Research Policy 40 (2011) 864–874

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

The impact of NIH postdoctoral training grants on scientific productivity

Brian A. Jacob[a], Lars Lefgren[b,*]

[a] Gerald R. Ford School of Public Policy, University of Michigan, 735 S. State Street, Ann Arbor 48109-3091, United States
[b] Department of Economics, Brigham Young University, 130 Faculty Office Building, Provo, UT 84602-2363, United States

ARTICLE INFO

Article history:
Received 11 June 2010
Received in revised form 5 April 2011
Accepted 5 April 2011
Available online 7 May 2011

Keywords:
Training grants
NIH
F32
R&D
Scientific productivity

ABSTRACT

In this paper, we estimate the impact of receiving an NIH postdoctoral training grant on subsequent publications and citations. Our sample consists of all applications for NIH postdoctoral training grants (unsuccessful as well as successful) from 1980 to 2000. Both ordinary least squares and regression discontinuity estimates show that receipt of an NIH postdoctoral fellowship leads to about one additional publication over the next five years, which reflects a 20% increase in research productivity.

1. Introduction

In recent years, the size of PhD cohorts in the life sciences has increased sharply. According to the Survey of Earned Doctorates,[1] between 1987 and 2007 the number of individuals earning life science doctorates nearly doubled even as the overall population in the United States increased only by one quarter over the same period. Despite the large influx of young researchers into the life-sciences, the median age of first-time recipients of NIH research grants has increased from 37 in 1980 to 42.[2]

The increased investment in the production of life-science PhDs, combined with aging of the population of NIH grant recipients, underscores the difficulty that individuals increasingly have in making the transition from doctoral student to independent researcher. Pion (2001) provides direct evidence regarding this transition. She reports that 20% of biomedical PhDs who graduated in 1993 or 1994 no longer worked in a research position by 1995.[3] Similarly, only about 40% had applied for a NIH or NSF grant within 10 years of completing their PhD.

The difficulty of this transition is troubling for several reasons. First, graduate training represents a tremendous financial investment by taxpayers, universities, and students. In 2009 the average cost of tuition, fees, and living expenses associated with graduate study of the biological sciences in a high quality program was approximately $51,000 per year.[4] In 2006, 8537 biomedical researchers completed their PhDs. If the average length of time to complete a biomedical PhD is five years, the annual cost of training new biomedical researchers is close to $2.2 billion. Given this enormous financial investment, it is imperative that we maximize the social return by ensuring that new PhDs make a seamless transition to a productive research career. Second, Stephan and Levin (1989) report that the research productivity of life scientists is greatest prior to the age of forty. Delays in beginning an independent research career may therefore reduce the productivity of new scientists, and create a mismatch between the period where NIH funding is available and when the scientist can make the greatest use of the funding. Finally, the difficulty of this transition may discourage potentially excellent researchers from pursuing graduate training in the biomedical field.

The NIH has taken steps to ease the transition from graduate school to an independent research career. For example, in 2007 the NIH instituted a numerical quota for the number of awards

* Corresponding author. Tel.: +1 801 422 5169; fax: +1 801 422 0194.
E-mail addresses: bajacob@umich.edu (B.A. Jacob), l-lefgren@byu.edu, lars_lefgren@byu.edu (L. Lefgren).

[1] See Doctorate Recipients from United States Universities Selected Tables 2007.
[2] See Kaiser (2008). The median age of all grant recipients has likewise increased.
[3] According to Pion “Research career positions are defined as either: (a) holding a faculty position in an institution with one or more biomedical doctoral programs ranked in the 1995 Research Doctorate Study; (b) working in a nonacademic job for which research is the primary responsibility; or (c) being in a postdoctoral training appointment”.

[4] This was calculated by averaging the reported tuition, fee, and living expenses of top 25 graduate programs in the biological sciences. We used the program rankings reported by US News and World Report in 2007. See http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-biological-sciences-programs/rankings accessed on January 21, 2010.

0048-7333/$ – see front matter
doi:10.1016/j.respol.2011.04.003

made to new researchers in order to help new researchers pursue an independent research agenda. The NIH also makes career development awards (K awards) to researchers who have recently completed their graduate training. Finally, the NIH awards postdoctoral research fellowships (F32's) that are designed to facilitate the transition to a research career.

Unfortunately, the benefits of such programs are unclear. For example, if resources are allocated to those young researchers already well on their way to a productive career, the marginal impact of the interventions may be small. Alternatively, the size of a postdoctoral fellowship may be insufficient to meaningfully boost a young researcher's productivity. As a consequence, careful empirical evidence is required to assess the effectiveness of these interventions.

In this paper, we examine the impact of NIH F32 postdoctoral research fellowships on the subsequent career outcomes of recipients. Our sample includes all successful and unsuccessful applicants for F32 grants between 1980 and 2000. Unlike the NSF and most foundations, the NIH allocates research funding in a largely formulaic way on the basis of priority scores derived from independent scientific reviews. This results in a highly nonlinear relationship between a proposal's priority score and the likelihood it is funded. We use this nonlinearity to estimate the causal effect of funding on a variety of outcomes, including publications, citations, and future research funding. The intuition underlying our approach is that individuals on either side of the cutoff for funding will be extremely similar in all relevant respects, except that those applicants who score just above the cutoff receive a NIH postdoctoral fellowship while those just below the cutoff do not.

We find that NIH postdoctoral fellowships increase research output on a variety of dimensions. For example, receiving an NIH postdoctoral fellowship increases five-year publication rates from 4.6 to about 5.2, or roughly 20%. It also increases the likelihood the recipient has five or more publications by 24%. The effect is similarly large for first-author publications.

While our results suggest that NIH postdoctoral grants provide substantial benefits to the young scholars who receive the awards, these estimates will not capture benefits that accrue to other researchers, or to society at large. Our estimates reflect the effect of an NIH postdoctoral fellowship *relative* to the applicant's next best option. If this outside option is also very good, the observed effect of an NIH postdoctoral fellowship will appear small. Yet, assuming that the existence of NIH fellowships do not completely "crowd out" other fellowship opportunities, the F32 program increases the total number of research opportunities for recent graduates, and thus benefits even those applicants who do not receive a NIH fellowship themselves. To borrow terminology from economics, our estimates capture the partial, not general, equilibrium effects of the policy.[5]

The reminder of our paper proceeds as follows. In Section 2, we review the prior literature in this area. Section 3 provides important background information on the NIH funding process. Section 4 describes the data we use and how we construct our analysis sample. In Section 5, we outline our empirical methodology. In Section 6, we present our findings. In Section 7, we discuss the policy implications of our results and conclude.

