2017

Three Essays On The Costs Of Inducing Innovation

Kyle Roy Myers
University of Pennsylvania, myersky@wharton.upenn.edu

Three Essays On The Costs Of Inducing Innovation

Abstract

This dissertation explores the costs and benefits of inducing innovation at the level of markets, individuals and knowledge. First, I examine the extent to which the National Institutes of Health can direct its resources and generate production in specific areas of science. I find that it can, and that science funded through these directed efforts is significantly more productive than average. Second, I identify how willing individual scientists are to adjust the trajectories of their research in exchange for additional resources - the elasticity of direction. Estimated magnitudes suggest that the directional adjustment costs of biomedical science are large enough to warrant policy attention. Finally, in joint work with Mark Pauly, we explore the growing costs of R\&D in the pharmaceutical industry in order to identify how much might be efficient cost growth in response to a larger market. Almost all of the growth in R\&D spending can be attributed to demand, with no evidence that marginal investments have become less productive over the past thirty years.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Health Care Management & Economics

First Advisor

Matthew Grennan

Second Advisor

David Hsu

Keywords

economics, innovation, science

Subject Categories

Business Administration, Management, and Operations | Economics | Health and Medical Administration | Management Sciences and Quantitative Methods

THREE ESSAYS ON THE COSTS OF INDUCING INNOVATION

Kyle R. Myers

A DISSERTATION

in

Health Care Management & Economics

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

Matthew Grennan
Assistant Professor of Health Care Management & Economics

Co-Supervisor of Dissertation

David Hsu
Richard A. Sapp Professor, Professor of Management

Graduate Group Chairperson

Catherine Schrand
Celia Z. Moh Professor, Professor of Accounting

Dissertation Committee

Ashley Swanson, Assistant Professor of Health Care Management & Economics

Dan Levinthal, Reginald H. Jones Professor of Corporate Strategy

THREE ESSAYS ON THE COSTS OF INDUCING INNOVATION

Dedicated to my mother and father, whose love and support throughout my life has put me in the most fortunate positions. I cannot express my thanks enough.

ACKNOWLEDGEMENT

I am exceptionally grateful to my dissertation committee, Matt Grennan, David Hsu, Ashley Swanson and Dan Levinthal, for their wonderful insights and support over the past years. In particular, I must further acknowledge Matt Grennan; without his thoughts and encouragement this would certainly have been a less interesting dissertation, and I a less skilled researcher. Although not an official advisor, Mark Pauly has been both a supportive mentor and colleague - as evidenced by Chapter 4 of this dissertation. My discussions, research and teaching with John Kimberly also shaped a large part of how this dissertation evolved over time, certainly for the better.

My fellow Ph.D. students at Wharton provided uncountable amounts of help and conversations - both insightful and wonderfully off-topic - that have helped me grow as a student and scholar.

And it was only with the support of my friends and family that I was able to balance my pursuit of this dissertation with the rest of my life, which, thanks entirely to them, was not put on hold, but rather has easily been the best five years of my life. To John, Maegan, Katie, Corie-Mae, Tommy and most recently and importantly Kendra, thank you!

This dissertation would not have been possible without the hard work and support of Misty Heggeness who welcomed me and my ideas to the NIH. I am also grateful to Ryan Pfirrmann-Powell, Walter Schaffer, Sherry Mills, Kay Lund, Michael Lauer and the entire staff of the NIH's Office of Extramural Research for hosting and supporting me during my interim there. I am indebted to W. John Wilbur for taking the time to reconfigure his algorithm for use in this project. Funding from the Kauffman Foundation Dissertation Fellowship and the NBER program on the Value of Medical Research is gratefully acknowledged. The views and opinions expressed in this dissertation do not necessarily reflect those of the NIH or any other federal agencies.

ABSTRACT

THREE ESSAYS ON THE COSTS OF INDUCING INNOVATION

Kyle R. Myers

Matthew Grennan

This dissertation explores the costs and benefits of inducing innovation at the level of markets, individuals and knowledge. First, I examine the extent to which the National Institutes of Health can direct its resources and generate production in specific areas of science. I find that it can, and that science funded through these directed efforts is significantly more productive than average. Second, I identify how willing individual scientists are to adjust the trajectories of their research in exchange for additional resources - the elasticity of direction. Estimated magnitudes suggest that the directional adjustment costs of biomedical science are large enough to warrant policy attention. Finally, in joint work with Mark Pauly, we explore the growing costs of R&D in the pharmaceutical industry in order to identify how much might be efficient cost growth in response to a larger market. Almost all of the growth in R&D spending can be attributed to demand, with no evidence that marginal investments have become less productive over the past thirty years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : Introduction

The economic and social determinants of the rate and direction of innovation have long been of interest to scholars. These factors are relevant to many aspects of the economy from the management of specialists in laboratories and organizations, to national policies on the subsidizing and regulation of R&D. Initial work by Kenneth Arrow, Richard Nelson and Sidney Winter identified many of the major tensions in this field, but recent advents in data accessibility and the ability to quantify traditionally imprecise concepts, such as the "direction" of innovation, provide a unique opportunity to revisit and address these questions with robust empirical techniques.

This dissertation investigates how specialized individuals and firms respond to changes in the demand for particular types of innovations, emphasizing the magnitude of frictions that can constrain the extent to which a manager or market may be able to "induce" innovation.

To begin, Chapter 2 investigates the use of targeted research competitions at the National Institutes of Health (NIH). The rationale for the public funding of basic science is well understood - firms will underinvest in *certain types of science* because of limited appropriability (Nelson 1959; Arrow 1962). However, the ability of the public agencies to direct, and not just provide, funds for specific types of science has not received much scrutiny. This chapter provides causal evidence on the ability of targeted contests at the NIH to achieve this direction. An algorithm is used to classify the biomedical subjects of abstracts and measure the rate of science in each possible direction. This facilitates a straightforward difference-in-difference design at the level of these scientific subjects to identify changes in the rates of science targeted by the NIH, while accounting for the potential role of political capital. I find an increase in the rate of targeted science proposed and funded at the NIH on the order of 10%. Instrumental variable estimates indicate large productivity gains - projects funded through targeted requests generate almost twice as many publications per dollar.

Given evidence that new science is induced, Chapter 3 identifies a key parameter under-

lying the responsiveness of scientists to these targeted interventions: the elasticity of the redirection of research, ε. Using a novel dataset of targeted grant contests at the NIH and a validated measure of scientific similarity, I identify the parameter using two approaches: first as a compensating differential, and second as a fixed cost of entry. Conditional on applying to the NIH, scientists are indifferent between a 1 s.d. redirection (moving to less similar science) and an additional $1.6 million in expected grant value. Per the entry model, being 10% closer to a contest, in the scientific sense, increases entry probability the same as an $1.2 million (48%) increase in the grant funds available for competition. These magnitudes ($\varepsilon \approx 0.25 - 0.6$) suggest that the directional adjustment costs of biomedical science are large enough to warrant policy attention.

Looking at the determinants of innovation at a more aggregated level and within a private market of firms, Chapter 4 with Mark Pauly investigates the dramatic growth in the pharmaceutical R&D costs per new drug. We draw insight from David Ricardo's theory of demand-driven productivity in settings of variable scarce resources - here, valuable new drug ideas - and estimate the industry's responsiveness to changes in demand over the past three decades. In contrast to many critics of this industry, our results suggest that innovative new drugs have been brought to market remarkable consistency since the late-1980s, and rather than any major supply-side frictions, it appears that increased competition in drug development have effectively bid down excessive profits from decades past.

CHAPTER 2 : Managing the Rate and Direction of Public Science: Evidence from the NIH

2.1. Introduction

Since the arguments of Nelson (1959) and Arrow (1962), it has been held that the private sector will underinvest in the socially optimal amount of basic science - knowledge generation - given the inability to appropriate its full value. In order to address this friction, the government must not only *provide* funds, but they must also *direct* these funds to the specific activities underinvested in by the private sector. Typically, public organizations rely on peer-review panels to assess the value of funding opportunities. There is growing evidence from the National Institutes of Health (NIH) that such panels are able to identify promising lines of research (Li and Agha 2015) and draw on expertise without losses due to strategic incentives (Li 2017). However, it is still not clear a priori that the allocation of funds decided by scientists themselves will be optimal from society's perspective (Dasgupta and David 1994)[1]. So while related work has identified the productivity of publicly funded basic research in general (Jacob and Lefgren 2011; Azoulay et al. 2015c), this Chapter is an important complement as it examines the ability of the NIH, the largest public funder of science, to choose *where* in the spectrum of science production occurs.

While the NIH traditionally solicits any proposal for research that may "enhance health, lengthen life, and reduce illness and disability," roughly 25% of the annual budget (∼$8.2 billion in 2016) is allocated to requests for applications (RFAs) that focus on specific scientific subjects. The goal of this Chapter is identifying to what extent these allocations can induce new science that would not have otherwise occurred - an effect necessary to address the issues raised by Nelson and Arrow, but one lacking causal evidence as to its magnitude.

[1] To quote the authors, "This sort of non-congruence between private and social rankings of final outcomes creates fundamental grounds for suspecting that the research portfolio that would be, in effect, selected, for society by the self-governing community of scientists will be an inefficient one" (Dasgupta and David 1994, pp. 506). This incongruence stems from the fact that a prize-centric reward system - which is necessary in science given the inappropriability of information - will inherently cause directional distortions (Bryan and Lemus 2016)

Essentially, the NIH's objective is to generate a demand shock large enough to induce a corresponding supply increase. What might limit the NIH's ability to do so? First, since science is funded through a multitude of public and private sources, other funders may anticipate growth in particular funding streams at the NIH and reduce their own allocations. Such crowd-out is possible due to the limited appropriability of knowledge. Second, given the quasi-fixed supply of this specialized workforce (e.g. Goolsbee 1998) and the control relinquished to scientists during production (e.g. Aghion et al. 2008), it is unclear that a demand shock can be generated that is large enough to increase the supply of a particular type of science. Conversely, incentivizing new research pursuits may alleviate certain frictions that constrain scientists in their evaluation of new ideas (e.g. Boudreau et al. 2016a).

To investigate the aggregate effect of these forces, I utilize the NIH's administrative database to construct a novel dataset of roughly 900 RFAs, 8 years of grant applications from more than 80,000 scientists, along with all of the scientists' publications, regardless of funding source. This dataset facilitates two complementary analyses: (1) a difference-in-difference design at the level of scientific subjects (i.e. "neoplasms", "hormones") to test whether new science can be induced by RFAs, and (2) instrumental variable regressions to disentangle the extent to which productivity effects are driven by within-project differences (i.e. more science per publication) or across-project differences (i.e. more publications per project).

The data and algorithm help me tackle a major difficulty facing empirical work on the rate and direction of innovation: categorizing and indexing technologies. The state-of-the-art text algorithm from the National Library of Medicine enables me to extract information from abstracts of grant applications, publications, and RFAs. The algorithm identifies the unique set of biomedical subjects relevant to a body of text, and is based on a standardized and highly detailed dictionary providing clear boundaries and relationships between subjects.

Measurement issues aside, estimating the causal effects of these interventions from observational data requires that the NIH's decision to create each RFA is exogenous to factors that would otherwise influence scientists' preferences over the targeted science. However,

the NIH is a political institution who's own demand - which is influenced by Congress - is likely correlated with other forces of supply and demand in the market of science.

Consider the case of the recent Zika Virus outbreak. This event was followed by both Congressional requests for research on the topic as well as an RFA the same year[2]. Presumably, the outbreak also increased the inherent value of Zika virus research to scientists - individuals responsible for any breakthroughs would likely gain notoriety that would not have occurred in the absence of the outbreak[3]. Thus, estimating scientists' responsiveness in this scenario would confound both the NIH's targeted allocation with other, non-NIH demand-side changes, potentially overestimating the ability of the NIH to redirect research.

In order to account for these correlated demand- or supply-side events (e.g. new discoveries) I construct a proxy for the political capital of each subject - its appearance in Congressional appropriation bills - in order to separately identify effects for the science that is "caught up" in these contests, given that when creating these contests NIH staff must target a range of subjects. For example, in the case of the RFA requesting research on Zika, the research objectives also welcomed projects on "Lyme borreliosis" and "Mycobacterium tuberculosis" and I focus on these latter subjects.

Overall, my results indicate that with an average size of $11 million and 6-7 grants, the targeted contests induce a 8% (12%) increase in the rate of science proposed in all (successful) applications. The evidence also suggests that the newly proposed science is of roughly comparable ex-ante quality (per success rates) to counterfactual science, while being nearly twice as marginally productive in ex-post terms of publication mentions per application mention. These productivity gains at the level of science appear to be almost entirely due to differences *across* projects - compared to the traditional "investigator-initiated" mechanism, projects awarded via RFAs generate 1.8 times as many publications per dollar awarded.

[2]Requests for research specifically on Zika virus appear in the 2016 appropriation bills for the Department of Health and Human services, as well as in a 2016 targeted research contest at the NIH, see RFA-AI-16-034.

[3]In fact, they already have; see: S. Mukherjee, "The Race for a Zika Vaccine", *The New Yorker*, August 22, 2016.

These gains combined with a lack of any apparent substitution away from non-treated subject areas suggests that, so long as the marginal value of knowledge in the treated subjects is at least half as large as in the non-treated subjects, these RFAs are a cost-effective use of public resources committed to basic science. However, RFAs intended for small businesses do not exhibit the clear signs of inducement observed in the basic science grants. Thus, while these early-stage commercial grants have been seen to alleviate frictions unique to emerging firms (Howell 2015), these results highlight the difficulties that policymakers may face when redirecting scientists at this stage of development.

This Chapter is most closely related to the literature on public "inducement prizes". Williams (2012) provides an overview of the state of science regarding inducement prizes citing the major difficulties facing empirical work on these mechanisms, namely, identifying a valid counterfactual, which I am able to address with my data. Brunt et al.'s (2012) analyses of prizes awarded by the Royal Agricultural Society of England during the turn of the 20^{th} century is closely related to my paper. They find that prizes awarded for new, patentable technologies were effective and that efforts did not appear to substitute away from non-targeted technologies. They argue, as I do here, that the timing of the governments' intervention was plausibly exogenous to underlying trends in the supply or demand for these technologies that might have unobservably increased the benefits (or decreased the costs) of invention. Unlike this prior study, my data allows me to statistically test this assumption.

An earlier examination of the NIH by Hegde and Sampat (2015) focused on the private sector's ability to influence the NIH's allocations. Their result - that lobbying efforts can increase NIH award rates for rare conditions - indirectly supports the notion that the NIH has control over the direction of research. However, the authors' data prevents them from observing changes in application behavior at the NIH or constructing a valid counterfactual post-award, which are both necessary for a full understanding of these interventions.

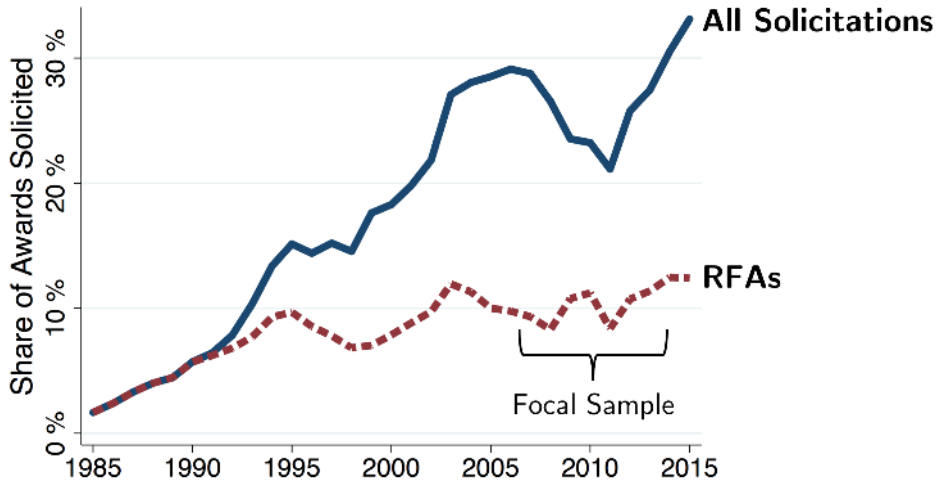
The outline of this Chapter is as follows: Section 2.2 describes the NIH setting and data; Section 2.3 outlines a simple supply and demand model that clarifies the analyses; Section

2.4 presents the difference-in-difference results on inducement; Section 2.5 presents the I.V. regression results on productivity; Section 2.6 concludes.

2.2. Managing Research at the NIH

Although typically billed as the world's largest funder of biomedical research, the NIH has increasingly taken the role of as one of the world's largest managers of research. In 2015, roughly 33% ($11B) of the research projects funded by the NIH were awarded via solicitations for specific types of research, compared to roughly 5% in 1990. Figure 1 plots the share of research project grants awarded via solicitations for specific science, as opposed to investigator-initiated or "open" applications, which comprise the remainder of awards. The "Focal Sample" on Figure 1 identifies the subset of Requests For Applications (RFAs) - a specific solicitation mechanism at the NIH - that I examine in this Chapter.

Figure 1: Trends in Solicitation at the NIH



Note: Data is based on the number of all research project grants awarded at the NIH. Solicitations are grants awarded through certain mechanisms (e.g. Program Announcements, Requests For Applications (RFAs)), which request specific types of science to be submitted. The "Focal" sample includes only the RFAs that are empirically examined in this Chapter.

The traditional, continuously available open grant competitions at the NIH award research grants three times per year. The application process is as follows. During each of these rounds, an application is received by an NIH Institute sorted to one of the nearly 200

standing or temporary peer review study sections of the NIH's Center for Scientific Review. Irrespective of the Institute that received the application, this sorting is determined by the type of science proposed. For example, the "Biomedical Computing and Health Informatics", "Instrumentation and Systems Development" and "Genetics of Health and Disease" study sections have all at some time reviewed applications from more than 20 separate NIH Institutes given their breadth of coverage. Within a study section, a peer review panel assigns a raw score to each application, which is then transformed into a "percentile score" that is intended to adjust for underlying differences across study sections as well as within study sections over time[4]. This is one of the primary ways through which the NIH exhibits a preference for equity across types of science. Final funding decisions each round are then based at the Institute level, of which there are 24, where applications are ranked per their study section percentile scores and funded as far down the percentile ranking as an Institute's budget permits[5].

It is extremely difficult to predict success in these open contests. The average application receives its percentile score based on the quality of 75.6 (s.d.=108.3) other applications in its study section, and competes with 707.9 (s.d.=843.4) other applications at the Institute level. As evidence, the variance in success rates *within* scientists is more than twice as large as the variance in average win rates *across* scientists. This is not driven by a bimodal distribution of always-winners and always-losers: of the the 28,315 scientists with 3 to 5 applications in my sample, only 66 were successful each time they applied. And the administrative costs of developing applications and maintaining administrative responsibilities are not trivial. A 2007 survey found that academic scientists spend upwards of 20% of their time preparing federal grant applications (Decker et al. 2007), with many academic institutions devoting significant resources to NIH grant writing workshops and full-time staff to aid in these proposals[6].

[4]Formally, each applications percentile score is the percent of applications in the three prior rounds of the same study section with raw review scores greater than the focal application's raw score. Econometrically speaking, this is analogous to using study section fixed effects as a control.

[5]This is an approximation of the official funding process, which are outlined here: https://goo.gl/blLuuU.

[6]A cottage industry of grant-writing consultants exists; see https://goo.gl/fpZ5zq.

Compared to these open grant competitions, RFA competitions occur only once and request a limited number of scientific subjects to be pursued and are thus a primary mechanism through which targeted research is requested by staff at the NIH[7]. In these one-time contests, funds from one or more NIH Institutes are set aside for a single grant competition related to the predefined area of science. The RFAs are typically announced four to eight months prior to the deadline for submissions. Applications are evaluated by a single peer-review panel and the awarding of grants is based on the number and quality of applications given the amount of money allocated to the RFA[8]. The average RFA includes $2.8 million to be awarded annually over 3 to 5 years for 6 to 7grants, with an average of 37 scientists (s.d. = 58) participating in each contest. I provide more detailed statistics and discuss the typical scope of these competitions in Section 3.3 below.

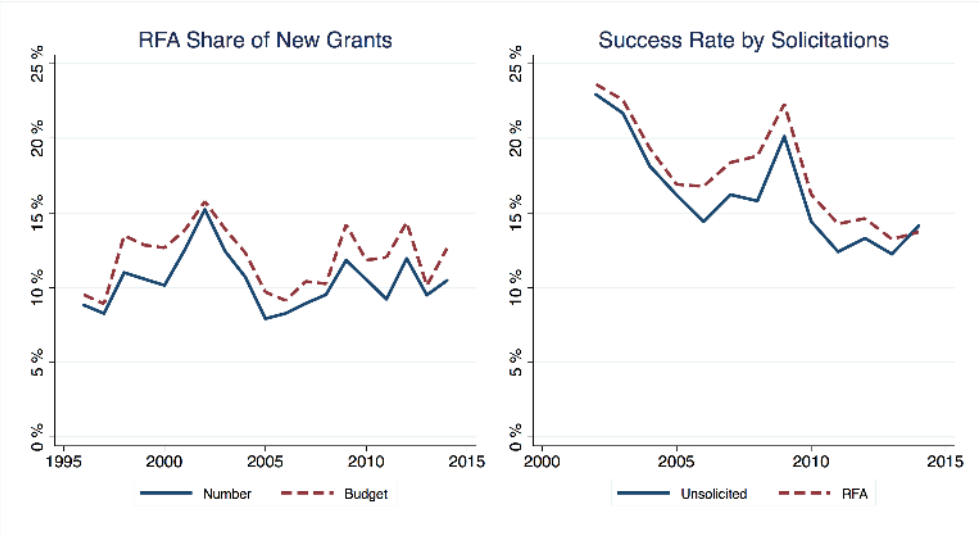
Figure 2 shows the share of the NIH's total budget allocated to RFAs, and their average success rates as compared to the unsolicited investigator-initiated grant mechanisms. To get a sense, in 2014 the NIH awarded $1.03 billion in new grants via RFAs and an additional $2.66 billion in continuing RFA grants for projects funded through multiple years. Section 2.2.2 below provides an overview of the reasons why this targeting occurs how forces are relevant for my analyses.

I focus my empirical analyses on four of the most common research grant types awarded. The set can be summarized by two distinct groups: basic and applied research. While delineating between basic and applied science is a notoriously difficult exercise, the NIH grant types provide relatively clear boundaries. Overall, the set of basic science grants are the traditional funding mechanisms designed to support "discrete, specified, circumscribed research projects". The differences amongst these grants lies largely in the size of the award

[7]The NIH also releases "Program Announcements" where certain types of science is solicited; however, these calls are not accompanied by set-aside funds made specifically available for competition and in practice vary widely in their format. I focus my analyses on RFAs because of their well-defined properties as funding mechanisms.

[8]The language typical to the grant announcements description of the number of awards is along the lines of "...the [sponsoring Institute] expects to award X to Y grants, contingent upon the submission of a sufficient number of scientifically meritorious applications and the availability of funds."

Figure 2: RFA-Open Comparison



Note: Data for all R-series grants studied in this Chapter (R01, R21, R41, R43), described at length in text. The share of budget and awards is with respect to all new R-series grants awarded. A new grant is an application submitted to fund a research project that has not previously received NIH funds in its current state. After the first year of a multi-year grant award, scientists must re-apply for the continuation of funds that were a part of the initial award, or to re-compete for additional years of funding; these continuing and re-compete applications and awards are not included here.

and the stage of research (e.g. feasibility studies versus large-scale). The difference between the two most common basic science grants, the R21 and the R01, is akin to the explore-exploit tradeoff commonly investigated in studies of R&D strategy. The larger R01 grants are intended to support full-scale research projects that have preliminary results and build on the scientists' current base of knowledge (exploit). The smaller R21 grants are intended to develop new lines of research and do not require any preliminary results (explore).

In contrast, the applied science grants are awarded through the NIH's Small Business Technology Transfer and Small Business Innovation Research programs. These mechanisms are designed to fund "cooperative research and development carried out between small business concerns and research institutions" (R41) or "research and development for for-profit institutions for ideas that have potential for commercialization" (R43), obviously more applied compared to the traditional grants of the NIH. The major difference between the two is that the R41 mechanism requires participation from both a private entity and a research

institution such as a University.

Within RFAs, the average total award size for the two basic grant types, R01 and R21, are $1.3 million and $375,000, respectively, with the more applied grants averaging roughly $200,000 for technology transfer R41 grants and $1 million for the commercial R43 grants. In all cases, applicants may pursue these grants in the rolling open competitions. But notably, the average size of these awards when awarded through the open contests is roughly 17-20% smaller than the awards in the RFAs. The majority of applications (90%) and award dollars (95%) are allocated via the basic science grants across both targeted and open contests. Still, the NIH awards roughly 750 new applied grants ($900 million) each fiscal year. And in light of recent evidence that federal grants awarded through a very similar program at the Department of Energy can cause significant growth for early-stage ventures by spurring investment from private sources (Howell 2015), understanding the extent to which these grants can direct the nature of science at these ventures is an open and important public policy question.

An important policy surrounding RFAs is that scientists are permitted to resubmit unsuccessful RFA applications to the open grant competitions, unlike submissions to the open contests. This rule enables inter-temporal spillovers in application rounds following the contest. Scientists may be induced by the RFA to propose a research idea, fail, but receive feedback useful in a future application. However, this also implies that RFAs may be a low-cost source of information regarding a proposal already in preparation, or simply provides a second chance for the applications. Thus, there may also be inter-temporal substitution from previous or future open application rounds as scientists delay or rush proposals already in intended for submission. I will also examine potential substitution and spillover effects across grant types (e.g. does soliciting R01 grants on subject A crowd-out R21 grants on subject A?) as well as across similar types of science (e.g. is an increase in research proposed on subject A associated with a decrease in research proposed on the similar subject B?).

