an upper bound of 2005 allows for a 10-year follow-on timeframe to trace post GRFP activity for the most recent cohort (data construction for this project began in 2016). This timeframe produced a sample of 25,317 GRFP proposal-year observations with 10,124 listed as awardees and 15,193 listed as honorable mention observations. Table A1 reports the field distribution for the full GRFP sample over this timeframe.

Restrict Sample to Life Sciences Division

Second, we focused on proposals submitted under the broad division of life sciences—formally named at NSF as the Directorate of Biological Sciences. The use of one division improves the data collection process given the challenges associated with building a large individual-level panel data set where we triangulate information across multiple external sources (this is elaborated in Step Two). This approach follows with the existing literature focused on R&D investments in the life sciences (e.g., Azoulay, Graff Zivin, & Manso, 2011; Jacob & Lefgren, 2011a; Jacob & Lefgren, 2011b).

Within the life sciences we restricted the sample to the top three most active sub-fields of GRFP activity. These are: Ecology & Evolutionary Biology (Ecology); Biochemistry, Biophysics, & Structural Biology (Biochemistry); and Biology, Integrated Biology, Integrated Biomedical Sciences, & Kinesiology (Biology). Each sub-field has a similar distribution of awardees and honorable mentions across the life sciences and full samples. These three sub-fields include 3,896 proposal-year observations. We report descriptive statistics of the distributions by sub-field in Table A2 panel A.

[51] The level of data is available at the proposal—not student—level. Students are eligible to apply more than once so long as they have not received the award and have not exceeded 12 months of graduate training. Based on the proposal record, we are able to identify re-applicants.

Table A1. Distribution of GRFP sample by sub-field (1995 to 2005).

NRC program field	Freq.	Percent
Life Sciences		
Animal Sciences	703	2.78
Biochemistry, Biophysics, and Structural Biology	1476	5.83
Biology/Integrated Bio/ Integrated Biomedical Sciences	888	3.51
Cell and Developmental Biology	557	2.20
Ecology and Evolutionary Biology	1532	6.05
Entomology	73	0.29
Forestry and Forest Sciences	25	0.10
Genetics and Genomics	492	1.94
Immunology and Infectious Disease	117	0.46
Microbiology	240	0.95
Neuroscience and Neurobiology	718	2.84
Nutrition	10	0.04
Pharmacology, Toxicology and Environmental Health	45	0.18
Physiology	109	0.43
Plant Sciences	201	0.79
Physical Sciences and Mathematics		
Astrophysics and Astronomy	718	2.84
Chemistry	1415	5.59
Computer Sciences	1011	3.99
Earth Sciences	737	2.91
Mathematics	790	3.12
Oceanography and Atmospheric Sciences and Meteorology	299	1.18
Physics	852	3.37
Statistics and Probability	116	0.46
Engineering		
Aerospace Engineering	324	1.28
Biomedical Engineering and Bioengineering	868	3.43
Chemical Engineering	970	3.83
Civil and Environmental Engineering	770	3.04
Computer Engineering	542	2.14
Electrical and Computer Engineering	1180	4.66
Engineering Science and Materials	168	0.66
Materials Science and Engineering	682	2.69
Mechanical Engineering	1,092	4.31
Operations Research, Systems Engineering, and Industrial Engineering	83	0.33
Social and Behavioral Sciences		
Agricultural and Resource Economics	1	0.00
Anthropology	904	3.57
Economics	637	2.52
Geography	112	0.44
Linguistics	429	1.69
Political Science	707	2.79
Psychology	1674	6.61
Public Affairs, Public Policy and Public Administration	59	0.23
Sociology	433	1.71
Arts and Humanities		
History	112	0.44
Unable to classify	446	1.76
Total	25,317	100.00

Table A2. Comparison of sub-fields to GRFP sample.

Panel A	Ecology	Biochemistry	Biology	Life sciences division
Proposal-year observations	1,532	1,476	888	7,186
Share of full GRFP sample (25,317)	6.05	5.83	3.51	28.38
Share of life sciences sample	21.32	20.54	12.36	
Awardees (10,124)	601	527	299	2,608
Share of own sample (39.99 percent)	39.23	35.70	33.67	36.29
Honorable mentions (15,193)	931	949	589	4,578
Share of own sample (60.01 percent)	60.77	64.30	66.33	63.71
Panel B	**Ecology**	**Biochemistry**	**Biology**	**Sample**
Proposal-year observations	1,378	1,328	801	3,507
Share of own sample lost to duplicates	10.05	10.03	9.80	9.98
Awardees	540	469	274	1,283
Share of own sample	39.19	35.32	34.21	36.58
Honorable mentions	838	859	527	2,224
Share of own sample	60.81	64.68	65.79	63.42

Notes: Panel A reflects the sample distribution with the timeframe 1995 to 2005 as the only restriction. Number and relative share (percent) of full GRFP sample are listed in parentheses. Panel B further reflects the restriction of the name disambiguation (discussed in the following section).

Remove Common Last Names

Third, we restricted the sample by removing observations with common last names given the challenges of name disambiguation in tracking individuals over time. Pragmatically, we dropped observations if their last name has more than five duplicates in the data set. This eases the process of verifying whether the correct individual is being tracked over time and reduces the presence of false positives in the analysis. Of note, the presence of measurement error in the dependent variable would reduce the statistical power of the analysis.

By limiting the sample of individuals to those with less common names, we reduce this concern. Further, we do not expect that individuals with more common names would behave or be treated differently in their research careers, so this additional restriction should not bias the results. This restriction removed 389 observations leaving 3,507. As shown in panel B of Table A2, the name disambiguation does not affect one sub-field more than another, with each losing roughly 10 percent to common names. Further, the share of awardees and honorable mentions remains nearly identical to the previous distributions and to the total life sciences sample.