2. Prior literature

A few earlier studies have examined the effect of NIH funding provided on the career outcomes of participants in training programs. Pion (2001) compared the career outcomes of participants in NIH sponsored pre-doctoral training programs to two other groups of doctoral students—students in the same university as the NIH-sponsored fellows but who did not receive an NIH fellowship and students at institutions that did not receive NIH support. She finds that NIH pre-doctoral support had little impact on an individual's research productivity (as measured by publications and citations) or success at applying for NIH and NSF research support. However, due to questions regarding the similarity of the comparison groups, it is hard to interpret these results.

In their evaluation of NIH career development awards, Carter et al. (1987) compare successful versus unsuccessful applicants, controlling for a linear measure of the applicant's priority score. They find that the award may increase future grant funding slightly, but that it does not appear to increase publication-based measures of research productivity. This strategy leverages the intuition behind a regression discontinuity analysis and thus, in theory, should eliminate selection concerns. However, this particular analysis appears to have suffered from several potential shortcomings. Specifically, the authors do not provide evidence that career grants were, in practice, awarded strictly on the basis of priority scores, nor do they include more flexible controls for the priority score to account for any underlying non-linear relationship between score and productivity.

Arora and Gambardella (2005) estimate the impact of receiving a National Science Foundation grant on subsequent research productivity of a sample of economists. They find generally small impacts of grant receipt but present some evidence that the impacts might be larger for young researchers.

3. Background

NIH post-doctoral fellowships (F32 grants) are awarded to young researchers who have just completed their graduate training. The grants are awarded for up to three years and include a stipend for the individual and a payment made to the sponsoring institution. In 2008, NIH awarded approximately 648 postdoctoral fellowships (not including ongoing awards) with annual stipends averaging roughly $50,000. These fellowships are likely to comprise the primary or only source of funding for individuals during this period in their career. A primary goal of the fellowship is to steer recipients into a research career.

Applications for F32 grants are accepted three times per year. All applications are subject to peer review within Integrated Review Groups (IRGs) organized around topics or areas. Postdoctoral fellowship award applications are generally reviewed by a single body within the particular institute, and are therefore not transformed into percentile ranks.

Reviewers evaluate proposals on the basis of five criteria (significance, approach, innovation, investigator and environment) and assign each application a priority score on a scale of 1–5 (reviewers assign a score up to two significant digits, e.g., 2.2, with 1 being the highest quality). The average of these scores is calculated and multiplied by 100 to obtain the priority score. A certain fraction of the lowest quality applications (as determined by the reviewers) do not receive priority scores. Typically, half of all research program grant applications do not receive scores, whereas all fellowship and career applications receive scores, but this varies considerably across institutes.

Funding determinations for postdoctoral fellowships are made at the institute level, so that applications from different programs

[5] It is also interesting to consider the effectiveness of the NIH postdoctoral fellowship program relative to other ways in which NIH could use the same resources. For example, it is possible that shifting resources from the F32 program to the standard NIH research grant programs (i.e., those mechanisms that provide R01 grants) would have an even larger impact on the research careers of young scholars (for example, by funding principal investigators who employ recent graduates). Unfortunately, this analysis is beyond the scope of the current paper.

within the institute compete against each other for funding. The number of grants funded depends on the institute budget for the fiscal year. In practice, each decision-making unit is allocated a budget. Generally, grants are awarded solely on the basis of priority score. Researchers whose applications receive a poor score and do not receive funding have the ability to respond to the criticisms raised by reviewers and submit an amended application. Amended applications are treated in the same manner as new applications for the purposes of evaluation and funding.

4. Data and sample

Information on NIH applicants and applications, including priority scores, are drawn from administrative files that include records for all applications for research grants and fellowships. The records provide information on the principal investigator (name, department, home institution, etc.), the type of application (including the date the grant was considered, the grant type or mechanism as well as the institute and program area to which it was submitted), the priority score received by the application, whether the application was funded and how much funding it received.

The outcomes we examine include publications, citations and future NIH funding. The NIH files utilize a unique individual identifier so that we are able to match applicants in any given year, institute and mechanism to past and future funding information. We match NIH applicants to publications using last name and first initial. Of course, this will likely result in a number of false positives. Therefore, we utilize a variety of different strategies to minimize the incidence of bad matches.[6] To minimize the impact of extreme positive outliers in the outcome measures, we recode all values above the 99th percentile to the 99th percentile value for all publication, citation, and NIH funding variables.

We start with all applications for postdoctoral fellowships (F32s) submitted to NIH between 1980 and 2000. To minimize measurement error in matching researchers to publications, we focus on the 44% of F32 applications in which the applicants have uncommon names, defined as those whose last name was associated with 10 or fewer unique NIH applicants during our time period. Since name frequency is unlikely to be correlated with whether an individual is just above or below the funding cutoff (conditional on flexible controls for their priority score), this restriction will not influence the consistency of our estimates.[7] We also exclude institutes with fewer than 100 applicants for the entire sample period and a small number of applicants who had large amounts of NIH funding prior to their postdoctoral fellowship application. Finally, we focus our analysis on individuals with IRG scores within 100 points of the cutoff. This increases the comparability of rejected and accepted applicants.

Our analysis sample has 13,426 observations reflecting 12,189 unique individuals over 20 years and 16 different institutes. Table 1 shows summary statistics of all F32 applicants along with applicants in our analysis sample. The summary statistics for these two groups tend to be quite similar. Focusing on our analysis sample, researchers from the biological sciences constitute 82% of the sample while researchers from physical science, social science and other miscellaneous departments constitute 9, 6 and 4% of applicants, respectively. The institutes receiving the largest number of applicants are General Medicine; Heart, Lung and Blood; Cancer; and Neurological Disorders. Roughly 46% of applicants receive a fellowship on any given application, with an additional 4% obtaining a fellowship on a later application.

5. Methodology

If fellowships were randomly allocated, one could identify the causal effect of an award by simply comparing the research output of successful and unsuccessful applicants. However, the data suggest that more qualified applicants are more likely both to receive NIH postdoctoral fellowships and to have high numbers of future publications and citations.[8] To the extent that this is true, naïve comparisons of successful and unsuccessful applicants may be biased upward, reflecting both the causal impact of receiving a fellowship as well as differences in latent scientific productivity (or interest). In this section, we describe the empirical strategies that we use to address these concerns, illustrating the intuition behind our approaches as well as outlining some of the details regarding statistical estimation.

5.1. Controlling for selection on observable characteristics

As a first attempt at identifying the causal impact of postdoctoral funding, we use the abundant information available to us regarding the quality of the grant application and prior productivity of the applicant to control for the expected productivity of the researcher in the absence of the grant. More specifically, we estimate a regression of the following form:

$$productivity_{it+1} = \beta\ funded_{it} + f(n_{it}) + X_{it}B + \varepsilon_{it+1} \quad (1)$$

where $productivity_{it+1}$ is the research productivity of individual i in period $t+1$, $funded_{i,t}$ indicates whether the researcher's application was ultimately successful, n_{it} is the priority score of the researcher's application normalized relative to the grant funding cutoff (described in more detail below), $f(\ )$ is a smooth function, X_{it} is a vector of researcher-level covariates, and ε_{it+1} is a mean zero residual. As covariates we include NIH institute, region, year, discipline, and organization type fixed effects; demographic controls including marital status, number of children, age, gender, and type of degree; quadratic functions of the normalized priority score, name frequency, age, institution rank, graduate institution rank; and productivity measures in the five years prior to grant application including quadratic measures of publications, citations, research direction, NIH funding, and NSF funding. Our hope is that by controlling adequately for the priority score of the grant application, researcher characteristics, and prior productivity we can identify the approximate causal effect of grant receipt on subsequent research productivity.