One other policy relevant for my analyses is the NIH's New Investigator program[9]. In order to prevent grant funds from being concentrated among older experienced investigators where there are potentially less efficient (Levin and Stephan 1991; Freeman and Van Reenen 2009), the NIH explicitly labels individuals who have yet to successfully compete for a research grant as New Investigators. When making funding decisions (after peer review) for open applications, these proposals are typically evaluated separately in order to maintain comparable success rates among New and experienced investigators. However, this policy does not hold in the targeted contest. While the theory of scientific choice described below will suggest that these younger, less experienced scientists may face lower costs when adjusting the trajectory of their work and hence be more likely to compete in the targeted contests, the lack of this feature may also prove a significant barrier. The raw data suggests this preference for New Investigators in the open contests is valuable: only 14% of RFA applications are from New Investigators, compared to 24% in the open contests.

2.2.1. Data Sources

Data on all grant applications to the NIH from 2006 to 2014 are from the NIH's confidential administrative database. The data contains all information surrounding each application: abstract of the proposed research, review scores, funding decisions and project outcomes (i.e. publications). Because the NIH's protocol for review scores changed in the middle of my sample period, I construct a new "grade" for each application based on the standardized distributions of review scores before and after the protocol change. This grade variable ranges from 0 to 1 and indicates the percent of all applications with lower review scores. Controls introduced in my empirical analyses will capture any variation over time across all applications in the propensity to award certain grades that may be unrelated to the true quality of applications. The percentile score mentioned earlier is not useful for my analyses because it is not utilized in RFA contests, and more importantly, while it would be possible to impute these percentiles, I am interested in the *raw* application quality. Using the

[9]See https://goo.gl/Ynd0W3 for the specific details of this program.

unadjusted review scores allows me to implement my precise science-level controls without limitations from the organization of NIH review groups or Institutes.

Details of each RFA from 2006 to 2014 was scraped from the NIH announcement website[10], and formatted as a dataset including each contest's timing, text of the research objectives, type of grants requested and the number of awards and dollars allocated.

Each scientist's publication history, regardless of whether the project was NIH-funded, are constructed using the disambiguated version of the PubMed scientific article database developed by Torvik and Smalheiser (2009). To address the notorious difficulties of accurately matching publications to scientists given the lack of standardized identifiers in this data, the authors developed a maximum likelihood based agglomerative algorithm for computing clusters of articles that belong to the same inferred author. Thus, I am confident in the fidelity of my data, which is essential for the algorithms described below. To approximate the quality of publication outcomes, Journal Citation Reports from Thompson Reuters is used to obtain the impact factor and other citation-based metrics associated with each journal.

2.2.2. Targeted Contest Generation

Why do RFAs occur? For my empirical analyses to be generalizable to science not treated with RFAs, it is important to identify the extent to which the creation of these contests might be correlated with unobservable time-varying changes to the value associated with each subject. To be clear, one might be concerned that events such as changes to the prevalence of a disease or a scientific breakthrough might occur at the same time of the NIH's decision to request a particular type of science. In this case, estimating the effect of RFAs would confound both the NIH's intervention with this underlying change in the value of that science to researchers, as the staff have effectively selected on these unobservable dimensions of science. In this section, I discuss each impetus in terms of its relevance to my empirical estimates - i.e. is selection of concern?

[10] Available at https://goo.gl/LuaBOQ

Discussions with NIH staff responsible for the creation and management of these solicitations identified two main reasons: political capital and portfolio analyses[11]. The portfolio analyses involve staff within each Institute evaluating the distribution of supported research in order to identify "gaps" in the portfolio of funded work. The preference for equity in the distribution of NIH-funded research suggests that these gaps likely exist because of *lower* scientific fertility or demand associated with these subjects. Thus, to the extent that RFAs are created by NIH staff in order to create a more uniform distribution of science, my estimates of inducement will be biased downward as a lower bound relative to the true average effect across all subjects.

Interviewees also mentioned that the likelihood of an RFA being developed is often tied to unplanned shocks to the budget. For reasons outside the control NIH program officials, an Institute may have an unexpected amount of funds remaining after prior obligations are met. This notion of RFAs being tied to budgetary slack is apparent in the aggregate data: sharp increases in the number of targeted contests are apparent from 1997-2003 and in 2009, during Congress' "budget doubling" initiative and the American Reinvestment and Recovery Act, respectively, where large surges to Institute budgets occurred (see Figure 2). Because the NIH's budget is Congressionally allocated at the very broad level of Institute, it is unlikely that these Institute-wide budgetary shocks are correlated with the underlying value of the specific biomedical subjects eventually requested by any RFA made possible by these budgetary shocks.

To the notion of political capital, prior work has found that Congressional requests can influence NIH funding allocations (Hegde and Sampat 2015). In fact, Hegde and Sampat (2015) find that this influence is mediated largely via the targeted contests I study here. To quote the authors, "Specific types of NIH grants - for example, 'requests for application' (RFAs) and 'program announcement' (PAs), which solicit research proposals in particular

[11] As part of an NIH-sponsored internship, I worked on-site for roughly one month at the NIH Office of Extramural Research where all funding opportunities and grant programs are approved. Informal interviews were conducted with staff involved in the management of RFAs and other policies governing the granting process.

areas of research rather than unsolicited investigator-initiated research - are most strongly related to soft earmarking" (Hegde and Sampat 2015, pp. 2282). Conducting an analysis similar to Hegde and Sampat (2015), but using my data on subject occurrences in both RFAs and appropriation bills, I am able to explicitly connect the soft earmarks for scientific subjects to their appearance in the text of targeted funding opportunities. Subjects that appear in these soft earmarks at least once in my sample are roughly 12 percentage points (> 2 times) more likely to also appear in the text of the research objectives of the RFAs I study, which occurs for only 5% of all subjects.

Overall these findings are in line with prior research on the determinants of the direction of publicly funded research. Azoulay et al. (2015c) explore the impact of NIH grants around random discontinuities that arise in the NIH's budgeting process, but find little evidence of "strategic" funding of grants with higher downstream potential[12]. Similarly, Williams' (2013) investigation of gene patents finds that publicly funded genetic sequencing efforts did not appear to select on ex-ante more valuable genes. Certainly, public research efforts do not occur at random. But taken together, there is good evidence to suggest that the uncertainties of science and the influence of political interests appear to vastly overwhelm any information that public officials may have about the most fruitful lines of research.

Still, the political capital associated with certain scientific subjects may be correlated with the underlying value of scientific field. Recall the example of the recent Zika virus outbreak in the Introduction. This (exogenous) occurrence drastically increased the inherent value of research on this topic, particularly for US-based biomedical scientists. While my empirical methods will control for fixed time-invariant differences across types of science, it is this sort of shock that is problematic. In the case of Zika, Congress did in fact request research from the NIH, and the NIH did in fact create an RFA soliciting research on the virus. In order to identify the extent to which selection such as this is problematic, I construct a proxy for

[12]"We add increasingly detailed fixed effects in each successive column; interestingly, our estimates remain relatively stable. One explanation for this is consistency is that, at the time it makes funding decisions, the NIH may not be able to anticipate which [areas of science] have greater future innovative potential.," Azoulay et al. (2015c), pp. 21-22.

the political capital of each subject based on whether or not the term appears in any of the Congressional appropriations bills from 2006-2013. I then rely on the fact that when creating this contests, NIH staff must use more of a "shotgun" approach, rather than a "rifle" approach to targeting specific areas of science. In the case of the RFA targeting Zika, it is actually titled a request for "Countermeasures against Select Pathogens", to include the Zika virus, as well as vaccines with small market potential such as "Coccidioides spp." infections. By separately examining effects only for these subjects *not* requested by politicians but still caught up in the creation of these contests, I assume that the RFA treatment is uncorrelated with time-varying shocks to the value of these particular subjects. Statistical tests built into my empirical specification will test this assumption.

Furthermore, by comparing differences in effects for the two sets of subjects (mentioned and not mentioned by Congress) I can provide some insight as to the role of political capital in the selection of government programs. This is an important empirical exercise given the large number of studies in the social sciences that examine public programs and assume that the decisions of government actors do not select on expectations of outcomes in a systematic way[13].

2.2.3. Data Construction and Relevant Facts

A classic difficulty facing empirical studies of trends in science or innovation is accurately identifying the unit of analysis. The problem lies in creating stable definitions of products, markets or technologies and separating these units into control and treatment groups. Accurately and consistently defining these boundaries while also quantifying the proximity (or similarity) between separate groups often requires strong assumptions when the unit of analysis is not obvious or evolves over time. Well-defined boundaries are important for making clear statements about changes in outcomes over time, and understanding the proximity between units is important for identifying spillovers or substitution effects, which are likely

[13]For example, the parallel trends assumption of many district-based policy analyses implies that the policy-makers who chose to enact a certain program or law were not able to select on features of their population known to be positively correlated with the treatment effect or outcomes of interest.

to come from "nearby" units when distance, be it physically or scientifically speaking, is an important mediating factor.

Consider the RFAs I focus on in this study. How might one use the raw data to evaluate their effect on the rate of science funded by the NIH? One method for categorizing applications at the NIH used in prior studies is to combine the study sections and Institutes (discussed in the prior section) to form "Disease-Science" units that can be traced over time (Azoulay et al. 2015c). However, because RFAs are one-time mechanisms they are not directly associated with any particular study-sections and can be sponsored by multiple Institutes. Clearly identifying treatment and control groups would be challenging. Furthermore, the grouping of study sections and Institutes is based on the organizational structure of the NIH and does not necessarily reflect the organization of science[14]. So for example, if an RFA is seen to receive ten applications it is not clear how to use this information to test whether or not these ten applications - or more importantly, the specific types of science proposed by these applications - would have occurred in the counterfactual scenario where no research was targeted.

Thankfully, the hierarchical nature of information in the life sciences and the two algorithms I utilize allow me to (1) accurately transform research projects into well-defined units: scientific subjects; and (2) quantify the distance, what I term scientific similarity, between any pair of research projects. This enables me to build panels of data at the level of science and scientist that facilitate difference-in-difference research designs. The remainder of this section outlines how my main datasets are constructed and highlights patterns in the data relevant to my research design. Here I succinctly discuss how the text algorithms operate referring the interested reader to the Appendix for a detailed description of the algorithms and their output.

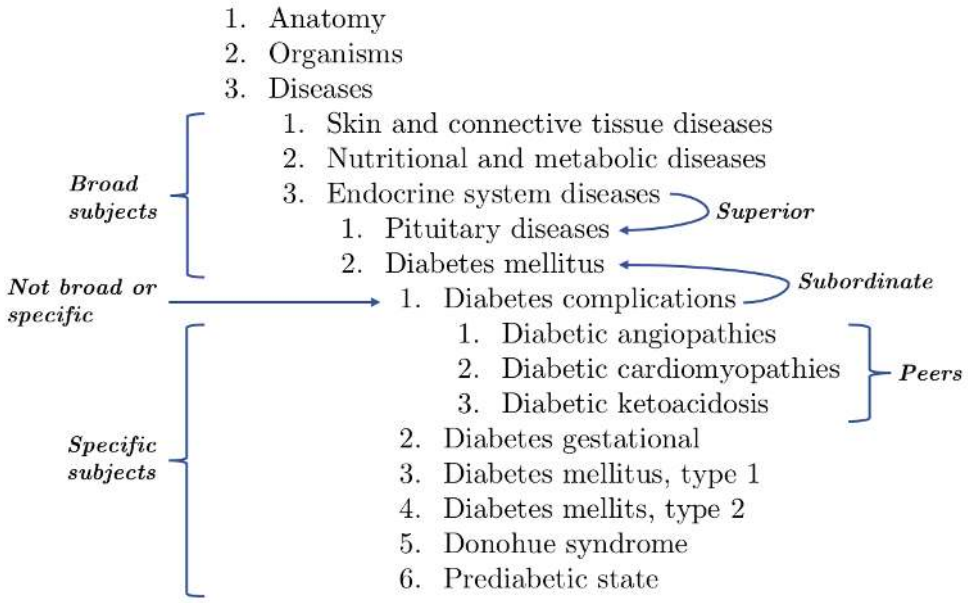
[14]For example, some Institutes are organized around diseases, National Cancer Institute, some around science, National Institute of General Medical Sciences, and some around populations, National Institute of Aging.

2.2.4. Rate of Science

Indexing Science: MeSH and the MTI

The MeSH hierarchy is a vocabulary of biomedical subjects maintained by the National Library of Medicine (NLM). The tree of terms has sixteen categories at its highest level, for example "Diseases [C]" and "Chemicals and Drugs [D]" with up to twelve levels of detail below, for example "Spinal Neoplasms [C04.588.149.828]" and "Antigens, Fungal [D23.050.202]". Some terminology warrants definition: a *superior* is a term at a level above a given point, as in closer to the highest sixteen categories, with a *subordinate* referring to the opposite; *peers* are a set of terms that share the same direct superior one level above. Figure 3 provides a sample portion of the hierarchy and illustrates these concepts.

Figure 3: MeSH Structure and Terminology



Note: This is a very restricted section of the MeSH tree to demonstrate relevant terminology. Indentations indicate separate levels of the hierarchy. For the most current, full MeSH tree visit https://goo.gl/zyeQ5z

The MeSH tree provides the well-defined boundaries for my units of analyses; however, as one might intuit in Figure 3, there are obvious differences in the breadth of subjects depending on their level within the tree. "Diseases" is a *broad* scientific subject that encompasses many

specific subjects such as "Diabetes mellitus, type 1". Intuitively, the location of a subject within the hierarchy is related to the breadth of the subject. However, the organization of the tree varies widely across the major categories, so simply using a subjects level would not be a consistent measure of subject specificity[15]. Thus, I define broad subjects as the highest two subordinates of the tree and specific subjects as those with no subordinates (see Figure 3). Because changes to the rate with which subjects occur in my data will depend largely on the breadth of a subject, my empirical analyses below estimates *percent* changes in my outcomes. I will present specifications that separately examine broad and specific subjects. Furthermore, each MeSH term also has a corresponding date of creation. With this information, I have a proxy for how "new" a scientific field, an aspect of heterogeneity that I will explore in the empirics. In total my set of MeSH subjects include 23,896 subjects, 94% of which are specific[16].

As mentioned earlier, the possibility of spillovers or substitution effects is most often mediated by the similarity between two units. When testing for these effects within subjects, my data allows me to look over time and across grant types. When testing for these effects across subjects, I focus on the peer subjects given their obvious proximity in the scientific sense. Referring again to Figure 3, it is reasonable to assume that an RFA requesting research on "Diabetic angiopathies" would most likely substitute for or spillover into research related to other diabetic complications.

Because it is commonplace for multiple words to refer to the same fundamental MeSH term (e.g. cancer, tumor and neoplasm) accurately matching free-form text to a unique MeSH term is a difficult task, particularly for large datasets. Previous work that has utilized the MeSH vocabulary as a means of indexing science has relied individual librarians to hand-code text (Boudreau et al. 2016a). To overcome this challenge at the scale of my data, I

[15]For example, "Human" has ten superior terms and "Bronchiolitis, Viral" has only two superior terms while both are equally specific subjects in the sense they have no subordinates

[16]The online Appendix outlines how I arrived at the final sample of MeSH terms, essentially, by removing subjects obviously less valuable to the progress of science, such as "Tape Recording" or "Architectural Drawings".

employ the Medical Text Indexer (MTI) developed at the NLM (Mork et al. 2013) in order to index abstracts automatically. The MTI utilizes a machine-learning approach to parse and index words in order to identify the unique set of MeSH terms relevant to a body of text[17]. For example, the sentence "Clustered regularly interspaced short palindromic repeats are segments of prokaryotic DNA containing short repetitions of base sequences" is identified as being related to "base sequence", "CRISPR", "DNA" and "genetics". Note that while the word "genetics" does not appear in the sentence, the MTI is able to identify its relevance.

One may be concerned that the appearance of certain terms within, say a grant application, is not necessarily indicative as to the actual science proposed. For example, many applications mention the potential implications of their findings in the closing sentence of their abstracts (e.g. "our findings on X may be relevant to diseases A, B and C"). To adjust for the inclusion of terms likely to be less relevant to the true nature of the study, I conduct robustness tests of my analyses where I restrict my sample to subjects identified by the MTI as being most relevant to the input text[18]. Furthermore, when examining publication outcomes, I directly process the abstracts of articles funded by the grants and do not rely on the application abstract. This allows for the possibility that the true nature of a project are not accurately captured by the proposed abstract.

Dataset & Patterns: Science

The first dataset I construct is based on the MTI algorithm and is used to identify the aggregate effects of RFAs on the rate of science proposed, funded and published at the NIH. Each abstract from grant applications and publications (connected to successful applications) is processed by the MTI to identify the MeSH terms associated with each project, successful or not. The count of these terms - what I refer to as the rate of science - is collapsed to the

[17]The MTI is accessible at https://ii.nlm.nih.gov/MTI.

[18]The MTI ranks and scores the relevance of each term identified based on its number of occurrences and the confidence of the machine-learning algorithm. By limiting the output to the words with higher rank or score, per some cutoff, can identify the subjects most relevant. Conversations with Dr. James Mork of the NLM team who maintains the MTI (Mork et al. 2013) assisted in the use and interpretation of the MTI's output.

grant type (e.g. R01, R41) by application round level. The resulting dataset is comprised of 23,896 subjects for each of the 4 grant types spanning 25 application rounds from 2006 to 2014. The average rate of occurrence in each round for broad subjects for the three measures - all applications, successful applications and resulting publications - is 105.3 (sd=441.0), 15.5 (sd= 64.6) and 21.8 (sd=126.0), respectively. Likewise, the average rates for specific subjects are 3.6 (sd=29.7), 0.53 (sd=4.3) and 0.88 (sd=8.9), respectively.

Next, the "Research Objectives" section of each RFA announcement is processed by the MTI in order to identify the scientific subjects requested by the contest. In total, I examine 854 RFAs of which 85.7% utilize basic grant types with the remainder requesting commercial grant types. The average number of subjects identified (treated) in each RFA is 90.7 (sd=37.2) with the share of subjects requested per their major MeSH categories as follows: 8% anatomy; 6% organisms; 12% diseases; 16% drugs or chemicals; 23% techniques or equipment; 17% psychology or psychiatry; and 18% phenomena or processes.

Table 1: Summary Statistics of RFAs by Grant Type

	Basic		Commercial	
	R01	R21	R41	R43
Awards Announced	7.68 (5.09)	15.4 (11.3)	4.89 (1.01)	6.00 (2.13)
Awards Actual	5.92 (5.63)	6.05 (6.51)	0.92 (1.26)	3.37 (5.06)
Applications Received	36.8 (51.2)	41.5 (49.0)	4.54 (3.87)	22.6 (66.1)
Total $M FY Announced	3.48 (3.61)	2.13 (1.36)	1.20 (0.76)	1.74 (1.05)
Total $M FY Actual	3.03 (3.39)	1.35 (1.49)	0.19 (0.28)	0.71 (1.08)
Avg. grant duration	4.06 (1.05)	2.10 (0.55)	1.16 (0.37)	1.30 (0.46)
N	531	272	39	99

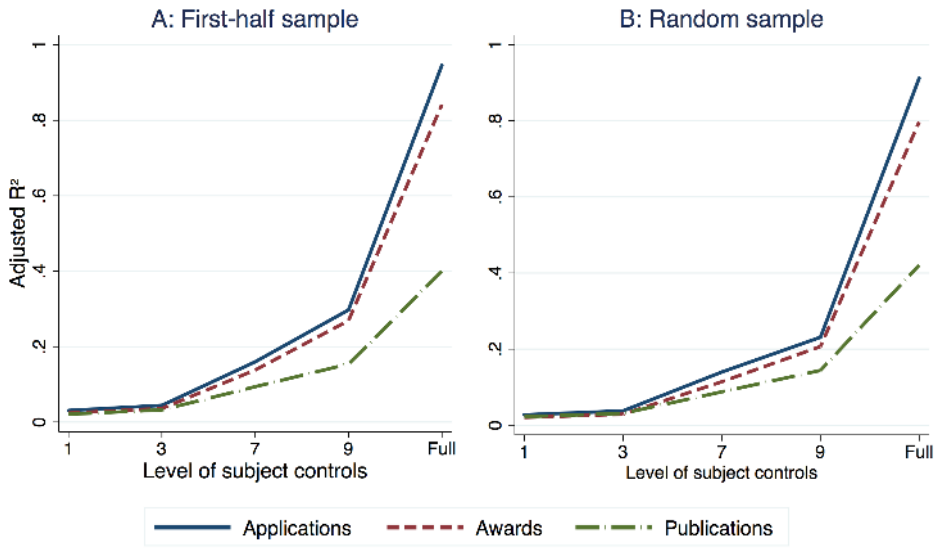
Note: Standard deviations in parentheses. Total $M FY is the total value of all grants awarded in the first fiscal year of the RFA in millions. Announced statistics are from texts of the RFA announcements, actual statistics are based on the final awards. Duration is the number of fiscal years a grant is planned for at time of awarding. Grant types are defined in text.

Table 1 summarizes the explicit parameters of each RFA (e.g. award number and size) as well as the realized number of applications and awards. In general, these contests tend to receive two to five times as many applications as awards, and award fewer grants than announced. There is significant variation across the types of grants in terms of responsiveness. For example, the most common basic science grant (R01) receives nearly five applications for every award announced. Conversely, the technology transfer grant (R41) RFAs receives roughly the same number of applications as awards announced on average. This is initial evidence that even those applications that are received are quite often of lower quality. These technology transfer grants are not solicited in RFAs very frequently (N=39); however, these low response and award rates are indicative of significant difficulties in inducing research that bridges academic and commercial ventures. Still, the raw statistics cannot separately identify whether or not this is because the NIH typically targets scientific subjects that are inherently difficult to translate, or the act of inducing this translation is inherently difficult.

Constructing variables to perfectly capture the nature of each RFA would be a difficult task given the number of parameters associated with each contest (e.g. number of expected awards, total award purse, limitations on grant size, limitations on grant length, total scope of the RFA, sponsoring Institute, etc.). To simplify the analyses I instead create a single indicator variable for whether or not a subject is requested in an RFA. There is not a tremendous amount of variation in the number or size of awards for RFAs using the same grant type, so this simplification does not remove an excessive amount of information and facilitates a simpler interpretation in the empirical analyses: what is the average change in the rate of science for a subject solicited by the average RFA?

To investigate variation within and across these rates, Figure 4 plots the adjusted R^2 from OLS regressions of the rate of subject occurrences on increasingly detailed levels of fixed effects for each subject using each of the three main outcomes in this dataset. By levels, I refer to the subject groupings superior to each subject. For example, in Figure 4 the level 1 control for "Diabetes mellitus, type 1" is "Diseases", the level 2 control is "Endocrine system

Figure 4: Out-of-Sample Variation Explained by Subject Fixed Effects



Note: The rate of science is the number of times a subject occurs in the text of abstracts from each source (all NIH applications, funded NIH applications and publications connected to funded NIH applications). The levels of the fixed effects are based on the MeSH structure with higher levels indicating more specific fixed effects and are described in detail in the text. The adjusted R^2 is from OLS regressions of the rate of science in each application round on different levels of the subject fixed effects estimated using two different hold-out samples. In Panel A, the fixed effects are estimated in the first half of the sample (2006 to mid-2010) and used as explanatory variables in the second half (mid-2010 to 2014). In Panel B, the fixed effects are estimated on a randomly selected half of the sample, and used as explanatory variables on the remainder of the sample.

diseases" and so on. In Panel A, I first estimate the fixed effects using the first half of the data and regress these estimates on the rates for the second half of the data. Panel B replicates this exercise using a random sample for the first estimation. In both cases, the full set of time-invariant fixed effects explain the vast majority (>80%) of variation in the rate of science that occurs in applications and nearly half of the variation in publication outcomes. Given that this data spans eight years, it is evidence that significant shocks to the opportunities within scientific subjects do not occur frequently. Conversely, it is also implies that any of the NIH's efforts to influence these rates, at least within applications, have not had any dramatic effects.

This notion that technological trajectories are relatively stable and incremental in nature

(Dosi 1982) until infrequent punctuated equilibrium events spur major changes (Levinthal 1998) has been appreciated for some time, and these patterns tell a similar story. Furthermore, the large discrepancy in explanatory power between the simpler and most refined categorizations indicates that controlling for these fixed differences is empirically relevant.

2.3. Theoretical Motivations

This section outlines a simple supply-demand model that describes the effects to be estimated, emphasizes how frictions could prevent price shocks from inducing quantity changes, and highlights identification concerns relevant to the empirical analyses. A selection model is also referenced in order to predict heterogeneity in supply curves across science and scientists.

2.3.1. Supply and Demand of Science

To fix ideas and clarify the main effects estimated in this Chapter, consider an economy with a single scientific subject that may be pursued by a fixed number of scientists who generate new knowledge on the subject, one public consumer $i = 0$ and many private consumers $i = (1, 2, ..., I)$ of knowledge. Given the large number of consumers and scientists, both will be treated as price-takers[19].

There are two, connected markets that facilitate (1) the funding of ideas, and then (2) the implementation of ideas to generate new knowledge. This two-part production occurs since the consumers cannot themselves directly identify production possibilities and the scientists face large input constraints, but consolidation is not feasible given information asymmetries, preference differences and the contractual frictions they create (Aghion et al. 2008; Lacetera 2009).