One possible concern with this restriction, however, is whether common names are disproportionately drawn from a specific race or ethnicity. Diversity and socioeconomic factors may account for research production differences stemming from variation in access to educational training. While we do not have access to race and ethnicity data for the sample of GRFP recipients, we have reviewed the distribution of common names across the *entire* NSF GRFP program. Table A3 lists the 30 most common names with 30 or more occurrences in the GRFP database. One theme that does emerge from this analysis is that 21.6 percent of the most common names are of Asian heritage (denoted by italics in Table A3). This level exceeds the national Asian-American population in 2000, the middle year of our

Table A3. Common names in GRFP database.

Smith (116); Lee (105); Johnson (104); Brown (86); Miller (76); *Chen* (68); Williams (64); Anderson (63); Davis (63); Jones (62); *Wang* (58); Thompson (48); Wilson (48); *Chang* (45); Moore (44); Green (41); Thomas (39); *Kim* (36); Young (36); *Liu* (35); Martin (34); Jackson (33); Nelson (33); *Wong* (33); Lewis (32) *Lin* (30); Roberts (30); Allen (29); Evans (29); *Yang* (29)

Notes: Derived from NSF GRFP complete award database (1995 to 2005). Italicized names denote names with predominantly Asian-American heritage. The number in parentheses reports the number of occurrences with duplicate last names.

Table A4. Full sample distribution by year.

	Biochemistry		Ecology		Integrated biology	
Year applied	Awardees	Honorable mentions	Awardees	Honorable mentions	Awardees	Honorable mentions
1995	18	16	1	4	7	9
1996	12	16	11	12	6	12
1997	16	11	16	14	21	15
1998	15	16	16	9	10	9
1999	12	10	16	12	18	12
2000	13	9	17	9	21	12
2001	11	12	21	13	16	9
2002	17	10	10	17	12	16
2003	12	19	14	14	10	16
2004	9	15	13	17	17	19
2005	15	12	14	19	12	13
Average	13.64	13.27	13.55	12.73	13.64	12.91
Total	150	146	149	140	150	142

Notes: Values reflect full sample of 877 individuals.

sample (4.2 percent).[52] This suggests that this approach of removing more common names disproportionately removed Asian-American graduate students. We recognize this as a tradeoff; however, we argue that the increased likelihood of identifying false positives through the multiple triangulation efforts presented in Step Two leads to inaccurate data. We view this to be a greater concern.

Random Sampling

Finally, we randomly sampled 150 awardees and 150 honorable mentions (two groups), which were selected without replacement, from each sub-field (3). This yields a sample of 900 proposal observations ($2 \times 3 \times 150 = 900$). Twenty-three students appeared twice in the data set. In instances with duplicate student records, we kept the most recent record. This yielded a total of 877 unique students. Table A4 shows the distribution of the sample across the timeframe of 1995 to 2005 by sub-field and award status.

[52] 2000 U.S. Census Statistics.

Step Two: Data Construction—Matching to Third Party Data Sources

Once we selected the sample of GRFP awardees and honorable mentions, we gathered annual metrics on graduate training activity; professional placement following the completion of graduate training; and research production outputs in terms of peer-reviewed publication activity. We relied on the following data sources: ProQuest; online web searches; Scopus; National Research Council's (NRC) decennial survey of U.S. research doctoral programs; the NSF Higher Education Research and Development (HERD) survey; and the National Center for Education Statistics (NCES) Integrated Postsecondary Education Data System (IPEDS).

ProQuest—Graduate Training Information

NSF publishes the student's *current institution* and *proposed graduate institution*; however, this does not confirm that the student actually enrolled in the listed institution or completed the degree.[53] Thus, we relied on ProQuest, which serves as the largest repository of online theses and dissertations[54] to identify the dissertation (or Master's) thesis for the sample of students. Although ProQuest does not share user agreements with every academic institution, they purport that most research-active institutions make it a requirement for students to upload their research documents prior to receiving their final degree. This accounts for the ProQuest's extensive scale and scope.

To match the sample of individuals to ProQuest's database, our team of research assistants conducted an advanced Author-AU search by name. In addition, we relied on the listed institution and sub-field of study from the GRFP database to assist in identifying the individual. By drawing upon a random sample of life science graduate students in a designated timeframe, this also served as a useful boundary condition for delimiting the set of potential matches. In instances where there were discrepancies or multiple results, we relied on additional online searches to determine the correct match. We identified the ProQuest Dissertation ID match for 711 individuals (81 percent match rate on the eligible sample). We then sent ProQuest the list of Dissertation IDs for them to scrape their database. This yielded a match for 697 individuals (79.4 percent match rate on the eligible sample). We received the following information for each match: Subjects; Keywords; Author; Title; Date, School Code; Advisors, School; Department.

Although ProQuest offers the most complete set of degree completion records, some schools (including Stanford) have opted not to share this information with ProQuest. Thus, for some individuals, we conducted supplementary online searches to acquire and confirm this information. Most of the information was confirmed in the student's advisor's webpage.[55] We report the match rate for this additional online effort in the subsequent section.

Online Searches—Professional Placement Information

After supplementing the data with information from ProQuest, we were better equipped to collect data on post-graduate professional placement. We focus on the

[53] We find that 73 percent of the students matriculated into the program and university listed in the GRFP data set. In most cases, students switched between peer-ranked institutions (i.e., from Cal Tech to MIT or Harvard to Yale). Among those who switched only 36 students changed institutions with notable differences in rank (i.e., proposed = University of New Mexico and graduate = Princeton).
[54] See http://www.proquest.com/about/who-we-are.html.
[55] Most life science faculty with research laboratories list their current laboratory members and alumni. For alumni, they state the year they graduated and the degree conferred.