5.2. A regression discontinuity approach

In addition to the controlling for a rich set of observable applicant characteristics, we use a second approach which relies upon the fact that F32 fellowships are awarded on the basis of observable priority scores, and that there is a highly nonlinear relationship between this score and the probability of funding. This strategy

[6] For a more complete description of the statistical methodology utilized in this analysis, see the online appendix.

[7] However, if this group of researchers is different than the overall pool of applicants in important ways, this strategy may change the interpretation of our estimates. For example, if researchers with uncommon names are more likely to be immigrants or come from relatively small ethnic groups, our estimates will reflect the impact of grant funding on these groups. If such researchers use grant funding either more or less productively than other individuals, our estimated treatment effects will not generalize to the broader population of researchers. In order to assess the external validity of our estimates, we compared NIH applicants with common and uncommon names on a variety of observable characteristics. The comparison suggests that those with uncommon names are quite comparable to those with more common names. See the online appendix for a more detailed discussion.

[8] For example, applicants with more publications prior to the application date are more likely to receive a fellowship and have higher numbers of subsequent publications.

Table 1
Summary statistics.

	Analysis sample			
	All applicants	All	Applicants scoring just below the cutoff	Applicants scoring just above the cutoff
Normalized score	15.07	−1.30	−24.00	23.83
Awarded	0.41	0.46	0.72	0.15
Ever awarded	0.46	0.51	0.74	0.25
Applicant's background				
Female	0.38	0.38	0.38	0.40
Age	32.26	31.94	31.55	32.38
Married	0.50	0.49	0.49	0.50
Divorced	0.28	0.28	0.30	0.24
Number of Dependents	0.29	0.27	0.25	0.28
Name frequency	174.62	2.12	2.08	2.10
Has PhD	0.76	0.77	0.78	0.78
Has MD	0.20	0.19	0.19	0.16
Has PhD and MD	0.05	0.05	0.05	0.03
Rank of graduate institution in terms of NIH funding	130.83	129.37	116.39	148.54
Rank of current institution in terms of NIH funding	69.46	68.73	67.20	71.14
Department				
Biological sciences department	0.82	0.82	0.82	0.81
Physical sciences department	0.09	0.09	0.09	0.11
Social sciences department	0.04	0.04	0.04	0.04
Other department	0.05	0.05	0.05	0.05
Productivity measures				
Years 1–5 prior to the application				
Any NIH funding		0.02	0.02	0.02
Amount of NIH funding (/$100,000)		0.28	0.30	0.18
Any publications		0.77	0.79	0.76
Number of publications		3.54	3.53	3.42
Years 1–5 following the application				
Any NIH funding		0.23	0.25	0.18
Amount of NIH funding ($/100,000)		1.42	1.48	1.39
Any publications		0.91	0.93	0.90
Number of publications		6.07	6.10	5.69
Any citations		0.91	0.93	0.89
Number of citations		254.67	276.46	219.72
Years 6–10 following the application				
Any NIH funding		0.30	0.33	0.22
Amount of NIH funding ($/100,000)		5.11	5.15	5.12
Any publications		0.77	0.81	0.73
Number of publications		8.15	8.32	7.58
Any citations		0.74	0.77	0.68
Number of citations		246.57	258.54	203.38
Sample size	36,302	13,426	5752	3611

Notes: Sample includes applicants with uncommon names (name frequency ≤ 10) who scored within ±100 points of the funding cutoff for F32s and ±200 points for R01s. The unit of observation is a grant application (n = 13,462). Estimates of amount of NIH or NSF funding includes zeroes for those who received no funding.

is based on the regression discontinuity (RD) design, which has become increasingly popular in economics research in recent years, and has been used to successfully evaluate a variety of programs. The intuition behind the RD design is that if one compares applicants just above and just below some pre-specified cutoff, there will be little, if any, difference in unobservable determinants of productivity but a large difference in the likelihood of receiving funding.[9]

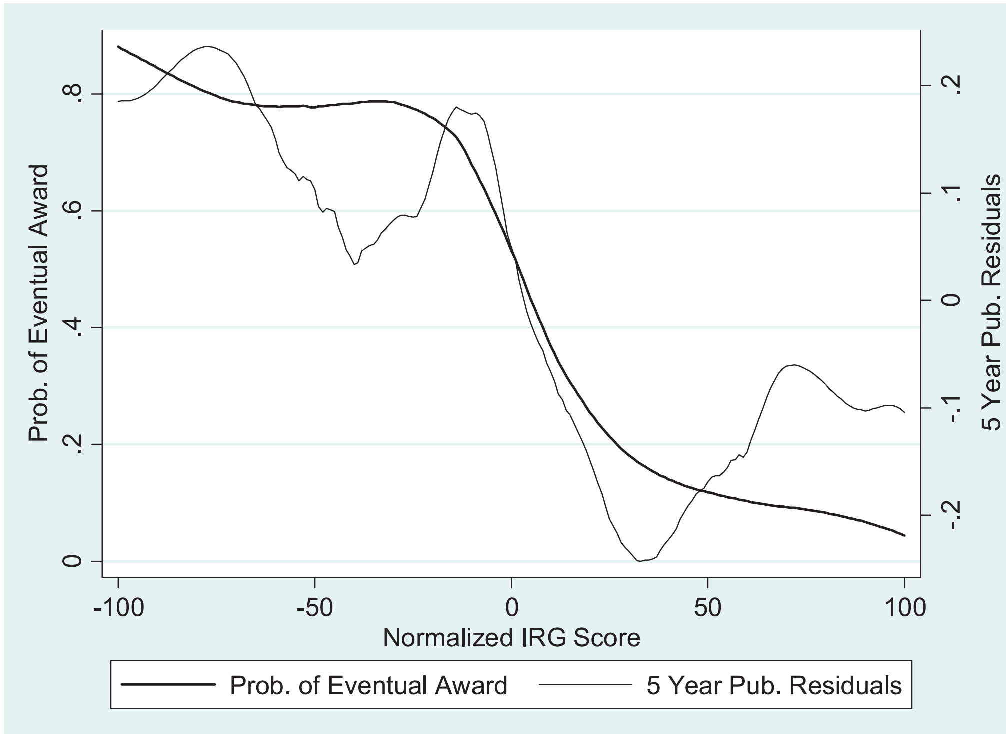
There is no pre-determined cutoff for funding that applies universally across the NIH. Instead, the realized cutoff in each situation depends on the level of funding for a particular institute, year and mechanism, along with the number and quality of applications submitted. This political reality provides significant advantages for our identification since it essentially establishes dozens of different cutoffs that we can exploit, and reduces the concern that a single cutoff might coincide with some other factor that is correlated with research productivity.

