THE IMPACT OF POSTDOCTORAL FELLOWSHIPS ON A FUTURE INDEPENDENT CAREER IN FEDERALLY FUNDED BIOMEDICAL RESEARCH

Misty L. Heggeness
Donna K. Ginther
Maria I. Larenas
Frances D. Carter-Johnson

The Impact of Postdoctoral Fellowships on a Future Independent Career in Federally Funded
Biomedical Research
Misty L. Heggeness, Donna K. Ginther, Maria I. Larenas, and Frances D. Carter-Johnson
NBER Working Paper No. 24508
April 2018
JEL No. J24,O3,O38

ABSTRACT

The Ruth L. Kirschstein National Research Service Award (NRSA) program is a major research
training program administered by the National Institutes of Health (NIH) with funds appropriated
each year by Congress. This study examines the impact of NRSA postdoctoral fellowships on
subsequent research-related career outcomes using NIH administrative records on applicants who
applied for a fellowship between 1996 and 2008. We find that postdoctoral fellowships increased
the probability of receiving subsequent NIH research awards from 6.3 to 8.2 percentage points
and of achieving an NIH-funded R01 award, an indication of an independent research career,
from 4.6 to 6.1 percentage points. Our findings demonstrate that the NRSA postdoctoral
fellowship awards have the potential to promote retention of scientists in NIH-funded research
and in the biomedical workforce pipeline.

Misty L. Heggeness
U.S. Census Bureau
misty.l.heggeness@census.gov

Donna K. Ginther
Department of Economics
University of Kansas
333 Snow Hall
1460 Jayhawk Boulevard
Lawrence, KS 66045
and NBER
dginther@ku.edu

Maria I. Larenas
CLAPES UC
Centro Latinoamericano de
Políticas Económicas y Sociales
Av Libertador Bernardo O'Higgins 440
Piso 13, Santiago
Chile
mlarenas@gmail.com

Frances D. Carter-Johnson
National Science Foundation
2415 Eisenhower Avenue
Alexandria, VA 22314
frajohns@nsf.gov

INTRODUCTION

The National Institutes of Health (NIH) has advanced the frontier of biomedical research and innovation since its inception in the 1930s. The NIH mission includes funding and training the next generation of scientists. Since 1974, the NIH has formally committed to training high-potential, early-career scientists to carry out the nation's biomedical research agenda through the congressionally mandated Ruth L. Kirschstein National Research Service (NRSA) Award. While the NRSA programs are subject to periodic review (National Research Council 2011), few researchers have specifically examined the impact of NIH's investment in the NRSA F32 postdoctoral program on subsequent receipt of NIH funding for research. Using NIH administrative records on applicants and awardees of NRSA F32 fellowships between 1996 and 2008, this study provides evidence that an F32 fellowship substantially increases the likelihood of obtaining subsequent NIH research funding, including the probability of an NIH R01 grant, which is a major indicator of transition to an independent biomedical research career.

We examine two questions: (1) *Does the NRSA F32 fellowship affect the likelihood of future NIH funding*? These outcomes are measured by whether a fellowship applicant or awardee applies for or receives a Research Project Grant (RPG) award,[5] the number of RPG awards, and whether they appear as an NIH RPG applicant four years or more from the time they applied for the NRSA F32 award; (2) *Does the NRSA F32 award contribute to an NIH-funded independent research career?* This outcome is measured by the probability of receiving an NIH R01 award, considered a milestone that establishes an independent research

---

[5] In recent years, Research Project Grant awards include the following grant mechanisms: R00, R01, R03, R15, R21, R33, R34, R35, R36, R37, R50, R56, R61, RC1, RC2, RC3, RC4, RF1, RL1, RL2, RL9, P01, P42, PM1, PN1, RM1, UA5, UC1, UC2, UC3, UC4, UC7, UF1, UG3, UH2, UH3, UH5, UM1, UM2, U01, U19, U34, DP1, DP2, DP3, DP4, DP5. For more information, see: https://grants.nih.gov/grants/glossary.htm#ResearchGrants. While the definition of RPGs changes slightly from year to year, these mechanisms give a general overview of the mechanisms generally included in research project grant awards.

career, four years or more after applying to the NRSA F32 award (Science Careers Editors 2007, Rockey 2014). We use the scoring of fellowship applications and matching techniques that allow us to analyze the outcomes of individuals who received the fellowship award from those whose application scores and observable characteristics are similar, but who did not receive an award.

We find that the average treatment effect of an NRSA F32 postdoctoral award (for all applicants) increases the future probability of receiving NIH research funding by anywhere from 6.3 to 8.2 percentage points and the probability of receiving NIH R01 funding by anywhere from 4.6 to 6.1 percentage points on average. We find that the benefits are slightly larger for the average treatment effect on the treated (for those who received the award), increasing the probability of receiving NIH funding by 7.0 to 8.8 percentage points and the probability of receiving an NIH R01 award by 4.9 to 6.5 percentage points. Overall, this study provides evidence demonstrating the positive impact of the F32 program and informs policymakers regarding the value of future investments in the program.

The paper proceeds as follows. First, we review the literature on the effect of particular postdoctoral and graduate appointment types on career outcomes. Second, we introduce the NRSA F32 fellowship award and scoring process. Third, we discuss the data. Fourth, we examine the fellowship award process and our matching identification strategy. Fifth, we evaluate the results and robustness checks. Finally, we conclude with our discussion of the impact of the NRSA F32 award on career outcomes.

BACKGROUND

The NRSA program has multiple award mechanisms that fund research training in biomedical science (NIH 2001; NIH 2002; NIH 2003). These include institutional training programs for undergraduates, graduate students and postdocs and individual fellowships for graduate students pursuing PhD students or postdoctorates. Historically, NRSA fellowships have been highly competitive. In 2008, the NIH allocated around $751.2 million in funding to NRSA programs to support research training for 16,370 undergraduate students, graduate students and postdoctoral researchers (https://report.nih.gov/nihdatabook/NIH Office of Budget, 2017).[6] Of these NRSA awardees or appointees 1,487 (around 9.1% percent) received the NRSA F32 postdoctoral fellowship award. Between 1998[7] and 2008, the number of F-series fellowship awards funded by the NIH increased by 15.2 percent, as shown in panel A of Figure 1. While NIH increased the total number of F-series awards, it also diversified the F-mechanism by funding more positions in other F-series awards, causing the total number of NRSA F32 postdoctoral fellowship awards funded in any given year to decrease over the period by 21.9 percent (see Figure 1, panel A).

Our analysis focuses only on the NRSA F32 postdoctoral award, which is an award applied for or received prior to or during a transition toward research independence. Also important to note is that even though overall the F-series fellowship awards have experienced growth over the period of our study, they grew much more slowly than the number of PhDs conferred in biological sciences during that time. According to the National Science Foundation, the number of new PhDs in biomedical sciences grew annually from 5,838 in 2000 to 7,797 in 2008, an increase of 33.6 percent between 2000 and 2008 (Fiegener, 2011).

---

[6] We report information through 2008 since that is the final year of our analysis.
[7] Data unavailable for 1996 and 1997

The Effect of Postdoctoral Study on Career Outcomes

There are a limited number of studies that have examined the impact of a postdoctoral appointment (postdoc) on later career outcomes. Those that do have combined many academic fields making it difficult to generalize the results to fields like biomedical science (Su 2013, Levitt 2010, National Science Board 2014, Nerad & Cerny 1999). Several studies investigated the impact of a postdoc on subsequent placement in academic careers. Su (2013) found that longer duration postdocs are not associated with improved academic placements. Levitt (2010) examined the association between the 1992-94 cohorts of NRSA F32 postdocs and their subsequent biomedical careers. He found significant gender differences in outcomes; women were more likely to leave scientific research careers than men. Career trajectories of women appeared to be affected by measures related to their mentor's quality. For female postdocs, the higher the mentor's h-index, the more likely the postdoc was to receive an NIH grant. Levitt concludes his study by raising the possibility that his results may not be extrapolated to more recent NRSA cohorts.

Some studies indicate a negative impact of postdoctoral and graduate appointments. Kahn and Ginther (2017) found that individuals who took a postdoc in biomedical fields were paid substantially less in future employment than those who skipped the postdoc. Although the postdoc was useful for obtaining an academic tenure-track research position, Kahn and Ginther found that the likelihood of achieving this goal dropped considerably over time. Blume-Kohout (2016) used the Survey of Earned Doctorates to examine whether graduate research assistantships or grants were more effective at launching research careers than graduate student support by NIH training grants or fellowships. Surprisingly, she found that graduate students with research assistantships had more successful research outcomes in terms

of research-focused jobs (Blume-Kohout 2016) although this study did not address postdocs funded by different mechanisms. The evidence on the impact of postdoctoral and graduate training on later career outcomes in general is mixed, and the evidence on the impact of NRSA postdoctoral fellowships using an adequate methodology in particular is limited, prompting this study. That said, studies assessing the impact of federal investments in training the biomedical research workforce, including evaluation studies commissioned by the federal government, show that individuals participating in NIH training programs experience higher rates of remaining in scientific research in comparison to their counterparts who did not receive formal NIH training (Pion 2001; Mantovani et al 2006).