In the first market for ideas, demand is a function of the price of ideas p^1, exogenous features of the knowledge these ideas may generate ϵ^D (e.g. disease burden if the subject is

[19]Despite the fact that the NIH is the single-largest funder of biomedical research, it still only accounts for roughly 20% of basic research in the U.S.

an illness), and the budgets of each consumer $x_i = X_i(\cdot)$. Importantly, the ultimate good produced - information - is, essentially, a public good. Therefore, purchasing an idea does not exclude other consumers from utilizing the knowledge it may generate. Thus, given the incentives not to duplicate investments, each consumer can strategically set their budget as a best response to the vector of budget decisions made by other consumers, denoted by $\mathbf{x}_{-i}$, and their own private demand shocks. This introduces the crowd-out (or crowd-in) problem common to most public subsidies[20]. Ideas are supplied based on their price, as well as exogenous technologies ϵ^S (e.g. random discoveries)[21]. This gives supply and demand in the market for ideas at time t,

$$q^{D1} = D^1\Big(p^1\,,\,\sum_i X_i\big(\mathbf{x}_{-i}\,,\,\epsilon^D_{it}\big)\,,\,\epsilon^D_t\Big) \tag{2.1.1}$$

$$q^{S1} = S^1\Big(p^1\,,\,\epsilon^S_t\Big) \tag{2.1.2}$$

In the final market for knowledge, demand is a function of the price of knowledge p^2 and the exogenous demand-side features of the subject ϵ^D. The supply of knowledge depends on price, exogenous technologies ϵ^S and the amount of inputs committed to production $y = Y(p^{1*}\,,\,q^{1*})$ per equilibrium outcomes in the first market. That y is simply not $p^{1*} \times q^{1*}$ reflects the contractual issues highlighted by Aghion et al. (2008) - funders will relinquish some control over inputs in order to avoid hold-up in production.

The specific amount of inputs generated by each consumer y_i can be viewed as consumer-level equilibrium outcomes in the initial idea market $(p_i^{1*}\,,\,q_i^{1*})$, which equal either the equilibrium values $(p^{1*}\,,\,q^{1*})$ if the consumer funded an idea, or $(0\,,\,0)$ otherwise. This gives supply and

[20] These resource choices are analogous to the income of consumers in a traditional market, except the consumers here can choose how much of their "income" to allocate to science, given the value of the outside option. This implies that ϵ^D is scaled by the value of investing resources elsewhere - not science.

[21] Certainly, most discoveries are not random, but rather the endogenous outcomes of previous investments in ideas. However, the goal of this model is to emphasize the static effects of consumer behaviors since data limitations prevent empirically exploring long-term outcomes.

demand in the market for knowledge at time t,

$$q^{D2} = D^2\left(p^2\,,\,\epsilon_t^D\right) \tag{2.2.1}$$

$$q^{S2} = S^2\left(p^2\,,\,\sum_i Y_i\left(p_i^{1*}\,,\,q_i^{1*}\right)\,,\,\epsilon_t^S\right) \tag{2.2.2}$$

and the equilibrium knowledge production, q^{2*}, occurring when supply and demand in both markets are equivalent[22].

When the public manager chooses to increase the resources available for scientists ($\partial x_0 > 0$), the implicit assumptions are that (1) after accounting for other consumers' strategic responses, the aggregate demand for ideas will increase (i.e. no full crowd-out); (2) the equilibrium price and/or quantity of ideas will increase (i.e. supply of ideas is elastic); (3) the additional inputs will be committed to successful production (i.e. supply of knowledge is elastic). Figure 5 demonstrates the two markets and this potential cascade of effects.

In the context of this model and given data limitations, the three main treatment effects I estimate empirically are:

1. $\partial q_0^{1*}/\partial x_0$: change in public ideas for the subject
2. $\partial p_0^{1*}/\partial x_0$: change in public funds allocated to the subject
3. $\partial q^{2*}/\partial x_0$: change in total knowledge on the subject

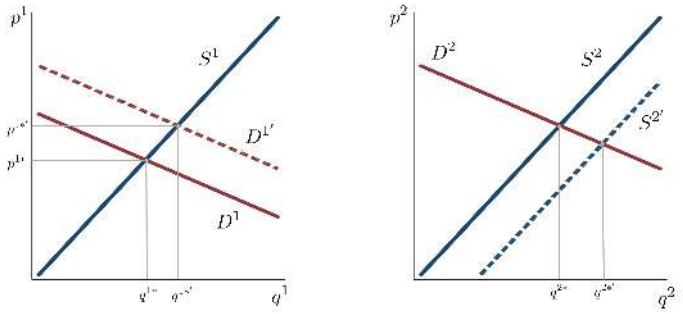
Since private investments are rarely disclosed, changes at the first two stages are only observable for the NIH, and thus only shifts in this subset of the idea supply curve will be identified. But given the dominant role of the NIH in the funding of biomedical science, these effects are still of policy interest. And since my final output measure - peer-reviewed

[22]Standard assumptions about functions D, S, X and Y can be made to ensure a unique equilibrium exists. A detailed characterization of this equilibrium is beyond the scope of this Chapter, since the objective is to simply test if the public manager can change the equilibrium quantity of knowledge supplied. That at least a stable equilibrium *should* exist is supported by the fact that the consumer shares (e.g. across public and private) of investments in biomedical research have remained relatively constant over the past decades (Moses et al. 2015).

Figure 5: Two-Stage Markets of Science

Panel A: Market for Ideas

Panel B: Market for Knowledge



Note: In the market for ideas, consumers fund production possibilities for the market for knowledge. Equations (2.1 - 2.2) detail the supply and demand functions. The treatment evaluated in this Chapter is the potential demand shift in the first market ($D_1 \rightarrow D_1^{'}$) that may be induced by the NIH's targeted allocations, given other consumers' responses. The ultimate treatment effect of interest is the quantity change in the second market ($q_2^* \rightarrow q_2^{*'}$).

publications - are observable regardless of funding source, I can still estimate the ultimate effect on total quantity. Thus, in terms of the canonical supply-demand model, this third effect can be interpreted as the income elasticity of supply.

To recall Arrow and Nelson's point, solving the underinvestment problem necessitates that for each specific scientific subject j, $\frac{\partial q_{jt}^{2*}}{\partial x_{0jt}} > 0$. Allowing for multiple subjects in the model introduces the possibility of substitution or complementarities at any stage of production. Thus, when estimating the three effects it will be important to separate the extent to which changes are on the extensive margin - no science - or intensive margin - different subjects.

In terms of generalizability, my estimates of these elasticities will be linear approximations local to the magnitudes of ∂x_{0jt} I observe at the NIH. Although drawn as linear functions in Figure 5, there is no reason to assume this is the case for the full distribution of prices and quantities. Still, the major goal of this Chapter is to test the null hypotheses that $\frac{\partial q_{jt}^{2*}}{\partial x_{0jt}} = 0$ given the lack of causal evidence to the contrary.

2.3.2. Frictions of Science

That a price shock would influence quantities in a neoclassical economy is certainly trivial. However, given the particulars of this market, it far from clear that price shocks can be generated by the public manager that are salient and sizable enough to induce a quantity response. The main complications of public science and scientists are: (1) that the public consumer's allocations occur amongst other (private) strategic consumers; (2) the indirect connection between allocation and production that is moderated by competitions (e.g. grant contests); (3) scientists suffer from similar appropriability limitations as private firms and (4) the supply of scientists is highly constrained both in terms of total number and their ability to pursue new subjects.

2.3.3. Treatment Magnitude & Heterogeneity

Treatment Magnitude

My empirical specification will formalize treatment - ∂x_{0jt} - as subject j being solicited in a targeted grant contest. Per the contest selection model of Appendix 2.3.4, the magnitude of this treatment is based on differences in the size and scope across the targeted and open contests. The approximate differences are as follows[23]: the average open contests award 85 new grants, $30 million and utilize 50 different scientific review sections (a proxy for the scientific scope of competition); the average RFA awards 7 new grants, $3 million and utilizes a single review section. Loosely speaking, RFAs are 10% the size and 2% the scope of the standard contests utilized at the NIH.

Predicted Heterogeneity

As highlighted in Appendix 2.3.4, the types of science and scientists that should be most responsive to the targeted contests are those where the correlation in outcomes are the

[23]Generating precise estimates of these parameters is challenging given the complexity of the NIH's policies. The assumptions used to arrive at these values are in line with the "rules of thumb" used by NIH applicants based on discussion with NIH staff.

highest - the costs of redirection are the lowest. How might this play out across the hierarchy of scientific subjects? Intuitively, inducing redirections across broader subjects at the top of the hierarchy (e.g. Diseases or Organisms) would necessitate larger levels of redirection compared to the changes necessary to transition from one specific subject to another (e.g. Diabetes mellitus, type 1 or Diabetes mellitus, type 2). Should we expect any variation in redirection costs as a function of the "age" of a subject? In other words, is it more or less difficult to redirect scientists towards new areas of science? Intuitively, newly discovered subjects will have a smaller base of information from which scientists could source ideas from. However, following the arguments of Jones (2009), pushing the frontiers of knowledge forward might also be easier in the case of new subjects that lack a great depth of existing knowledge. Which, if either, of these countervailing forces dominate in this setting is an empirical question.

What types of scientists should have the lowest costs of redirection? It has been shown that the incentives for scientific effort in general decline over time (Levin and Stephan 1991). Furthermore, to the extent that one's prior research endeavors shape and constrain the search and evaluation of new ideas (Nelson and Winter 1982; Levinthal 1997; Gavetti and Levinthal 2000), more experienced scientists would face greater costs of redirecting their work. Thus, we would expect less experienced researchers to be most responsive to the targeted contests. Data limitations prevent me from evaluating specific dimensions of experience (e.g. age, years active), but one observable variable is the NIH's "New Investigator" designation - assigned to individuals who have not yet successfully competed for an NIH research project grant. While this policy selects on both dimensions of experience, age as well as ability to compete at the NIH, it provides a clear and policy-relevant way of delineating applications from those who I expect to be less constrained in their redirections.

2.3.4. Selection Model of Contests

This section reframes the targeted-open contest entry decision in the format of the Roy-Borjas (1951; 1987) models of self-selection. I consider a population of specialized individuals

who may choose to compete for prizes in a contests offered by the manager, or receive their outside option $\overline{u}_i$. In these contests individuals submit a single new idea of a certain type t. Two contests k exist, an "open" contest $k = 0$ that permits ideas of any type $t = (1, 2, ..., T_0)$ and a "targeted" contest $k = 1$ that permits a subset of types, such that $T_1 \leq T_0$. Thus, the "scope" of a contest is given by T_k. The manager's objective in creating the targeted contest is to increase the rate at which individuals pursue types the types made eligible in this contest. This is analogous to the investigator-initiated (open) and RFA (targeted) contests of the NIH.

I assume the expected value of the contest is given by the contest average μ_k (e.g. prize structure), a type average within either contest ν_{kt} (e.g. the relative ease with which a subject is pursued) and an individual-type performance component ϵ_{ikt} that captures all individual-level variation from the contest and type means (e.g. individual's abilities or preferences) such that the expected payoff of each contest is summarized by:

$$U_{ikt} = \mu_k + \nu_{kt} + \epsilon_{ikt} - \overline{u}_i \tag{2.3}$$

where $\epsilon_{ikt} \sim N(0, \sigma_k^2)$. Individuals entering the targeted contest must incur fixed costs, C, which may be negative. Let t^*_{ik} be the type with the highest payoff for each individual-contest pair. Following Borjas, C can be adjusted to include the opportunity costs by scaling C by the value of the default contest: $\pi_i = \frac{C}{U_{i0t^*}}$. Individuals know their ϵ_{ikt}, π_i and $\overline{u}_i$. Each individual is extremely small relative to the full set of potential entrants[24], so while individuals may form strategic expectations, I assume any resulting general equilibrium effects are negligible. Individuals will enter the targeted contests when

$$(\mu_1 + \nu_{1t^*} - \mu_0 - \nu_{0t^*} - \pi_i) + (\epsilon_{i1t^*} - \epsilon_{i0t^*}) > \overline{u}_i \tag{2.4}$$

It is straightforward to show that this model predicts increasing entry as the relative value of

[24]The number of unique applicants to the NIH during my eight-year time period is roughly 80,000 scientists. Using a narrower definition of potential entrants, each of the 24 Institutes receive applications from roughly 1,700 unique scientists each year.

the new contest grows ($\mu_1+\nu_{1t^*}-\mu_0-\nu_{0t^*}$), or fixed costs C decline, importantly, irrespective of the average correlation in individual's performance across the contests, $\rho = \frac{\text{cov}(\sigma_1,\sigma_0)}{\sigma_1\sigma_0}$. My empirical analysis in Section 3.4 treats the prize structure of the targeted contests as a constant - treatment is binary - so I do not empirically explore variation in this regard. However, this model also implies that the targeted contests should see larger entry when the set of subjects targeted have the lowest relative payoffs in the open contest. In this case, because competition is relative across types per ν_{kt}, the rate of entry into the targeted contest will be increasing in $\frac{\sum_{t_1}\nu_{1t}/T_1}{\sum_{t_0}\nu_{0t}/T_0}$. Essentially, if competition is restricted to types that, on average, fair worse in competitions amongst the full set of types, the relative value of the restricted competition will be larger. Here, the narrower scope essentially "levels" competition, with the value of this leveling being greatest for those types that have the smallest ν_{0t}. In Section 3.4, I proxy for ν_{0t} with the success rate of each type prior to the announcement of the targeted competition. In line with this prediction, the RFAs have the largest effects for scientific subjects with prior success rates in the lower quartile. The analyses in Section 3.5 sheds light on the sign and magnitude of C. Interestingly, the results suggest that, conditional on differences in μ_k, the targeted contests may in fact be preferred; or in the context of this model, that is to say, C is negative.

To consider how individuals might sort between the two contests, we need to consider differences in the distributions of random variables per σ_k and their correlation in the population, ρ. How might the difference in scope influence the distributions or correlations of individuals' returns, or in other words, what is the partial derivatives $\frac{\partial\sigma_1}{\partial T_1}$ and $\frac{\partial\rho}{\partial T_1}$?

This will depend on the extent to which individuals are constrained in pursuing ideas of different types. For example, assuming that individuals can costlessly pursue an idea of any type, then restricting competition to any sized subset of types will not influence ρ for any particular individuals or types. Specifically, both of the partial derivatives will be zero, and the effect of the targeted contest will only depend on $\frac{\mu_1+\nu_{1t^*}}{\mu_0-\nu_{0t^*}}$ and C.

However, if individuals are constrained in their ability or willingness to pursue different

types of ideas, then it is implied that $\frac{\partial \rho}{\partial T_1} > 1$, where smaller scopes lead to a reduction in the population average ρ. Furthermore, if this reduction in ρ occurs differentially across the population, as it would if ρ_i was say, a function of the similarity between an individuals' past and chosen type, then $\frac{\partial \sigma_1}{\partial T_1} < 1$. Consider the limit cases; as $T_1 \nearrow T_0$ all individuals will face the same returns in either contest, conditional on μ_k, and therefore $\sigma_1 \searrow \sigma_0$. Conversely, as $T_1 \searrow 1$ (scope narrows), the individuals for whom ρ_i decreases the least will see relatively higher returns in the targeted contest and $\sigma_1 \nearrow \infty$. Therefore, the types of science and scientists that would be most responsive to a targeted contest will be the individuals or types for which ρ is reduced *the least* given the scope reduction and growth in σ_1. Whether or not this reduced scope can induce an overall increase in the rate at which targeted types are pursued will then depend on the distribution of individuals and how they sort across contests.

In this case where it is assumed that the scope reduction implies that $\sigma_1 > \sigma_0$, sorting may occur via "positive" or "inverse"[25] sorting where either the most able in the open contest (highest ϵ_{i0t^*}) or least able (lowest ϵ_{i0t^*}) participate in the targeted contest, respectively. The direction of this selection will depend on whether ρ is greater or less than minimum ratio of the returns, $m = (\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1})$. If $\rho > m$, then the average returns across contests are still sufficiently correlated following the scope reduction, and positive sorting will occur; vice versa for inverse sorting. At the NIH, the direction of this sorting will determine whether or not RFAs can induce new applications where individuals pursue new ideas (inverse sorting) or if the RFAs simply crowd-out funding that would have been acquired in the open contests (positive sorting).

Thus, to reframe the objective of the manager, the premise of targeted contests is to create inverse sorting - managers raise the incentives associated with particular lines of work in hopes of inducing new entrants that would not have otherwise pursued the targeted work. However, if when this inverse sorting occurs the individuals select from "too low" in the

[25] Borjas uses the term "refugee" to describe inverse sorting.

distribution of returns, then the quality of individuals induced by the new contest may be low enough that the manager does not value the additional work given the costs. Conversely, if positive sorting is too strong, then to the extent that the high quality individuals already exist in the targeted field, they will capture the targeted resources and the manager will not solicit any new work. This balance of structuring targeted prizes such that they induce new work without crowding out existing work closely mimics the results of low-powered incentives being optimal in many multi-task settings (Holmstrom and Milgrom 1991)[26]

In summary, the effect of providing a narrower contest amidst the presence of an "open" contest will depend on the relative sizes of the contest, and to the extent that individuals are constrained in their selection of types, the relative scopes of the contest.

2.4. Aggregate Analysis: Rate of Science

2.4.1. Motivating Theory & Empirics

The NIH's objective is to increase the rate at which scientists successfully compete for research grants related to certain scientific subjects. To do so, an RFA is created, consisting of a certain amount of grant funds to be competed for only by proposals related to those subjects. Importantly, the RFAs occur amidst the continually available investigator-initiated "open" contests at the NIH, where any subject may be pursued. Thus, the RFAs operate conditional on the features of this key outside option[27]. How might the creation of these targeted contests influence the overall rate at which targeted subjects are pursued, or successfully pursued, at the NIH?

In general, there are four scenarios that may occur: either there is no change from the counterfactual rate of science because no new applications are induced and scientists completely

[26] Aghion et al. (2008) explore this sort of tradeoff precisely in the setting of academic scientists.

[27] Certainly, the landscape of funding opportunities for biomedical scientists includes numerous other public and private entities, which may selectively fund certain subjects, e.g. the American Cancer Society. I will assume that conditional on the controls in my empirical specification, which include subject fixed effects and time dummies, any changes in the value of these outside-the-NIH options are uncorrelated with RFA generations.

substitute the RFA for the open contest, or complete substitution does not occur and (1) no new applications are induced, so the RFA funds the marginally worse proposal, per ex-ante quality, that would have been unsuccessfully submitted otherwise; (2) new applications are induced; however, these new proposals are of too low quality to be successful and the targeted allocations again fund those that would have been submitted otherwise; and (3) new applications are induced *and* the new proposals are of high enough quality that, at least some, are successful. Table 2 outlines these scenarios clearly.

Table 2: Inducement Scenarios

	Counterfactual		(1)		(2)		(3)	
q	Apply	Win	Apply	Win	Apply	Win	Apply	Win
100	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
95							**N**	**N**
90	**X**		**X**	**N**	**X**	**N**	**X**	
85	**X**		**X**		**X**		**X**	
80					**N**		**N**	
75					**N**			

Note: This Table represents the three general outcomes possible in response to a targeted contest. In the counterfactual scenario, the focal scientific subject is not treated with any inducement prizes, proposals are awarded based on their relative quality (q) with existing resources enabling the funding of the single best proposal. In columns (1) - (3), a targeted contest is held that also awards a single proposal of highest quality. In these scenarios, new proposals or successful proposals made possible by the targeted contest are indicated by **N**.

From the standpoint of what I can test in the data, distinguishing between outcome (1) and the other scenarios is a straightforward question: do application rates increase for the targeted subject? Although a simple test, it is far from obvious that such inducement would occur given the aforementioned constraints facing scientists. Distinguishing between scenarios (2) and (3) is important because it will indicate the extent to which the NIH can induce new, quality ideas. However, with my data of realized outcomes it is impossible to determine whether a specific application would have been submitted in absence of the

RFA[28]. We can get a sense of the extent to which the RFA induces new proposals that are in fact successful by comparing the size of the application effect to the size of the award effect. If the induced science is of very low quality (as in Scenario 2), then I should observe a larger change in the number of awards mentioning the targeted science relative to the change in total applications. Conversely, if the new science is significantly higher in quality compared to the counterfactual science (as in Scenario 3), then the change in total applications should be larger. To the extent the induced applications are of equal quality as the counterfactual applications, then the two effects will be equivalent.

To inform the conditions under which these different scenarios may occur, Appendix 2.3.4 reframes Borjas's (1987) model of self-selection as the choice between RFAs and the open contests. The model implies that the value of these targeted grant competitions will depend on their relative size and, to the extent that scientists are constrained in their abilities to pursue different types of science, their relative scope. Exploring the magnitude of these constraints is the focus of Section 3.5.

Empirical Specification

To recall, the dataset for this analysis is a panel of observations at the subject (s) grant type (j) 4-month application round (t) level. My outcomes of interest (y) are the rate of subject occurrences in (1) all applications, (2) funded applications (awards) and (3) publications from funded applications. In order to identify the effect of being soliciting in an RFA, I estimate the expected rate of science within each subject as:

$$\mathbb{E}[y_{sjt}|\beta_b, \gamma_s, \tau_{jt}] = \exp\Big(\sum_{b=t-3}^{t+3} \beta_b \cdot \mathbf{1}\{\text{RFA}_{sjt}\} + \gamma_s + \tau_{jt}\Big) \tag{2.5}$$

where the independent variable of interest $\mathbf{1}\{\text{RFA}_{sjt}\}$ is a binary variable equaling 1 if an RFA requests subject s with grant type j to be funded in application round t. The main treatment effect of interest, described previously, is given by $\beta_{t=0}$.

[28]This is because applications mentioning targeted subjects can still be submitted to the open contests.

I allow for dynamic effects of the RFAs in order to capture any inter-temporal spillovers or substitution effects during the announcement phase $b = (t-1)$ and in the application rounds immediately after the targeted grants are awarded $b = (t+1, t+2, t+3)$[29]. The subject fixed effects γ_s control for all time-invariant differences across each subject and the application round-grant type fixed effects τ_{jt} control for any time-varying differences across all observations in each round for each type of grant (e.g. R01).

My main assumption is that the change in the rate of science for subjects *not* treated with an RFA in a given round are valid estimates of the counterfactual change in rate of science for treated subjects had they not been targeted with an RFA, conditional on their mean differences. Letting ϵ_{sjt} be the error term in the estimation, then the major threat to this assumption is a correlation between the creation of RFAs and the time-varying subject-grant specific factors ϵ_{sjt}; this is the aforementioned concern regarding the NIH's selection of which subjects to target. As discussed in Section 3.2, RFAs are not generated at random but do appear to occur for reasons likely unrelated to the changes in the unobservable value of particular subjects. The coefficients $\beta_{b=t-2,t-3}$ provide a statistical tests for this assumption where rejecting the null would indicate the presence of pre-existing trends, evidence that my assumption may be invalid.

I use the Poisson specification, where estimated coefficients capture the *percent* changes in the outcomes, for two reasons: first, as discussed previously, the change in rate of science within subjects will depend largely on factors unique to each subject such as the breadth of science it entails; and second, more practically speaking, the Poisson model handles the high degree of positive skewness, the non-negativity of outcomes and frequent zero values (approximately 30%) without requiring the transformations necessary to estimate linear models or the simultaneous estimation of extensive and intensive margin effects. Standard errors are clustered at the subject (j) level in all specifications to account for possible serial-

[29]The application rounds are roughly four months in length and RFAs are announced six months prior to the deadline on average, which extends from the focal application round ($b = t$) to the prior round ($b = t-1$).

correlation in the error terms within subjects[30].

Table 3: Summary Statistics of Subject-Level Rate of Science Outcomes

	All Subjects	Treated Subjects	Diff. (Treat.-All)
Applications	22.79	47.20	24.41***
	(55.84)	(72.31)	
Awards	3.28	6.77	3.48***
	(8.310)	(10.83)	
Publications	5.29	11.05	5.76***
	(20.81)	(31.04)	
Ever RFA	17.0%	100%	-
Obs.	592,800	100,850	

Note: Rows (1) - (3) indicate the count of abstracts referencing each of the scientific subjects. Mean values at the subject-application round level, standard deviations in parentheses. Awards are successful applications, publications are all subsequent publications in the top 500 biomedical journals (per Journal Impact Factor) generated by awards in a given application round. *** p<0.01.

Table 3 presents the mean values of the three main outcomes (applications, awards, and subsequent publications) at the subject-level. Notably, the rate of subject occurrence in all outcomes is roughly twice as large for those subjects treated by RFAs. However, using the non-RFA treated subjects as the counterfactual control groups in my estimation will be valid so long as these differences do not persist conditional on the the subject-level and time-trend fixed effects.