To aggregate across institutes and years, we define the cutoff in institute j in year t,c_{jt}, as the score of the last funded application in the counterfactual case that no out-of-order funding had occurred. Denote p_{ijt} as the priority score received by researcher is application in institute j in year t. We then subtract this cutoff from each priority score to obtain a normalized score, which will be centered around the relevant funding cutoff: $n_{ijt} = p_{ijt} - c_{jt}$.[10]

In the case of a "sharp" discontinuity, situations in which the probability of receiving a treatment is determined completely by performance relative to a predetermined cutoff, one can estimate a RD model by simply comparing the outcomes of those immediately above and below the cutoff. This identifies the causal effect of treatment for individuals in the neighborhood of the cutoff. In this spirit, columns 3 and 4 in Table 1 show summary statistics for applicants between 0 and 50 points below the cutoff and between 1 and 50 above the cutoff. Prior research output and

[9] For a formal treatment of RD designs, see Hahn et al. (2001). For empirical examples, see Jacob and Lefgren (2004a,b), Thistlewaite and Campbell (1960), Berk and Rauma (1983), Trochim (1984), Black (1999), and Angrist and Lav (1999).

[10] Ideally, one would like to create the theoretical cutoff score taking into account the amount of funding associated with each application. Unfortunately, the NIH files do not contain any information regarding the *requested* funding amounts for the unfunded applications. Note that if all applications requested the same amount of funding, both approaches would yield identical cutoff scores. This is an excellent approximation for our sample, however.



Notes: Data is smoothed using a lowess estimator with a bandwidth of .25. Publication residuals are calculated by regressing five year publication rates on researcher demographics and prior productivity measures.

Fig. 1. Relationship between normalized IRG score, eventual award, and publications. Notes: data is smoothed using a lowess estimator with a bandwidth of 0.25. Publication residuals are calculated by regressing five year publication rates on researcher demographics and prior productivity measures.

demographic characteristics across these two groups are mostly balanced. While these groups seem quite similar on the basis of observable characteristics, the probability of eventual treatment is quite different—74% for the group below the cutoff and only 25% for the group above. The eventual publication rates also differ. Applicants below the cutoff had 6.1 publications on average in the five years after grant application compared to only 5.7 publications for those applicants above the cutoff. If this difference is attributable entirely to the difference in probability of grant receipt, the implied effect of an F32 postdoctoral fellowship is 0.8 publications over five years.[11]

In our case, however, the relationship between priority score and receipt of an NIH postdoctoral fellowship does not follow a sharp discontinuity. The bold line in Fig. 1 shows the probability that a grant application is funded as a function of the normalized priority scores for applicants in our sample. Note that while the probability of funding is a highly nonlinear function of the normalized application score, there is clearly evidence of out-of-order funding. Some successful applicants ultimately decline the award in favor of alternative employment, further attenuating the relationship between normalized score and grant receipt. Roughly 7% of individuals who scored above the cutoff received the grant, while 26% of those below the cutoff did not receive a grant or declined the award.

Because the unobserved productivity of applicants may be related to the application score, we might expect publication rates to vary with the IRG score even in the absence of a causal effect of grant receipt. If, however, receipt of a postdoctoral fellowship is associated with a large improvement in productivity, we would expect the relationship between the normalized IRG score and publication rate to be strongly negative in the vicinity of the granting cutoff.

And, indeed, this is what we observe. The thin line in Fig. 1 shows the relationship between the normalized priority score and publication rate in the five years following submission of the grant application. To represent this relationship most clearly, we report the average publication residuals from a regression of the five-year publication rates on demographic and productivity measures that predate the grant application. In the range of normalized IRG scores in which the probability of grant receipt is declining, there is also a marked reduction in the research productivity of grant applicants. Along with simple comparison of individuals above and below the cutoff, this suggests a positive causal impact of grant receipt on the subsequent publication rate.

5.3. Instrumental variables estimation

We formalize the intuition underlying Fig. 1 by conducting an instrumental variables (IV) regression in which the nonlinear relationship between the priority score and the probability of a fellowship award serves as the excluded instrument. Specifically, we estimate a simple two-stage least squares (2SLS) regression procedure. Our first stage equation is given by

$$funded_{it} = \gamma below\_cut_{it} + g(n_{it}) + X_{it}\Gamma + \eta_{it+1} \tag{2}$$

where $below\_cut_{it}$ is a binary variable indicating that the normalized score was below the imputed funding cutoff and the other

[11] This equals the difference in five year publication rates between the two groups divided by the difference in the probability of grant receipt 0.84 = (6.10–5.69)/(0.74–0.25).

variables are as described earlier. The second stage equation is still given by Eq. (1). The identifying assumption is that, having controlled for a smooth function of the normalized application score, any further change in research productivity associated with being below the cutoff is attributable to receiving an F32 grant. Note that our first stage equation takes advantage of variation attributable only to the *observed* priority score. Because of this, the estimated treatment effect will be unbiased even if administrators fund out of order based on the unobserved aspects of the applicant or research idea. This approach works even though there is not a strict discontinuity as long as being below the cutoff is predictive of the eventual award status. It is potentially superior to the simple comparison of individuals above and below the cutoff because we are able to control for the application score along with other covariates.

5.4. Assumptions of RD analysis and threats to identification

While an RD analysis can overcome non-random sorting in many contexts, it does make several important assumptions. One assumption is that agents cannot manipulate the measure used to determine treatment—in this case, the priority score. If agents had the ability and incentive to manipulate this measure, individuals ending up just below the cutoff may differ systematically from those just above. In our situation, however, applicants all have an incentive to write strong grant applications yet have no direct control over the score their application receives. Additionally, the cutoff is unknown until all applications have been received and evaluated.

A more substantial concern is that the funding agency may endogenously decide to choose a funding cutoff based on where the quality of applications begins to drop off rapidly. If this were the case, we would expect the observable characteristics, including prior productivity, of the application to be systematically different across the funding cutoff, even after controlling for the normalized priority score. We can test this by conducting a falsification exercise in which we employ our IV strategy to identify the "impact" of grant receipt on pretreatment characteristics and productivity. If our instrument is uncorrelated to the pre-application characteristics of the applicant, we would expect the coefficient on grant receipt to be zero. We later show that our instruments have little correlation to pre-application characteristics of the applicant, providing support for our analysis strategy.