Our work is most closely related to studies by Jacob and Lefgren (2011a, 2011b). Using a regression discontinuity design (RDD), these authors found that receiving an NIH postdoctoral fellowship increased publications by 20 percent in the next five years relative to non-awardees (Jacob and Lefgren 2011a). In addition, their study found that NIH postdoctoral fellows received higher dollar amounts in subsequent NIH funding compared to non-awardees. Our analysis below indicates that RDD has limitations for estimating the impact of the NRSA F32 award because not all NIH institutes and centers fund NRSA F32 proposals in the order of the score received. A detailed discussion of the F32 application, review and award process is discussed below.

The NRSA F32 Postdoctoral Proposal Process

A thorough understanding of the NRSA application and award process is necessary for the development of our empirical approach. The goal of the NRSA F32 award is "to enhance the research training of **promising postdoctoral applicants** who have the **potential to become**

8

productive, independent investigators in scientific health-related fields relevant to the missions of the participating NIH Institutes and Centers (IC)." In defining a postdoctoral position for the purposes of this paper, we follow the guidance of the National Institutes of Health and the National Science Foundation who in 2007 defined a postdoc as "…an individual who has received a doctoral degree (or equivalent) and is engaged in a temporary and defined period of mentored, advanced training to enhance the professional skills and research independence needed to pursue his or her chosen career path" (NSF, 2007). One mechanism by which NIH accomplishes this goal is through funding a postdoctoral fellowship position in a mentor's scientific laboratory. The NRSA F32 fellowships are allocated based in part on a score assigned during peer review which considers the merit of an applicant's research proposal and the quality of proposed training and career development support available through the institution and mentors.[8] In addition, before offering an F32 award program staff of the NIH institute or center (IC) perform a thorough review to ensure the validity of research idea, quality of the training environment and whether the research proposed matches well with the institute or center's needs and priorities.

Figure 2 shows the number of NRSA F32 applications by year. The number of new applications ranged from around 1,500 to 2,500 per year in the mid-2000s. During the NIH Doubling (1998-2003), the number of new applications decreased from around 2,100 in 1998 to a low of approximately 1,500 in 2002. This may reflect the increase in the NIH budget during the doubling that likely increased postdoctoral slots on NIH-funded research grants. After the NIH

---

[8] During our study timeframe, the NIH assigned priority scores that ranged from 100 to 500. Starting in 2009, the NIH assigned impact scores that ranged from 10 to 90. Throughout we refer to the scores assigned to proposals during peer review as the review score.

doubling, the number of new applicants once again began to rise and peaked in 2006. In 2008, one F32 award cost the NIH, on average, around \$48,998.[9]

We posit that applicants who meet the basic requirements for the NRSA F32 award are willing to apply if the benefit of winning is higher than the cost of applying. Applicants are required to have a research doctorate or professional degree by the time of the award, be a citizen or a permanent resident of the United States, and have a sponsor (who functions as a mentor), and a sponsoring institution. All of these requirements limit the size of applicant group and make it more homogeneous, limiting the impact of unobserved characteristics on the probability of receiving an award.

Selection to receive an NRSA F32 fellowship award depends on multiple measures: First, each new application is evaluated for scientific merit by an NIH Scientific Review Group (SRG), which, for the most part, is developed and run by the NIH Center for Scientific Review (CSR).[10] CSR is an independent body that provides institutes services related to peer review, as well as documenting the evaluation and scoring of applicants. NIH SRGs for NRSA F32 fellowships, like SRGs for other funding mechanisms, include accomplished scientists and experts as external reviewers. The SRG evaluation considers observable characteristics of the applicants, A (the applicant's past academic and research records, publications); observable characteristics of the sponsor S (the sponsor's general qualifications); research potential, Pot, (references, the applicant's research goals); research environment E (training environment); the research proposal, P, plus the perception of the reviewer in that period ε_{qt} . Each reviewer, r, assigned to an application, i, for institute or center,

---

[9] For more information and additional details on the total cost of an F32 by institute and individual, see https://projectreporter.nih.gov).

[10] In a handful of instances, some NIH institutes and centers will conduct their own internal review on topic specific areas of research. The decision to conduct reviews in house or through CSR is a function of resources, time, and feasibility.

j, in a particular year, t, and council round, q, gives a preliminary reviewer overall review

score, PS_{ijrtq}, for each application defined as:

$$PS_{ijrtq} = f\left(A, S, Pot, E, P, \varepsilon_{rtq}\right) \tag{1}$$

$$PS_{ijtq} = \frac{\sum_{n=1}^{r} PS_{ijrtq}}{\sum_{n=1}^{r} N_{ijrtq}} \tag{2}$$

Equation 2 gives a final overall review score, PS_{ijtq}, for each discussed application, i,

who applied to an institute or center, j, in year, t, in council round, q, and is simply determined

by the mean score of all SRG members' preliminary scores. NIH review panels score proposals

from best to worst, where the lowest scores are considered the best. Until 2009, proposal review

scores ranged from 100 (high impact) and 500 (low impact). Starting in 2009, the process called

Enhancing Peer Review changed the scale of evaluation to 10-90 where the final overall impact

scores now range from 10 (high impact) through 90 (low impact) (NIH, 2011a).

Prospective awardees are evaluated in council round, q, in a calendar year, t. At the end

of the scoring process, the SRG reports the review scores to the institute or center's selection

process. At the end of the first selection, those applications with the lowest scores receive a

secondary level of review by NIH institute and center staff. This step is a key determinant of

the funding decision. NIH program staff examines the applications' summary statements and

review scores, and weighs these factors against the needs and priorities of the institute or

center. The institute or center director or their delegate makes the final decision as to whether

to offer an award.

Applicants who do not receive an award typically explore multiple options, which

include appointment to a position as postdoctoral researcher on a mentor's research grant,

appointment to a T32 training grant if a slot is available, application to other funding sources

such as foundation grants, or, if feasible, resubmission of their application. Other scenarios are to leave academic research and explore other career paths. The true value of the NRSA F32 award is its power of developing and conducting a research project and the freedom it gives the postdoc to explore that line of research under the supervision of their mentor. Therefore, we do not argue that one does not receive additional training if one does not receive the NRSA F32 award. We are testing whether funding to develop and pursue a research project early in one's career has an effect on the probability of receiving future in NIH research funding.

A sharp regression discontinuity design is appropriate if, NIH institutes and centers fund the majority of proposals in the order of the score. Aberrations in funding of proposals with review scores near the cutoff would constitute a fuzzy regression discontinuity design. However, if institutes skip several (competitive) low-scoring proposals and reach for higher scoring proposals that are more consistent with institute priorities or because of other administrative hurdles on low-scoring applications, then an RDD has limitations. We turn now to the data before evaluating the appropriate methods for determining the impact of the F32 award on subsequent career outcomes.

DATA AND METHODS

We use administrative records from the NIH's Information for Management, Planning, Analysis, and Coordination (IMPACII) system from 1996 to 2008. NIH matches IMPACII records to data from the National Science Foundation's (NSF) Survey of Earned Doctorates (SED), an annual census of doctoral recipients from U.S. institutions.[11] The NSF SED contains information on individual demographics, characteristics of graduate study, and future career

---

[11] Under existing Memorandums of Understanding (MOUs) and interagency agreements, we use an ID variable that tracks respondents over time to link individual-level NIH (IMPACII) administrative records with individual-level data from the Survey of Earned Doctorates (SED) (NSF 2013).

plans. By linking these datasets, we are able to obtain missing data and add additional individual-level covariates on our sample. We used demographic variables before or at the point of PhD completion extracted from the SED including age at PhD completion, gender, race/ethnicity, marital status at PhD completion, PhD field of study, and type of doctorate education funding. We used the two data sources to construct one large panel dataset for analysis, limiting our sample to those who applied for NRSA F32 funding between 1996 and 2008, and then observing these individuals' NIH application and funding patterns through 2015.

We use detailed fellowship and subsequent NIH grant application and funding information, including application review score, funded or non-funded status, timeframe and institute/center (IC) receiving applications or funding the award, as well as previous grant funding or training affiliations. This information comes from IMPACII for NRSA F32 postdoctoral fellowship applicants from 1996 to 2008. We further queried IMPACII for subsequent applications for NIH funding and awards for these individuals. Similar to Jacob and Lefgren (2011a), we define our outcome variables to identify research award application or receipt four or more years out from the individual's NRSA F32 application year.

Table 1 shows descriptive statistics for all fellowship applications from 1996 to 2008 by award status. Awardees and non-awardees differ across a number of observable characteristics. Awardees are younger, more likely to be married and white. Awardees are significantly less likely to be black or Hispanic. Individuals with MD degrees are less likely to receive fellowship awards whereas PhDs are more likely. Individuals with biomedical or social science degrees are more likely to receive fellowship awards compared with those whose PhD field is not reported. Individuals who have had T32 predoctoral traineeships are more likely to receive awards. NRSA F32 awardees are significantly more likely aspire to and to receive subsequent NIH funding as

measured by the number of RPG applications and awards, the probability of an RPG award, and the probability of an R01 award. As expected, awardees have significantly lower (better) scores on their last observed F32 application. In the next section, we examine whether NIH institutes and centers fund proposals in the order of the score, and whether a regression discontinuity design (RDD) is warranted.

Our analysis sample is a subset of the full sample. We drop all applications that are higher than the 60th percentile in each council round because scores for these applications are not consistently saved in the reporting database and practically none of them get funding. Some institutes and centers have too few applicants for our preferred analysis method, so we drop applicants from seven institutes and centers for this reason. Finally, the dataset we use is incomplete for some council rounds prior to fiscal year 1996 and are dropped. Our final analysis sample contains 14,276 individuals, and descriptive statistics are reported in Table 2. In our analysis sample, awardees and non-awardees no longer differ in terms of age at application, marital status, or likelihood of having a prior T32 traineeship.