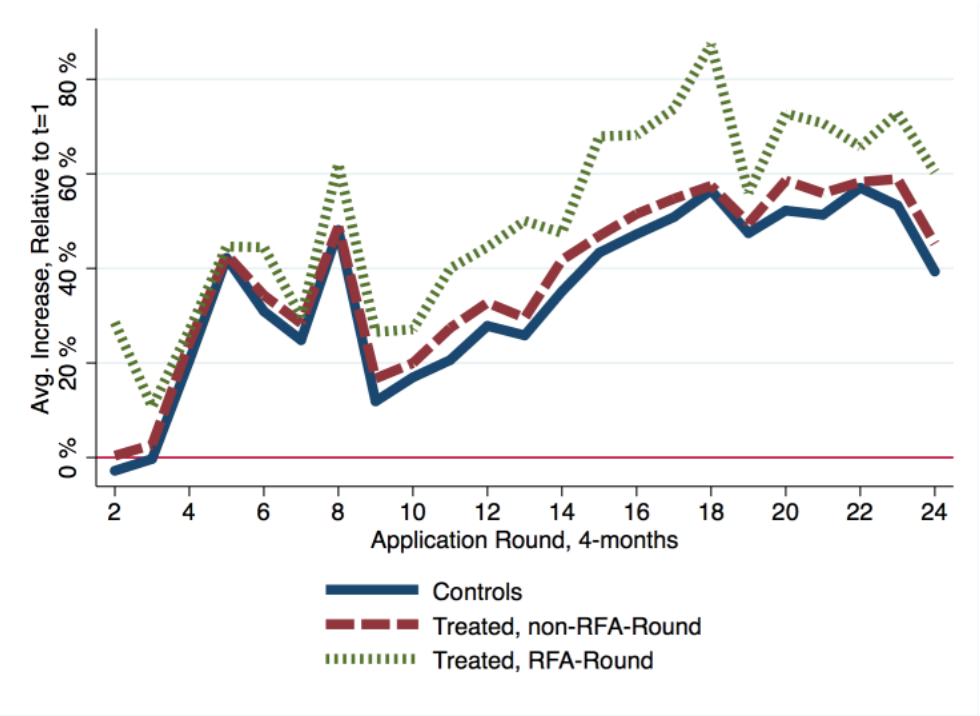
2.4.2. Results: Main Treatment Effects

To make the difference-in-difference results clear, it is useful to first observe the raw trends in the rate of science for the treatment and control subjects in the sample. A straightforward pre- post- plot is not possible given the staggered treatment occurrence where a subset of treated subjects are requested in RFAs each application round. Figure 6 accommodates

[30]Thus, this level of clustering spans multiple panels (each grant type) and also makes the estimation of the standard errors robust to violation of the equidispersion assumption underlying Poisson regressions. To be clear, this assumption is relevant only for consistent estimates of the standard errors and not the coefficients.

this by plotting the average percent growth in the rate of science for (1) control subjects, (2) treated subjects in non-RFA-treated rounds, and (3) treated subjects in RFA-treated rounds[31]. Clearly, when not requested in an RFA the treated subjects behave very similarly to the control subjects. This implies that both pre-treatment trends are likely similar across the two groups, and interestingly, that post-treatment (the rounds following an RFA) spillovers are likely not significant[32].

Figure 6: Raw Difference-in-Difference of Rates of Science



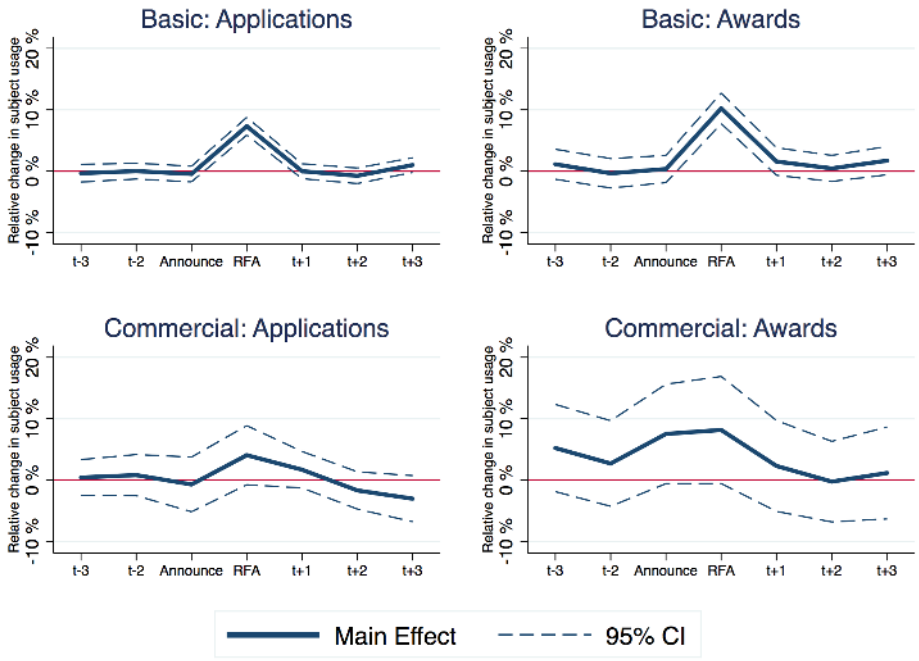
Note: Rate of Science is the count of applications mentioning each subject in each 4-month application round. Percent growth is relative to the first round in the data.

Figure 7 plots the β_b coefficients estimated in Equation 2.5 for the basic and commercial science grants using the two application outcomes, total and successful (awards). Given the political nature of the NIH allocation system discussed in Section 3.2, these main results are based on subjects not found to have political capital per my proxy. Below, I test for the

[31] Displaying average percent growth in this manner effectively incorporates the subject-level fixed effects into this plot.

[32] Separating the treated subject non-RFA-round trend into pre- and post-RFA subsets reveals no significant difference in the levels.

Figure 7: RFA Treatment Effects, Application Outcomes

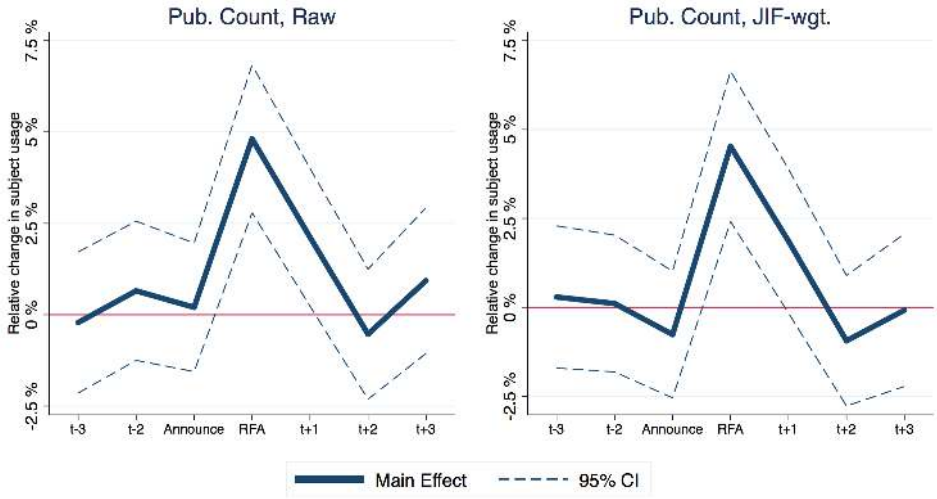


Note: β_b coefficients from estimating Equation 2.5 with full set of fixed effects; zero indicating the scientific subject's mean rate of occurrence in applications submitted each round, conditional on time trends across all subjects. Awards are successful applications. Application rounds are approximately four months. Grant types outlined in text.

presence of any selection bias associated with this political capital. The $\beta_{b=t-2,t-3}$ estimates indicate that prior to the announcement of the RFA, the rate of science for each subject was not significantly different from zero, conditional on their mean rates (subject fixed effects) and temporal trends (time fixed effects), providing support for the main assumption of this specification. The magnitudes of these coefficients indicate that being requested in an RFA induces a one-time increase in the rate of targeted subjects in applications and awards by roughly 8-10%, but only for the basic science grants. Examining the commercial science grants, Figure 7 identifies no significant increase in the rate at which targeted subjects are either pursued in all applications, or successful awards.

Given the evidence for an effect amongst the basic science grants specifically, Figure 8 presents the results for the publication outcomes associated with these grants. The coef-

Figure 8: RFA Treatment Effects, Subsequent Publications from Basic Science Awards



Note: Results from estimating Equation 2.5 with full set of fixed effects only using the basic science grants (R01, R21); zero indicating the scientific subject's mean rate of occurrence in publications generated by applications funded each round, conditional on time trends across all subjects. 95% confidence intervals for each outcome in dashed lines and respective colors. Application rounds are approximately four months. "Raw" measures publications as the count of publications in top-500 biomedical journals per Journal Impact Factor (JIF) mentioning the targeted subject. "JIF-wgt." publications are multiplied by the journal's JIF. Grant types described in text.

ficients indicate that the number of subsequent publications mentioning the targeted subjects increases by approximately 5%, using both the unadjusted and Journal-Impact-Factor weighted publication counts. Interestingly, there is also an increase in publication outcomes for grants awarded the round after the RFA, roughly 2%. This is somewhat counterintuitive when combined with the fact that no new grants appear to be awarded during this period (See Figure 7). How might these "spillovers" arise? Discussions with NIH staff suggest that this pattern is in line with scientists holding off from assigning publications to a prior grant in expectation of success in the RFA, failing to win a grant in the RFA, and then assigning those publications to a grant obtained in the following round. Examining this substitution of publications across grants is beyond the scope of this Chapter, and the magnitude does not appear to be a first-order concern for my subsequent productivity analyses.

Of note, publication outcomes based on the raw count of articles in any top-500 biomedical

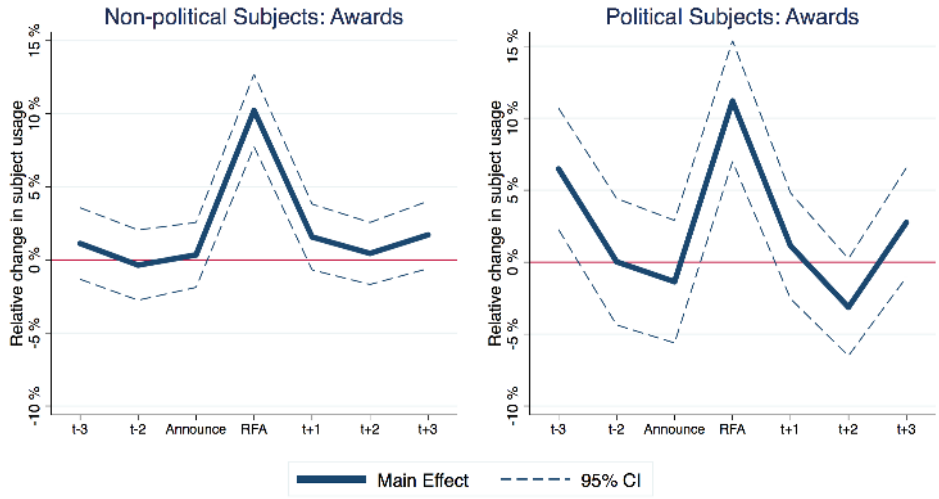
journal, versus quality-adjusted counts both reveal a very similar treatment effect. Had the new science been of much lower ex-post quality, I should have identified larger effects for the raw count measure relative to the quality-adjusted measure. Section 2.5 explores this measure of productivity - publications per applications - in more detail.

Does Political Capital Bias the Effects?

The previous results support the main assumption that these contests do not occur during significant, but unobservable, shocks to their value that would otherwise bias my estimates. These results were based on subjects that had not been specifically referenced by Congress in appropriation bills, my proxy for political capital and an indicator of subjects being "at risk" of experiencing an unobservable shock from my perspective. By comparing the pattern of effects for scientific subjects with and without political capital, I can test the extent to which this proxy may capture the outcomes of such selection, should it occur. Figure 9 plots the main effects, estimated only for awards through the basic science subset of grants, for treated subjects with and without political capital. In both cases, I still include the full set of untreated subjects as the control groups.

The treatment effects for subjects with political capital reveals two important findings. First, there is a significant pre-existing trend for treated subjects at time $t = -3$ where the rate of science is roughly 6% above average. Second, despite this pre-existing trends, the magnitudes of the main treatment effects at $t = 0$ are statistically equivalent. This result suggests that *at least within the set of treated subjects*, the bias of selecting on political capital shocks (per my proxy) is not significant. Examining the mean effects over time, the pattern for political subjects indicates there may be substitution away from future application rounds to the RFA - $\beta_{t+2} = 3.1$ (N.S.). While not statistically significant, this "hurrying" effect is visually apparent when examining the dynamic treatment effect, suggesting that scientists may rush their applications to the RFA, and the density of applications is such that these gaps in are not replaced.

Figure 9: RFA Treatment Effects, Role of Political Capital



Note: β_b coefficients from estimating Equation 2.5 with full set of fixed effects for basic science grants; zero indicating the scientific subject's mean rate of occurrence in successful applications each round, conditional on time trends across all subjects. Application rounds are approximately four months. "Political Subjects" are identified by their occurrence in Congressional appropriation bills for the NIH during 2008-2013.

Heterogeneity Across Science and Scientists

Table 4 explores predictions related to the types of science and scientists that might be most responsive to RFAs. First, in columns (1) and (2) I calculate the rate of science for "New" and experienced investigators separately. Although it was suggested that the inexperienced scientists would be the most responsive to RFAs, the raw data indicated a much lower level of participation relative to the investigator-initiated contests. As the two columns indicate, these scientists do appear to be more responsive to RFAs with an estimated effect for both outcomes (applications and awards) roughly twice the size of the estimates observed for experienced investigators.

Columns (3) and (4) of Table 4 test the prediction that for subjects which it is inherently harder to get funded in the traditional open contests - as indicated by a lower success rate prior to the RFA - we should see larger effects. In line with this prediction, column (3) indicates that subjects with success rates below the 25^{th} percentile are significantly more

Table 4: Heterogenous Effects Across Scientists and Subjects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Apps.							
RFA	0.145***	0.074***	0.077***	0.079***	0.081***	0.061**	0.109***
	(0.0198)	(0.0076)	(0.0061)	(0.0066)	(0.0066)	(0.0289)	(0.0112)
RFA × Win% $\leq 25^{th}p.$			0.265***				
			(0.0367)				
RFA × Win% $\geq 75^{th}p.$				0.022			
				(0.0152)			
RFA × $\text{MeSH}_{Post'99}$					0.008		
					(0.0162)		
Panel B: Awards							
RFA	0.286***	0.161***	0.110***	0.120***	0.106***	0.107**	0.139***
	(0.0291)	(0.0162)	(0.0109)	(0.0122)	(0.0119)	(0.0498)	(0.0196)
RFA × Win% $\leq 25^{th}p.$			0.409***				
			(0.0641)				
RFA × Win% $\geq 75^{th}p.$				-0.004			
				(0.0264)			
RFA × $\text{MeSH}_{Post'99}$					0.059**		
					(0.0276)		
Sample	New Invest.	Exper. Invest.	Full	Full	Full	Broad Subjects	Specific Subjects
Obs.	723,125	1,012,225	1,060,600	1,060,600	1,060,600	27,322	481,194

Note: Standard errors in parentheses, clustered at the subject level. * p<0.10, ** p<0.05, *** p<0.01. New Investigators are scientists yet to have received a research project grant award from the NIH, all scientists are otherwise Experienced. All dependent variables are binary indicators. Win% $\leq 25^{th}p. = 1$ indicates that in the time prior to the RFA the subject's mean rate of success in the open contests was below the 25^{th} percentile, and likewise for Win% $\geq 75^{th}p.$ $\text{MeSH}_{Post'99}$ is an indicator variable set to 1 if the subject was added to the MeSH hierarchy after 1999. Broad and Specific subjects are made clear in Figure 3.

responsive to RFAs. In fact, the effect for this subset is roughly 3-4 times larger than the mean effect estimated for the full distribution of treated subjects. This indicates the value of restricted competitions for "difficult-to-fund" subjects, where scientists pursuing these subjects do not have to compete with "easy-to-fund" alternatives. Conversely, the mean effect for subjects in the upper quartile of success rates prior to RFA treatment are equally responsive as the average subject with respect to changes in applications and awards.

Column (5) of Table 4 examines heterogeneity in effects for "old" and "new" science. Here, I define new science as MeSH subjects who were added to the vocabulary post-1999[33]. The motivating theory for this analyses was ambiguous. The results show no differential treatment effect across these two types of science for total applications, but there does appear to be a significantly larger increase in successful applications when the science targeted is relatively new. This indicates that while new science is equally responsive in terms of inducing new applications, these new applications are more likely to be funded, suggesting they are of greater ex-ante quality.

The presence of redirection costs also suggested that there should be smaller effects when requesting "broad" subjects given the larger magnitudes of redirection necessary for scientists to pursue new broad topics. Columns (6) and (7) of Table 4 provide estimates for broad and specific subjects, respectively. The estimated coefficients support the theoretical predictions with the specific subjects seeing larger relative increases in their rate of usage in both total and successful applications.

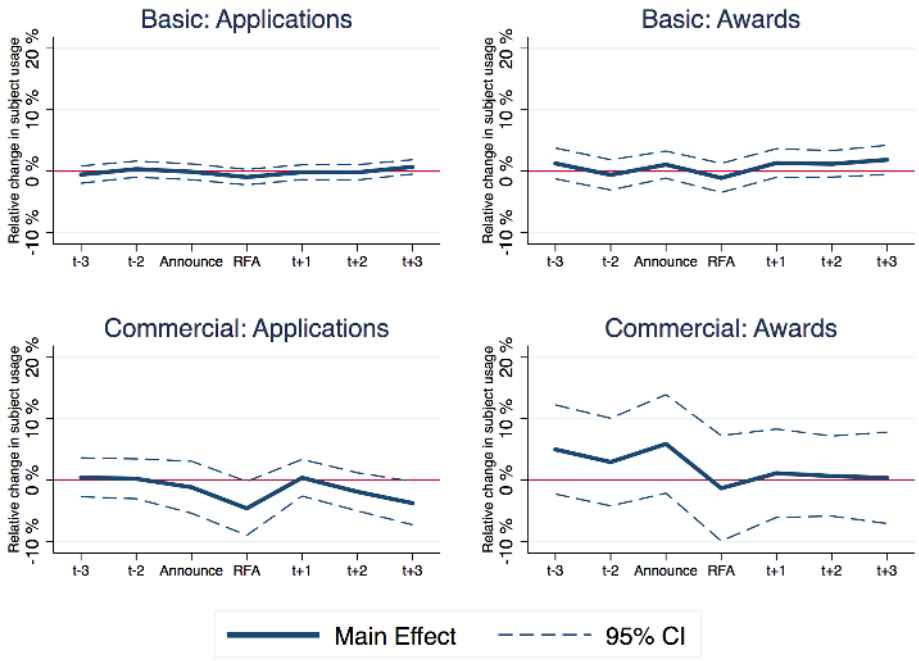
Substitutions and Spillovers

Four avenues of substitution or spillovers are possible in this setting: over the targeted (RFA) and open (investigator-initiated) contests at time of treatment; over time from future open contests; over grant types (i.e. R01, R41, etc.); and over science. Although the main effects are positive, it is possible that some of the applications to these RFAs were scientists substituting the RFA contest for the open contest. Conversely, it is possible that the shocks associated with RFAs may incentivize scientists to pursue a targeted subject in the open contests[34]. Figure 10 re-estimates Equation 2.5, but only including applications to the investigator-initiated open contests.

[33] In 1999 there was a large restructuring of the MeSH hierarchy that included significantly more (5 times) than the annual average of new term additions. This cutoff prevents these mechanical additions from being labeled as "new"; robustness tests where the delineating year is made more recent are qualitatively similar.

[34] Scientists may observe the RFA, anticipate a "gap" in competition it would create in the open contest and pursue this opportunity. Or, information shocks associated with the targeted subjects may also spur pursuits.

Figure 10: Substitution or Spillovers from Investigator-Initiated Contest



Note: Results from estimating Equation 2.5 with full set of fixed effects using applications only from the investigator-initiated "open" contests at the NIH; zero indicating the scientific subject's mean rate of occurrence in applications submitted each round, conditional on time trends across all subjects. Application rounds are approximately four months. Compare to Figure 7, which plots these estimates using all applications to the NIH (i.e. "open" and RFA contests).

For the basic science grants, there is no evidence that any of the successful applications are a result of scientists substituting away from the open contest during or after the RFAs. For the commercial science grants, there is a significant decline in applications referencing the targeted subjects during the round of the RFA, as well as three application rounds post-RFA. These declines are not statistically significant amongst successful applications; however, a decline in targeted subject occurrences during the RFA-round is qualitatively apparent in the pattern of main effects.

The two other potential types of substitution or spillovers possible in this setting are across grants and subjects. To the first, it is possible that when requesting research on a subject using a certain grant type (e.g. R01), that scientists simply substitute away from proposing

Table 5: Substitution or Spillovers Across Grant Types and Peer Subjects

	(1) Apps.	(2) Awards	(3) Apps.	(4) Awards
RFA: other grant types, same subject	-0.000262 ((0.00688)	0.0121 (0.0135)		
RFA: peer subjects			0.0415*** (0.00590)	0.0538*** (0.0108)
Unit	Subject	Subject	Peers	Peers
Obs.	1,238,000	970,300	142,550	129,425

Note: Standard errors in parentheses, clustered at the subject level. * p<0.10, ** p<0.05, *** p<0.01. For the columns (1) and (2), subject units are the same unit of analyses as all other estimations and the treatment is whether or not the subject was requested in an RFA using a different grant type than the focal panel. For example, if "Diabetes mellitus, type 1" is requested in an RFA for R01 grants, then this variable equals 1 for the panels of "Diabetes mellitus, type 1" with all other grant types. For columns (3) and (4) peer units are groupings of subjects that share the same superior subject in the MeSH hierarchy with a treated subject (See Figure 3), but are not treated themselves. The rates of the treated subject are excluded from this measure and the panels are collapsed to the level of peer groupings, hence the change in observations.

the requested subject in other grant types. Table 5 columns (1) and (2) test for this effect by estimating the rate of science in subject-grant pairs that are not requested in an RFA as a pair, but the subject is requested in another grant type. Neither coefficient is significantly different from zero, implying such substitution does not occur on a meaningful level.

Looking across subjects it is possible that when responding to these contests scientists are substituting away from similar subjects not requested in the RFA. On the other hand, it is also possible that these similar subjects might see an increase themselves because (a) practically speaking, the method through which I identify subjects as being treated may not perfectly capture the full set of subjects eligible for the RFA[35], or (b) the information shocks associated with these contests might induce scientists to consider subjects similar, but not explicitly requested. In order to test for these effects, I remove treated panels from the dataset and collapse the subjects based on their peer groupings (see Figure 3) while

[35]My methodology requires the subject explicitly appears in the text of the RFA. Whereas it may be that a range of subjects is requested, and my method only identifies the boundaries of that range.

noting the timing of the removed panel's RFA treatment. Thus, I can test for changes in the total rate of science for these subjects most proximate to the treated subject at times when the RFAs occur. Columns (3) and (4) of Table 5 indicate that the spillover forces dominate any amount of substitution at this level of subject groupings. Specifically, treating a subject with an RFA is associated with a 4-5% increase in the total rate of science for the peers of these treated subjects. It is still possible that scientists are substituting away from areas less proximate to the treated subjects; however, beyond this test of peers it is very difficult to a priori identify where these substitutions may come from.

2.5. Productivity: Science and Scientist

2.5.1. Science-Level Results

The main results indicate that RFA treatment is associated with an increase in the occurrence of targeted subjects in publications resulting from basic science (i.e. R01, R21) NIH awards. In order to more specifically explore the connection between new awards and their resulting publications - productivity - I estimate the rate of subject s occurrences in publications from grant types j at time t as the following linear equation:

$$\text{Pubs}_{sjt} = \alpha + \beta \times \text{Awards}_{sjt} + \gamma_{sj} + \tau_t + \epsilon_{sjt} \tag{2.6}$$

where γ_{sj} and τ_t are the same fixed effects from the main specification, and Awards_{sjt} is the rate of subject occurrence in successful applications. I examine both unadjusted and quality-adjusted (per Journal Impact Factor) publication measures (Pubs_{sjt}). Note that this is not investigating productivity at the level of project, but rather within projects at the level of science. Formally, β indicates the number of publications on a subject for every awarded project that is related to that subject. I first estimate the model using OLS, in order to identify the marginal productivity of endogenously submitted NIH applications. I then estimate the model using the RFA treatment as a binary instrumental variable in order to identify the marginal productivity of science induced by these targeted contests. I estimate

linear models in order to simplify interpretation, and because we are mainly concerned with the *relative difference* between the OLS and IV estimates of β, and not the magnitudes.

Table 6: Productivity of Awarded Applications: Endogenous vs. RFA-Induced

	Pub. Count, Top-500 Journal			Pub. Count, JIF-Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Awards	1.341***		2.742***	13.51***		28.31***
	(0.0370)		(0.420)	(0.359)		(4.316)
RFA		1.882***			19.43***	
		(0.300)			(3.097)	
Observations	1,483,825	1,483,825	1,483,825	1,483,825	1,483,825	1,483,825
Spec.	OLS	OLS	IV	OLS	OLS	IV
FE	Y	Y	Y	Y	Y	Y

Note: Standard errors in parentheses, clustered at the subject level. * p<0.10, ** p<0.05, *** p<0.01. Estimated using Equation 2.6 Outcomes are the number of publications mentioning a subject, restricting the sample of publications to only include those in a journal ranked in the top 500 per Journal Impact Factor (JIF), or weighting all publications by the JIF. The set of Fixed Effects (FE) included are subject by grant-type pairs, and application round. Columns (1) and (4) present endogenous mentioning publications per mentioning award; Columns (2) and (5) present the reduced form treatment effect of a subject being requested in an RFA; and Columns (3) and (6) present the 2SLS estimate using RFA treatment as an instrumental variable for the number of mentioning awards.

Table 6 presents the results of these OLS and IV estimates, along with the reduced form estimates of the RFA treatment effect on publications[36]. Relative to the marginal occurrence of subjects submitted to the default NIH contest, the IV estimates indicate that when induced via RFA, the marginal science is roughly twice as productive in terms of publication outcomes, even when adjusting for publication quality.