In cases such as ours in which the relationship between an index score and treatment is not discontinuous, one must make additional assumptions to identify an unbiased causal estimate. Specifically, this approach assumes that one can accurately model the relationship between the index variable and the outcome in the absence of the intervention. In our analysis, for example, we assume that a low-order polynomial in the normalized application score captures the baseline relationship of this variable with future research output. To the extent that we fail to control adequately for the baseline relationship between application score and future productivity, our instrument may capture the residual relationship between these two variables. This would lead to inconsistent estimates of the parameter of interest.

Fortunately, there are several ways to test this modeling assumption. The falsification exercise in which we look at the "impact" of grant receipt on pretreatment characteristics sheds light on the appropriateness of the model assumptions. In addition, we can test the sensitivity of our results to our functional form assumptions by estimating models with more flexible controls for the priority score and to use samples restricted to be within varying distances from the cutoff. In the sections below, we show our results are robust to wide range of modeling choices.

5.5. Estimating models with binary outcomes with non-classical measurement error

While examining continuous output measures (e.g., number of publications or citations) is useful, it is also interesting to explore how receipt of a postdoctoral fellowship increases the probability that a researcher surpasses various output thresholds. Nearly all applicants in our sample have at least some publications. By focusing on a certain threshold (e.g., at least five publications), we can gain insight into whether the fellowship represented a gateway to a research career or simply affect the output of individuals who would have enjoyed a research career regardless.

For outcomes that we can measure with certainty, such as subsequent NIH funding, it is straightforward to estimate these type of threshold effects through the use of standard regression models appropriate for binary outcome data such as a Probit model. However, as we discussed briefly in Section 4 (and in more detail in online appendix), the publication and citation measures we use have non-classical measurement error because of false positives in the matching of bibliometric data with NIH records.

In the presence of such non-classical measurement error in our outcome variable, the standard Probit model will yield biased estimates. For example, suppose we want to examine how receiving an F32 grant affects the probability that a researcher has at least one publication. This threshold could be met because the researcher actually has at least one publication, the researcher has at least one false match, or both. To see why this is problematic, consider the case in which an individual has a false match. We will observe that this individual has surpassed the cutoff regardless of her true productivity, so that any factors that increase her actual productivity will appear to have no impact. In other words, for the subset of individuals with false matches, the coefficient on grant receipt will be zero by construction. For this reason, conventional estimation techniques for binary outcomes (e.g., Logit or Probit) will yield attenuated coefficients.

To adequately address concerns regarding non-additive measurement error, we develop and estimate a simple model that accounts for the fact that the false matches can push individuals across a particular productivity threshold. For reasons of statistical precision, in these models we treat grant receipt as exogenous, conditional upon grant and applicant characteristics. We model the probability of observing a particular realization given that if a researcher crosses a threshold it could have occurred either because of actual output or measurement error.[12] Assuming that a latent index of output is a linear function of funding status, a set of control variables, and a normally distributed residual, the probability that actual productivity exceeds a particular threshold can be written: $\Phi(\beta\, funding_{it} + X_{it}B)$. In the productivity equation, we use a slightly more parsimonious specification than in our linear models. We include fixed effects for institute,[13] year, type of institution, discipline, and type of degree. We also control for marital status, number of children, quality of current and graduate institution, and second order polynomials in prior publications funding. We model the latent index of false productivity as a linear function of variables indicating last name commonness, W_i and a normally distributed residual. Thus the probability that false productivity surpasses the

[12] In cases where the threshold is greater than one, it is possible that the threshold was exceeded due to a combination of actual *and* false matches. This greatly complicates the modeling of the threshold effects. For simplicity, we abstract from this possibility and assume that the threshold is met only with true matches, only with false matches, or that both the false and true matches were sufficiently high to surpass the threshold.

[13] We include fixed effects only for institutes representing at least 5% of grant applications. The smaller institutes are implicitly grouped together as the reference category.

threshold can be written: $\Phi(W_i\Pi)$. Given these two probabilities, it is simple to write each observation's contribution to the likelihood function. The probability that a particular observation does *not* surpass the threshold is given by the following equation.

$$Pr(below\_threshold) = [1 - \Phi(\beta funding_{it} + X_{it}B)][1 - \Phi(W_i\Pi)] \quad (3)$$

The probability an observation surpasses the threshold is given by:

$$Pr(above\_threshold) = 1 - [1 - \Phi(\beta funding_{it} + X_{it}B)][1 - \Phi(W_i\Pi)] \quad (4)$$

We identify the parameters of this statistical model using maximum likelihood.

6. Findings

6.1. Ordinary least squares (OLS) estimates

To provide a baseline for understanding the relationship between NIH funding and future productivity, Table 2 presents a series of OLS estimates for a variety of professional outcomes. For all outcomes, row 1 shows unconditional estimates and rows 2–4 add in progressively more controls. Robust standard errors that cluster by researcher are shown in parenthesis beneath the estimates.

Consider the first column, which shows the effect of a postdoctoral fellowship on publications in the five years following grant application. The unconditional estimates reveal a positive association between receipt of the fellowship and subsequent research productivity. For example, in column 1 we see that individuals who receive a fellowship have roughly 0.83 publications more in the five years following the grant application compared with their peers who did not receive a fellowship. Once we control for researcher background characteristics and the application score, the point estimate drops considerably but is still statistically significant. To judge the relative magnitude of the effect, consider that the mean and standard deviation of true publications among unsuccessful applicants are 4.57 and 4.53, respectively. Hence, the effect of 0.65 shown in row 4 reflects a 0.14 standard deviation increase in the number of publications. Finally, note that the point estimate drops considerably from row 1 to row 2, but does not change significantly as additional covariates are added. This indicates that postdoctoral fellowships are awarded almost entirely on the basis of IRG scores, as the formal NIH funding process intends.

Column 2 shows the impact of grant receipt on the number of publications 6–10 years after grant application.[14] Controlling for applicant characteristics and normalized IRG score, we see that grant receipt is associated with 0.47 additional publications in the 6–10 years after grant application. This suggests that the effects of an NIH postdoctoral fellowship on research productivity last far beyond the period of fellowship itself. Columns 3 and 4 show the effect of grant receipt on first author publications. These results imply that an NIH postdoctoral fellowship increases not only research participation but also the amount of independent research conducted in both the short and long run. While the effect of grant receipt on the number of publications is positive, the estimated impact is small and statistically insignificant for total citations and subsequent NIH funding.[15]

6.2. Instrumental variables (IV) estimates

The OLS estimates suggest that the receipt of an NIH postdoctoral fellowship increases publications. However, the selection on observables that is evident in Table 2 raises concern that selection on unobservable characteristics may also be present, and may bias the estimates. To address this concern, we calculate instrumental variables (IV) estimates that exploit the plausibly exogenous variation in grant receipt generated by the nonlinear relationship between priority score and the likelihood of funding. As discussed in Sections 5.2 and 5.3, the intuition behind this approach is to compare applicants who scored just above the funding cutoff with those who scored just below the funding cutoff.