Table A5. Degree completion and first professional placement classifications (FOC and SOC).

FOC	1. Completed degree (Yes/No) 2. Degree level (masters/doctorate/neither) 3. Dual professional placement (sole/consecutive/concurrent) 4. Type of organization of first professional placement (academic – higher education/public/private/non-profit/other) 5. Field of employment (life sciences/ (non-life) science/non-science) 6. Placement information confirmed (Yes/No)
SOC	1. Academic position (post-doc/tenure-track faculty, non-tenure-track researcher/non-tenure-track instructor/professional schooling/graduate student/other, administrative/non-academic) 2. Non-academic position (post-doc/researcher/start-up, founder/other, administrative, academic)

Notes: FOC, first-order conditions; SOC, second-order conditions. The information listed in parentheses denotes the categorical options for each field.

student's *first* job placement after graduate school. We define first placement as the first employment position a student holds for at least one year within five years of graduating. Of note, this excluded instances where students held temporary summer or semester-long positions immediately following their degree completion. The professional placement verification process went through seven rounds of searching, coding, vetting, and cleaning. We concluded the review when additional verification re-confirmed the data.

We sought to collect the following information on the first professional placement: (1) position/employment title; (2) name of employment organization; and (3) name of department, lab, or advisor (if available). We also relied on web searches to supply supplementary individual-specific information and, when possible, to either corroborate or complete graduate school-related data (especially for those individuals for whom ProQuest data was not available).[56] In addition to placement, we also sought information on the following fields: (4) gender; (5) graduate school institution; and (6) degree status, type, and completion year.

To fill in these fields, we relied primarily on the Internet, searching by first name, last name, and institutional affiliation (baccalaureate and graduate). We drew heavily upon LinkedIn profiles, personal websites, academic department/laboratory websites, organization mentions, media mentions, etc. For each individual, we retained a record of HTML links to all relevant information. We also downloaded all CVs and resumes when available (133 out of 877). Whenever web searches yielded information incongruous with ProQuest data for the student, we reviewed the ProQuest searches, working iteratively to adjust graduate education or placement information, as necessary. This only happened for a handful of cases.

We used information from the online search in conjunction with ProQuest data to identify degree completion and the first professional placement classification. The list of first order and second order conditions (FOCs and SOCs, respectively) and respective variable options are contained in Table A5 below.

We obtained degree completion and first professional placement information for 800 unique students for a match rate of 91.2 percent.

[56] Of note, ProQuest does not have full data for all schools after 2010. Most of the students in the sample complete their graduate training before this point; nevertheless, we rely on additional online review to address this limitation with the ProQuest data.

Scopus—Research Productivity

We considered a series of bibliometric databases to identify peer-reviewed publications for the sample of individuals—specifically PubMed, Web of Science, and Scopus. After consulting with representatives from each source, we discovered that Scopus includes a more comprehensive set of publication data for our timeframe of interest. This database purports to serve as the "largest abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings."[57] Thirty-two percent of their publications are from health sciences, 29 percent from physical sciences, 24 percent from social sciences, and 15 percent from life sciences.[58] Scopus provides proprietary data on the reference of publications and details of authors, journal, edition, and cumulative citations.

To match the list of individuals (877) to Scopus, we drew upon the set of information already retrieved. Most importantly, we relied on iterations of the author's name from the NSF GRFP, ProQuest, and online sources. We identified 77 name changes in our efforts. This is an important consideration as some of the individuals in the sample, especially females, are likely to experience life events—notably marriage—during this timeframe.

We relied on a python script to systematically scrape records with plausible matches. Then, we manually vetted the accuracy of this scrape by triangulating with our other data on the student's graduate institution (ProQuest and online searches), research focus (ProQuest, as listed by keywords), institutional affiliations—baccalaureate and graduate institutions and professional placement (NSF GRFP, ProQuest, and online searches), graduate advisor co-author (ProQuest), and publications listed on LinkedIn and CV (online searches). While there are currently extensive efforts to present publication data by ORC-ID,[59] in Scopus all authors do not necessarily have a unique ID. We manually identified 971 matches, representing 699 unique individuals.

With this set of 971 matches, we ran a subsequent python script to scrape the publication activity. This returned information for 16,174 publication records. Given that we are primarily interested in peer-reviewed publication activity, we restricted the sample by publication type. This removed a significant number of publications that included conference proceedings, editorial reviews, books, and reports. Moreover, we dropped duplicate publication records.

Given the scope of our research design, we also removed publication activity that falls outside the standardized 16-year timeframe for each student. This timeframe is set by the year the student applied for the GRFP and includes the five years leading up to the GRFP acknowledgment, the year the student applied for the GRFP, and the 10 years following. After cleaning the data, we identified 5,900 unique peer-reviewed publication records for 677 individuals.

NRC, HERD, and IPEDS—Higher Education Institutional Controls

To augment the student-level data, we drew upon a series of higher education variables to account for organizational factors that impact graduate training. Specifically, we matched program, division, and university-level variables from the NRC, HERD, and integrated postsecondary education data system (IPDES) data sources. Among the 877 in the original data set, we identified the program (defined by

[57] See https://www.elsevier.com/solutions/scopus.
[58] See Elsevier: https://www.elsevier.com/solutions/scopus/content.
[59] See https://www.elsevier.com/authors-update/story/innovation-in-publishing/new-orcid-id-aims-to-resolve-authorship-confusion.

university-field[60]) for 90.1 percent of the sample. Those without this identification either did not have complete information or completed graduate training outside of the United States.