An Analysis of NIH Institute and Center Award Behavior

Each institute and center has discretion in deciding who is offered a fellowship based on the current research priorities of the institute. Institutes and centers vary in their process for awarding a fellowship. For this study, we interviewed staff at various institutes and centers who provided a representative view of the variation in procedures across NIH. Through our interviews, we learned that the process for awarding fellowships is complex.

The review and scoring of fellowship proposals and NIH institute and center budgets create scoring thresholds that theoretically would support the use of a regression discontinuity

14

design (RDD) like the one used in Jacobs and Lefgren (2011a). Decisions of who to fund are made at the year, institute,[12] and council round. Jacobs and Lefgren only use year and institute to estimate an RDD model, which we demonstrate below is not an optimal method for evaluating training fellowships given the fellowship selection process at NIH.

If awards are based solely on the review score allocated from best (lowest) score to worst (highest) score until the institute budget allocations for the fellowship are exhausted, the percent of applicants selected out of order would be zero or close to zero. If this were the case, a sharp RDD would be valid because, in this scenario, the institute follows the guidance of the review score for the best (lowest) scores. If, however, an institute has funded a majority of applicants with meritorious scores and some discretion is used regarding awards to applications with scores near the pay line, we would observe minimal disorder in funding. In this example, the institute uses discretion to potentially skip some applicants close to the pay line in order to fund applications with slightly worse scores but that best fit within their scientific priorities and where the institute staff believes the applicant has the best-case scenario for future success. If this were the case, a fuzzy RDD would be appropriate. Additionally, institutes could just chose to use a significant amount of discretion when selecting proposals for funding in order to meet institution goals related to scientific priorities or, perhaps, diversity of the workforce. For institutes who engage in this behavior, no real cutoff exists and an RDD is not appropriate.

Our interviews with NIH staff indicated that the funding process is multifaceted. The institute receives an application's review score, which, in most cases, is defined by the study section through the NIH's Center for Scientific Review (CSR). Once the institute receives the scores, the staff assesses the full application, including the summary statement from peer

---

[12] For the rest of this paper, "institute" will be used to refer to NIH Institutes and Centers.

review, the quality of the training institution, and the alignment of the research proposal with the institute's research priorities. Institute staff, specifically program officers at each institute then generally participate in a team meeting in which they defend the proposals that best match their defined priorities. Together, the program officers, the training director and other institute staff make a joint decision for recommendations to the institute director. Either the institute director or his/her delegate makes the final decision and signs off on which proposals to fund. Institute directors vary in terms of direct involvement in the consideration and final approval of proposals. Before a final decision is made and the candidates are informed, the budget office reviews and signs off on the final list of candidates, primarily making sure sufficient funds are available for the recommended awards.

Decisions related to fellowship funding at NIH use discretion. Figure 3 illustrates why RDD is not the appropriate design for evaluating fellowship awards. It demonstrates the range of discretion used in fellowship decision making for the full sample over the period of our study. In a given council round, anywhere from around 6.9 percent to 20.9 percent of applicants received a decision that was not based on their ranked score. Either the institute's funded the applicant even though other applicants had better scores or the institute's final decision was not to fund the applicant even though they had better scores than others that received funding.

What does this mean for our study? Figure 3 illustrates the level of non-compliance for an RDD method and the strong tendency for discretion in decision making within each council round. While the level of discretion has decreased in recent years, over the entire sample, around 12.3 percent (1 in 8) of all applications within a given council round received a decision that was not in line with the peer review score. Given the large proportion of

discretion used within council rounds, we argue that RDD is an invalid method for evaluating the true impact of early career training on later career outcomes as it relates to the NRSA F32 award.

Using Matching to Identify the Causal Effect of the NRSA F32 Award

As described above, the NIH's multi-step selection process for NRSA fellowships and research grants generates a review score via a systematic peer review process. Given this, and the fact that the groups of individuals applying to the awards are relatively homogenous given the total population, we use matching techniques, specifically Propensity Score Match (PSM) and Nearest Neighbor Match (NNM). While we understand that any unobserved characteristics that are different between funded and unfunded confound our results, we argue that matching is a feasible approach for the following reasons. First, we can control for unobserved differences by institute and council round by controlling for these factors. Additionally, selection is made at the institute level, and any unobserved differences among applicants are also unobserved by the institute and, therefore, not a driving component of the selection process. In short, there is no self-selection of fellowship award offerings. Finally, the groups of individuals who apply for funding are relatively homogeneous within the institute. They all clearly excel in academics, have been encouraged to apply by their mentors (which means their mentors believe they have a chance of getting the award), and are typically intensely interested in biomedical research. Given the unobservable variation that could exist, we argue that within this select group of applicants, it is minimal.

We use the potential outcomes framework employed in econometric analysis to estimate the causal effect of fellowship awards on subsequent NIH funding outcomes (Rubin 2004). To

fix ideas, let $T_i = 1$ be the treatment when an individual's fellowship application is funded, and let $T_i = 0$ if the application is not funded. In the potential outcomes framework each individual has two potential outcomes of subsequent NIH funding: $Y_i(1)$ if the individual receives the award treatment and $Y_i(0)$ if the individual is not treated. For each individual, the causal effect of the award on subsequent NIH funding is defined as the difference in potential outcomes $Y_i(1) - Y_i(0)$. However, each individual is only observed when they receive the award or they do not, and we must estimate the counterfactual outcome, in this case, using matching methods.

In order to implement matching methods, we assume that treatment is independent of the outcome conditional on covariates, $T_i \perp \left(Y_i(0), Y_i(1) \right) | X_i$. This is the unconfoundedness assumption, which means that the treatment is conditionally independent of the outcome after conditioning on observable characteristics. Given unconfoundedness, we can define the average treatment effect in terms of potential outcomes as the expected value of potential outcomes:

$$ATE = E[Y_i(1) - Y_i(0)]$$

We can define the average treatment effect on the subsample of the treated as:

$$ATT = E[Y_i(1) - Y_i(0) | T_i = 1]$$

We use two matching methods to identify the ATE and ATT. First we employ propensity score matching, defining the propensity score as the probability of receiving treatment conditional on observed characteristics $e(X) = \Pr(T_i = 1 | X_i = x)$. In order to implement propensity score methods, the propensity scores for the treated and untreated in our sample must overlap such that $0 < e(x) < 1$. Although the unconfoundedness assumption cannot be tested directly, we can examine whether the propensity score has a causal effect on a pseudo outcome

18

that was determined prior to the treatment. If the estimated effect of the treatment on the pseudo outcome is significant, then unconfoundedness has likely been violated (Imbens 2015). We can evaluate the overlap assumption directly by assessing the balance of the covariates in the treated and untreated groups as well as visually inspecting the overlap in the propensity scores.

Propensity score matching has been widely used in economics and other social sciences (Imbens 2015). However, King and Nielson (2016) and Imbens (2015) note that propensity score estimates break down if the propensity score model fits to the data too well. As a result, we cannot use the review score to estimate the propensity score related to fellowship funding. Thus, we use the coarsened exact matching (CEM) algorithm (Blackwell et al 2009) to improve the balance of the data, and nearest-neighbor methods to facilitate matching on the review score. We use both propensity score matching and nearest neighbor matching after reducing the data using the CEM algorithm.

RESULTS

We begin the analysis by estimating the probability of receiving an NRSA F32 award in the full and analysis samples using probit models; Table 3 reports marginal effects. In addition to covariates listed in the table, all models include controls for institute and council round. In the first column, using the full sample, several covariates predict the likelihood of receiving an NRSA F32 award. Several age dummy variables are statistically significant as well as an indicator for sex being missing in the data. Blacks and Asians are significantly less likely to receive an award, but those indicating Other race are significantly more likely to receive the award. Having an MD, PhD, or MD/PhD degree increases the probability of the award as does having a field in biomedicine or social science. Those applicants who have been predoctoral

19

trainees are significantly more likely to receive an award, but those who have been postdoctoral trainees are not. We include the score in the second column, and many observable characteristics lose statistical significance. Proposals with lower review scores are significantly more likely to receive funding. Older applicants (age greater than 37) are less likely to be funded, as are applicants with sex missing, those who are black, and NRSA T32 postdoc trainees. PhD and MD/PhD applicants are about 9 percentage points (ppt) more likely to receive fellowship awards than individuals who hold other medical professional degrees such as DVMs or DDSs.

The analysis sample deletes observations that have review scores above the 60th percentile or those that have applied to smaller institutes. As Table 2 indicated, the treatment and control groups in the analysis sample are more closely related than in the full sample. After controlling for review score, institute, and council round, only the sex missing variable has a significant negative impact on the likelihood of receiving a fellowship award. We tested the joint hypothesis that the institute and council round fixed effects were significantly different from zero, and rejected that hypothesis (p<.000) in both cases. Thus, our argument that NRSA F32 applicants are relatively homogeneous is supported by the analysis, conditional on the review score, council round, and institute. The probit models in Table 3 are the basis of the propensity score estimates used in our matching models below.