For each outcome, row 1 of Table 3 shows the OLS estimate of receiving a postdoctoral grant, which were reported earlier in Table 2. Row 2 shows the corresponding IV estimate and row 3 shows the difference between the OLS and IV estimates. Looking across outcomes, we see that the IV estimates are very similar to the OLS estimates in all cases. This suggests that our OLS estimates are consistent and controlling for application score and applicant characteristics may be sufficient to identify the causal effect of grant receipt.[16] It is interesting to note that these IV estimates are virtually identical to the impact implied by our simple unadjusted comparison of applicants just above and below the cutoff, shown in columns 3 and 4 of Table 1. This implies that the controlling for the normalized IRG score and other covariates has little impact on the estimated impact of grant receipt.

6.3. Estimates for binary outcome measures

It is also informative to examine binary measures of research productivity in order to explore whether fellowships have a substantial effect on the extensive as well as intensive margin. This analysis provides insight into whether an F32 award represents a gateway to a research career, or simply affects the productivity of individuals who would have enjoyed a research career regardless. To do so, we estimate the binary choice model discussed earlier. The results are presented in Table 4.

Our results suggest that receiving an F32 fellowship has a statistically significant impact on a variety of important career productivity "thresholds," including 5 or more publications in years 1–5 and 6–10 after grant application, 5 or more first author publications over the same periods, more than 200 citations in the 10 years after grant applicant, and more than $200,000 in NIH funding within 10 years of application. At first glance it may seem odd that grant receipt is associated with significant increases in the probability of crossing citation and funding thresholds, but has an insignificant effect on the total number of citations and NIH funding. This is consistent with grant receipt improving outcomes at lower levels but having little impact on high achievers.

[14] For researchers who applied after 1995, we do not observe their productivity for the full period of 6–10 years after publication. For these researchers we inflate their publications by a factor of 5 divided by the number of years we observe them. Mean productivity measures are very similar for the cohorts which we performed this correction. When examining the probability of achieving a particular productivity threshold, this procedure works less well. As a consequence, when examining binary outcomes further than five years out, we exclude grant applications from 1998 or later.

[15] The fact that publications rise and citations do not suggests that the marginal publications may be less influential than the average. Its not clear why this would be the case.

[16] We also explored whether the impact of NIH postdoctoral fellowships differ by researcher characteristics. We found no statistically significant differences in the impact of fellowships by time period, gender, age, degree type, or discipline.

Table 2
OLS estimates of NIH postdoctoral fellowships on research productivity.

Independent variable: binary indicator for NIH grant receipt Specification	Publications in years 1–5 (1)	Publications in years 6–10 (2)	First author publications in years 1–5 (3)	First author publications in years 6–10 (4)	Total citations in years 1–10 (5)	NIH funding in years 1–10 (/\$100,000) (6)
(1) No controls	0.83** (0.10)	1.09** (0.15)	0.50** (0.04)	0.45** (0.05)	73.68** (11.37)	0.617** (0.074)
(2) Quadratic priority score + institute and year fixed effects	0.67** (0.12)	0.57** (0.18)	0.39** (0.05)	0.31** (0.05)	6.90 (13.56)	0.026 (0.097)
(3) = (2) + applicant characteristics	0.59** (0.12)	0.41** (0.18)	0.36** (0.05)	0.27** (0.05)	0.46 (13.14)	−0.038 (0.093)
(4) = (3) + Measures of prior publications, funding	0.65** (0.10)	0.47** (0.16)	0.38** (0.05)	0.27** (0.05)	4.41 (12.82)	−0.019 (0.092)
Control group mean (S.D.)	4.57 (4.53)	5.20 (7.22)	2.02 (2.00)	1.32 (2.06)	361.84 (567.65)	1.37 (2.42)
R-squared from model in row 4	0.35	0.25	0.18	0.14	0.22	0.15
Number of obs.	13,426	12,749	13,426	12,749	12,749	11,311

Notes: Each cell in rows 1–4 of this table represents the coefficient (S.E.) from a separate OLS regression where the dependent variable is shown at the top of the column and the set of control variables are described in under "Specification" in the first column. The unit of observation is an application. In each case, the estimate shown is the coefficient (S.E.) on a binary indicator for eventual NIH grant receipt. The sample sizes in columns 3 and 7 are smaller than the others because researchers with zero publications do not have a value for the research direction variable. The sample sizes in column 6 are smaller because we only have NIH funding information through 2003, and so we cannot calculate measures for later years. For observations late in our sample period, we extrapolate some values as described in the online appendix. The control variables include fixed effects for institute, year of award, name frequency, name frequency squared, age and age squared at time of award, a binary indicators for female, married and divorced, a linear measure for the number of dependents, binary indicators for region (West, Central and South, with East omitted), binary indicators for degree type (MD, and MD/PhD with PhD as the omitted category), binary indicators for field (social sciences, physical sciences and other, with biological sciences as the omitted category), binary indicators for organization type (research institute, hospital with university as the omitted category), binary indicators for unit within organization which only applies to universities (hospital, arts and sciences, school of public health, institute, or other with medical/dental school omitted), linear and quadratic terms for the rank of the applicant's current and graduate institutions where rank is measured in terms of amount of NIH funding received in prior years, and linear and quadratic terms for a host of prior productivity measures including number of publications in years 1–5 prior to application, number of publications in years 6–10 prior to application, research direction in years 1–5 prior to application, research direction in years 6–10 prior to application, amount of NSF funding in years 1–5 prior to application, amount of NSF funding in years 6–10 prior to application, amount of NIH funding in years 1–5 prior to application, and amount of NIH funding in years 6–10 prior to application. The control group means and standard deviations for the publication and citation columns are adjusted to account for the presence of false positive matches, as described in the online appendix. Standard errors are clustered by researcher.
* Significance at the 10% level.
** Statistical significance at the 5% level.

Table 3
The effect of NIH postdoctoral fellowships on research productivity.

	Second-stage estimates					
	Publications in years 1–5 (1)	Publications in years 6–10 (2)	First author publications in years 1–5 (3)	First author publications in years 6–10 (4)	Total citations in years 1–10 (5)	NIH funding in years 1–10 (/\$100,000) (6)
OLS estimates	0.65** (0.10)	0.47** (0.16)	0.38** (0.05)	0.27** (0.05)	4.41 (12.82)	−0.02 (0.092)
IV estimates	0.86** (0.41)	1.18* (0.64)	0.34* (0.20)	0.43** (0.21)	3.58 (50.04)	−0.16 (0.31)
Diff: IV-OLS	0.22 (0.32)	0.72 (0.67)	−0.03 (0.17)	0.16 (0.21)	0.83 (54.78)	−0.14 (0.35)
Control group mean (S.D.)	4.57 (4.53)	5.20 (7.22)	2.02 (2.00)	1.32 (2.06)	361.84 (567.65)	1.37 (2.42)

Notes: The estimates are derived from specifications (1) and (2) in the text. Normalized publications are calculated by dividing each publication by the total number of authors on the publication prior to summing across years. Each regression includes the full set of control variables described in the notes to Table 2. The control group means and standard deviations for the publication and citation columns are adjusted to account for the presence of false positive matches, as described in the text. Standard errors are clustered by researcher.
* Significance at the 10% level.
** Statistical significance at the 5% level.