2.5.2. Scientist-Level Results

The magnitude of difference between the IV and OLS estimates in the previous section could be driven by two factors: (1), the new projects induced by RFAs are more productive at the project-level, that is, more publications per project, or (2) the marginal appearance of subjects in new applications submitted to RFAs are most indicative of the ultimate direction of a proposal, that is, more science per project.

[36]This amounts to a linear specification Equation 2.5.

In order to disentangle these forces, I exploit a discontinuity in win probability to identify causal estimates of project-level production in both RFA and Open projects. This approach circumvents the fact that the awarding of grants to projects at the NIH is not random in general.

I begin by supplementing my data with all publications for each scientist, using the Authority database, regardless of whether the publication cites an NIH grant. My empirical approach extends the "simulated threshold" identification strategy of Jacob and Lefgren (2011), and relies on the fact that randomness arises due to differences in quality rank (absolute merit) and funding order ranks (relative merit) in the NIH budgeting process (See Azoulay et al. (2015c) for further discussion). This is empirically useful since (1) the simulated thresholds vary over absolute quality scores, so it is less likely to be biased by any unique discontinuities in the quality distribution; (2) these thresholds do not actually exist in practice and are not public information, so they are less subject to traditional concerns for endogenous selection around the thresholds; and (3) the LATE estimated using these thresholds in an instrumental variables framework identifies the treatment effect for precisely the projects that would be affected by reallocation decisions at the NIH.

Figure 11: Win Probability Discontinuities

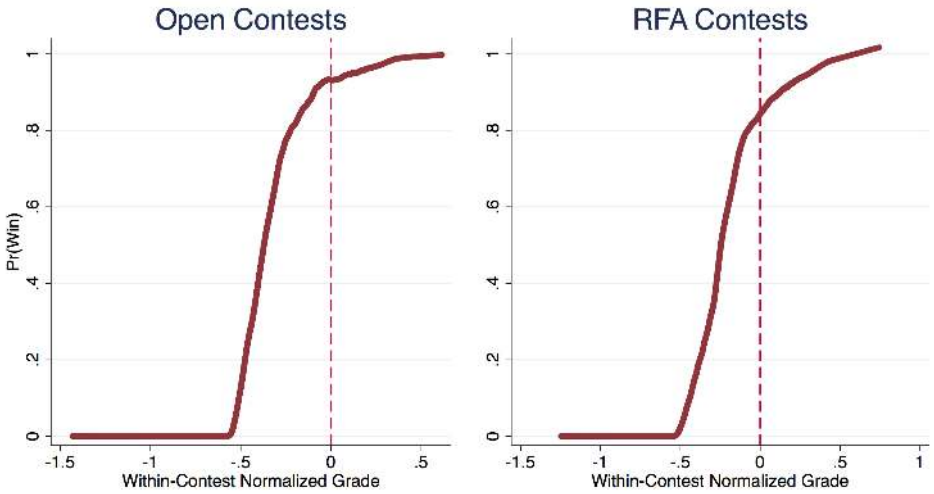
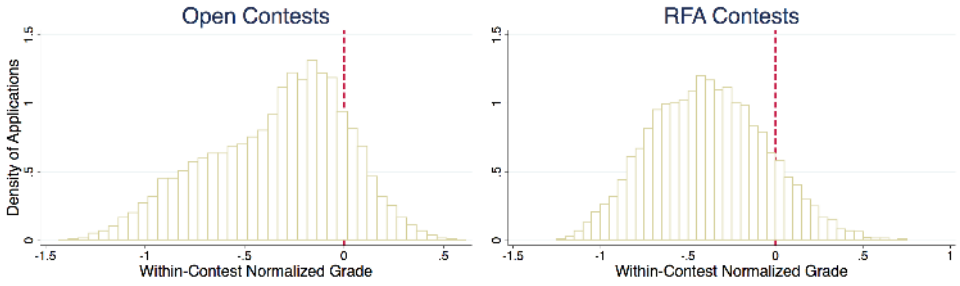


Table 7: Simulated Threshold Construction

Quality Score	Quality Rank	Win	Above Threshold
100	1	Y	Y
99	2	Y	Y
98	3	Y	Y
97	4	Y	Y
96	5		Y
95	6	Y	Y
94	7	Y	Y
93	8		Y
92	9		
91	10	Y	
90	11	Y	
89	12		
88	13		

Table 7 displays how the simulated thresholds are generated. Figure 11 demonstrates the discontinuities in win probability across the thresholds for both RFA and Open contests. Within 0.5 s.d. of the thresholds, the win difference is roughly 23 p.p. (33%) in the Open contests and 47 p.p. (110%) in the RFA contests. Figure 12 plots the density of applications for both contests and shows no evidence of "bunching" or discontinuities around the thresholds, that might have been suggestive of endogenous selection issues.

Figure 12: Distribution of Application Grades



The estimating equations for this fuzzy regression discontinuity approach are as follows: for scientist i submitting application a to contest k at time t, their publication production post-application (win or not) is

$$\begin{aligned} \text{Pubs Post}_{iakt} = &\beta_w \text{Win}_{iakt} + g(\text{Grade}_{iakt}) \\ &+ \mathbf{X}_{iakt}\beta_x + \kappa_k + \tau_t + \mu_{iakt}, \end{aligned} \tag{2.7}$$

where I utilize the simulated threshold (T) to instrument Win_{iakt} per

$$\begin{aligned}\text{Win}_{iakt} = &\gamma_1 \text{Above T}_{iakt} + \gamma_2(\text{Above T}_{iakt} \times \text{Grade}_{iakt}) \\ &+ g(\text{Grade}_{iakt}) + \mathbf{X}_{iakt}\beta_x + \kappa_k + \tau_t + \mu_{iakt}\end{aligned} \tag{2.8}$$

where, $\text{Above T}_{iakt} = 1$ if $\text{Grade}_{iakt} > \text{T}$ and γ_2 permits a discontinuous effect of Grade on Pr(Win) at the threshold. The key assumption of this approach is that conditional on the controls, an application's position, per its grade, relative to the threshold is uncorrelated with expectations about the publication outcomes of that project.

Table 8 indicates that on average, the receipt of an average-sized NIH grant leads the scientist to publish 0.3 or 0.6 additional publications in the following years for Open and RFA grants, respectively. This discrepancy between Open and RFA grants is not driven by differences in the size of the grants awarded, since these estimates combined with sample averages suggest that the marginal publication costs the NIH approximately $4.1 million when awarded through Open contests compared to only $2.1 million in the RFA contests.

Table 8: IV Results: Project-Level Production

	IV Full Sample		**IV 0.5 Bandwidth**	
	Open	RFA	Open	RFA
Win_{ikt}	0.323***	0.707***	0.312***	0.572***
	(0.061)	(0.13)	(0.11)	(0.17)
Prior Pubs_{it}	0.074***	0.096***	0.077***	0.096***
	(0.0027)	(0.0047)	(0.0032)	(0.0057)
N	25,331	10,214	18,814	6,932
k , t FE; $\mathbf{X}$; g(Grade)	Y	Y	Y	Y
Avg. $ / Pub	$3.91M	$1.72M	$4.06M	$2.13M
Avg. $ / %Pub	$126K	$82K	$135K	$105K

Note: Standard errors in parentheses, clustered at the contest level per estimating Equations 2.7 and 2.8 jointly in 2SLS. * p<0.10, ** p<0.05, *** p<0.01. Avg. $ / Pub and Avg. $ / %Pub are predicted using the coefficients on Win_{ikt} and the sample averages for amounts awarded and publication rates.

Returning to the science-level I.V. estimates of the previous section, these project-level

estimates can plausibly explain the entire gap in science-level production identified. That is to say, it appears that RFA-induced projects generate equal amounts of new science per publication, but are responsible for approximately twice as many publications (per dollar). So long as the science targeted within these contests are at least half-as valuable as untargeted science, these suggest indicate that RFAs are likely welfare-increasing.

2.6. Discussion

Given the features of a highly specialized workforce, it is not clear a priori that the grant competitions I study should induce any new research; scientists have strong preferences, face large fixed costs and are constrained in their ability to pursue topics less scientifically similar to their prior work.

This Chapter explores the responsiveness of scientists to the NIH's attempts to overcome these constraints and induce new research with targeted grant competitions - RFAs. I use data on all applications to the Institutes for eight years along with two novel text algorithms in order to create two complementary datasets on the rate of science and the redirection of scientists. Using difference-in-difference research designs I find that RFAs can induce and fund new science, but only amongst those soliciting basic science. RFAs utilizing small business research grants intended to fund applied research appear to fund science that would have been successful otherwise. A study of these small business grants at the Department of Energy concluded that such subsidies can have large positive impacts on the growth of emergent firms (Howell 2015); however, these results highlight the difficulties of determining precisely which types of firms should experience these gains.

Amongst the basic science grants, important sources of heterogeneity are apparent across different types of science and scientists. In line with a selection model, science with the lowest success rates prior to the RFA and scientists who have yet to receive an NIH research project grant are the most responsive. By restricting the scope of competition, the RFAs provide a more valuable avenue to pursue grant funds for these types. I also find that

"specific" subjects (lower in the hierarchy) are more responsive than "broad" subjects. This is consistent with the presence of constraints on scientists' ability to redirect their research trajectories, since moving from one specific subject to another requires smaller levels of adjustment.

Examining the productivity of the new basic research awards indicates that the science induced via RFAs is about twice as productive in terms of publication output per science awarded. Interestingly, this appears to be driven by a growth in the amount of *science published per science proposed*, and not the funding of projects that are any more productive in terms of *science published per dollar*. Thus, while the RFA awards are roughly 20% larger in dollar amounts, the redirection of science-level productivity without any change to project-level productivity suggests these larger investments are worthwhile from the NIH's perspective.

One phenomenon I do not investigate, that is certainly applicable to many scientists outside the NIH, is the choice of project partners. While the majority of proposals submitted to the NIH are done so by a single primary investigator (PI), roughly 14% are submitted by two or more co-primary investigators. Interestingly, the rate of co-PIs is almost twice as large within the RFAs investigated in this study. Thus, this may also be a useful setting to explore the determinants and outcomes of collaboration in scientific endeavors[37].

[37]See Catalini (2015) for a review of the literature on collaboration in science and the impact of randomly assigned neighbors - in the geographic sense.

CHAPTER 3 : The Elasticity of the Direction of Science

3.1. Introduction

A fundamental duty of the manager is to determine the allocation of resources across people and projects - the direction of work. But when the optimal direction is uncertain, information asymmetries can dictate that managers relinquish some control to their workers (Aghion and Tirole 1997; Dessein 2002). In the case of science - the production of new knowledge - this agency-information tradeoff is especially relevant (Aghion et al. 2008; Lacetera 2009), and managers often balance this tradeoff through processes that resemble contests[1]: resources are allocated to specific topics (allowing the manager to express preferences), while scientists are permitted to self-select pursuits (leveraging their private information). In these cases, resources are both inputs and *"pull" incentives* that must be valued enough to overcome the costs of changing directions given the high degree of specialization that characterize scientists (e.g. Jones 2009). This Chapter provides evidence to the cumulative magnitude of these scientific adjustment costs by asking: to what extent do scientists require more resources in order to change their research trajectories some amount? In other words, what is the elasticity of the direction of science?

The search for new directions has long been viewed as central to growth across markets (Schumpeter 1934), organizations (Nelson and Winter 1982), and individuals (Fleming and Sorenson 2001). And prior work has identified a number of forces that influence scientists' choices of direction including their preferences (Stern 2004), social interactions (Stuart and Ding 2006), incentive structures (Azoulay et al. 2011; Sauermann and Cohen 2010), and supportive institutions (Furman and Stern 2011), in addition to highlighting how cognitive constraints can impede optimal search (Levinthal 1997; Boudreau et al. 2016a). However,

[1] A contest (or tournament) simply defined here by entrants incurring sunk costs in exchange for a reward (e.g. Tullock 1980). Although not explicitly labeled as such, internal competitions for resources are a chief component of many organizations' allocation procedures as highlighted by Bower (1970), Williamson (1975), and Stein (1997). And to be clear, the focus here is on the use of contests as sorting mechanisms, as opposed to the effort incentives made possible via contests as compared to traditional piece-rate wages highlighted by Lazear and Rosen (1981), Wright (1983) and Moldovanu and Sela (2001), among others.

there is no causal evidence to date on the total costs of incentivizing new directions *as a function of the magnitude of the change in direction*, likely due to the empirical demands of identifying this relationship.

In order to estimate the elasticity of direction, this Chapter examines the choices of academic biomedical scientists in response to a series of targeted grant contests at the National Institutes of Health (NIH), a setting both empirically useful and policy relevant. Administrative data and a novel text algorithm allow me to observe a large number of individuals (∼100K scientists), consider a large number of new opportunities (∼700 targeted contests) with well defined payoffs (grant dollar amounts) and, importantly, measure the "scientific similarity"[2] between two abstracts (e.g. a prior publication and a new grant proposal), and thereby a change in direction. Furthermore, the NIH has increasingly become an active manager of the biomedical workforce, and much like many public and private organizations that utilize similar mechanisms to manage high-skilled workers, their optimal research policy will depend critically on how responsive - i.e. elastic - their workforce is. As the costs of redirection grow, so to does the lower bound at which managerial intervention is worthwhile.

I begin by framing this parameter as a compensating differential; essentially, what is the expected value of a grant competition such that a scientist is indifferent between it and a given level of redirection? Ideally, to identify this differential, an experiment would be conducted where individuals must choose their next project from a random menu of redirection-reward pairs. However, such a research design is practically infeasible amongst this sample, given the large costs of biomedical research. Instead, my analysis utilizes the fact that scientists seeking funding from the NIH can either (1) submit a proposal of any scientific project to the default "open" contests that are continuously available and accept proposals of any type, or (2) undertake some redirection and submit a specific proposal to one of the "tar-

[2]This concept has also been labeled "intellectual distance" (Boudreau et al. 2016a) with distance being the inverse of similarity, and "technical marginality" (Jeppesen and Lakhani 2010) where an individual's marginality refers to how similar their technical skills are to the focal area. My choice of terminology stems from the fact that the algorithm used to quantify this concept is based on estimating the similarity in scientific word usage between two pieces of text.

geted" contests. Here, instead of random menu of redirection-reward choices, scientists face a random menu of redirection-contest choices[3]. The research design is then based on the following assumption: if scientists can costlessly adjust the trajectories of their research then both types of contests, open and targeted, should see competition to the point that their expected values are equivalent. Otherwise, if the (measurable) redirections necessitated by the targeted contests are costly, then any wedge in (unconditional) payoffs between the two contests can be related to the average amount of redirection the targeted contests require[4].

I use a simple choice model to inform a difference-in-difference estimation that separates this differential from other sources of variation in the data. More specifically, I examine redirection within the two main grant types that embody the explore-exploit tradeoff of R&D in order to condition out fixed features of these targeted contests. To account for potential selection or unobservable variables biases, I control for the grant proposal review score, the key (typically unobserved) determinant of funding outcomes, and also include specifications with scientist-fixed effects. Conditional on the decision to apply to the NIH for funding at all, I find that scientists are indifferent between a 1.6 s.d. (≈$1.7M) increase in the expected value of a grant competition and a 1 s.d. decrease in the scientific similarity - pursuing new directions - needed to compete for that payoff (elasticity ≈ 0.6). Importantly, I find that this differential is driven entirely by scientists "experienced" with the NIH, evidence that older, more successful scientists are less willing to adjust the trajectories of their work.

Notably, these first estimates of the elasticity of direction condition on the decision to seek funding from the NIH, and are therefore likely an upper bound of population mean elasticity[5]. To identify estimates that are not conditioned on the apply-or-not decision, I conduct a second analysis that examines entry into the targeted contests by creating a

[3] In Section 3.2 I discuss how these targeted contests are generated by NIH staff, and in refer to earlier work on this project that does not suggest NIH management is selecting on unobservable factors that would significantly bias this approach.

[4] This is approach has been used in the industrial organization literature where one could infer differences in the fixed costs of entry between two industries if, holding all else constant, the profit margins were not equal.

[5] Still, the first estimates are policy relevant given the large number of applications received annually.

dataset of scientist-contest pairs and utilize the methods developed in Bajari et al. (2010) to account for endogenous competitive effects. I identify an elasticity of 0.25, which, when combined with the first set of estimates, suggests substantial variation in this parameter both across and within individuals over time.

Estimates of these kind are necessary for any model of input allocation where the supply of workers is fixed (at least in the short run) and control over direction is delegated to the workers to some degree[6]. And while contests like these used by the NIH have been proposed as useful "low-powered incentives" through which managers can delegate only a portion of authority (Aghion et al. 2008), understanding the magnitude of directional adjustment costs is necessary to ensure these incentives are not too low-powered.

This Chapter connects the literatures on scientific choice (discussed above) and "inducement contests", whereby managers solicit specific types of innovations with targeted rewards[7]. A growing empirical literature has found strong support for theoretical predictions about behaviors within contests (i.e. Boudreau et al. 2011; Boudreau et al. 2016b) by leveraging exogenous variation in the assignment of individuals to certain contests. This Chapter is a valuable complement to this work as it is precisely this selection issue - how do individuals choose to enter contests - that the NIH setting allows me to study.

The Chapter proceeds as follows: Section 3.2 describes the NIH setting and data; Section 3.3 contains the compensating differential approach and results; Section 3.4 contains the entry model approach and results; and Section 3.5 highlights the implications of these findings.

[6]Solving the efficient allocation of R&D funds amidst these sorts of adjustment costs is beyond the scope of this Chapter. But see, for examples in the biomedical setting, Lichtenberg (2001), while noting that adjustment costs when re-allocating funds are not considered.

[7]See Williams (2012) for a review of the state of science.

3.2. Setting and Data

3.2.1. Open and Targeted Funding Opportunities at the NIH

Broadly speaking, the NIH's chief objective is to award roughly $25 billion annually in research grants to extramural scientists based at universities, medical centers and other research institutions. While the NIH has long been the largest funder of biomedical research in the world, it has become increasingly involved in determining how these funds are allocated. The key mechanism through which the NIH attempts to steer these funds, and thus the direction of science, is Requests For Applications (RFAs). But before detailing RFAs, it is useful to understand the traditional investigator-initiated, or what I will refer to as "open", grant contests.

Essentially, the open contests comprise of a large set of channels through which a grant application is funneled for scientific review and then rank-ordered to determine funding priority[8]. Following Azoulay et al. (2015c), it is useful to think of the NIH application process as consisting of three levels of competition: Disease, Science and Time (D-S-T). Applications receive their review scores from one of roughly 200 review panels organized around scientific disciplines (Science), but then compete for funding with those scores at the Institute (e.g. National Cancer Institute) it was submitted to (Disease), with total allocations depending on Congressionally-determined Institute-level budgets each fiscal year (Time). Thus, each application to the open contests can be characterized by a unique Science-Time pair where it receives it's review score and a unique Disease-Time pair where it competes for funding with that score, against applications from other-Science same-Time channels.

Importantly, there are no explicit restrictions on the types of science that may be submitted to the open contests. In other words, the review process is almost entirely aimed at identifying the *vertical* quality of a given application, so long as it fits within the broad objectives

[8]This is an approximation of the official funding process, which are outlined here: https://goo.gl/blLuuU.

of the NIH[9].

To contrast, in an RFA, funds from one or more NIH Institutes are set aside for a single grant competition related to the predefined area of science[10]. The RFAs are typically announced four to eight months prior to the deadline for submissions and applications are evaluated by a single peer-review panel. Using the D-S-T terminology, RFAs consist of a single Disease-Science-Time contest, where entrants face 28.6 (s.d.=38.2) competitors on average.

Table 9: RFA-Open Levels Comparison

	Occurrence	**Size/Scope Competition**	**Outcomes**
Open Institute-Level	Ongoing N[a] = 24	85 new awards $110M 50 science panels	Win Rate: 15% Avg. size: $1.3M
RFA	One-time 6-8 month window N[a] ≈ 24	7 new awards $11M 1 science panel	Win Rate: 19% Avg. size: $1.6M

[a]N is the number of active, unique contests available each 4-month application round. For the open contests, this set of 24 is fixed over the sample period. For the RFAs, roughly 10 new contests are announced each application round.

Table 9 summarizes the differences in the scale and scope of the open and RFA mechanisms, proxying for the scientific "scope" of each with the number of unique scientific review panels involved and describing "size" per the total amount of research funds available for disbursement. On average, an open competition is roughly 10 times larger than RFAs in terms of the funds to be awarded. However, the narrower range of science eligible within the RFAs creates a contest with a larger scope-per-scale ratio - twice as many dollars per review panel. Conversations with multiple NIH applicants confirm that these marked differences in the features of the RFA and open contests are of first-order concern to them, suggesting that any empirical analysis should take them into account.

[9] *"The NIH's mission is to uncover new knowledge that will lead to better health for everyone. Simply described, the goal of NIH research is to acquire new knowledge to help prevent, detect, diagnose, and treat disease and disability."*

[10] The NIH also releases "Program Announcements" where certain types of science is solicited; however, these calls are not accompanied by set-aside funds made specifically available for competition and in practice vary widely in their format. I focus my analyses on RFAs because of their well-defined properties as funding mechanisms.

Given the smaller and well-defined level of competition unique to the RFAs, it is reasonable to assume that applicants to these contests may have different information as to their chances of funding, a factor I explicitly incorporate into the choice model of Section 3.3. However given the wedge in expected value between the two contest types described above, it is not clear if this information is valuable to scientists considering an application. If, for example, this information reduced the uncertainty of competing in the targeted contests, then risk averse individuals would place a greater value on them, be more likely to enter compared to open and success rates should be lower on average given the larger awards. Although Table 9 presents summary statistics, the fact that the difference in levels is the *opposite* suggests this may not be a large effect or countervailing forces lower the expected value of these targeted contests.

Section 3.3 discusses relevant details about the allocation of funds within these contests as it related to the empirical design.

3.2.2. Grant Types

I focus my empirical analyses on the two most common research grant types awarded, the R01 and the R21. These two types capture alternate ends of the exploit-explore continuum commonly investigated in studies of R&D strategy. The larger R01 grants are intended to support full-scale research projects that have preliminary results and build on the scientists' current base of knowledge (exploit). The smaller R21 grants are intended to develop new lines of research and do not require any preliminary results (explore). The average total lifetime award sizes for R01s and R21s, are $1.3 million and $375,000, respectively.

3.2.3. RFA Generation and Relevant NIH Policies

Because my empirical approaches below will not perfectly control for the entirety of underlying differences and trends in the supply and demand of science, it is important to clarify how RFAs are generated and the extent to which these motivations may influence the empirical results. One should be concerned that if, for example, RFAs were created to

target especially promising areas of biomedical research, then it would not be clear to what extent scientists are responding to the grant incentives versus these unobservable changes in the demand for the targeted science. However, in related work on this setting, I evaluate the causal effect of RFAs on science-level outcomes (e.g. what happens to the rate of research on disease A or organism B when targeted in an RFA), and reports no statistical evidence for a significant selection bias, even amongst those scientific topics most likely to be selected on unobservable features. This is largely due to the fact that the preferences of NIH staff responsible for the development of RFAs largely revolve around maintaining equality in their distribution of research funding. This strong preference in turn leads to the creation of most RFAs for "underserved" areas of science per the NIH's preferences, not potential applications. Intuitively, these areas of science are underserved because either the supply of knowledge or demand for advances in those fields, as interpreted by the scientists themselves, are too low to warrant pursuit. Thus, throughout the analyses I will maintain the assumption that the features of these contests are orthogonal to unobservable features of the science and scientists targeted.

An important policy surrounding RFAs is that scientists are permitted to resubmit unsuccessful RFA applications to the open grant competitions, unlike original submissions to the open contests. This provides yet another useful feature of the RFAs that, in addition to the narrow scope of science highlighted before, motivate an empirical design that can account for the fixed value of these particular grant contests.

One other policy relevant for my analyses is the NIH's New Investigator program[11]. In order to prevent grant funds from being concentrated among older experienced investigators where there are potentially less efficient (Levin and Stephan 1991; Freeman and Van Reenen 2009), the NIH explicitly labels individuals who have yet to successfully compete for a research grant as New Investigators. When making funding decisions (after peer review) for open applications, these proposals are typically evaluated separately in order to maintain

[11] See https://goo.gl/Ynd0W3 for the specific details of this program.

comparable success rates among New and experienced investigators. However, this policy does not hold in the targeted contest. While the theory of scientific choice described below will suggest that these younger, less experienced scientists may face lower costs when adjusting the trajectory of their work and hence be more likely to compete in the targeted contests, the lack of this feature may also prove a significant barrier. The raw data suggests this preference for New Investigators in the open contests is valuable: only 14% of RFA applications are from New Investigators, compared to 24% in the open contests.

3.2.4. NIH Data Sources

Data on all grant applications to the NIH from 2002 to 2011 were collected from the NIH's confidential administrative database. The full data contains the following: application identifier, review score; funding decision; Institute (for Disease); review section (for Science); and fiscal year (for Time). Because the NIH's protocol for review scores changed in the middle of my sample period, I construct a new "grade" for each application based on the standardized distributions of review scores before and after the protocol change. The grade variable ranges from 0 to 1 and indicates the percent of all applications with lower review scores. For applications submitted on or after 2006, the data also contains the abstract and title of the proposed research for both funded and non-funded applications.

For the compensating differential analysis, data on all RFA and open applications for R01 or R21 grants between 2006 and 2011 were compiled. This dataset contains 122,333 applications from 72,421 unique scientists to either a standing open contest or one of 727 unique RFAs hosted during this timeframe.

For the entry analysis, data was collected for 247 of the RFA announcements between 2002 and 2009 soliciting some combination of R01 and/or R21 research grants. Details on the timing, administration, research objectives and funds allocated were scraped from the NIH announcement website[12].