Propensity Score Matching

Table 4 presents the propensity score matching (PSM) estimates of the ATE and ATT for the analysis sample. As mentioned previously, if we include the review score in the propensity score estimates, the propensity score becomes too precise, and the matching algorithm breaks

down (King and Nielson 2016, Imbens 2015). Thus, our propensity score estimates include

institute and council round fixed effects and the covariates listed in column 3 of Table 3. The

first row of Table 4 shows the ATE of the NRSA F32 award on the number of RPG awards,

number of RPG applications, the probability of an RPG, the probability of an R01 and the

probability of never applying for additional funding.   The ATE estimates indicate that the

fellowship award increases the number of RPG awards by 0.15; increases the number of RPG

applications by 0.73; increases the probability of an RPG award by 8.2 ppt; increases the

probability of an R01 award by 6.0 ppt; and decreases the likelihood of never applying for

subsequent NIH funding by 11.2 ppt. The ATT estimates are remarkably similar in size and

magnitude.

In order for the estimates to be considered valid, there should be considerable overlap in

the propensity score estimates of the awardees and non-awardees. Figure 4 Panel A shows this

PS overlap for the analysis sample. The mass of the kernel density estimate of the propensity

score for receiving an award lies to the left of that for not receiving an award. In order to

improve the overlap, Imbens (2015) recommends trimming the tails of the propensity score

distribution below .1 and above .9. The second row of Table 4 presents the ATE estimates after

this trimming. The estimated effects of the fellowship award on the ATE and ATT fall

somewhat after trimming but remain statistically significant. Figure 4 Panel B indicates that the

overlap in propensity scores appears quite similar. Finally, Smith and Todd (2005) recommend

trimming the propensity score in order to optimize the common support. We trimmed the

propensity scores below .33 and above .67. Figure 4 Panel C indicates that the overlap has

improved considerably after this trimming. Our estimated effect of the ATE and ATT remain

significant but are smaller. The ATE estimates after the .33 and .67 trimming indicate that the

fellowship award increases the number of RPG awards by .11, increases the number of RPG applications by .67, increases the probability of an RPG award by 6.7 ppt, increases the probability of an R01 award by 4.8 ppt, and decreases the likelihood of never applying for subsequent NIH funding by 10.2 ppt. As before, the ATT estimates are remarkably similar in size and magnitude.

Although we cannot test the unconfoundedness assumption directly, Imbens (2015) recommends using PSM on pseudo outcomes that occur prior to the award treatment. Given the SED data, we evaluate whether the fellowship award predicts the probability that an applicant has a PhD degree, the applicant's field of highest degree is in biomedicine, and the applicant's doctoral funding was from a fellowship or scholarship. Table 5 presents these results and finds no significant impact of the fellowship award on these pseudo outcomes.

Nearest-Neighbor Matching

Given the sensitivity of the PSM to the propensity score specification and recommendations to use alternative methods by King and Nielson (2016), we used nearest-neighbor matching for the analysis sample. Throughout we use the Mahalanobis distance for nearest-neighbor matching. Let X^T be the set of control variables for those receiving the fellowship award and X^C be the set of control variable for non-awardees. The Mahalanobis distance is given by:

$$D = (X^T - X^C)'V^{-1}(X^T - X^C)$$

and V is the covariance matrix of X. Mahalanobis distance will reduce differences between the treated and control groups by an equal percentage for each covariate in the matrix X.

We experiment with a variety of specifications of the covariate list and matching approach for the nearest neighbor matching. First we match on the review score only since it is a significant determinant of fellowship funding and we could not include it in the PSM approach. Second, we match on the review score, institute and council round. Third, we match on review score and council round, and impose exact matching on institute. Fourth, we match on all of the variables in the fourth column of Table 3 and impose exact matching on the institute. We drop about 1,000 observations that do not have an exact match leaving us with 13,653 in the analysis sample. We concentrate on the estimates of the ATE in the last column of Table 6 where the number of covariates is the largest. The nearest neighbor matches increase the number of RPG awards by .10, the number of RPG applications by .73, the probability of an RPG award by 6.3 ppt, the probability of an R01 award by 4.6 ppt, and reduces the probability of never applying for subsequent NIH funding by 9.8 ppt. These estimates are somewhat smaller than the PSM estimates but still statistically significant.

Blackwell et al (2009) recommend coarsened exact matching (CEM) for sample selection as a way to improve the balance of the matching estimator by exactly matching on categorical values of the data. They show that the coarsened data have significantly improved balance and balances nonlinearities and interactions in the sample. We begin by using the CEM algorithm to perform exact matches by institute and fiscal year of funding. Next we use the nearest neighbor matching and same covariates to estimate the ATE, and these estimates inherit the balance properties from the CEM procedure. These estimates appear in the bottom of Table 6. Despite the sample size reduction from 13,653 to 8,630, the nearest-neighbor matches after the CEM algorithm are remarkably similar to the full analysis sample.

We also use the CEM sample to estimate propensity score matches in Table 7. The first row of Table 7 repeats the ATE estimates of the effect of fellowship awards on our outcomes. The second row reports the ATE after employing the CEM algorithm to improve the balance. The estimated effects of NRSA F32 awards are virtually identical despite the significant sample size reduction. Figure 4 Panel D indicates significant improvements in the overlap of the propensity score estimates.

DISCUSSION

Overall, our findings provide evidence that both the NRSA F32 award and the process used to make funding decisions matter for future engagement with NIH as an independent researcher. The NRSA fellowships for postdoctoral scientists increase the probability that awardees will receive NIH research funding later in their career. These results provide strong evidence to the science policy community and leaders that using mechanisms like the NRSA F32 fellowship to prepare and sustain a biomedical research workforce does achieve the goal of keeping scientists in NIH-funded research careers. These results are also consistent with the recommendations of NIH's Biomedical Workforce Working Group, which suggested that the number of postdoctoral students receiving fellowships should increase (NIH 2012).

A limitation of our study is that it does not allow us to examine the extent to which funding from other agencies or research organizations affects those that do and do not receive NRSA F32 funding. However, if individuals were able to receive support from other sources in early career and then develop their NIH-funded research, we would expect this to bias our estimates downwards. Thus, our results may be considered a lower-bound on the impact of fellowships on subsequent NIH funding and an independent research career.

In Fiscal Year 2016, one NRSA F32 award cost the federal government, on average, around $60,000 per individual. If a fellowship usually lasts, on average, two years, then the total cost to NIH of funding training via the NRSA F32 mechanism is around $120,000 per individual. Our study shows that making that kind of investment today in young scientists increases the chance that they remain in a scientific research career.

Congress used to require the National Research Council to conduct a periodic review of the impact of the NRSA training program (National Academies 2011). Furthermore, several researchers and policymakers have noted an increasing length of time spent in training as graduate students or postdocs (Alberts et. al. 2015, Tilghman & Rockey 2012). These reports have called for policies designed to promote researcher independence. One viable option would be to expand the number of NRSA F32 fellowship awards available.

CONCLUSION

During the time of our study, the total number of NRSA F32 awards granted decreased; however, their impact as an early career training mechanism to keep individuals in NIH-funded, future, independent science careers was robust. Of interest to note, the number of NRSA F32 applications during the NIH doubling years decreased. The decrease in applicants occurred when funding for postdoctoral researchers on R01 and RPG research grants became more available. Postdoctoral appointments on RPG and R01 grants do not require postdocs to develop a research proposal distinct from their mentors' research in order to be appointed, and this may influence the direction of their careers.

Our study is related to work where the NRSA F32 award was modeled as a regression discontinuity and the peer review scoring mechanism was used as a cut off to determine the

causal effect of NIH funding at the margin (Jacob and Lefgren 2011a). We found that these methods have significant limitations. NIH institutes and centers take their job seriously with staff devoting time, energy, and effort to selecting candidates who have the right mix of institutional support, innovative ideas, and alignment with institute scientific priorities. Candidates are not solely selected based on the review score. Proposals with competitive scores are skipped over for administrative reasons as well, either the package is not complete or additional review is required because of human subjects research requirements or other complexities that may delay the application. This means, however, that at least for the NRSA F32 fellowships, regression discontinuity is not an optimal method for this analysis.

Using matching methods, we find robust results that overall the NRSA F32 keeps postdoctoral researchers in NIH-funded science at higher rates than they would otherwise experience. Along the margin of being funded or not, the NRSA postdoctoral fellowship award mechanism significantly improves the probability of receiving subsequent NIH funding and launching an independent research career.

Our findings provide that targeted and focused NIH training programs associated with independent research ideas can have a significant impact on keeping individuals in academic science and engaged in NIH-funded research. The F32 programs can be seen as a viable mechanism for this purpose. Future research should address the issues of cost-benefit and whether the amount of investment in these trainings programs matches the return.

REFERENCES

Blackwell, M. Iacus, S., King, G. and Porro, G. (2009) CEM: Coarsened Exact Matching in Stata. The Stata Journal. 9: 524-546.

Carnevale, A.P., Rose, S.J., & Cheah, B. (2011). The college payoff: Education, occupations, lifetime earnings. Georgetown University Center on Education and the Workforce, Washington, D.C.

Couzin, J. & Miller, G. (2007). Boom and bust. *Science*, April 20, http://science.sciencemag.org/content/316/5823/356.full.

Fiegener, Mark (2011). Numbers of doctorates awarded in the United States declined in 2010. NSF Report No. 12-303, National Science Foundation, Arlington, VA. https://www.nsf.gov/statistics/infbrief/nsf12303/.

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

Imbens, G. (2015). Matching Methods in Practice. Journal of Human Resources, 50(2): 373-419.