Moreover, the magnitudes of the effects are substantial in most cases. For example, postdoctoral trainees are 7.3 percentage points (24%) more likely to have five or more publications in the five years after grant application. The effect of receiving a postdoctoral grant on having more than five publications in years 6–10 after grant application is a bit smaller at 4.5 percentage points (16%). Post-doctoral fellowship receipt is also associated with increases in the probability of achieving at least five first author publications both

Table 4
The effect of NIH postdoctoral fellowships on binary measures of career success.

	5 or more publications in years 1–5 (1)	5 or more publications in years 6–10 (2)	5 or more first author publications in years 1–5 (3)	5 or more first author publications in years 6–10 (4)	200 or more citations in years 1–10 (5)	More than \$200,000 NIH funding in years 1–10 (6)
ML point estimate	0.215** (0.035)	0.137** (0.041)	0.165** (0.041)	0.160** (0.049)	0.124** (0.032)	0.082** (0.034)
Marginal effect—percent points	0.073** (0.011)	0.045** (0.013)	0.032** (0.007)	0.026** (0.007)	0.041** (0.010)	0.020** (0.008)
Marginal effect—percent of mean	0.244** (0.055)	0.160** (0.058)	0.327** (0.104)	0.351** (0.142)	0.113** (0.033)	0.141** (0.065)

Notes: The estimates shown above are derived from the specifications outlined in Appendix A. Each column represents a separate model. In each case, the dependent variable is a binary productivity measure. Standard errors are clustered by researcher.
* Significance at the 10% level.
** Statistical significance at the 5% level.

Table 5
Effects on pre-treatment outcomes.

Specification	Control group mean (S.D.) (1)	Coefficient (S.E.) (2)
Pre-treatment number of publications	2.68 (3.31)	0.50* (0.29)
Pre-treatment index of research relevance	1.18 (1.60)	0.08 (0.16)
Pre-treatment NSF funding ($/100,000)	0.00 (0.12)	−0.01 (0.01)
Pre-treatment NIH funding ($/100,000)	0.00 (0.09)	0.00 (0.01)
Name frequency	2.12 (2.58)	−0.25 (0.24)
Female	0.39 (0.49)	0.01 (0.05)
Age	32.02 (4.03)	−0.38 (0.44)
Married	0.50 (0.50)	−0.01 (0.05)
Divorced	0.28 (0.45)	0.02 (0.04)
Number of dependents	0.28 (0.45)	0.01 (0.05)
Has PhD	0.76 (0.43)	−0.02 (0.04)
Has MD	0.17 (0.37)	0.04 (0.03)
Rank of graduate institution	134.48 (294.17)	−75.98** (32.71)
Rank of current institution	68.17 (88.82)	−10.05 (8.84)
Biological sciences department	0.82 (0.39)	0.04 (0.04)
Physical sciences department	0.10 (0.30)	−0.04 (0.03)
Social sciences department	0.03 (0.18)	0.00 (0.02)
Research institute	0.11 (0.31)	0.00 (0.03)
Hospital	0.06 (0.24)	−0.01 (0.02)
Arts and sciences	0.31 (0.46)	0.00 (0.04)
School of public health	0.01 (0.09)	0.00 (0.01)
Hospital	0.00 (0.06)	0.01 (0.01)
Institute	0.02 (0.13)	−0.02 (0.01)
Other	0.10 (0.30)	0.01 (0.03)

Notes: These specifications were estimated using IV in which we control only for institute and year fixed effects and linear and quadratic measures the normalized IRG. The instrument is whether the IRG score was below the cutoff. Standard errors are clustered by researcher.
* Significance at the 10% level.
** Statistical significance at the 5% level.

in the short and long run. While the increase is only about 3 percentage points, these represent an over 30% increase relative to the baseline probability. Grant receipt increases the probability of receiving more than $200,000 in NIH grant money by 2 percentage points (14%) as well. Collectively, these results suggest that receiving an NIH postdoctoral fellowship has an important impact on the probability that a biomedical PhD becomes a successful researcher.

6.4. Robustness checks

As mentioned in Section 5.4, we may be concerned about bias arising due to endogenous choice of the cutoff by program officers within the various NIH institutes. If institutes chose the cutoff strategically to accept better applicants, then being just below the cutoff (relative to just above the cutoff) would be systematically correlated to prior productivity and other observable characteristics. To examine this possibility, we estimate the "impact" of a postdoctoral fellowship on *pretreatment* outcomes using our IV strategy. We control only for a second order polynomial in normalized IRG score, institute, and year fixed effects. Examining Table 5, we see only one coefficient that is statistically significant at the 5% level, which is about what one would expect due to chance.

To further minimize concerns regarding function form assumptions, we check the robustness of our results to the use of more flexible controls of the normalized score and to the use of samples restricted to be within varying distances from the cutoff. Table 6 shows these alternative specifications for publications 1–5 years after grant receipt. The results are all qualitatively similar to our baseline. We obtain similar results for our other outcome measures (results available upon request).

An additional concern with our specification is that the outcome measures are by their nature non-negative integers. Hence, it might be more appropriate to use a negative binomial specification which allows for count data and overdispersion. In row 10 of Table 6, we examine a negative binomial specification analogous to our baseline OLS regression with the full set of controls. The coefficient suggests that receiving a grant increases the number of publications over the following five years by approximately 14%, which is very similar to baseline estimates shown in Table 3.[17] The implied confidence interval is also similar to our OLS estimates.

7. Discussion

The fundamental premise underlying the NIH postdoctoral fellowship program is that market forces alone fail to provide adequate resources for the production of basic science, and that, on the margin, a scientist engaged in basic research generates a greater benefit to society than the same individual engaged in an alternative activity. Our results suggest that the program does indeed appear to increase the amount of health science research and the number of individuals engaged in a biomedical research career. In particular, we find that for applicants in the neighborhood of the funding cutoff receipt of an NIH postdoctoral fellowship significantly increases the probability that a new PhD will successfully make the transition to a research career and the number of articles published in the 10 years following grant receipt. These results shed light on the likely increase in output for similar applicants who would be funded if the program were expanded. The findings imply that among marginal applicants, an F32 grant represents a better gateway to a research career relative to the applicants' next best professional option. These results are significant from both a statistical and policy perspective and highlight the importance of

[17] This negative binomial specification shows the percentage increase in observed, including false, publications. Consequently, the percentage impact is potentially biased downwards because we are measuring a percentage effect on a baseline level that is too high. However, assuming that the treatment had no increase in the number of false publications, we can still measure the absolute number of additional publications implied by the negative binomial coefficient. Among unsuccessful applications in our analysis sample, the average number of actual and false publications is 5.78. Hence our estimate implies that receiving a grant increases the number of publications by [exp(.14) − 1]*5.78 = 0.87.