First, we pulled data from the NRC for annual average level of faculty publication, program rank, and size indicators. From 2005 to 2006, the NRC surveyed over 5,000 doctoral programs that span 62 academic fields from 212 universities (Ostriker et al., 2011). This survey is cross-sectional and reports data from program activity between 2000 and 2006. Although the data are available at the program-level, we computed statistics based on the division (life science) to improve the match rate. The match rate between this data source and the sample of graduate students with program identification is 94.6 percent.

Second, we pulled panel data from the HERD survey to include measures of federal research funding. This data point is only available at the division-level during the timeframe of interest (life sciences). The match rate between this data source and the full sample of students with program identification is 95.6 percent.

Third, for university-level variables, we drew upon the individual's graduate institution. We pulled data from the NCES IPEDS database. This provides data on Carnegie classifications—corresponding to research rankings—and indicators of institutional governance (public/private). The match rate for the full sample of graduate students with university identification is 96.2 percent. Across these three sources, we have complete data for 765 student-PIs (87.2 percent).

Samples

Primary Sample—With Complete Data (562 students)

The primary sample comprises research-active graduate students—specifically, students with any publication activity over their respective standardized 16-year timeframe. This reduces the set of GRFP students from 877 to 677. Among this set, we have data from the complete set of additional external sources for 562 (83 percent of the research-active sample)—296 are awardees and 266 are honorable mentions. We use the primary sample for the primary difference-in-differences (DD) model estimations. Moreover, we draw from this sample for the stratified estimations discussed in the fourth section of the article.

Extended Sample—With Complete Data (697 students)

This includes the primary sample in addition to the set of students *without* any publication activity over their standardized 16-year timeframe, yet complete data from the additional set of sources. We use this sample to estimate the primary model (equation 1) and fixed effects model (equation 3). The results from the former are presented in Table 4, column 1; the results from the latter are presented in Appendix B, Table B4, column 2.

Ever Publish Sample—Without Complete Data (677 students)

This includes the set of students with any publication activity over their standardized 16-year timeframe irrespective of if they have complete data from the additional set of sources. We use this sample for the fixed effects estimation given that the

[60] Field is analogous to graduate program department.

individual fixed effect accounts for time-invariant factors. The results from the fixed effects model using this sample are presented in Appendix B, Table B4, column 3.

Full Sample—Without Complete Data (877 students)

This includes the full sample of students from the initial GRFP sampling approach (refer to Step One). We have complete information from the GRFP program for this set of students. For the set of students that we were unable to identify in Scopus, we assume that they do not have any publication activity over their standardized 16-year timeframe. We use this sample for the fixed effects estimation given that the individual fixed effect accounts for time-invariant factors. This is discussed in the fifth section of the article. The results from the fixed effects model using this sample are presented in Appendix B, Table B4, column 4.

APPENDIX B: ALTERNATIVE OUTCOMES AND METHODS

Sensitivity Analyses

Adjusted Timeframe

First, to provide a linear approximation we estimate the DD using only two time periods. Columns 1 and 2 in Table B1 present these results. The full sample ranges from year -5 through 10. However, column 1 restricts the sample to year -1 and year 5 (the year just prior to the GRFP and five years after), while column 2 restricts the sample to year -1 and year 10. For the former, the interaction is of a smaller magnitude but statistically insignificant. For the latter with the more extreme time contrast, the interaction is larger and statistically significant (1.3 publications). These provide a reference to a traditional two-period DD.

Second, column 3 provides an abbreviated timeframe. We reduce the 16-year timeframe to a 10-year period starting two years prior to the GRFP and spanning until seven years after. The interaction is positive and statistically significant but smaller than the full estimation.

Publication Activity in Logarithm Functional Form

We also adjust the functional form of the productivity outcome variable to the logarithmic form of the publication level and re-estimate the primary model. Table B1 presents these results using an OLS estimation. We calculate the logarithmic value by taking the natural log of the level of publications plus 0.1, as you cannot take the log of zero.[61] Column 4 reports the results from the primary sample. Column 5 reports the results for the subsample of those with publications prior to applying for the GRFP and column 6 reports for those without prior publications.

For column 4, the effect size is approximately a 19 percent increase over the 10-year timeframe. Moreover, there is a smaller positive and statistically insignificant effect on the sample with prior publications. The effect size is slightly larger for the sample without prior publications and statistically significant (22 percent).

The DD extension with time trends (equation 2) is statistically insignificant for each sample with more than two time periods. While we find an effect for this estimation for the primary sample and sample without prior publications using the level of publications as the outcome, the results are not robust to this sensitivity specification of the outcome variable.

Robustness Checks

Coarsened Exact Matching

The central design threat for the analysis is endogeneity of researcher quality due to non-random assignment of the GRFP award. Column 1 in Table B2 provides detail on the comparison between the two groups—awardees and honorable mentions—for the pretreatment measures with the primary sample. These include indicators for

[61] As an additional sensitivity check, we also adjusted the level of publications by one prior to taking the natural log. The results are robust. However, this calculation adds more noise to the data given the number of student-year observations with zero or one publication in our sample.

Table B1. Adjusted timeframe and functional form of outcome variable.