Jacob, R., & Zhu, P. (2012). A practical guide to regression discontinuity. Manpower Demonstration Research Corporation (MDRC), Washington, D.C. Retrieved May 9, 2017 from, http://www.mdrc.org/sites/default/files/regression_discontinuity_full.pdf.

Jacob, B.A., & Lefgren, L. (2011a). The impact of NIH postdoctoral training grants on scientific productivity, Research Policy, 40(6), 864-874.

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

Kahn, S., & Ginther D.K. (2017). The impact of postdoctoral training on early careers in biomedicine, Nature Biotechnology, 35, 90-94.

King, G. & Nielson R. (2016). Why propensity scores should not be used for matching. Available online at http://j.mp/2ovYGsW.

Lee, D.S., & Lemieux, T. (2010). Regression discontinuity designs in economics, Journal of Economic Literature, 48, 281-355.

Levitt D.G. (2010). Careers of an elite cohort of US basic life science postdoctoral fellows and the influence of their mentor's citation record. BMC Medical Education, 10(80), 1-7. Retrieved May 9, 2017 from, http://www.biomedcentral.com/1472-6920/10/80.

Manski, C.F., & Wise, D.A. (1983). College choice in America. Cambridge, Mass: Harvard University Press.

Mantovani, R., Look, M.V., & Wuerker, E. (2006). The career achievements of National Research Service Award postdoctoral trainees and fellows: 1975-2004. (Prepared for the National Institutes of Health, Office of the Director, Office of Extramural Programs). ORC Macro, Bethesda, MD.

Mendoza, P., Villarreal III, P., & Gunderson, A. (2014). Within-year retention among Ph. D. students: The effect of debt, assistantships, and fellowships. Research in Higher Education, 55(7), 650-685.

National Institutes of Health. (2001). Senate report language associated with the authorizing legislation at 42USC288 (RTF - 13 KB). Bethesda, MD. Retrieved May 9, 2017 from, https://researchtraining.nih.gov/resources/policy-notices.

National Institutes of Health. (2002). TITLE 42. The public health and welfare chapter 6a. The public health service national research institutes awards and training. Bethesda, MD. Retrieved May 9, 2017 from, https://researchtraining.nih.gov/resources/policy-notices.

National Institutes of Health. (2003). Ruth L. Kirschstein national research service awards for individual postdoctoral fellows (F32). NIH Guide Publication No. PA-03-067. Bethesda, MD. Retrieved May 9, 2017 from, http://grants.nih.gov/grants/guide/pa-files/PA-03-067.html.

National Institutes of Health (NIH). (2011a). Enhancing peer review. Bethesda, MD. Retrieved May 9, 2017 from, https://enhancing-peer-review.nih.gov/index.html.

National Institutes of Health (NIH). (2011b). Grants policy statement. Bethesda, MD. Retrieved May 9, 2017 from, http://grants.nih.gov/grants/policy/nihgps_2011/nihgps_ch11.htm#_Toc271265032.

National Institutes of Health (NIH) Office of Budget (2017). History of congressional appropriations, fiscal years 2000 – 2016. Bethesda, MD. Retrieved May 9, 2017 from, https://officeofbudget.od.nih.gov/pdfs/FY16/Approp%20History%20by%20IC%20FY%202000%20-%20FY%202016.pdf.

National Research Council (US) (2011) *Research Training in the Biomedical, Behavioral, and Clinical Research Sciences* (National Academies Press, Washington, DC).

National Science Board (2014). Science and engineering indicators 2014. National Science Foundation 14-01, Arlington VA.

National Science Foundation (2007). Correspondence to Alyson Reed, Executive Director of the National Postdoctoral Association. Arlington, VA. Retrieved May 9, 2017 from https://grants.nih.gov/training/Reed_Letter.pdf.

Nerad M., & Cerny J. (1999). Postdoctoral patterns, career advancement, and problems. *Science*, 285, 1533–1535.

Pion, G.M. (2001). The early career progress of NRSA predoctoral trainees and fellows. (Prepared for U.S. Department of Health and Human Services, National Institutes of Health, NIH Publication No. 00-4900).

Rockey, S. (2014). A look at programs targeting new scientists. *NIH Office of Extramural Research Rock Talk Blog*, April 29, https://nexus.od.nih.gov/all/2014/04/29/a-look-at-programs-targeting-new-scientists/.

Science Careers Editors. (2007). The NIH R01 toolkit. *Science*, July 27, http://www.sciencemag.org/careers/2007/07/nih-r01-toolkit.

Smith, J.,and P. Todd (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?," *Journal of Econometrics*, 125(1-2), 305–353.

Su X. (2013). The impacts of postdoctoral training on scientists' academic employment. Journal of Higher Education, 84(2), 239–265.

Teitelbaum, M.S. (2014). Falling behind? Boom, bust, and the global race for scientific talent. Princeton, New Jersey: Princeton University Press.

Tilghman, S., & Rockey S. (2012). Report of the Biomedical Research Workforce Working Group of the Advisory Committee to the NIH Director. Retrieved on August 15, 2014 from, http://acd.od.nih.gov/bwf.htm.
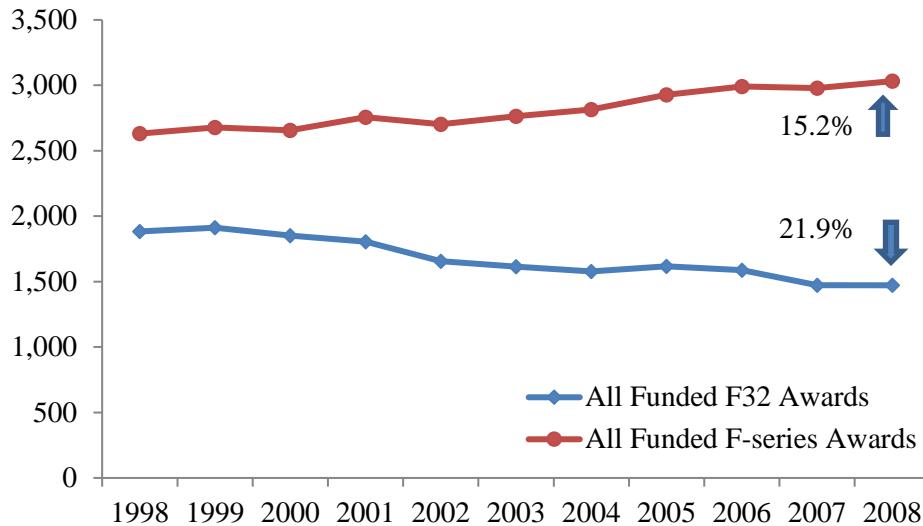
Van Der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. International Economic Review, 43(4), 1249-1288.

Webber, K.L. (2013). Cultivating the future of graduate education? Factors contributing to salary for recent doctoral degree recipients. Planning for Higher Education, 41(4), 1-7.

Zolas, N., Goldschlag, N., Jarmin, R., Stephan, P., Owen-Smith, J., Rosen, R.F., Allen, B.M., Weinberg, B.A. & Lane, J.I. (2015). Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients. Science, 350(6266), 1367-1371.

Figure 1. The Number of Total NRSA F-series Awards and F32 Awards Funded by the National Institutes of Health by Year, 1998 to 2008