Table 6
Alternative specifications and samples.

Specification	Pubs in years 1–5 (1)
(1) Baseline	0.86** (0.41)
(2) No covariates	1.26** (0.50)
(3) Narrower range	1.01** (0.51)
(4) Wider range	0.66** (0.33)
(5) Include only linear term in the rating	0.86** (0.41)
(6) Including 3rd order polynomials in the rating	1.45* (0.77)
(7) Including 4th order polynomials in the rating	1.46* (0.79)
(8) Including a linear term in the rating but allowing it to differ above vs. below the cutoff	0.88** (0.41)
(9) Including 2nd order polynomials, and allowing both terms to differ above vs. below the cutoff	1.91 (1.25)
(10) Negative binomial regression of baseline specification—grant receipt exogenous	0.14** (0.02)

Notes: The specifications are identical to those in Table 3 except as indicated. In specification 10, we employ a negative binomial regression but assume grant receipt is exogenous. The coefficient is approximately equal to the percentage increase in productivity associated with grant receipt. The absolute implied productivity increase is approximately 0.87 additional publications over 5 years.
* Significance at the 10% level.
** Statistical significance at the 5% level.

NIH grant policy in facilitating the transition from graduate school to research career.

When examining the benefits of the program, it is also important to consider the costs associated with the program. Over the life of a typical two-year F32 grant, the NIH spends approximately $100,000. Our results suggest that receipt of an NIH postdoctoral fellowship leads to a 7 percentage-point (24%) increase in the likelihood that an individual will pursue a career as an active (publishing) researcher. This implies that the program spends roughly $1.4 million to produce one additional research scientist.[18]

However, this back-of-the-envelope calculation represents a very conservative estimate of the social payoff associated with the NIH F32 program. First, as discussed earlier, our estimates capture only the impact of receiving an NIH postdoctoral fellowship *relative* to the next best option, which will, in turn, depend on the nature of the labor market for young researchers. To the extent that many organizations offer fellowships to high caliber graduates, we might expect the treatment effect to be small due to the quality of the outside option. To the extent that NIH postdoctoral fellowships expand the total supply of postgraduate research options (which will be the case unless there is complete crowd-out of non-NIH fellowships), the F32 program benefits even those applicants who do not receive a NIH fellowship themselves.

Second, our estimates will not capture any spillover benefits of postdoctoral fellowships, including the benefits accrued to the principal investigators or institutions in which the postdocs are employed. Third, our estimates do not include the potential benefit associated with encouraging other students to enter the biomedical field due to the larger number of research opportunities. Fourth, our publication and citation measures do not capture other important outputs associated with a research career such as patents and teaching.

Finally, data limitations also make it impossible for us to identify the precise mechanism through which the effect of a postdoctoral fellowship operates. We know that receiving an F32 grant increases the probability a researcher receives future NIH funding. Additionally, a fellowship might put the young researcher in contact with high ability colleagues, limit teaching obligations and/or increase his or her visibility in the profession. Future researchers may want to focus on more fully characterizing the total benefits associated with the F32 and exploring the mechanisms through which the fellowship influences recipients.

Given that the federal government is likely to continue to fund basic science research opportunities for young graduates in one form or the other, perhaps the most immediately relevant policy question involves the effectiveness of the F32 program relative to alternative grant mechanisms that also fund postgraduate research and training. This is particularly true given that fewer than 10% of new biomedical PhDs will receive F32 grants.[19] For example, NIH T32 grants are explicitly designed for both the pre and postdoctoral training of scientists. Funds from other grants, including R01 grants, can also be used to provide research support for new PhD's. The fact that F32 grants produce better outcomes, on average, than a researcher's next available options suggests that the program does have important benefits. A comprehensive examination of the effectiveness of F32 grants relative to alternative funding mechanisms would require a rigorous evaluation of the alternatives, which is beyond the scope of this analysis but should be a high priority for future research.

Acknowledgements

We would like to thank Andrew Arnott, Andrew Canter, Jessica Goldberg, Lisa Kolovich, J.D. LaRock, Stephanie Rennae and Thomas Wei for their excellent research assistance. We thank Richard Suzman, Robin Barr, Wally Schafer, Lyn Neil, Georgeanne Patmios, Angie-Chon Lee, Don McMaster, Vaishali Joshi and others at NIH for their assistance. We gratefully acknowledge support provided by NIH Express Evaluation Award 263-MD-514421. All remaining errors are our own.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.respol.2011.04.003.

References

Angrist, J.D., Lav, V., 1999. Using Maimonides rule to estimate the effect of class size on scholastic achievement. Quarterly Journal of Economics 114, 535–575.
Arora, A., Gambardella, A., 2005. The impact of NSF support for basic research in economics. Annales d'Economie et de Statistique 79/80, 91–117.
Berk, R.A., Rauma, D., 1983. Capitalizing on nonrandom assignment to treatments: a regression-discontinuity evaluation of a crime-control program. Journal of the American Statistical Association 78 (381), 21–28.
Black, S, 1999. Do better schools matter? Parental valuation of elementary education. Quarterly Journal of Economics 114, 577–599.

[18] The calculation is $100,000/0.07 = $1.4 million. This cost naturally varies depending on whether we adopt a more or less stringent definition of research scientist as well as the cost of the particular F32 grant.

[19] In 2006, 703 new F32 fellowships were awarded (see http://report.nih.gov/nihdatabook/ accessed on 11/15/2010) compared to 8537 individuals who completed a life science PhD.

Carter, G.M., Winkler, J.D., Biddle, A.K., 1987. An evaluation of the NIH research career development award. RAND Corporation Report. R-3568-NIH.
Hahn, J., Todd, P., Van der Klaauw, W., 2001. Identification of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–209.
Jacob, B., Lefgren, L., 2004a. Remedial education and student achievement: a regression-discontinuity analysis. Review of Economics and Statistics 86, 226–244.
Kaiser, J., 2008. Two strikes and you're out, grant applicants learn. Science 322, 358.
Pion, G.M., 2001. The early career progress of NRSA predoctoral trainees and fellows. NIH Publication Number 00-4900.
Stephan, P., Levin, S., 1989. Age and research productivity of academic scientists. Research in Higher Education 30, 531–549.
Thistlewaite, D., Campbell, D., 1960. Regression-discontinuity analysis: an alternative to the ex-post facto experiment. Journal of Educational Psychology 51, 309–317.
Trochim, W., 1984. Research Design for Program Evaluation: The Regression-Discontinuity Approach. Sage Publications, Beverly Hills, CA.