[12] Available at https://goo.gl/LuaBOQ

Each scientist's full publication history (including non-NIH-funded work) were constructed using the disambiguated version of the PubMed scientific article database developed by Torvik and Smalheiser (2009). To address the notorious difficulties of accurately matching publications to scientists given the lack of standardized identifiers in this data, the authors developed a maximum likelihood based agglomerative algorithm for computing clusters of articles that belong to the same inferred author. Thus, I am confident in the fidelity of my data, which is essential for the algorithm described below.

3.2.5. Scientific Similarity: pmra *Algorithm*

Fundamental to the notion of redirecting scientists is "how much" their course of work is adjusted. The task of moving a scientist from working on topic A to topic B will depend largely on the *scientific similarity* between A and B. Do they make use of the same knowledge? Do experiments use the same inputs such as chemicals or organisms? Can the scientists' laboratory equipment perform the necessary tests? Scholars have long pondered the costs and benefits of of scientists operating in a new field "outside" their prior specialty. But measurement issues have limited empirical analyses for some time (Gieryn and Hirsh 1983), and data availability has limited studying the *choice* to enter new scientific fields before outcomes are known - precisely the focus of this Chapter.

Formally, I define scientific similarity as the overlap in scientific terminology between two sets of information, here, abstracts of publications, grant applications and/or contest objectives. Intuitively, if two abstracts describing research projects both use the same scientific terms, especially if those terms are rare in general, it is likely the underlying science is more similar. Estimating the similarity between bodies of text has been a longstanding focus in the field bioinformatics where researchers have devised a number of methods to establish relationships between text pairs. The most widely used similarity estimation algorithm is the PubMed related articles (*pmra*) algorithm, developed by Lin and Wilbur (2007) at the NLM. *pmra* is currently employed by the NLM as the algorithm underlying the "Similar articles" feature of PubMed and has become a benchmark within the field of bioinformatics for measuring

similarity.

Recent work studying the movement of scientists has utilized *pmra* in order to trace their trajectories through topics over time (Azoulay et al. 2015a; Azoulay et al. 2015b; Azoulay et al. 2016). A novel feature of my implementation of the *pmra* algorithm is that I can generate similarity scores between published journal articles and user-defined text[13]. Thus, I can compute the similarity between a scientists' prior publications and (a) the research objectives of a contest, or (b) their application abstracts, whether or not the application resulted in any publications. Being able to explore the similarity in the requested or proposed science whether or not the project resulted in publications is a major advantage of this Chapter.

Because *pmra* requires publication abstracts as a part of its input, I make the assumption that all of a scientists' prior knowledge and skills are embodied within the journal articles they have published previously. Indeed, the entire purpose of publications is to disclose to the public the information each scientist has generated. Thus, the number of similarity scores generated for each scientist will equal their number of prior publications. To simplify this vector a single value for each scientist-abstract pair to I use the maximum score returned by *pmra* for each scientist at the time of application[14]. My intuition is as follows; if each scientists' set of publications defines the boundaries of their set of knowledge, then the maximum similarity score between any single publication and the target abstract captures the shortest proximity between the new science and the scientist. Certainly, the density of one's knowledge may vary within these boundaries, a new abstract may be "close" to a single publication but "far" from the majority, but but relying simply on this maximum score my measure captures only variation in similarity in information rater than depth of information. Robustness checks in which I instead use the mean and median statistics show no significant differences in the relative magnitudes of the results.

[13]The code for my implementation of *pmra* was very kindly developed by W. John Wilbur of the NLM who wrote the initial code for the algorithm.

[14]For any applications listed with co-principal investigators I use the maximum of the two scientists scores

3.3. Identification via Compensating Differentials

3.3.1. Scientific Choice Model

In order to clarify the parameters of interest, it is useful to model the decision of scientists seeking resources in a setting of contests that solicit ideas and award resources to pursue those ideas. The standard model of a contest's ex-ante payoff is the expected probability of success (receiving resources) times the value of the award (size of resources and the production made possible) minus the costs of participation (generating the idea). I begin by outlining the major forces at play for each of these three sets of parameters, and then describe how they can vary for "open" - all types of ideas are permitted - or "targeted" contests - a restricted set of ideas are permitted.

Entry Costs: Each scientist i can submit one idea of type s and quality q to a contest k. Costs of developing the idea C depend on the scientist's ability α_i and contest-specific factors ω_k. Furthermore, I assume that variable costs are incurred per the *scientific similarity*, D, between a scientists current type s_i' (i.e. specialty) and the type they pursue in the contest s_{ik}. In summary, the costs of applying to contest k are given by $C(q_{iks} \mid \alpha_i, D(s_i', s_{ik}), \omega_k)$. This D function captures two important phenomena related to these changes in direction: the transferability of resources as well as the ability to *predict* the value of these transfers. First, scientists' knowledge and resources may not be perfectly transferrable across types of science. Human capital theory has long appreciated the limitations of specialized knowledge (e.g. Becker 1962), while others have argued that this limited transferability may actually lower costs of innovation when the problems at hand are particularly uncertain or novel (Jeppesen and Lakhani 2010). Scientists will also face more tangible adjustment costs as they must purchase or modify inputs to the research process. In the biomedical sciences, these fixed costs of adjustment are substantial; individual pieces of lab equipment routinely cost in excess $100,000. Thus, repeated adjustments may be inefficient simply because these large fixed costs cannot be spread over multiple research projects. Additionally, because the decision to pursue a research project must be made before, often years before, the true value

of the project is revealed, scientists' ability to accurately predict the value of pursuing a subject may depend on their existing knowledge base on that subject. In line with the view of bounded rationality, scientists may be biased in their ability to form expectations about new opportunities less "local" to their prior endeavors (Nelson and Winter 1982; Levinthal 1997)[15].

Competitive Expectations: While these contests are certainly competitions, to simplify the model I assume that conditional on the quality of a submission, q_{iks}, these competitive interactions are "fixed" in the sense that they vary only across contests per ω_k and scientists per α_i. In other words, ω_k is a reduced form summary of the competition and information surrounding contest k (i.e. prize structure, expected entrants), and the scientist's fixed-effect α_i is a reduced form summary of their ability to compete on average (i.e. form expectations about competition). Essentially, I am defining expected outcomes, both in terms of success and the size of the award conditional on success, to be solely a function of idea quality and fixed differences across contests and scientists. This simplification of competitive interactions is reasonable given how "small" each scientist relative to the field. As discussed in Section 2, these contests regularly involve anywhere from 40 to 700 entrants, amongst an potential entrant pool in the tens of thousands[16].

Award Value: Conditional on success, the size of the award Award_{iks} is also a function of idea quality. And since the awards in these contests are resources to pursue the idea submitted, I allow individuals to value both the resources themselves as well as the production they enable, $P(\text{Award}_{iks})$, conditional on the quality of the idea, which may complement or substitute for awards. To the inherent value of the resources, scientists may have "empire-building"

[15]Should we expect, for example, a molecular biologist who specializes in parasitic nematodes to conduct the mental calculus of predicting how good of an epidemiological study on diabetes they might be able to perform? Perhaps not. A similar prediction of this aversion to less local ideas can be generated in a model of rational expectations where information acquisition is a costly function of the locality of the new idea and individuals are risk-averse. Here, scientists prefer local ideas not because they are unbiased, but because they are less uncertain.

[16]The number of unique applicants to the NIH during my 8-year time period is roughly 80,000 scientists. Using a narrower definition of potential entrants, each of the 27 Institutes receive applications from roughly 1,700 unique scientists each year.

preferences or, as is the case for academic scientists, are required by their institutions to acquire a certain amount of grant funds each year. Additionally, they may value the signal associated with these resources. Given the uncertainties of production in basic research, the ex-ante quality of a scientists' ideas - which largely influence the amount of resources they can acquire - can provide additional information about their underlying ability. Indeed, it is commonplace for academic scientists to publicize the number and size of grants they have received.

Taking the three proceeding factors together, the payoff from a contest is given by

$$\begin{aligned} U_{iks} =& \mathbb{E}\Big[\underbrace{\mathbf{1}(\text{Award}_{iks} > 0)}_{\text{Win probability}} \times \underbrace{V\big(\text{Award}_{iks}, P(\text{Award}_{iks})\big)}_{\text{Award \& production value}} \mid \underbrace{q_{iks}\,,\, \alpha_i\,,\, \omega_k}_{\substack{\text{Quality \&} \\ \text{competition}}} \Big] \\ & - \underbrace{C\big(q_{iks} \,|\, \alpha_i\,,\, D(s_i', s_{ik})\,,\, \omega_k\big)}_{\text{Entry costs}} - \underbrace{\bar{U}_i}_{\substack{\text{Outside} \\ \text{option}}} \end{aligned} \tag{3.1}$$

We are interested in how scientists respond to extra resources being made available (here, in an RFA) given the costs of scientific redirections necessary to pursue those resources. The elasticity of redirection is then based on (1) the costs of changing one's research direction, $\frac{\partial U}{\partial D(s_i', s_{ik})}$; and (2) the expected value of resources to be competed for, $\frac{\partial U}{\partial \mathbb{E}[\mathbf{1}(\text{Award}_{iks}>0)\times V]}$. As the ratio of costs to expected value grows (the elasticity of redirection decreases) managers will be forced to use larger prizes to induce their desired level of redirection. If it is very costly to redirect scientists, then managers must place a very large value on a potential goal to justify targeting resources towards it.

Ideally, to identify these parameters we would generate a random menu of potential grant competitions that vary in terms of both the level of redirection they necessitate as well as the nature of competition and rewards (expected payoff) they entail. If we could then force scientists to choose a competition, their choices could be used to identify the point of indifference when $D(s_i', s_{ik})$ and $\mathbb{E}[\mathbf{1}(\text{Award}_{iks} > 0)\times V]$ are equivalent and we could quantify the expected dollar amount necessary to induce a given level of scientific redirection. If

managers know this parameter along with the distribution of scientists, they can evaluate whether the costs of redirection are worth incurring given their preferences for the outcomes of these redirections.

Conducting this experiment is practically prohibitive, but these targeted contests provide a setting where scientists must make this precise tradeoff in a real-world setting. The NIH creates a (plausibly exogenous) menu of RFAs that allocate a number of grants and funds to certain types of science. These RFAs force scientists to ask themselves whether they should compete in the open contests with the best idea of their choosing, or in the targeted contest with the best idea that meets the objectives of that contest.

If I can empirically estimate the difference in expected value between the RFA and open contests that arises from the NIH's exogenous allocation decisions and scientists' endogenous responses, as well as the average level of redirection that RFAs induce beyond what is observed in the open contests ($\partial D(s_i', s_{ik})$), then I can relate these two parameters. In this way, I am phrasing the elasticity of redirection as a compensating differential: what is the expected value of a grant competition such that scientists are indifferent between it and a given level of redirection? This approach is in the same vein as Stern's (2004) analysis of job offers to biologists considering employment in industry and academia. Stern tests whether scientists are willing to take lower salaries in exchange for the right to publish - "do scientists pay to be scientists?" Here, I want to test whether scientists are willing to take on competitions with lower expected value in exchange for the right to direct their own work - do scientists pay to choose their science?

To fix ideas, consider a world where scientists can costlessly adjust the trajectories of their research. That is to say, they do not care about the nature of their research, they generate ideas of the same quality regardless of the topic or their existing knowledge, and can easily acquire any inputs necessary. In this case, all contests, regardless of whether they are targeted or not, should see competition to the point that their expected values are equivalent. If a contest exists with larger prizes (e.g. grant funds, or expected publication outcomes)

and/or an increased likelihood of success, then these "free-range" scientists would simply enter this contest and compete down the expected value until it equates with the alternatives. This world embodies the zero-profit nature of perfectly competitive markets with free entry.

Previously, I highlighted a number of forces that may limit scientists willingness to conduct certain types of science, and thus prevent scientists from competing away rents that might exist in targeted contests. This is analogous to economic models of markets where large fixed costs prevent new entrants from competing away profit margins. Here, the fixed costs of concern are the amount of redirection the targeted contests necessitate. As indicated by Table 9, the RFA contests appear to be associated with both an increased likelihood of success as well as larger grants conditional on success. This suggests that rents may remain within these contests, but scientists are unable (or unwilling) to compete for them.

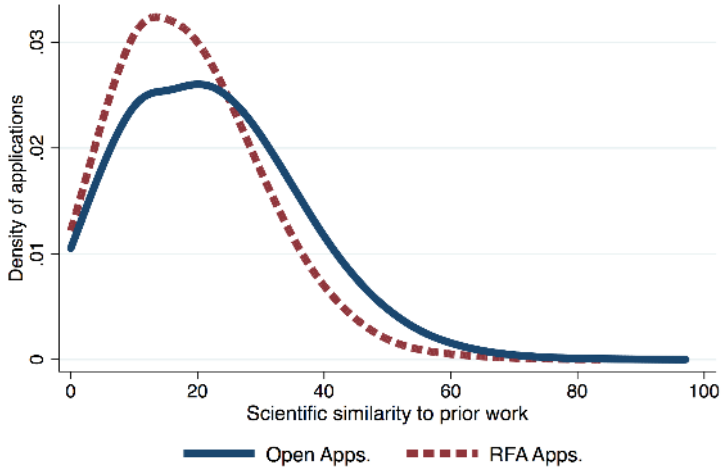
3.3.2. Data Construction

The two parameters of interest are the redirection effect and the expected payoff differences, which will amount to three outcomes in my data that I formalized as follows: (1) the scientific similarity of the application relative to the scientists' prior publications; (2) the probability of receiving a grant conditional on applying; and (3) the dollar amount of the grant, conditional on receipt.

For the first outcome, I use *pmra* to calculate the standardized scientific similarity scores between the applicants current application and prior publications to measure the degree of redirection undertaken. Figure 13 separately plots this distribution for applications submitted in the open (investigator-initiated) and RFA (targeted) contests. The fact that the distributions for RFA applications is shifted to the left indicates that the new projects proposed in these contests are less similar to scientists' prior work compared to what scientists' typically propose when left to their own devices.

The second outcome is a binary variable equaling 1 if the application is awarded a grant. And for the third outcome, the value of the grant conditional on success, I construct the

Figure 13: Scientific Similarity Distributions: Open vs. RFA



Note: Scientific similarity score is calculated by the *pmra* algorithm for each application based on scientists' prior publications. Larger scores indicate the science proposed in the application is more scientifically similar to the scientists' prior work.

lifetime value of each grant awarded, including all future funding years and extensions. Table 9 summarizes the unconditional win probabilities and award sizes, and summary statistics of all outcomes for the estimation sample are presented in the results tables.

3.3.3. Empirical Specification

The goal of the analysis is to identify differences between open and RFA contests with respect to these three outcomes. As a first pass, I could regress these outcomes for each individual i's application a to Disease-Science-Time contest dst on an indicator variable for whether the application was submitted to an RFA as a simple test for the difference in means between the two as follows:

$$y_{iadst} = \alpha + \beta \mathbf{1}\{\text{RFA}_{iadst}\} + \epsilon_{iadst} \tag{3.2}$$

where $\mathbf{1}\{RFA\}$ is an indicator for whether or not the application was submitted to an RFA contest. The main problem with identifying the parameters using this approach is that, as indicated in Equation 3.1, it is impossible to disentangle whether observed changes

in the outcomes are *because of* the RFAs and not any of the other sources of variation: application-specific quality (q_{ijk}); scientists-specific factors (α_i) and outside options ($\bar{U}_i$); or contest-specific differences (ω_k) that may vary across Institutes n at the NIH.

As a start, I could estimate Equation 3.2 with individual scientist, Science-Time and Disease-Time effects to remove all of the variation across individuals and most of the variation across contests. This approach would identify an RFA-specific treatment effect, which is a policy-relevant parameter for NIH administration. But if, for example, scientists place a significant value (positive or negative) on features inherent to an RFA, such as relinquishing control of their research or facing a narrower scope of competition, Equation 3.2 would not allow me to separate this RFA-specific effect from the redirection effect. In order to make a more general statement about the costs of scientific redirection, I need to remove any of the costs and benefits unique to these targeted contests ($\omega_{k=t}$).

I do so by exploiting variation in outcomes across both the targeted and open contests as well as within two of the most common grant types, the R01 and R21. To recall, the R01 grant is the most common grant type at the NIH and is intended for "exploitation" research. The R21 grant is intended to fund more "exploratory" research, and on average is about 65% smaller than R01 grants in dollar terms. Using variation in outcomes between these two grant types enables a difference-in-difference approach, where I can condition out the mean effects of scientists, grant type (R01 or R21) *and* RFAs. Because I can only observe outcomes for individuals that chose to apply, my results are applicable to scientists who have made the decision to apply to the NIH for funding and are faced with the choice of which mechanism to enter. Scientists fixed effects and application quality measures provide controls for selection by conditioning on a scientists' underlying propensity to apply to the NIH (scientists fixed-effects) as well as the possibility that those with the highest quality ideas are more likely to apply.

To make the research design clear, consider again Figure 13. Essentially, when comparing differences in redirection per my scientific similarity metric, I am testing whether or not

the difference between the RFA and Open distributions for R01 grants is different than the difference between the RFA and Open distributions for R21 grants. I estimate the following equation for each outcome:

$$\begin{aligned} y_{iadst} = &\alpha_i + \beta_j \mathbf{1}\{\text{R21}_{iadst}\} + \beta_k \mathbf{1}\{\text{RFA}_{iadst}\} + \beta_{jk} \mathbf{1}\{\text{R21, RFA}_{iadst}\} + \\ &\gamma_i \mathbf{X}_{iadst} + \delta_{dt} + \tau_{st} + \epsilon_{iadst} \end{aligned} \tag{3.3}$$

where $\mathbf{1}\{\cdot\}$ is an indicator for whether or not the application was submitted to an R01 grant competition, an RFA contest or both. The coefficients β_j and β_k are the first-difference effects of submitting an R21 or an RFA application, respectively. The focal coefficient β_{jk} captures the difference-in-difference of outcomes between the two grant and contest types. Scientist (α_i) Disease-Time (δ_{dt}) and Science-Time (τ_{st}) fixed effects control for variation across scientists and contests. The set of time-varying controls $\mathbf{X}$ depends on the outcome and are described below.

While the scientist-fixed effects can condition on any stable differences across individuals, it is also likely that scientists form expectations about the quality of each specific proposal and their likelihood of success. To the extent they can do this accurately, scientists may be more willing to undertake costly redirections when they predict their application is high quality. Thankfully, because I can directly observe proposal quality, whether the application is funded or not, such selection is not problematic[17].

Another selection concern relates to the fact that in a grant application, scientists must request a specific dollar amount for their project, which plays a large role in determining the final award amount, if successful[18]. Thus, with respect to a selection bias, scientists might be more likely to request larger amounts of funds when their expectations of success are high. Also, more practically speaking, these fund requests are very indicative of the

[17] Unless there is a behavioral bias present that some how generates a positive or negative correlation between beliefs and actual quality. Discussions with NIH applicants suggest that this is very unlikely, as expectations for each application are almost always set at the average success rates observed across the full set of applications.

[18] The correlation between realized and requested grant sizes, conditional on award, is 0.89.

scope of work proposed. It would be difficult to compare the value of a $1 million grant relative to a $1 thousand grant if the only reason for the difference was that the former study required a $999 thousand dollar piece of equipment. From the scientists' perspective, it is not necessarily the total size of the grant that is valuable, but rather the fraction of the amount they requested that is actually awarded - the grant surplus - that is valuable. If a project is proposed and the scientist requests $1 million, then I assume that they value the amount of funds they receive *conditional on this request*. This approach both addresses the selection concern, and makes it possible to equate award sizes across projects that request different funding levels. Therefore, I include the logged requested dollar amount as a control in the regression.

My main assumption throughout these analyses is that the change in the outcomes associated with the decision to pursue an R21 grant and/or an RFA contest are valid estimates of the counterfactual change for scientists that pursued the alternative grant (R01) and/or contest (open) types instead. Because these are one-time decisions, the traditional concern for pre-trends in this research design is not warranted. Instead, the major threat to identifying the focal effect β_{jk} for each outcome is a correlation between the outcomes and unobserved application-specific factors ϵ_{iadst}. For example, if scientists would have made the observed level of redirection *conditional on the fixed effects, observed grade and funds requested* regardless of whether or not they applied to the RFA contests and systematically choose to enter RFA contests when making redirections larger than average, then I would overstate the effect of entering these contests. In related work, analyses of science-level effects of RFAs suggest that they do induce new science that would not have occurred otherwise, so I am confident in the main assumption, which posits that, conditional on the rich set of controls, such selection into these contests does not occur[19].

[19] Anecdotal interviews with scientists who have participated in RFA competitions indicated that in all cases scientists incorporated features into their proposed studies that they would not have otherwise and that their submission to the RFA was at least in part induced by the RFA announcement itself.

3.3.4. Results

Table 10 presents the results from estimating Equation 3.3 for the four outcomes. The main effects (β coefficients) can be combined to estimate the elasticity of redirection, denoted by ε, as follows: $\varepsilon \equiv \frac{\beta^{SS}}{(1+\beta^{WP})*\beta^{GV}}$, where the superscripts SS, WP and GV denote the outcomes of scientific similarity, win probability and grant value, respectively. In other words, what is the change in scientific similarity induced by a certain change in expected value of a contest.

Table 10: Compensating Differentials per Redirections

	(1) Scientific Similarity	(2) Pr(Win)	(3) Award $	(4) ε
R21 × RFA (β_{jk})	0.170*** (0.0160)	-0.0424*** (0.0145)	-0.265*** (0.0442)	0.62
RFA (β_k)	-0.258*** (0.0105)	0.0750*** (0.00972)	0.236*** (0.0314)	1.01
R21 (β_j)	-0.213*** (0.00612)	-0.0170*** (0.00625)	-0.0924** (0.0388)	
Grade, $ Controls	Y	Y	Y	
Scientist F.E.	Y	Y	Y	
D-T, S-T F.E.	Y	Y	Y	
Obs.	122,094	56,718	24,223	
R^2	0.04	0.57	0.81	
mean(DV)	24.09	0.199	$1,323,760	
sd(DV)	13.45	0.40	$1,004,383	

Note: Dependent variables in columns (1) and (3) are standardized. Standard errors in parentheses, clustered at the scientist level. * p<0.10, ** p<0.05, *** p<0.01. Grade and $ controls include review grade for models (1) - (3), standardized dollars requested for models (1) - (3). Column (4) presents the first- (Row 1) and second-difference (Row 2) estimates of the elasticity of direction, ε.

Looking first to the RFA-specific treatment effect (β_k) that is estimated in first-differences (a la Equation 3.2), the second row of Table 10 indicates that, on average, applications to RFA contests are 0.25 s.d. less scientifically similar than when applications are submitted as investigator-initiated proposals. In terms of magnitude, this is a substantial effect. For comparison, Row 3 of Table 10 indicates that the difference in scientific similarity between R01 (exploitative) and R21 (explorative) grants in general is slightly lower at 0.21 s.d. Thus, it appears these contests solicit work that is substantially different, in the scientific sense,

than the average application submitted in the open contest. With the included controls there is also a significant increase in win probability between the two contests of 7.5 percentage points (36%), and the RFA awards are significantly larger (0.23 s.d. $\approx$ \$240,000) than their open contest counterparts. Thus, the wedge apparent in the raw data persists. These three estimates together imply that for each 1 s.d. increase in the expected value of a grant competition, the RFA contests are able to induce roughly the equivalent (1.08 s.d.) magnitude of redirection. However, as discussed earlier, this does not separate RFA-specific costs and benefits that may be unique to these mechanisms from the true costs of redirection that is of interest more generally.

Row 1 of Table 10 reports the difference-in-difference estimates for the main outcomes. Column 1 indicates that, compared to R01 grants, R21 grants submitted to RFA contests are 0.17 s.d. more scientifically similar (less redirected) than their open contest counterparts. Columns 2 and 3 indicate that, compared to R01 grants, R21 grants submitted to RFA contests are 4.2 percentage points (22%) less likely to be successful and 0.26 s.d. smaller than their open contest counterparts. Taken together, these estimates imply that each 1 s.d. increase in the expected value of a grant competition can induce a scientists to pursue a research project 0.65 s.d. less scientifically similar than they would have otherwise.

To get a sense of the economic magnitude of this effect, consider a scientist who has conducted preliminary research on subject A and is prepared to submit a \$1 million R01 application for a large-scale study on this subject. Now what if instead the NIH would prefer that this scientist pursues subject B, which would involve submitting the average R21 application to begin this "new" line of inquiry? Per my estimates and the sample means, the NIH could incentivize this level of redirection if they held an RFA soliciting subject B and awarded R21 grants roughly 180% larger than is customary: \$450,000 as opposed to the standard \$250,000. Note that these magnitudes include the fixed costs and benefits that an RFA provides (i.e. are based on the first differenced result). Looking beyond the RFAs at the NIH, the amount of funds necessary to induce this level of redirection via this funding

incentive alone is roughly $650,000.

How should we interpret the discrepancy between the elasticity estimates from the first-differenced (Row 2, $\varepsilon \approx 1$) and second-differenced (Row 1, $\varepsilon \approx 0.6$) results? Aside from requiring scientists to redirect their work, RFAs may involve a number of other fixed costs and benefits. Because the second-differenced elasticity estimate, which remove these RFA-fixed effects, is smaller than the first-differenced estimate it is implied that the fixed benefit of these contests is significantly greater than any fixed costs (beyond the change in scientific similarity). Although I cannot separately identify the magnitudes of these effects, it does suggest that the narrow scope of competition relative to the size of the prizes is valuable in this setting.