Panel A. Total funded awards in by award type and fiscal year

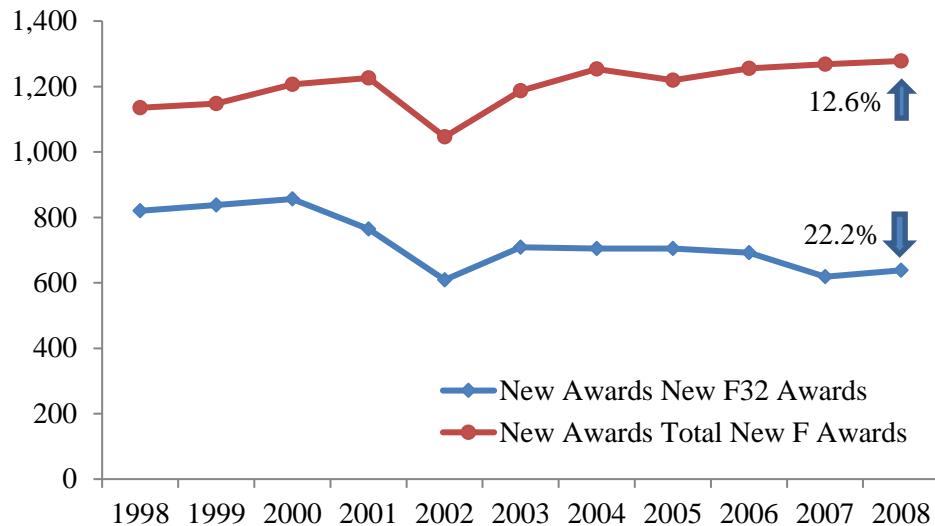Panel B. Total newly funded awards by award type (all F-series and F32) and fiscal year
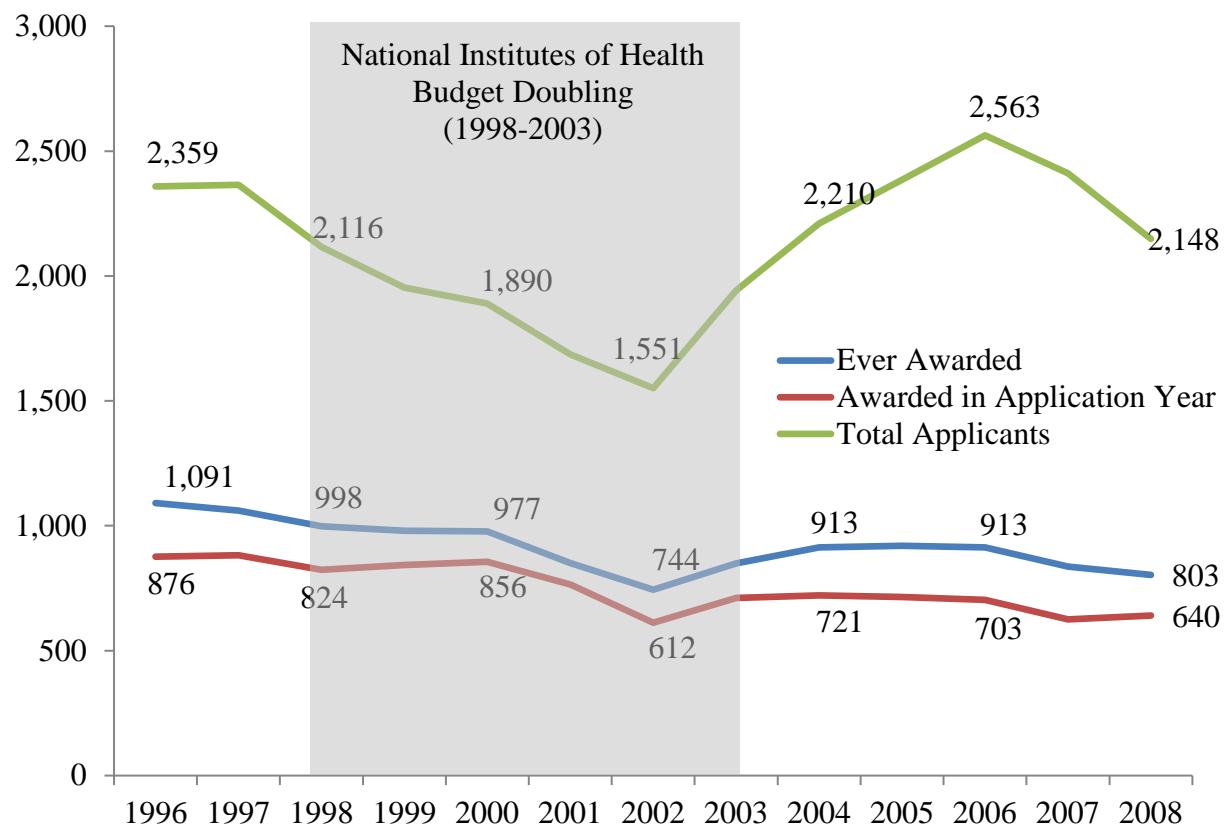
Source: NIH Research Portfolio Online Reporting Tool (RePORT): http://report.nih.gov/fundingfacts/fundingfacts.aspx (accessed on November 19, 2017)

Note: In any given year, NIH agrees to fund additional years of training from previously accepted awards (non-compete awards within that year), competitive new awards that are newly awarded, and other awards related to administrative extensions and other factors. F32 fellowship
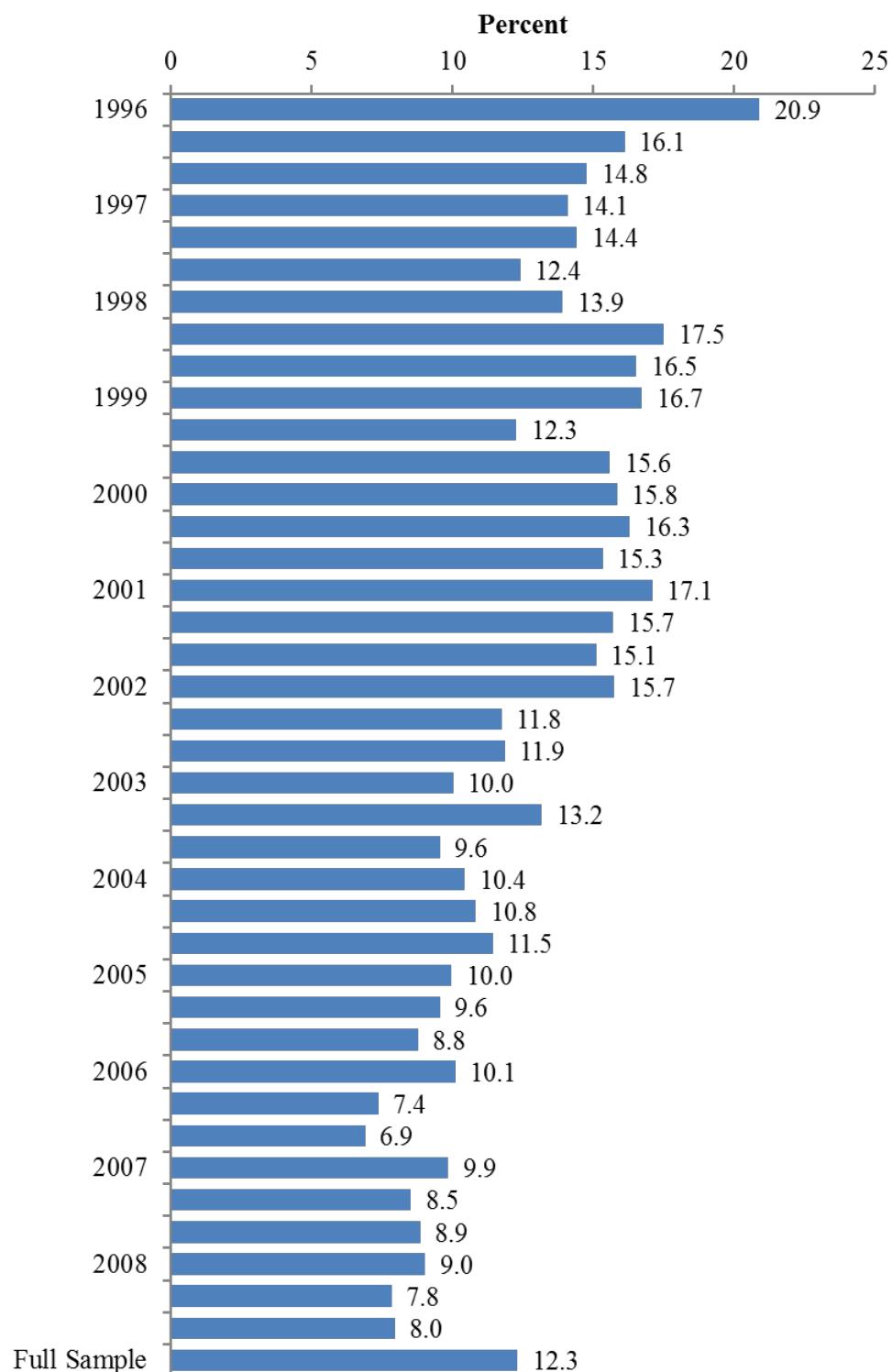
awards are generally given for two to three years. The first year of the award is considered a new, competitive awards and all following years of funding are classified as non-compete awards. Panel A shows all awards, regardless of competitive status. Panel B shows only the number of new competitive grants awarded in each year. Years 1996 and 1997 are not shown because they are unavailable in RePORT.

Figure 2. Total new NRSA F32 applicants, individuals ever awarded, and individuals awarded in same application year by fiscal year of application, 1996 to 2008

Source: Authors' calculations using NIH IMPACII, 1996-2008

Figure 3. Percent of Applicants Selected Out of Order by Year-Council Round, 1996 to 2008

Percent

Year	Value
1996	20.9
	16.1
	14.8
1997	14.1
	14.4
	12.4
1998	13.9
	17.5
	16.5
1999	16.7
	12.3
	15.6
2000	15.8
	16.3
	15.3
2001	17.1
	15.7
	15.1
2002	15.7
	11.8
	11.9
2003	10.0
	13.2
	9.6
2004	10.4
	10.8
	11.5
2005	10.0
	9.6
	8.8
2006	10.1
	7.4
	6.9
2007	9.9
	8.5
	8.9
2008	9.0
	7.8
	8.0
Full Sample	12.3
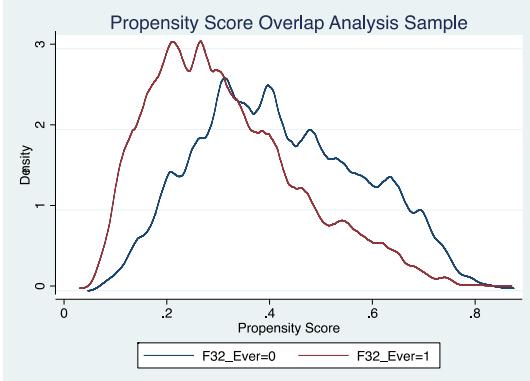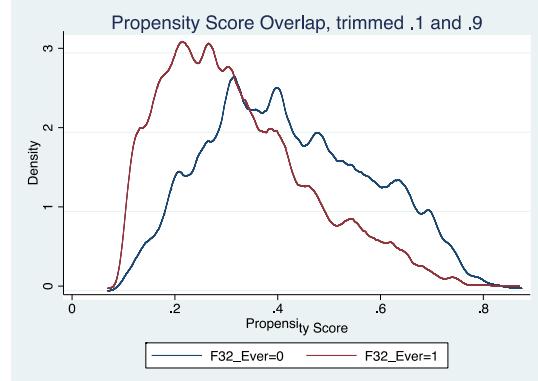
Source: National Institutes of Health IMPACII, 1996 to 2008
Note: Figure excludes council rounds with an N<20.