	Publication count			Logarithm publications		
	(1) Primary sample years—1 and 5	(2) Primary sample years—1 and 10	(3) Primary sample years—2 to 7	(4) Primary sample	(5) Prior publications	(6) No prior publications
Award * Post	0.378	1.261^{**}	0.381^{*}	0.186^{**}	0.134	0.223^{**}
	(0.273)	(0.582)	(0.216)	(0.080)	(0.145)	(0.095)
Award	0.055	0.128	0.046	0.008	−0.030	0.009
	(0.069)	(0.142)	(0.083)	(0.038)	(0.121)	(0.030)
Post period	3.267^{***}	7.560^{***}	2.676^{***}	2.732^{***}	2.578^{***}	2.782^{***}
	(0.204)	(0.397)	(0.159)	(0.060)	(0.114)	(0.070)
Female	-0.776^{***}	-2.078^{***}	-0.952^{***}	-0.263^{***}	-0.271^{***}	-0.245^{***}
	(0.141)	(0.286)	(0.167)	(0.049)	(0.082)	(0.061)
Any prior publications	2.018^{***}	1.653^{***}	2.049^{***}	0.994^{***}		
	(0.176)	(0.328)	(0.210)	(0.050)		
Constant	−0.317	0.504	−0.342	-2.564^{***}	-0.682^{*}	-2.698^{***}
	(1.074)	(2.379)	(1.304)	(0.364)	(0.397)	(0.407)
Observations	1,124	1,124	5,058	8,430	2,280	6,150
Graduate students	562	562	562	562	152	410
Adjusted R^2	0.445	0.462	0.254	0.590	0.623	0.552
Clustered by student	Yes	Yes	Yes	Yes	Yes	Yes
Year applied and field FE	Yes	Yes	Yes	Yes	Yes	Yes
Program and uni. controls	Yes	Yes	Yes	Yes	Yes	Yes
DD with time trends	N/A	N/A	NS	NS	NS	NS

Notes: Standard errors in parentheses; outcome is level of total publications for columns 1 through 3 and publications in logarithm form for columns 4 through 6. We adjust the level by an increase of 0.1 prior to estimating the logarithm of total publications. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$; NS, not statistically significant. Uni. refers to university.

Table B2. Comparison of means and distribution of pretreatment variables for primary and coarsened samples.

	(1) Primary sample ($n = 562$)					(2) Coarsened sample ($n = 477$)
	Honorable mentions mean	Awardees mean	p-value (t-test)	Distribution difference	p-value (KS-test)	Mean
Female	0.466	0.524	0.174	0.06	0.74	0.492
Any prior publications 5 years before GRFP	0.162	0.203	0.208	0.04	0.97	0.150
Public institution	0.553	0.493	0.160	0.06	0.71	0.469
University-field pubs per faculty quartiles						
Q1 (lowest)	0.312	0.220	0.014	0.09	0.18	0.181
Q2	0.211	0.243	0.355	0.03	1.00	0.252
Q3	0.173	0.203	0.367	0.03	1.00	0.209
Q4 (highest)	0.305	0.334	0.448	0.03	1.00	0.358
Program rank tercile						
1st (highest)	0.793	0.872	0.013	0.08	0.36	0.913
2nd	0.158	0.098	0.035	0.06	0.70	0.075
3rd (lowest)	0.049	0.030	0.266	0.02	1.00	0.012
Field indicators						
Biochemistry	0.350	0.331	0.644	0.02	1.00	0.331
Ecology	0.320	0.348	0.476	0.03	1.00	0.335
Biology	0.331	0.321	0.803	0.01	1.00	0.335
Observations	266	296				
Coarsened observations: awardees						254
Coarsened observations: honorable mentions						223

Notes: Honorable mentions are the control group and awardees are the treated group. The Kolmogorov–Smirnov (KS) test estimates the equality of distributions between the treated and control groups. We incorporate CEM-weights when estimating the t-test for the coarsened sample.

the following: female PI, any prior publications prior to applying for GRFP; quartile rankings of university-field publications per faculty; program rank terciles; and three life sciences field dummy indicators. We report detail on the comparison of means (*t*-test) and comparison of distributions (Kolmogorov–Smirnov test) between awardees (treated group) and honorable mentions (control group). While the difference in means and distributions are statistically insignificant along most of the pretreatment variables, the results show that awardees are in higher ranked graduate programs more than honorable mentions. Conversely, honorable mentions have a higher rate of placing in graduate programs with less research-active faculty (as measured by faculty publication activity).

To improve the balance between these two groups, we use coarsened exact matching (CEM). This approach utilizes a matched sampling technique to coarsen the sample so as to create strata of treatment and control observations that have substantively indistinguishable values (Iacus et al., 2011, p. 8). The approach eliminates imbalances *ex ante* (which effectively removes unmatched units) and offers estimates of the sample average treatment effect.

Operationally, we relied on the complete set of pretreatment variables available for this data set to coarsen the sample. We reduce the imbalance between the two groups along the full set of pretreatment variables such that they are statistically indistinguishable. Column 2 in Table B2 presents the mean statistics for the coarsened sample. Importantly, the value of each covariate is *equivalent* between the treated and control groups. While one tradeoff with this approach is a loss in efficiency, the loss in this case is minimal. With the coarsened sample, we estimate the primary model with 85.8 percent of the treated group and 83.8 percent of the control group.

Equation (A1) estimates the differential effect of the GRFP award on total publications, where i denotes the individual graduate student and t denotes the time lag post GRFP proposal year (5 or 10, respectively) for the coarsened sample. Given that we coarsen the sample by drawing upon matches at the student level (rather than student-year), we estimate the model in wide form.

$$Total\ Publications_{it} = \beta_0 + \beta_1\ Award_i + \beta_z X + \ \varepsilon_i. \tag{A1}$$

β_1 captures the effect of the GRFP Award. We include a set of research-related variables as controls[62] (equation A1). These include detail on graduate degree completion metrics, professional placement indicators, and GRFP measures. With the level outcome, we treat publication activity as a count variable and estimate a negative binomial model. With the outcome in logarithm form, we estimate an OLS model.