Table 11: Compensating Differentials for Scientist Subsets

	(1) Scientific Similarity	(2) Pr(Win)	(3) Award $	(4) ε
Pane A: New Investigators				
R21 × RFA	0.140***	0.0450	0.0198	$>1.14^a$
(β_{jk})	(0.0416)	(0.0432)	(0.112)	
Obs.	37,654	16,471	7,432	
R^2	0.0400	0.550	0.880	
mean(DV)	24.60	0.19	$1,386,899	
sd(DV)	14.46	0.39	$817,495	
Pane B: Experienced Investigators				
R21 × RFA	0.186***	-0.0614***	-0.317***	0.55
(β_{jk})	(0.0185)	(0.0170)	(0.0544)	
Obs.	84,440	40,247	16,791	
R^2	0.0400	0.570	0.810	
mean(DV)	23.87	0.19	$1,295,811	
sd(DV)	12.97	0.39	$1,075,642	

Note: Dependent variables in columns (1) and (3) are standardized. Standard errors in parentheses, clustered at the scientist level. * p<0.10, ** p<0.05, *** p<0.01. Grade and $ controls include review grade for models (1) - (3), standardized dollars requested for models (1) - (3) and standardized dollars awarded for model (4). All models include first-difference controls (RFA and R21) and the full set of controls as in Table 10. Experienced investigators have received at least one NIH research project grant prior to the focal application. Column (4) presents the first- (Row 1) and second-difference (Row 2) estimates of the elasticity of direction, ε.

[a]The elasticity is not identified since no statistically significant differential in expected payoffs exists.

Table 11 re-estimates the differential separately for "experienced" and "New Investigators", reporting only the second-difference coefficients. As expected, the younger, yet-to-win-at-the-NIH scientists are significantly more elastic than their search for funding. In fact, it appears that the differential estimated in the full sample is driven entirely by the experienced scientists. While the redirection levels are similar in magnitudes for the two subsets, there is no statistically significant difference in the expected payoffs for the New Investigators. Thus, I cannot identify their mean elasticity beyond a lower bound that reconciles the observed redirection with a zero difference in expected payoffs.

3.4. Identification via Entry

3.4.1. Entry Model and Empirical Specification

Because the elasticity estimates of Section 3.3 are conditioned on the decision to seek funding from the NIH, they are likely overestimates of the population mean. This is because the NIH application decision is a function of a scientists budget constraints, which are unobservable, where scientists in greater need of grant resources are both more willing to apply to the NIH and undertake redirections for lower expected payoffs. In order to identify an estimate that is not conditioned on this decision, this section estimates the elasticity of direction based on scientists decisions to enter (or not) an RFA given the size of funds available and their own scientific similarity to the research objectives.

Consider an entry model where scientists must incur some fixed costs to develop a research proposal and enter a contest where a pre-specified amount of funds are allocated per the number and quality of proposals submitted[20]. All potential entrants N observe each grant contest k, which are characterized by the total amount of funds available (Purse_k), the expected number of competitors ($\hat{n}_k \in (1, 2, ..., N)$) and a vector of other contest characteristics ($\mathbf{x}_k$; e.g. year, NIH Institute)[21]. Entry costs are a function of the scientists' similarity

[20]This simultaneous entry and production feature is reflective of the fact that these contests are essentially auctions, where scientists prepare research proposals as their "bids".

[21]This setup maintains the same structure as Equation 3.1 in a format similar to the homogenous firm entry model described in Berry and Reiss (2007). But here, potential entrants are homogenous conditional

to the contest objectives d_{ik}. A variable "profit" function V is then based on both contest and scientist characteristics[22], such that each individual's utility of entry is given by

$$U_{ik} = V(\text{Purse}_k\,,\,\hat{n}_k\,,\,\mathbf{x}_k\,,\;d_{ik}\,,\,\alpha_i) - C(d_{ik}) \tag{3.4}$$

where α_i is a stable, scientist-specific effect. The elasticity of redirection is then $\varepsilon \equiv \frac{\partial U_{ik}}{\partial \text{Purse}_k} / \frac{\partial U_{ik}}{\partial d_{ik}}$. This definition is different from that presented in Section 3.3 where instead of trading off direct resource compensation for redirections, scientists are now trading off the "market size" (Purse_k) for redirections given their competitive expectations.

In order to identify the partial derivatives necessary, I make the simplifying assumption that V is a linear function of its parameters yielding my main estimating equation[23]

$$\Pr(\text{Entry}_{ik}) = F(\text{Purse}_k) + G(\hat{n}_k) - C(d_{ik}) + \mathbf{x}_k\beta_x + \alpha_i + \mu_{ik} \tag{3.5}$$

where μ_{ik} are i.i.d. error terms that contain both random (mean-zero) noise from scientists' decisions and, importantly, unobservable and potentially non-random contest features. Note that although d_{ik} enters Equation 3.4 through both V and C, it only enters Equation 3.5 through C. Therefore C should be viewed as encompassing the aggregate effects of what are traditionally thought of as adjustment costs (e.g. new equipment), as well as the costs of producing a competitive research proposal[24].

The estimate of the elasticity of direction is then $\frac{F'}{C'}$. For the purposes of a straightforward and interpretable estimation, I assume that F, C, and G are log-linear functions (i.e. $C(d_{ik}) = \beta_c \times \log(d_{ik})$).

on the individual-fixed effect α_i, entry cost factors are observable and production occurs at entry.

[22] This function also describes the rules by which the purse is allocated amongst entrants.

[23] This is an admittedly strong assumption about both the nature of competition within these contests as well as scientists' utility over the outcomes. However, since the goal of this exercise is to estimate how scientists trade off redirection costs for expected resources - and not to simulate complex counterfactuals - these assumptions provide a transparent method for estimating the entry elasticity of direction for the marginal entrant while controlling for first-order competitive effects.

[24] That is to say, the $C'(d_{ik})$ estimated from Equation 3.5 is actually $\frac{\partial V}{\partial d_{ik}} - C'(d_{ik})$ per Equation 3.4. This prevents me from disentangling the extent to which the elasticity of direction is driven by tangible costs versus limitations on the production function (i.e. the substitutability of knowledge across topics).

3.4.2. Approach to Endogenous Competition

The difficulty in estimating Equation 3.5 is that, instead of competitive expectations ($\hat{n}_k$), only the realized number of entrants (n_k) is observed. Now, if there are unobservable and valued features of these contests not captured by $\mathbf{x}_k$ (e.g. scientific potential, downstream demand), then each scientists' likelihood of entry will be positively correlated. This will generate an endogeneity problem if Equation 3.5 is estimated with n_k instead of $\hat{n}_k$, which will bias estimates of G upward and then, potentially, F downward[25].

To address this issue, I utilize the procedure for estimating static strategic interactions proposed by Bajari et al. (2010)[26]. Essentially, $\hat{n}_k$ can be estimated if a variable exists that influences each individual's strategic choice *directly* but only influences others' choices via the *indirect* effect of those strategic choices. In what amounts to an instrumental variables approach, estimates of $\hat{n}_k$ can be obtained using variables that satisfy this "strategic" exclusion restriction. This approach is well-suited a setting where fixed costs of entry are observable and orthogonal to post-entry production (or effort) costs. This is precisely the case of the model outlined above with the scientific adjustment costs providing a valid instrument under the assumption that each scientists' similarity to an RFA does not directly influence any other scientists' behaviors[27].

The estimation approach is as follows:

1. Regress entry decision ($\text{Entry}_{ik} = \{0, 1\}$) on the vector of contest controls, purse size, individual-fixed effects and scientific similarity measure.

[25] Recall that I am assuming that conditional on the covariates and fixed effects scientists are homogeneous. This implies preferences for the unobserved contest features are the same across the population, which generates the positive correlation in entry probabilities and therefore a positive correlation between $\Pr(\text{Entry}_{ik})$ and n_k.

[26] While there may certainly be dynamics with respect to each scientist's decision to purse a particular contest (e.g. how would moving to topic A effect future research prospects?), the limited recurrence of RFAs and the massive scale of the default open competitions, which present a future option for funding, suggest that competitive dynamics within each RFA are likely not first-order concern.

[27] One mechanism that may invalidate this assumption is if scientists' likelihood of communicating with potential competitors about entry decisions is correlated with their similarity to an RFA. Anecdotal discussions with NIH applications who have competed in RFAs did not suggest high levels of communications between "rival" laboratories.

2. Calculate predicted entry probabilities, $\widehat{\Pr(\text{Entry}_{ik})}$

3. Sum predicted entry probabilities over each contest, minus each individual's own entry probability to estimate each individual's $\hat{n}_k$, given by $\widetilde{n}_{ik} = \sum_{j \neq i} \widehat{\Pr(\text{Entry}_{jk})} \; \forall \; i, j \in N$

4. Estimate Eq. 3.5 using $\widetilde{n}_{ik}$ in place of $\hat{n}_k$

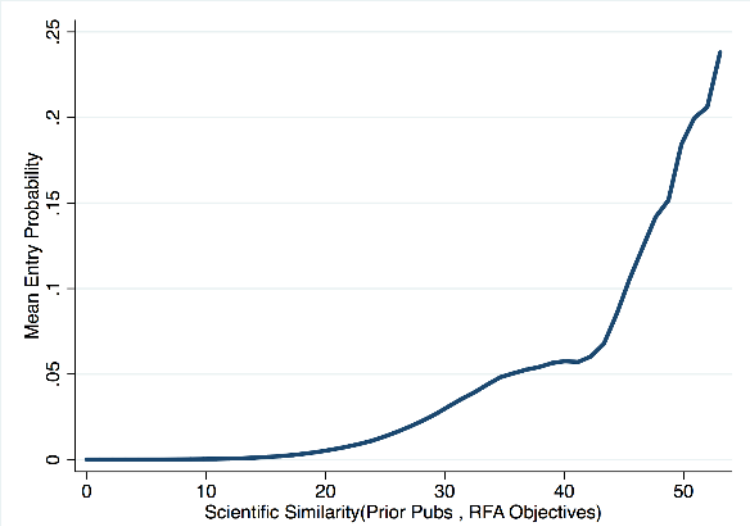
This approach has the added benefit of estimating a related parameter that I will refer to as the entry elasticity: the percent change in entry probability given a 1% change in the purse size, holding redirection costs and expected competition fixed. The magnitude of this elasticity can shed some insight on the more general costs of incentivizing scientists to pursue a particular direction purely through resource incentives.

3.4.3. Data Construction and Description

Estimating Equation 3.5 requires a dataset comprised of scientist-contest pairs containing (1) all potential entrants, and (2) each potential entrant's scientific similarity to the contest. To arrive at a close approximation to the full set of potential entrants, I include any individual that applied to the NIH from 2002 to 2009, totaling to 132,043 scientists. I then match each scientist (and their publication history) with each of the 247 RFA announcements between 2002 and 2009 that solicite some combination of R01 and/or R21 research grants and for which details on the timing, administration, research objectives and funds allocated were available. For each of these pairs, *pmra* was then used to calculate the scientific similarity between the scientists' prior publications and the research objectives of the contest.

Thus, this research design resembles a typical study of firm entry behavior, but, instead of a very small number of firms, potential markets and instruments such as geographic distance, this data is characterized by a large number of individuals and potential prizes, and will utilize the instrument of scientific distance. The relationship between entry and scientific similarity is made clear in Figure 14.

Figure 14: Entry per Scientific Similarity



Note: Local polynomial smooth plot. Scientific similarity score is calculated as the maximum *pmra* algorithm score between each scientists' set of prior publications and the research objectives of the RFA. Larger scores indicate the science requested in the RFA is more scientifically similar to the scientists' prior work.

3.4.4. Results

Table 12 presents the results of estimating Equation 3.5 using both the OLS approach based on observed competition n_k (odd columns), and the instrumental variables approach of Bajari et al. (2010) using $\widetilde{n}_{ik}$ (even columns). Clearly, OLS appears to result in a severe bias such that scientists appear to not value larger purses and prefer larger numbers of competitors. Certainly, given the unique preferences of scientists and the uncertainty and information asymmetries of science, it may be the case that these individuals behavior is as such. However, accounting for the endogeneity of these interactions presents a much more intuitive picture where scientists prefer larger purses and smaller sets of competitors.

With only contest-level controls, the estimates indicate an elasticity of direction of roughly 0.5 - close to the magnitude identified via the compensating differentials approach. But when conditioning on scientist-fixed effects, the magnitude ε is estimated to be about half as large at 0.24. This large discrepancy between the two specifications suggests that there is a significant fraction of scientists who are (A) both inherently more likely to enter RFA

contests and do so from a far distance, and/or (B) both inherently unlikely to enter RFA contests despite their close proximity.

Table 12: Entry Determinants and Elasticities

	(1)	(2)	(3)	(4)	(5)	(6)
log(Sci. Sim.$_{ik}$)	0.000384***	0.000418***	0.000395***	0.000442***	0.000658***	0.000906***
	(0.0000235)	(0.0000247)	(0.0000241)	(0.0000257)	(0.0000568)	(0.0000569)
log(Purse$_k$)	0.00000657	0.000186***	0.0000309*	0.000218***	0.0000301	0.000218***
	(0.0000138)	(0.0000292)	(0.0000164)	(0.0000265)	(0.0000216)	(0.0000241)
log(n_k)	0.000233***		0.000201***		0.000193***	
	(0.0000199)		(0.0000181)		(0.0000201)	
log($\widetilde{n_{ik}}$)		-0.000134***		-0.000184***		-0.000143***
		(0.0000345)		(0.0000313)		(0.0000250)
Contest Controls			Y	Y	Y	Y
Scientist F.E.					Y	Y
$\mathcal{E}$		0.44		0.49		0.24
Entry Elasticity		0.62		0.71		0.70
Share Avg. Grant for +1 Entrant		31.2%		27.1%		27.5%

Note: Mean entry probability 0.0002942. Standard errors in parentheses, clustered at the RFA contest level.

* p<0.10, ** p<0.05, *** p<0.01. ε is the elasticity of direction per $\frac{\beta_{\log(\text{Purse}_k)}}{\beta_{\log(\text{Sci. Sim.}_{ik})}}$. "Entry Elasticity" is the percent change in entry probability given a 1% change in the Purse size. "Share Avg. Grant for +1 Entrant" is the growth in Purse size necessary to induce the marginal entrant, in expectation, expressed as a percent of the average grant awarded.

Given the substantial differences in the elasticity of direction between new and experienced scientists identified in Section 3.3, Table 13 estimates the entry determinants for these two subsamples to see if the difference arises in this alternative approach[28]. Again, the New Investigators are more elastic with their direction; however, not to the degree observed once conditioning on the decision to apply.

Tables 12 and 13 also present the estimates for entry elasticities, and to facilitate interpretation, growth in Purse size necessary to induce one additional entrant, given as a percent of the average grant size for each specific sample. Under the assumptions of the empirical model, this latter measure represents the average total costs scientists face when preparing an NIH application (in terms of the corresponding grant value) - roughly 27% of a grant, or $95,000 on average. Interestingly, while the New Investigators appear to be more elastic with respect to the direction of their science, they also appear to be slightly more inelastic with respect to RFA entry, relative to their experienced colleagues. Why might this be the case? Recall that the aforementioned benefits of this designation are within the open

[28]In the specifications of Table 13, $\widetilde{n_{ik}}$ is still estimated using the full sample of scientists.

Table 13: Entry Determinants and Elasticities, by Scientist Experience

	(1) New Investigator	(2) Experienced
log(Sci. Sim.$_{ik}$)	0.00111***	0.00172***
	(0.0000847)	(0.000109)
log(Purse$_k$)	0.000284***	0.000378***
	(0.0000354)	(0.0000438)
log($\widetilde{n_{ik}}$)	-0.000211***	-0.000288***
	(0.0000414)	(0.0000453)
Contest Controls	Y	Y
Scientist F.E.	Y	Y
ε	0.26	0.22
Entry Elasticity	0.710	0.740
Share Avg. Grant for +1 Entrant	23.50%	15.20%

Note: Standard errors in parentheses, clustered at the RFA contest level. * p<0.10, ** p<0.05, *** p<0.01. ε is the elasticity of direction per $\frac{\beta_{\log(\text{Purse}_k)}}{\beta_{\log(\text{Sci. Sim.}_{ik})}}$. "Entry Elasticity" is the percent change in entry probability given a 1% change in the Purse size. "Share Avg. Grant for +1 Entrant" is the growth in Purse size necessary to induce the marginal entrant, in expectation, expressed as a percent of the average grants awarded to either applicant type.

contests, not RFAs. The New Investigator's relatively larger entry costs (24% versus 15% of a grant) are driven partly by this difference, but more so by their smaller eventual grant awards (85% the size of experienced investigator's awards on average).

3.5. Discussion

Managers face a unique set of issues when directing highly specialized individuals be they scientists, engineers or artists; they often lack information about who in particular should be assigned each specific task. To overcome this challenge organizations often rely on mechanisms, such as contests, that allow individuals to self-select into opportunities most suited to their skills and knowledge. The usefulness of these mechanisms depend critically on how individuals tradeoff the value of the prize (which are often resources to conduct work) and the costs of redirecting themselves.

These are precisely the issues facing the National Institutes of Health. This Chapter explores

the NIH's attempts to overcome the constraints of a high-skilled workforce and induce new research with targeted grant competitions. I use data on all applications to the Institutes for eight years along with a novel text algorithm in order to conduct two complementary analyses to identify the elasticity of direction - how much scientists tradeoff changes to their research trajectory in exchange for resources to conduct that research, or in other words, the adjustment costs of science.

My estimates range from 0.24 to 0.6, magnitudes that suggest relatively large amounts of grant resources are needed to induce directional changes in the scientific workforce[29]. In a similar vein, Stern (2004) finds that scientists are willing to take salaries 25% lower in exchange for the right to publish. So, it appears that scientists are willing to pay to be scientists, and some also to choose their own science. But in this case, the extent to which this elasticity is a function of scientists' preferences, real adjustment costs (e.g. equipment) or the substitutability of knowledge in the scientific production function is an important, unresolved question.

How generalizable are these findings? Certainly the types of scientists pursuing funding from the NIH are a particular set of innovators, often based at academic institutions and all pursuing biomedical research projects. But by and large they are not the prototypical "aloof" scientists who answer to none, have unlimited resources and pursue the most fundamental questions of nature with disregard for applicability. These individuals still report to senior figures at their institutions (i.e. Department Heads, Deans), are pursuing science with relatively short-term implications, and maintain close personal and professional ties with staff, colleagues and competitors, all while balancing a number scientific and non-scientific responsibilities. Thus, the tradeoffs I examine here are likely to be applicable to many scientists who operate on the edge of basic and applied research.

As in the case of the NIH, contests have become an increasingly popular mechanism to both

[29]Importantly, this is conditional on the current background infrastructure of support for science - the outside options - which, if drastically altered, could certainly influence these magnitudes.

identify innovative opportunities and direct the workforce towards them (Terwiesch and Ulrich 2009; Williams 2012). The usefulness of contests to induce effort in R&D settings has long been appreciated (Moldovanu and Sela 2001; Che and Gale 2003). And recent empirical work by Boudreau et al. (2011) and Boudreau et al. (2016b) provide stark empirical support for these predictions in a setting with psuedo-random assignment to contests, thereby removing issues of self-selection from their analyses. In this Chapter, it is precisely the nature of self-selection that I investigate. Thus, these results contribute to a growing literature on the use of resource allocations as sorting mechanisms and the particulars of incentivizing this sorting amidst high-skilled workers[30].

Measurement and empirical challenges have long prevented in-depth investigations of this parameter, despite the clear need for understanding its magnitude. Given the lack of evidence, it is not surprising that research policies, such as those designed to address the "bottleneck" in translational biomedical research (e.g. Collins 2011), have not traditionally considered that allocating funds to particular areas is not only necessary to *conduct* the research but it must also *incentivize the choice* of scientists' to redirect their research to that area.

[30]For other examples, see work on local governments crowd-sourcing inspection algorithms (Glaeser et al. 2016), hospitals hosting tournaments to improve the workplace (Blasco et al. 2016) and pharmaceutical firms adopting venture capital-like funding contests to identify promising therapies (Lerner 2012)

CHAPTER 4 : A Ricardian-Demand Explanation for Changing Pharmaceutical R&D Productivity

This Chapter is based on joint work with Mark Pauly

4.1. Introduction

Considerable criticism has been expressed about the research and development process for new drugs since at least the 1990s. Compared to those years, estimates of private R&D cost per new drug that makes it to market have been high and rising rapidly, more than tripling. The volume of new drugs was flat for much of the period. While it is hard to construct a price index for a mix of new products, there certainly have been complaints about the prices of those new products. The perception of limited introduction of truly novel drugs procured at dizzying research costs and sold at surprisingly high prices have led to feelings of malaise on the part of the industry and feelings of outrage by its critics[1]. Despite a little improvement recently, the drug development model has broken, people say, but there is no generally accepted account of why and how, and what if anything to do to put it back together again. Maybe it was the FDA's fault for restrictive and costly rules and procedures. Maybe it was the failure of firm management in managing the discovery process as well as it had in the past, as science changed in way that required adaptation. Maybe scientists were preoccupied with publication and grant financing. Maybe the number of good ideas in science were as numerous and attractive as ever, but there was some kind of glitch in translating them into clinically useful products (Scannell et al. 2012).

In this Chapter we try to explain what happened (and to some extent deflect this blame game) by using new disaggregated data on NIH and US private firm R&D investment across time and across therapeutic categories to establish the relationship between these two sources of inputs for drug discovery and development and the associated output in terms of New Medical Entities (NMEs): a new drug production function. Our best data are for the years

[1]See for example, Hewitt, Campbell & Cacciotti. *Beyond the shadow of a drought: the need for a new mindset in pharma R&D.* Oliver Wyman, 2011.

up to 2000, but we extend the analysis to 2013 with our empirical approach. We begin with a simple production function and use it to estimate the marginal NME product of each type of spending. We then tell a story about those estimates that draws on a demand side explanation for both investments in new products and their R&D costs.

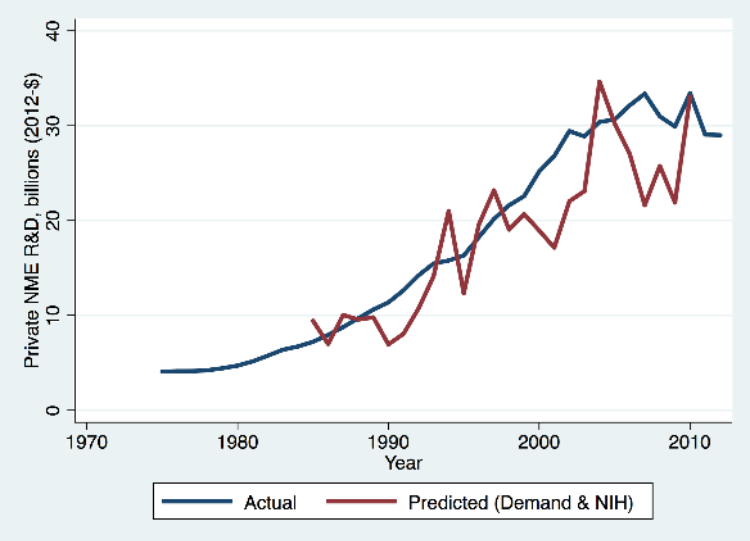
Disaggregated data on firm investments across therapeutic categories enables us to estimate the relationship between private R&D investment and (1) exogenous shifts in potential market size driven by demographic trends as suggested by Acemoglu and Linn (2004), and (2) prior investments in basic research at the National Institutes of Health (NIH). These two factors alone an explain a substantial proportion, nearly 90%, of the variation in R&D investments within each category, and predict a private marginal cost per new molecular entity (NME) in line with previous estimate for this time period.

We assume that there are two inputs into the production of NMEs of a particular category (such as cancer, cardiovascular, etc.): spending on research by NIH institutes, and spending on R&D by private firms. NIH spending may sometimes be thought of as a public good, developing information useful for many new products at near-zero marginal cost per additional user. But not all NIH research necessarily is a public good in the non-excludable sense, because NIH funding of university research can become intellectual property for the university or its assignees under the Bayh-Dole act, with no repayment to the taxpayers who funded the research. Conversely there is nothing to prevent a private firm from supporting basic research as well as applied research. Finally, not all spending in both sectors supports clinical research (rather than marketing or community outreach). While our data on these inputs may over- or under-estimate their true values for these reasons, we are mainly concerned with changes in the relationships within our model. Thus, bias will only arise if our measurement errors are correlated with other factors of the model and persist conditional on our controls.

We also explore the impact of private spending. If it is random and uncorrelated with public investments, it will not affect the estimated productivity of NIH spending and will

itself have an unbiased coefficient. But there are strong reasons to think that firms respond to NIH findings, and this we explore in considerable detail. In addition, we assume, in the spirit of Acemoglu and Linn (2004), that private investment responds to the "size of the market" (specifically, potential annual gross revenues) for particular categories of drugs in particular time periods. Managers will have expectations about the future of certain therapeutic markets, and will allocate their investments accordingly. After first treating private spending as an included exogenous variable, we then explore the consequences - both for measures of NIH productivity and private firm productivity - of treating private spending as endogenous. Along the way we consider the interesting question of whether NIH spending has a positive or inverse effect on private spending. However, for the most

Figure 15: Private NME R&D Investments: Actual and Predicted



Note: Actual spending data based on estimates of (1) total U.S. domestic R&D and (2) share of R&D for New Molecular Entities (NME) from PhRMA annual reports. Predicted R&D is based on the relationship between potential market size and NIH spending estimated with pre-2001 data.

policy relevant time period we would like to study, post-2000 compared to pre- we only have aggregated industry R&D data, which prohibits a more straightforward analysis. To overcome this limitation and attempt to diagnose the changes observed in this timeframe, we generate estimates of predicted private R&D spending as a function of future demand

and prior NIH spending. We then use those predicted values to estimate the production function for the second half of the period, where we are especially interested in changes in NIH and private R&D productivity. We interpret these descriptive empirical results in terms of a model of the supply of and demand for investment in new drugs, with special emphasis on the role of consumers' willingness to pay for new drugs.