Figure 4: Score overlap by F32 Award

Panel A.

Propensity Score Overlap Analysis Sample

Panel B.

Propensity Score Overlap, trimmed .1 and .9

Panel C.

Propensity Score Overlap trimmed .33 and .67

Panel D.

Propensity Score Overlap CEM Matching

Source: National Institutes of Health IMPACII, 1996 to 2008

Note: Panel A is the full analysis sample. Panel B is the analysis sample with score > .1 and score < .9. Panel C is the analysis sample with score > .33 and score < .67. Panel D is the CEM sample.

Table 1. Descriptive Statistics of NRSA F32 Applicants by Funding Status, 1996 to 2008

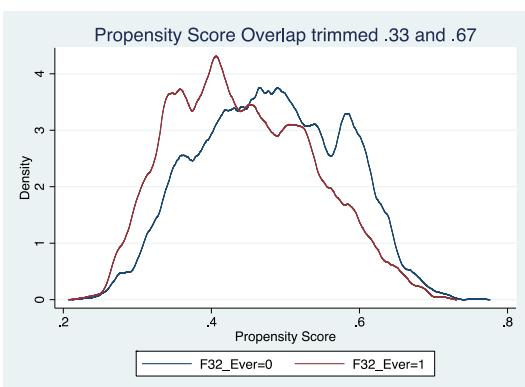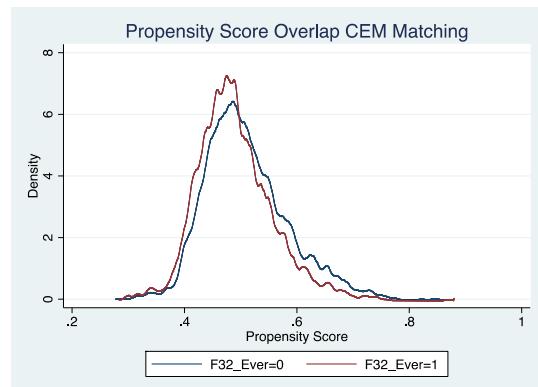|	All	F32 Awarded	No F32 Awarded	t-test	p-value
Review score	220.67	175.81	258.38	104.64	0.000
	(75.291)	(44.581)	(75.162)		
DEMOGRAPHICS					
Age at application	31.201	30.989	31.363	4.04	0.000
	(7.629)	(6.776)	(8.218)		
Age at application missing	0.042	0.033	0.048	6.00	0.000
	(0.200)	(0.180)	(0.214)		
Married at application	0.381	0.394	0.371	-3.97	0.000
	(0.486)	(0.489)	(0.483)		
Married at application missing	0.205	0.181	0.223	8.62	0.000
	(0.404)	(0.385)	(0.416)		
Female	0.417	0.417	0.416	-0.19	0.850
	(0.493)	(0.493)	(0.493)		
Sex missing	0.051	0.037	0.061	9.18	0.000
	(0.219)	(0.188)	(0.240)		
White, non-Hispanic	0.344	0.384	0.314	-12.17	0.000
	(0.475)	(0.486)	(0.464)		
Black, non-Hispanic	0.009	0.006	0.011	4.60	0.000
	(0.095)	(0.077)	(0.106)		
Asian, non-Hispanic	0.086	0.084	0.087	0.97	0.330
	(0.280)	(0.277)	(0.282)		
Other, non-Hispanic	0.002	0.003	0.002	-1.97	0.049
	(0.048)	(0.054)	(0.042)		
Hispanic	0.032	0.029	0.034	2.46	0.014
	(0.175)	(0.167)	(0.181)		
Race missing	0.544	0.512	0.569	9.40	0.000
	(0.498)	(0.500)	(0.495)		
EDUCATION and TRAINING					
MD	0.086	0.080	0.090	3.09	0.002
	(0.280)	(0.271)	(0.286)		
MD/PhD	0.032	0.034	0.031	-1.50	0.134
	(0.176)	(0.181)	(0.172)		
PhD	0.867	0.876	0.860	-3.95	0.000
	(0.340)	(0.330)	(0.347)		
Other Degree	0.016	0.011	0.020	5.93	0.000
	(0.125)	(0.103)	(0.139)		
Biomedical degree	0.594	0.619	0.575	-7.40	0.000

	(0.491)	(0.486)	(0.494)		
Physical Science degree	0.129	0.124	0.133	2.20	0.028
	(0.335)	(0.329)	(0.339)		
Social Science degree	0.069	0.074	0.065	-2.69	0.007
	(0.253)	(0.261)	(0.247)		
Prior T32 Predoc appointment	0.021	0.025	0.018	-3.74	0.000
	(0.144)	(0.156)	(0.134)		
Prior T32 Postdoc appointment	0.019	0.016	0.021	2.76	0.006
	(0.136)	(0.127)	(0.143)		
Prior NRSA Predoctoral Fellowship	0.001	0.001	0.000	-0.51	0.612
	(0.023)	(0.024)	(0.021)		
OUTCOME VARIABLES					
Number of RPG Awards	0.387	0.555	0.259	-22.95	0.000
	(1.071)	(1.243)	(0.894)		
Number of RPG Applications	1.940	2.664	1.388	-24.32	0.000
	(4.36)	(5.03)	(3.68)		
Probability of RPG	0.183	0.256	0.126	-28.01	0.000
	(0.386)	(0.437)	(0.332)		
Probability of R01	0.133	0.192	0.088	-25.56	0.000
	(0.339)	(0.394)	(0.283)		
Probability of Never Receiving an RPG	0.678	0.581	0.752	30.59	0.000
	(0.467)	(0.493)	(0.432)		
N	27,580	11,938	15,642		

Source: National Institutes of Health IMPACII and NIH/NSF Survey of Earned Doctorates, 1996 to 2008

Table 2. Descriptive Statistics of NRSA F32 Applicants by Funding Status Analysis Sample, 1996 to 2008

	All	F32 Awarded	No F32 Awarded	t-test	p-value
Review score	220.67	162.244	193.68	60.219	0.000
	(75.291)	[26.899]	[34.432]		
DEMOGRAPHICS					
Age at application	31.201	30.942	31.078	1.124	0.261
	(7.629)	[6.775]	[7.115]		
Age at application missing	0.042	0.034	0.037	0.977	0.328
	(0.200)	[0.181]	[0.189]		
Married at application	0.381	0.396	0.38	-1.841	0.066
	(0.486)	[0.489]	[0.485]		
Married at application missing	0.205	0.182	0.186	0.683	0.494
	(0.404)	[0.386]	[0.389]		
Female	0.417	0.412	0.424	1.3	0.194
	(0.493)	[0.492]	[0.494]		
Sex missing	0.051	0.038	0.055	4.9	0.000
	(0.219)	[0.190]	[0.228]		
White, non-Hispanic	0.344	0.386	0.319	-7.932	0.000
	(0.475)	[0.487]	[0.466]		
Black, non-Hispanic	0.009	0.006	0.01	2.643	0.008
	(0.095)	[0.075]	[0.098]		
Asian, non-Hispanic	0.086	0.085	0.079	-1.292	0.196
	(0.280)	[0.279]	[0.269]		
Other, non-Hispanic	0.002	0.002	0.002	-0.327	0.744
	(0.048)	[0.050]	[0.047]		
Hispanic	0.032	0.029	0.029	-0.068	0.946
	(0.175)	[0.168]	[0.167]		
Race missing	0.544	0.509	0.579	7.972	0.000
	(0.498)	[0.500]	[0.494]		
EDUCATION and TRAINING					
MD	0.086	0.084	0.08	-0.741	0.459
	(0.280)	[0.277]	[0.272]		
MD/PhD	0.032	0.035	0.03	-1.796	0.073
	(0.176)	[0.185]	[0.170]		
PhD	0.867	0.87	0.874	0.665	0.506
	(0.340)	[0.336]	[0.332]		
Other Degree	0.016	0.01	0.016	2.76	0.006
	(0.125)	[0.102]	[0.125]		

37

Biomedical degree	0.594	0.614	0.617	0.334	0.739
	(0.491)	[0.487]	[0.486]		
Physical Science degree	0.129	0.129	0.124	-0.721	0.471
	(0.335)	[0.335]	[0.330]		
Social Science degree	0.069	0.074	0.071	-0.666	0.506
	(0.253)	[0.261]	[0.256]		
Prior T32 Predoc Award	0.021	0.024	0.027	0.957	0.339
	(0.144)	[0.154]	[0.162]		
Prior T32 Postdoc Award	0.019	0.016	0.02	1.648	0.099
	(0.136)	[0.127]	[0.141]		
Prior NRSA Predoctoral Fellowship	0.001	0.001	0	-1.642	0.101
	(0.023)	[0.023]	[0]		
OUTCOME VARIABLES					
Number of RPG Awards	0.387	0.586	0.367	-10.237	0.000
	(1.071)	[1.281]	[1.095]		
Number of RPG Applications	1.940	2.752	1.698	-12.506	0.000
	(4.36)	[5.158]	[4.066]		
Probability of RPG	0.183	0.266	0.165	-13.692	0.000
	(0.386)	[0.442]	[0.372]		
Probability of R01	0.133	0.204	0.122	-12.245	0.000
	(0.339)	[0.403]	[0.328]		
Probability of Never Receiving an RPG	0.678	0.579	0.713	16.02	0.000
	(0.467)	[0.494]	[0.452]		
N	27,580	9,276	5,000		