Results

We follow Blackwell et al. (2009) and estimate equation (A1) using CEM weights derived from the coarsening procedure. Table B3 presents the results with two variations of functional form for the primary outcome measure of total publication activity. Columns 1 and 2 (3 and 4) estimate the measure in level (logarithm) form. Moreover, we adjust the follow-on timeframe and estimate five (columns 1 and 3) and 10 years (columns 2 and 4) following the GRFP.

The results are robust and the size of the effect is comparable. We estimate the differential effect of the award is associated with an increase of 1.2 total publications

[62] As noted in the main paper, we do not include subscripts for the vector of X in equation (1) given that some are time varying while others are time-invariant. Additionally, the level of the measure varies from the individual to the institution.

Table B3. Differential effect of GRFP award on research productivity (equation A1).

	Publication level		Logarithm publications	
	5 Years post (1)	10 Years post (2)	5 Years post (3)	10 Years post (4)
Award (binary)	0.277	1.162**	0.134	0.161**
	−0.234	−0.502	(0.107)	(0.079)
Completed a PhD (binary)	5.525***	7.004***	1.922***	1.133***
	−0.812	−1.345	(0.241)	(0.178)
Years from GRFP to degree (level)	−0.755***	−1.046***	−0.295***	−0.124***
	−0.097	−0.188	(0.037)	(0.027)
Research position (binary)	0.772	3.629***	0.028	0.490***
	−0.524	−1.114	(0.216)	(0.160)
Academic position (binary)	0.259	1.848	0.119	0.403**
	−0.667	−1.385	(0.265)	(0.196)
Academic-research position (binary)	0.059	−1.237	0.066	−0.219
	−0.73	−1.525	(0.300)	(0.221)
Post-doc position (binary)	0.454	2.451***	0.195	0.301***
	−0.297	−0.64	(0.137)	(0.101)
University-field GRFP two-year capacity (level)	−0.026	−0.161***	−0.024*	−0.029***
	−0.027	−0.059	(0.012)	(0.009)
GRFP stipend (deflated, logarithm)	0.38	0.635	0.323	0.247
	−0.503	−1.095	(0.234)	(0.173)
Constant			−3.004	−1.803
			(2.383)	(1.761)
Graduate student observations	477	477	477	477
Regression type	Neg. Bin.	Neg. Bin.	OLS	OLS
CEM weights	Yes	Yes	Yes	Yes
Pretreatment controls to coarsen sample	Yes	Yes	Yes	Yes

Notes: Key outcome variables—total publications following GRFP proposal year. Columns 1 and 3 (2 and 4) report the total publications within five (10) years following the GRFP award. Columns 1 and 2 estimate the primary outcome in level form. Columns 3 and 4 estimate the outcome variable in logarithmic form. Note we increase all original values by 0.1 before estimating the logarithm transformation to account for panel observations with 0 publications. Marginal effects are reported for the negative binomial models (columns 3 and 4). Standard errors are in parentheses (***$p < 0.01$; **$p < 0.05$; *$p < 0.1$). For all models, CEM weights are used. Pretreatment controls used to coarsen the data are reported in Table B2. Balance is achieved along all pretreatment measures before estimating the regression (column 2, Table B2).

10 years following the program. Given that we estimate this model with data in wide form, we are unable to directly compare the effect to the primary DD model reported in the article. However, this model most closely resembles the two-period DD model reported in Table B1, column 2. The effect size from CEM is consistent and robust to the two-period DD with an effect of 1.162 versus 1.261 publications, respectively.

Additional Considerations

CEM stands in contrast to propensity score matching methods (PSM), which have received criticisms not only for matching *post hoc*, but also for matching on the sample rather than within strata (Iacus et al. 2009, 2011, 2012). As an additional robustness effort to the CEM matching procedure, we estimate the model with PSM techniques as well. The results for PSM are robust and economically comparable

Table B4. Fixed effects estimation results.

	(1) Primary sample	(2) Extended sample	(3) Ever publish sample	(4) Full sample
Award * Post	0.636***	0.655***	0.604***	0.774***
	(0.119)	(0.108)	(0.115)	(0.100)
Constant	0.023	0.019	0.040	0.031
	(0.109)	(0.098)	(0.105)	(0.091)
Observations	8,430	10,455	10,155	13,155
Graduate students	562	697	677	877
Adjusted R^2	0.539	0.421	0.510	0.379
Individual FE	Yes	Yes	Yes	Yes
Year FE (range 1–16)	Yes	Yes	Yes	Yes

to the CEM results and primary model. Specifically, the average treatment effect on the treated estimates a weakly statistically significant increase of 1.4 in total publication activity 10 years following the GRFP. The effect size is comparable to the CEM results. Taken together, we find robust, comparable results to the DD design with the matching designs.

Fixed Effects

Due to concerns of underlying student quality, we also estimate the effect of the award using a fixed effects model (equation 3 in the article). The model controls for individual time-invariant factors through the individual fixed effect and annual factors with the year fixed effect. The latter represents the year in the timeframe ranging from five years prior to the GRFP decision and 10 years post (this is operationalized as 1 to 16). Table B4 shows the results across the multiple samples described in the sample section of Appendix A.

Column 1 matches the primary sample of those who ever publish and have complete data across the time invariant controls included in the primary model. Column 2 expands the sample to include those who may not publish over the 16-year timeframe, but still have complete data along the set of controls. Column 3 uses the sample of those who ever publish, whether or not they have complete data for the primary model. Finally, column 4 uses the full sample, including those who never publish and those with incomplete data. Across all four samples the results are robust to the primary DD model. Effect sizes range from 0.604 to 0.774 publications and are statistically significant.