Source: National Institutes of Health IMPACII and NIH/NSF Survey of Earned Doctorates, 1996 to 2008

Table 3. Probit Regressions on Ever Receiving an NRSA F32 Award, 1996 to 2008

	(1) Full Sample	(2) Full Sample	(3) Analysis Sample	(4) Analysis Sample
Review score		-0.006*** [0.000]		-0.008*** [0.000]
DEMOGRAPHICS				
Age (missing = <26)				
Age = 27	-0.107*** [0.024]	-0.033 [0.030]	-0.053 [0.036]	0.021 [0.039]
Age = 28	-0.04 [0.022]	-0.031 [0.027]	-0.034 [0.030]	-0.028 [0.036]
Age = 29	-0.056** [0.021]	-0.022 [0.026]	-0.05 [0.029]	-0.016 [0.035]
Age = 30	-0.061** [0.021]	-0.013 [0.026]	-0.041 [0.029]	0.002 [0.034]
Age = 31	-0.072*** [0.020]	-0.031 [0.026]	-0.035 [0.029]	0.024 [0.033]
Age = 32	-0.088*** [0.020]	-0.039 [0.026]	-0.060* [0.030]	-0.005 [0.034]
Age = 33	-0.086*** [0.021]	-0.017 [0.027]	-0.045 [0.030]	0.027 [0.033]
Age = 34	-0.105*** [0.021]	-0.031 [0.027]	-0.055 [0.031]	0.016 [0.035]
Age = 35 or 36	-0.149*** [0.020]	-0.035 [0.026]	-0.057 [0.031]	0.028 [0.034]
Age = 37 or 38	-0.181*** [0.020]	-0.068* [0.027]	-0.085* [0.034]	0.014 [0.036]
Age > 38	-0.248*** [0.017]	-0.104*** [0.026]	-0.137*** [0.034]	-0.019 [0.037]
Marital status (missing = not married)				
Married	0.009 [0.007]	0.014 [0.008]	0.016 [0.009]	0.018 [0.010]
Marital status missing	0.093 [0.062]	0.057 [0.090]	0.044 [0.085]	0.086 [0.092]
Sex (missing = male)				
Female	-0.01 [0.006]	-0.008 [0.007]	-0.009 [0.009]	-0.006 [0.009]
Sex missing	-0.120*** [0.014]	-0.098*** [0.017]	-0.115*** [0.023]	-0.103*** [0.026]

Race and ethnicity (missing = White, non-Hispanic)

Black, non-Hispanic	-0.156***	-0.099**	-0.146**	-0.103
	[0.030]	[0.036]	[0.055]	[0.055]
Asian, non-Hispanic	-0.038***	-0.021	-0.025	-0.016
	[0.011]	[0.013]	[0.016]	[0.017]
Other race, non-Hispanic	0.135*	0.167*	0.064	0.048
	[0.063]	[0.078]	[0.074]	[0.074]
Hispanic	-0.022	0.001	0.014	0.033
	[0.018]	[0.021]	[0.025]	[0.025]
Race missing	0.005	0.013	0.011	0.016
	[0.008]	[0.009]	[0.010]	[0.011]
EDUCATION AND TRAINING				
MD	0.088**	0.033	-0.009	-0.019
	[0.030]	[0.033]	[0.042]	[0.045]
MD/PhD	0.189***	0.098**	0.076	0.016
	[0.031]	[0.037]	[0.039]	[0.046]
PhD	0.126***	0.093***	0.06	0.049
	[0.024]	[0.027]	[0.039]	[0.041]
Biomedical science degree	0.145*	0.065	0.041	0.073
	[0.059]	[0.086]	[0.088]	[0.101]
Physical science degree	0.107	0.054	0.052	0.094
	[0.062]	[0.090]	[0.084]	[0.089]
Social science degree	0.133*	0.022	0.017	0.031
	[0.062]	[0.089]	[0.087]	[0.097]
Prior T32 Predoc Award	0.067**	-0.004	-0.024	-0.045
	[0.022]	[0.025]	[0.026]	[0.028]
Prior T32 Postdoc Award	-0.061**	-0.056*	-0.048	-0.033
	[0.022]	[0.025]	[0.031]	[0.033]
Prior NRSA Predoc Award	-0.002	0.128		
	[0.140]	[0.260]		
Observations	27,504	25,719	14,268	14,268

Source: IMPACII and NIH/NSF Survey of Earned Doctorates, 1996 to 2008

Note: Robust standard errors in brackets. *** p<0.001, ** p<0.01, * p<0.05 All specifications include controls for IC and council rounds.

Table 4. Propensity Score Matching Estimates

VARIABLES	Number of RPG Awards	Number of RPG Applications	Probability RPG	Probability R01	Probability Never RPG
ATE					
Full Sample	0.149***	0.731***	0.082***	0.060***	-0.112***
	[0.026]	[0.098]	[0.009]	[0.008]	[0.010]
Trimmed .1,.9	0.127***	0.584***	0.064***	0.048***	-0.098***
	[0.026]	[0.109]	[0.009]	[0.008]	[0.011]
Trimmed .33, .67	0.111***	0.671***	0.067***	0.048***	-0.102***
	[0.030]	[0.129]	[0.011]	[0.010]	[0.013]
ATT					
Full Sample	0.159***	0.775***	0.088***	0.065***	-0.118***
	[0.030]	[0.113]	[0.010]	[0.009]	[0.012]
Trimmed .1,.9	0.143***	0.605***	0.068***	0.050***	-0.100***
	[0.031]	[0.134]	[0.011]	[0.010]	[0.013]
Trimmed .33, .67	0.117**	0.642***	0.070***	0.049***	-0.108***
	[0.036]	[0.160]	[0.012]	[0.011]	[0.015]
Observations	14,273	14,273	14,273	14,273	14,273

Full sample = 14, 273 observations

Trimmed 1 = 14,021

Trimmed 2 = 6,516

Source: IMPACII and NIH/NSF Survey of Earned Doctorates, 1996 to 2008

Note: Robust Standard errors in brackets. *** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table 5. Counterfactual Treatment Effects with Pseudo Treatments

VARIABLES	(1) PhD Degree	(3) Biomedical Degree	(5) Fellowship or Scholarship PhD Funding
ATE	0.007 [0.007]	0.001 [0.010]	0.005 [0.010]
ATT	0.005 [0.008]	0.007 [0.012]	0.004 [0.012]
Observations	14,273	14,273	14,273

Source: IMPACII and NIH/NSF Survey of Earned Doctorates, 1996 to 2008

Note: Robust Standard errors in brackets. *** p<0.001, ** p<0.01, * p<0.05

Table 6. Nearest Neighbor Full Sample and CEM Sample ATE Estimates (Differ by specification of matching covariates)

VARIABLES	(1)	(2)	(3)	(4)
Number RPG Awards	0.025	0.072*	0.050	0.101**
	[0.028]	[0.036]	[0.037]	[0.031]
Number RPG Applications	0.428***	0.483***	0.366*	0.726***
	[0.102]	[0.135]	[0.146]	[0.108]
Probability RPG Award	0.040***	0.044***	0.041***	0.063***
	[0.009]	[0.012]	[0.012]	[0.010]
Probability R01 Award	0.026**	0.034**	0.031**	0.046***
	[0.008]	[0.011]	[0.011]	[0.009]
Probility of Never Applying	-0.077***	-0.079***	-0.073***	-0.098***
	[0.010]	[0.014]	[0.014]	[0.011]
Observations	13,653	13,653	13,653	13,653

Coarsened Exact Matching--Nearest Neighbor Estimates

	(1)	(2)	(3)	(4)
Number RPG Awards	-0.045	0.044	0.045	0.116***
	[0.034]	[0.037]	[0.037]	[0.030]
Number RPG Applications	0.156	0.378**	0.378**	0.777***
	[0.125]	[0.144]	[0.143]	[0.115]
Probability RPG Award	0.015	0.039**	0.039**	0.063***
	[0.011]	[0.013]	[0.013]	[0.011]
Probability R01 Award	0.010	0.032**	0.033**	0.051***
	[0.010]	[0.012]	[0.012]	[0.010]
Probability of Never Applying	-0.046***	-0.070***	-0.070***	-0.092***
	[0.014]	[0.016]	[0.016]	[0.012]
Observations	8,630	8,630	8,630	8,630

Source: IMPACII and NIH/NSF Survey of Earned Doctorates, 1996 to 2008

Note: Robust Standard errors in brackets. *** p<0.001, ** p<0.01, * p<0.05

Table 7. ATE Propensity Score Estimates, Full Sample and CEM Sample.

	Number of RPG Awards	Number of RPG Applications	Probability of RPG	Probability of R01	Probability of Never RPG
Full Sample	0.149***	0.731***	0.082***	0.060***	-0.112***
	[0.026]	[0.098]	[0.009]	[0.008]	[0.010]
Observations	14,273	14,273	14,273	14,273	14,273
CEM Sample	0.145***	0.767***	0.076***	0.061***	-0.099***
	[0.029]	[0.114]	[0.010]	[0.009]	[0.012]
Observations	8,630	8,630	8,630	8,630	8,630

Source: IMPACII and NIH/NSF Survey of Earned Doctorates, 1996 to 2008

Note: Robust Standard errors in brackets. *** $p<0.001$, ** $p<0.01$, * $p<0.05$