Additional Extensions to Primary Model

Alternate Standard Errors

We also estimate the primary DD model of publication activity with alternate clusters of standard errors. The primary estimation clusters standard errors by graduate student. However, given that publications are increasing over time, standard errors will increase over time, potentially biasing our standard errors. Thus, we estimate the model clustering first by year and then by using two-way clustering with year and graduate student. Both estimations produce robust results. The results are presented in Table B5 along with our primary results clustered by student.

Table B5. Alternate standard error clustering.

	Primary sample			Prior publications			No prior publications		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Award * Post	0.636**	0.636***	0.636**	0.147	0.147	0.147	0.709**	0.709***	0.709**
	(0.291)	(0.129)	(0.291)	(0.666)	(0.099)	(0.614)	(0.310)	(0.154)	(0.315)
Award	0.031	0.031	0.031	0.243	0.243***	0.243	0.034	0.034***	0.034
	(0.096)	(0.024)	(0.088)	(0.299)	(0.070)	(0.288)	(0.096)	(0.010)	(0.086)
Post period	4.029***	4.029***	4.029***	5.428***	5.428***	5.428***	3.577***	3.577***	3.577***
	(0.205)	(0.732)	(0.755)	(0.524)	(0.770)	(0.907)	(0.202)	(0.733)	(0.754)
Female	−1.265***	−1.265***	−1.265***	−1.846***	−1.846***	−1.846***	−0.937***	−0.937***	−0.937***
	(0.189)	(0.351)	(0.389)	(0.416)	(0.438)	(0.577)	(0.192)	(0.298)	(0.341)
Any prior publications	1.654***	1.654***	1.654***						
	(0.230)	(0.216)	(0.299)						
Constant	−0.037	−0.037	−0.037	0.446	0.446	0.446	0.952	0.952***	0.952
	(1.580)	(0.202)	(1.419)	(1.890)	(0.332)	(1.552)	(1.585)	(0.299)	(1.403)
Standard error clustered by	Student	Year	Student and year	Student	Year	Student and year	Student	Year	Student and year
Observations	8,430			2,280			6,150		
Graduate students	562			152			410		
Adjusted R^2	0.274			0.336			0.238		
Year applied FE	Yes			Yes			Yes		
Program and uni. controls	Yes			Yes			Yes		

Notes: Standard errors in parentheses; outcome for all estimations is total publications (level). ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Table B6. Life science journal descriptive statistics.

Journal Rank	Mean	Standard deviation	Min	Max	Total pubs	High-impact journals in sample	Percent of total high-impact journals
Top 1 percent	0.095	0.448	0	10	225	14	66.7
Top 5 percent	0.260	0.824	0	13	595	27	45.0
Top 10 percent	0.367	0.995	0	15	864	41	43.6

Additional Outcomes

Leadership and Quality Bibliometric Outcomes

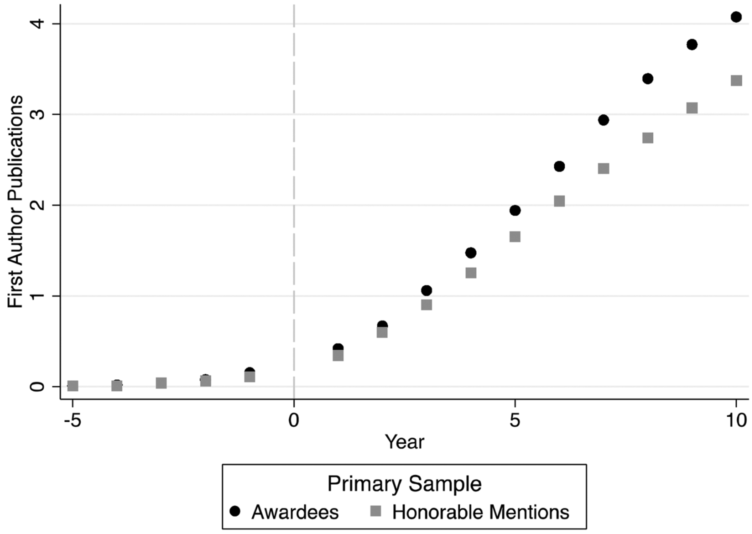
We estimate the primary model with additional bibliometric outcomes to account for research leadership and quality. The main results are presented in Table 5 in the article. For the former measure, we rely on the order of authorship. Within the field of life sciences, first authorship indicates that the graduate student led the research inquiry. Typically, all other authors or lab members are listed after with the lab director listed last (Dance, 2012; Tscharntke et al., 2007).

For the latter measure of research quality, we rely on the rank of the journal.[63] We rely on Scopus to define the set of publications published in "high impact" life sciences peer-reviewed journals. First, we reduce the set of journals available in the Scopus database (21,500 journals) to peer-reviewed journals in the life sciences (6,394 journals). Then, to determine the set of "high impact" life sciences journals, we rely on a series of journal rankings—SNIP, IPP, and SRJ.[64] For these three, we draw upon the journal ranks for the following years: 2000, 2005, 2010, and 2015. We identify the set of publications that exceed the top 1, 5, and 10 percent across all measures over the four panels. These yield 21, 60, or 94 journals for the top 1, 5, or 10 percent, respectively. We then computed journal publication levels by journal rank within the sample. Table B6 details the descriptive statistics for this measure. Among the highest ranked life sciences journals (Top 1 percent), 0.095 percent of students in the primary sample published 225 publications in 14 journals (or 66.7 percent of the possible Top 1 percent set).

Model Diagnostics for Leadership and Quality Bibliometric Outcomes

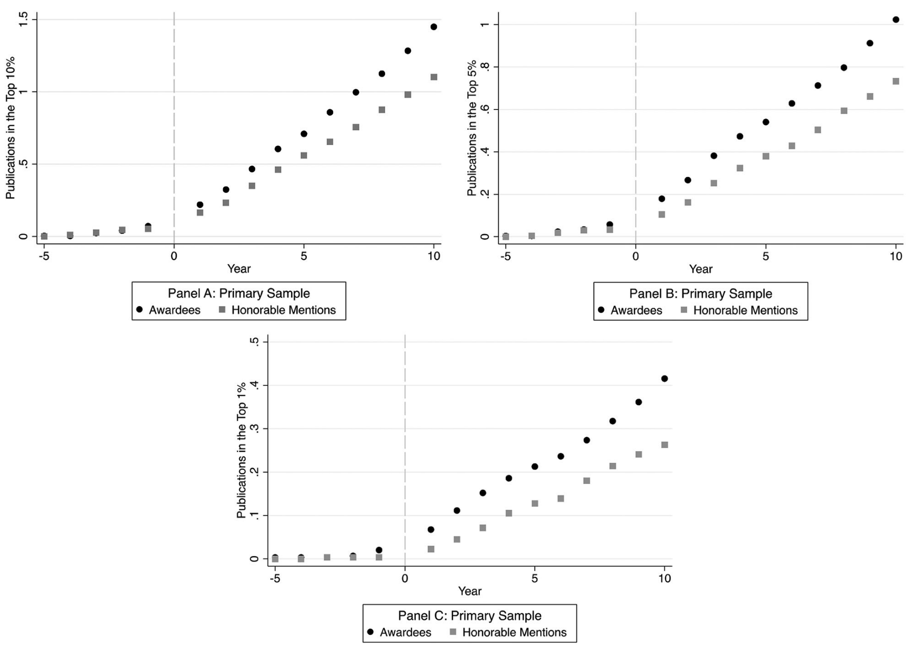
For each of these two broad additional bibliometric outcomes (first-author position and journal rankings), we provide figures with annual averages by award status (Figures B1 and B2). This follows the approach presented in the fourth section of the article. The figures show a similar pattern to that of total publication level, with a slowly increasing trend in the pre-period with awardees rising steadily above honorable mentions in the post-period. The notable exception is panel C of

[63] Initially, we estimated the impact on citations. However, due to data limitations, the citation measure available is a coarse, cumulative measure to the current date only.
[64] SNIP refers to the Source Normalized Impact per Paper and "measures contextual citation impact by weighting citations based on the total number of citations in a subject field"; IPP refers to Impact per Publication, this "measures the ratio of citations in a year to scholarly papers published in the three previous years divided by the number of scholarly papers published in those same years"; and SJR refers to SCImago Journal rank, this "is a prestige metric based on idea that not all citations are the same." Source: https://www.journalmetrics.com/about-journal-metrics.php.



Notes: Sample restricted to those who ever publish with complete case (primary sample as reported in Table 1) representing 562 students.

Figure B1. Annual Average First-Authored Publication Level by Award Status.



Notes: Sample restricted to those who ever publish (primary sample as reported in Table 1); each panel draws upon activity from 562 students. The panels A, B, and C, reflect the publications in the top 10, 5, and 1 percent of life science journals, respectively.

Figure B2. Annual Average Publication Levels in Top Tier Journals (10th, 5th, and 1st Percent Rank) by Award Status.

Table B7. Graduate training and professional placement outcomes.

	(1) Completed PhD	(2) Time to PhD	(3) Post-Doc Position	(4) TT Position
Award	0.063***	−0.124	0.112***	−0.027
	(0.022)	(0.186)	(0.037)	(0.019)
Female	−0.029	0.269	−0.071*	0.005
	(0.021)	(0.185)	(0.037)	(0.019)
Any prior Publications	0.031	−0.314	0.026	−0.013
	(0.036)	(0.261)	(0.053)	(0.031)
Top tercile program rank	0.042	0.130	0.066	0.057*
	(0.026)	(0.270)	(0.053)	(0.032)
Carnegie very high research institute	0.005	−0.279	−0.068	0.030
	(0.038)	(0.313)	(0.066)	(0.047)
Public institution	0.017	−0.088	−0.027	−0.003
	(0.023)	(0.203)	(0.040)	(0.020)
Years from proposal to degree			0.027**	0.007
			(0.011)	(0.005)
Observations	748	673	703	703
Model	Logit	Negative binomial	Logit	Logit
Year applied FE	Yes	Yes	Yes	Yes
Field controls	Yes	Yes	Yes	Yes

Notes: Marginal effects and standard errors are presented in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Binary outcomes estimated with logistic regression for columns (1), (3), and (4). Count outcome estimated with Negative Binomial distribution for column 2. Data in wide form (student level). For column 2, we estimate with a stratified sample of students who completed the PhD.

Figure B2, which reports on publications in the top 1 percent of life sciences journals. In this case, there is a clear divergence between awardees and honorable mentions in the pre-period, meaning that this is not a reasonable control group to assess the impact for this particular outcome. Future research should be conducted to assess the impacts of federal funding on high-quality publications.

Non-Publication Outcomes

We also examine impacts of the award on non-publication-related outcomes; these include degree completion, time to degree, and professional placement following graduate training. Table B7 presents the results. We estimate the outcomes for degree completion and professional placement following graduate training as binary indicators. For the time to PhD measure, we treat the measure as a count variable. The data are in wide form; in other words, there is one observation per graduate student.