Figure 15 shows the accuracy of our predictive model (described in detail below); it plots the sum of these predicted values alongside actual aggregate spending to illustrate the significant role of potential demand and basic science in generating private investments. The figure also highlights the scope of our analyses - we can only explore changes in the productivity of R&D dollars invested in response to the industry's major forces of supply and demand. Still, our results speak to the productivity of "real" R&D at the firm, irrespective of changes in regulations or other external industry-wide shocks that influence costs or production beyond the R&D phase. The accuracy of prediction clearly drops after 2000 (often under-predicting private R&D) but still remains reasonably close in terms of trend. That firms invested more than was predicted while the rate of NME approvals remained flat is suggestive of one type of productivity decline in this period (drugs per R&D dollar), however, these may still have been optimal investments from firms' perspectives if the growth in future demand suggested firms would remain productive in generating economic profits (returns per R&D dollar) even at lower rates of new drug productivity.

4.2. Theoretical Motivation: David Ricardo and Productive Ideas

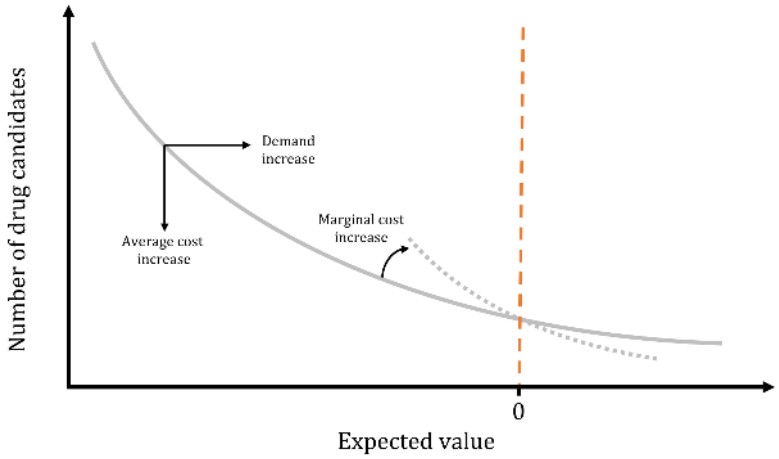
One of the major contributions of the economist David Ricardo was the distinction between price mechanisms for different kinds of production inputs. Ricardo argued the productivity of agriculture depended on the type of input. Most inputs (e.g. seed, animals) had their prices determined by the marginal cost of producing them, and in turn affected the price of crops. But the price of land (rent) was a different story; in this case, rather than the price of the final output being "cost determined," it was more correct to say that its price or cost per unit was determined by the price of the final product, itself a function of demand.

The simple notion is that the stock of arable land is exogenous (or at least fixed in the short run) but its price is determined by demand. His key insight for our purposes is that, as increases in demand bring more, less fertile, land into production, productivity will necessarily decline. That decline is an inevitable result of moving on to the set of next most profitable opportunities. Indeed, in this model with supply of opportunities fixed and growing demand, increases in demand must result in declining measured productivity with no person or institutional change being at "fault".

We propose a similar process determining the R&D cost per new drug. Our implied theoretical model is as follows: exogenous inputs, namely public investments in basic science and demographic-driven demand growth, determine the distribution of new drug ideas per their expected profitability. An increase in demand that may arise from growth in population size, income or insurance coverage will increase the return on investment across all potential drug opportunities, causing ideas with costs too high to justify investment under prior (lower) demand to become potentially profitable. But importantly, unless these demand shocks are accompanied by equal growth in the supply (or quality) of these ideas, these newly profitable ideas have the same lower expected productivity (higher costs) from when demand was lower.

That is, increasing the output of new drugs in response to demand increases has to lower the productivity of R&D when inputs are scarce - without anyone being at fault, without any evildoers creating glitches, scientists being distracted, and without any industry's R&D model breaking. Unless there is an equivalent offset in the number or quality of ideas - a scientific discovery or a surge in NIH investments - increasing R&D costs per new product is an expected fate in a world with growing demand. The movement to "high hanging fruit" is not bad luck; it is a result of firms' rational investment decisions. Figure 2 illustrates this distribution of R&D opportunities and how the optimal investment decision changes as demand and costs determine the expected profits associated with each potential new drug. The figure shows a distribution of new drug ideas by their expected value based on expected

Figure 16: Illustrative Count of Drug Candidates and the Role of Demand and Costs



revenues and costs (Expected Value = (Expected Revenues × Success Rate) - Expected Costs).

Our underlying intuition is that firms observe this distribution and select profitable ideas (Expected Value > 0) to commercialize. In practice, this amounts to monopoly rents being competed away post patent expiration. However, this distribution can be "restocked" in the sense that investments in basic research can change the number and/or quality of the ideas. Thus, the distribution represents the quality of ideas at a fixed point in time. Our empirical analyses will explore year-to-year changes in its shape. With information on the shape and evolution of this distribution one could predict both the number of new products (the number of ideas multiplied by the average success rate conditional on the Expected Value > 0) and expected net revenues (the sum of success rates multiplied by revenues for Expected Value > 0).

Holding costs constant, an increase in total demand will shift the distribution to the right as larger expected revenues increase the value of all ideas. Now, ideas previously unprofitable because of high costs relative to revenues are chosen for investment. Without any changes to the productivity of an industry increases in demand will spur higher spending on the

research and development of new products.

Conditional on demand, changes in the expected costs (success probability times actual costs) can affect the distribution in two important ways. Increases in the costs of the *marginal* project will shift the slope of the distribution downward, while increases in the *average* costs across all projects (which includes any fixed costs of having a non-zero number of projects) will shift the intercept of the distribution downward, and vice versa for decreases in these costs. Diagnosing the productivity of this industry is, in effect, identifying what (if anything) happened to the shape of this distribution.

To be clear, this figure depends on the productivity (profitability) of the marginal and average R&D *idea* and not the marginal or average productivity of each R&D dollar. However, project-level data of any industry is extremely difficult to obtain and in this Chapter our data on firm investments are aggregated to the therapeutic-class level. Thus, our empirical estimates will be based on changes in the total dollars invested within each class over time and will technically identify the average and marginal productivity of R&D dollars. But given the high fixed costs of pharmaceutical R&D that we are measuring, we will instead interpret our results as providing insight into changes over time in the productivity of the average and marginal ideas.

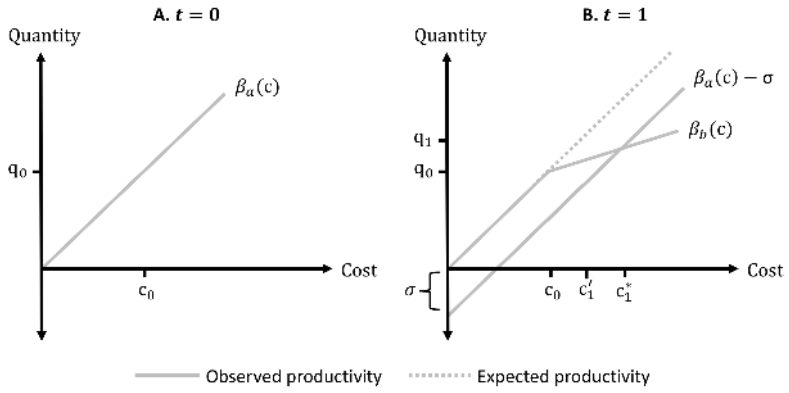
What drives differences in the costs associated with these ideas? Looking across ideas, recent work from Budish et al. (2015) highlights the importance of variation in the costs across projects for different types of cancers and find evidence of distortions away from research on early-stage cancers, which have a larger clinical trial costs[2]. Looking across time, increases in the length and complexity of clinical trials, which would decrease the average profitability of ideas, has been observed over the past decades. Regulatory burdens such as longer approval times and larger post-marketing surveillance efforts have also been cited as sources of larger total (average) costs in the industry (Kola and Landis 2004).

[2]Because the clinical trials must follow patients for much longer to reach clinical endpoints. The authors also discuss the role of "short-termism" whereby agency problems in the management process induces inefficient discounting that incentivizes the pursuit of shorter clinical trials (Budish et al. 2015).

In terms of the farmers first studied by Ricardo, comparing the productivity of the marginal and average new drug opportunity is analogous to evaluating whether increases in demand encouraged farmers to seek out land that was less fertile (decreasing productivity of the marginal idea) or incur higher fixed costs of transporting crops to market (decreasing productivity of the average idea).

This distinction of marginal and average costs is essential for understanding how certain policies will influence production in this industry. For example, R&D tax subsidies typically offset marginal R&D expenditures (since they're granted as a share of expenditures), while market exclusivity provisions and other prizes in effect reduce the average (fixed) costs of eligible trials. Some policies, such as the Orphan Drug Act, operate on both margins in that they provide prizes (exclusivity) and subsidies (tax breaks). Despite this importance, delineating between these two margins does not appear to occur regularly in policy debates or analyses of the pharmaceutical industry. In diagnosing this industry, we will make clear where our estimates indicate changes have occurred, on average or the margin. Figure 3

Figure 17: Identifying Production Shocks amidst Demand Growth



provides a visual depiction of the empirical analysis as it pertains to the distribution of ideas presented in Figure 2. Consider a monopolist firm who observes demand and sets production costs c which determine quantity produced q per $q = \beta(c)$. Demand at time $t = 0$ determines the optimal quantity q_0^*. As researchers, we observe c_0 and q_0, and then by assuming that firms optimally invested in R&D given their expectations of demand can

recover an estimate of the productivity parameter at this time, denoted β_a.

Now consider the next time period $t = 1$ where the optimal quantity to deliver is q_1^*, and we researchers observe c_1^* (Figure 3B). Because $c_1^{'}$ is less than c_1^*, we can infer that the distribution of ideas presented in Figure 2 has changed. Here, σ (the intercept) captures any level changes in the distribution, while the difference between β_a and β_b captures any change in the slope of the distribution. As drawn, the two production functions in $t = 1$ achieve q_1 for the same total (average) costs. While this is certainly a unique case, it demonstrates that when observed in the cross-section it is impossible to discern average versus marginal changes in the productivity of ideas, the goal of our exercise. This motivates our use of only within-drug category variation in order to condition out time invariant differences across drug classes.

4.3. Empirical Production Function

We model the number of NMEs approved by the FDA in therapeutic category j in time t as a reduced form conditional Poisson model as

$$\text{NME}_{jt} = \frac{\exp(\beta \mathbf{X}_{jt} + \sigma_t)}{\sum_{\tau=t} \exp(\beta \mathbf{X}_{j\tau} + \sigma_\tau)} + \epsilon_{jt} \tag{4.1}$$

where $\mathbf{X}_{jt}$ is a vector of category-time specific inputs to the production process and σ_t are industry-wide shocks to average productivity. This formulation accommodates the count nature of our dependent variable while conditioning out time-invariant differences across therapeutic categories (e.g. fixed costs of a clinical trial). These average differences appear to be substantial, accounting for roughly 30% of the observed variation in NME output and nearly 80% of the observed variation in private R&D spending across therapeutic categories[3].

We begin by assuming that our two main production determinants of $\mathbf{X}_{jt}$ are (1) lagged private R&D, and (2) lagged public basic science are exogenously determined, and estimate

[3]These estimate are from log transformed OLS models with therapeutic category fixed effects, since coefficients of these effects are not estimated in the conditional Poisson specification. Although these specifications are obviously biased, they provide a sense of the magnitude of these effects.

our main equation (1).

Importantly, some models will be presented without time period fixed effects to inform calculations of the implied marginal products per each input. Such "cost per NME" calculations, which are intended to inform managers and policymakers, should not control for industry-wide trends because the influence of such trends on the dependent variable is part of the main story. Any estimates of costs based on elasticities estimated in models demeaned by (statistically significant) time trends is likely to over- or underestimate productivity if aggregate productivity is decreasing or increasing, respectively. Moreover, at least as far as NIH spending is concerned, spending in one clinical category may have positive effects on productivity in other categories (e.g., the research that lead to Avastin a cancer drug also led to Lucentis an eye drug).

Private firms were likely not funding R&D projects at random even in the period of low productivity, but are rather responding strategically to prior investments in basic science, future expectations of demand and their consequences for profits in excess of normal economic profits. And because we cannot identify exogenous firm characteristics, estimating equation (1) including private spending along with the supply of science and ignoring the demand for certain types of drugs would confound firm-level and endogenous market-level characteristics, biasing our estimates of β.

Hence, we also write a two-stage model of investment decisions and production where private R&D in class j at time t is determined by market-level characteristics $\mathbf{SD}_{jt}$ that include our demand measure and NIH spending

$$\log(\text{R\&D}_{jt}) = \beta_0 + \beta_{\mathbf{SD}}\mathbf{SD_{jt}} + \delta_j + \mu_t + \varepsilon_j t \tag{4.2}$$

which in turn determines NME output within classes over time per

$$\text{NME}_{jt} = \frac{\exp(\beta_R \log(\text{R\&D}_{jt}) + \sigma_t)}{\sum_{\tau=t} \exp(\beta_R \log(\text{R\&D}_{j\tau}) + \sigma_\tau)} + v_{jt} \tag{4.3}$$

When disaggregated data on private spending is available, we can estimate a joint variant of these equations using instrumental variables via Generalized Method of Moments as outlined by Blundell et al. (2002)[4].

Given our data limitations of lacking post-2000 category specific investments (described below), we will also implement a two-step version of these equations under the assumption that the causal effects of supply and demand conditions estimated in the first stage ($\beta_{\mathbf{SD}}$) are persistent. Then, substituting predicted private R&D into the second equation, we can test for significant changes in β_R post-2000 which would indicate productivity changes. Furthermore, we can still test whether or not public investments in basic science have any direct effect on NME output beyond their ability to stimulate private investment.

It is worth noting that a key determinant of investment beyond forces of supply and demand (per $\mathbf{SD_{jt}}$) is firms' expectations about their own production functions, β_r and σ_t. Changes in expectations about these parameters would be captured in our model by the μ_t term in equation (3). For example, if firms knew in advance that was to change in future periods, perhaps due to frictions in the labor market, and responded by changing investments across all categories, then this would manifest in μ_t. However, we cannot estimate these time-period fixed effects for years we lack disaggregated investment data.

Therefore any changes in β_r or σ_t that we identify post-2000 are related to changes only in the productivity of the private R&D instrumented by our measures of supply and demand. We do not know whether there were unobserved industry-wide forces that prompted firms to invest additional amounts in R&D beyond what was predicted by the pre-2000 model. It is plausible there were unrealistically optimistic expectations of success. Supply side changes, like changes in FDA policies that made approval unexpectedly less likely, might also come have into play. If such changes in policy had been expected so that anticipated profits were reduced, that should have led to cutbacks in investment (or at least a slowing

[4]We kindly thank Timothy Simcoe for making a STATA implementation of this procedure available on his website at `http://people.bu.edu/tsimcoe/data.html`.

of growth relative to trend), something that did appear to have happened (see Figure 15). In fact, R&D investments grew at a rate greater than we might have expected given firms' estimated responsiveness to the NIH and consumers' demand. Later, we hypothesize reasons underlying this discrepancy, including a decline in abnormal industry profits and a feature of our demand measure that does not take the rising value of health into account[5]. Still, the variation we can explore is important because it speaks directly to real R&D dynamics induced by market-wide forces, regardless of changes to the industry-specific environment. Thus, although this is a study of pharmaceutical productivity, we are examining variation likely the most generalizable to other industries.

4.4. Data Construction

This section describes our main data sources and how they were utilized to construct our four key variables: private pharmaceutical R&D, exogenous potential demand (per Acemoglu and Linn (2004)), NIH investments in basic science, and New Molecular Entities approved by the FDA; as well as how we determined the structure of our sixteen therapeutic categories.

4.4.1. Drug Categories & FDA Approvals

Determining the specificity of drug categories for our empirical analysis is a tradeoff between the ability to control for important (more delineated categories) versus allowing for spillovers given the idiosyncratic nature of R&D (fewer broad categories). Although private R&D investments are only available for eight drug categories as detailed below, the demand and NIH data can more easily be decomposed into more specific categories.

Using the Anatomical Therapeutic Class (ATC) categorization scheme as a guide, we matched the eight industry R&D categories to the most obviously corresponding sixteen subgroups of the ATC hierarchy as shown in Table 14[6]. The real outcome we are concerned with is

[5]The potential demand measure holds the income elasticity of demand for new spending and new technology fixed at one. There is good evidence that this elasticity is in fact both greater than 1 and possibly increasing over time (Hall and Jones 2007). In fact, the estimates considered by Hall and Jones (2007) are large enough to plausibly explain the full gap in R&D investment we observe.

[6]Average cost estimates require that we aggregate mean inputs to the eight categories. The main results and directions of effects carry through if we conduct all analyses at the eight category only.

Table 14: Drug Category Crosswalk

Private R&D Category	**Matched ATC Category**
Anti-Infective	1. Anti-infective, non-viral [J01-J04]
	2. Anti-viral [J05]
	3. Parasitic [P]
Biological / Immunological	1. Biological / Immunological [L02-L04]
Cancer / Endocrine / Metabolism	1. Cancer [L01]
	2. Diabetes / Obesity [A08, A10]
	3. Hormonal [H]
	4. Endocrine, reproductive [G03]
Cardiovascular	1. Blood [B]
	2. Cardiovascular, non-blood-specific [C]
Central Nervous System / Eye	1. CNS [N]
	2. Eye [S01]
Gastrointestinal / Genitourinary	1. Gastrointestinal [A01-A07, A09, A11-A15]
	2. Kidney / Gynecological / Urological [G01, G02, G04]
Respiratory	1. Respiratory [R]
Dermatological	1. Dermatological / Musculoskeletal [D, M]

Note: Private R&D categories are the most disaggregated levels of research and development investments reported in historical PhRMA reports. Anatomical Therapeutic Chemical Classification System (ATC) categories are used to classify drugs in the drug approval and utilization (demand) data.

welfare changes as a result of new products released by the pharmaceutical industry. Because data necessary to calculate welfare is rarely available (i.e. drug-specific revenues and quality-adjusted life years generated), it has become common practice to evaluate the count of new drugs approved. Obviously a raw count of all drug approvals would place the same weight on all drugs whether they are revolutionary therapies such as statins or reformulations of age-old drugs such as aspirin. To alleviate some of this discrepancy it has also become common practice to restrict attention to approvals at the FDA that receive new molecular entity (NME) status. Newly approved drugs receive the NME status if the active moiety has not yet been approved by the FDA, thus providing a strong indication that the therapy has some potential to provide significant welfare improvements.

The count of annual approvals of new molecular entities (NMEs) from 1987 to 2013 was constructed from the Drugs@FDA database[7]. The database does provide information as to the sponsoring firm of each drug approval; however, it is prohibitive to track the "ownership" of any given drug over time given the prevalence of firm- and drug-level acquisitions, as well as licensing and manufacturing agreements. Thus, we do not restrict our sample to any set of firm sponsors. Drugs were matched to each therapeutic category per their primary ATC codes.

4.4.2. Private R&D Expenditures

Estimates of industry-wide investments in pharmaceutical R&D from 1970 to 2013 were constructed based on annual reports from the Pharmaceutical Research and Manufacturers of American (PhRMA), the industry's lead trade group. PhRMA conducts annual surveys of its member companies and reports summarized results for a number of relevant statistics. Using historical reports[8], we obtained PhRMA-wide estimates of: total industry R&D, the share of R&D allocated to new/innovative product lines (e.g. NMEs), and the share of R&D allocated across eight major therapeutic categories as outlined in Table 14.

[7] Available at `http://www.fda.gov/Drugs/InformationOnDrugs/ucm079750.htm`

[8] Reports prior to 2002 are not publicly available, but PhRMA representatives kindly shared copies of annual reports from 1990 to 2001 containing data from as early as 1972.

The share of R&D allocated to NME-type research is not decomposed by therapeutic categories and is reported in only a select number of annual reports 1998-2000. The average reported allocation is 78.3% and we choose to make the assumption of 80% for the entirety of our sample period and across each category.

The largest caveat to this data is that therapeutic category-specific investments are not available post-2000. Hence the projection methodology described in the proceeding section. Additionally, because PhRMA underrepresents the industry as a whole we will likely underestimate real industry-wide investments and therefore also underestimate the average cost per NME. So long as any level of misrepresentation is fixed over time, it will not bias our estimates of marginal productivity.

Our final estimate of industry investment must account for the notoriously long development times in this industry- investments in any given year may be related to drugs anywhere from 1 to 15 years away from approval. Based on an average development time of ten to twelve years and an average approval time of at least one year, we sum within-category expenditures from years $t-11$ to $t-1$ to generate our preferred measure. All private investments are deflated using the Biomedical Research and Development Price Index to account for inflation unique to this sector.

4.4.3. Exogenous Potential Demand

In their initial analyses of market size and innovation in the pharmaceutical sector, Acemoglu and Linn (2004) develop a plausibly exogenous measure market size they term "potential demand". This measure utilizes demographic trends to remove the influence of innovation on demand (consumers buying more new products because they are valued) in order to separately identify the influence of demand on innovation (firms developing new products because they expect them to be valued by consumers). The exogenous expected demand D

for drug category j at time t is given by

$$D_{jt} = \sum_{a} I_{at} \times S_{aj} \tag{4.4}$$

where a is a set of 5-year age bins, I_{at} is the aggregate national income of individuals in age bin a at time t, and S_{aj} is the average share of group a's income spent on drug category j.

Acemoglu and Linn (2004) provide evidence that within-group drug expenditure shares are relatively constant over time while the country's demographics are not. Intuitively, illnesses and the medications used to treat them often effect humans at certain ages (e.g. very few people under 45 take statins, but this share increases dramatically with age), so as the income of individuals of certain ages grows, so does the demand for drugs that differentially afflict their age group. Thus, by fixing the expenditure share and only allowing variation in demographics to cause changes in demand, the measure provides a plausibly exogenous measure.

Following Acemoglu and Linn (2004), the income component I_{at} is constructed from the Current Population Survey (CPS) March supplement and the expenditure component S_{aj} is constructed from the Prescribed Medicine Files of the Medical Expenditure Panel Survey (MEPS). Drug categories are determined based on matching drug names reported in MEPS to their Anatomic Therapeutic Class (ATC) codes[9] outlined in Figure 2. ATC codes are structured based on a combination of a drug's mechanisms and the illnesses it is approved to treat. The MEPS data is only available from 1996 to 2013, therefore we extend the S_{aj} measure calculated during this timeframe to our full timeframe of 1985 to 2013. CPS data is available for the full length of our study with all data deflated using the CPI.

4.4.4. NIH Investments in Science

Data on NIH extramural grant awards from 1965 to 2013 was constructed based on data available from the NIH ExPORTER data files as well as a Freedom of Information Act

[9] See `http://www.whocc.no/atc_ddd_index/` for more information.

request for data prior to 2000 that is not readily available in the ExPORTER files[10].

In line with prior research on the allocation of NIH funds (Toole 2012) we employ a keyword based approach to categorizing grants to each of the 16 ATC classes, and further delineate whether or not the grant is obviously "applied" research such as clinical trials, as opposed to basic research aimed at generating new knowledge. By definition, truly "basic" research has the potential to influence all of our therapeutic categories; however, the NIH requires that any grant have some potentially relevant public health implication. So when we refer to "basic" research, we are focusing on research at the NIH that lacks an immediate means of commercialization. This delineation is possible based on keywords as well as examining the grant "Activity Code", which is unique to certain types of grant programs intended to stimulate more applied research, such as the Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programs. This applied research accounted for roughly 15% of NIH spending annually. However, these investments were never found to have a significant relationship with either firm R&D or drug approvals, so all results below are based on the basic science-only measure of NIH investments.

Much like firm investments, we can imagine that NIH investments in research may take time to influence firm investment decisions or new drug approvals directly. Toole (2012), who used the same data but only the 8 major categories identified the most statistically significant lag-flow construction to be the sum of NIH investments 17 to 24 years prior. Using our sixteen categories, we identify significant correlations between lagged flows anywhere in timeframe of 10 to 20 years prior, and based on measures of fit and statistical significance including t-statistics, log-likelihood and Wald chi-square statistic, we identify the flow of NIH spending 15 to 20 years prior as the best predictor of our outcomes of interest. All NIH investments are deflated using the Biomedical R&D Price Index, which accounts for cost inflation unique to basic biomedical research.

[10] Available at `http://exporter.nih.gov`. Prior to 2000, the total funding amount for each project is absent in the ExPORTER files, hence the FOIA request was necessary.

4.5. Results

4.5.1. Main Findings

Table 15 explores very simple production models where NMEs are the outputs and private and public investment are the lagged inputs. The models are estimating using data from the timeframe for which disaggregated private spending is available, 1985-2000. Independently, and ignoring industry-wide trends, both measures of spending are significantly correlated with the approval of NMEs. At the sample means of private R&D and NIH spending flows and assuming a constant marginal effect, the coefficients from columns (1) and (2) of Table 15 suggest an average marginal cost per NME of $273 million and $976 million (in real 2012 dollars) for private R&D and NIH basic science spending, respectively. The private cost per NME is very close to accounting-based studies, which estimated an average cost between $200-400 million (DiMasi et al. 1991; DiMasi et al. 2003).

Table 15: NME Production Per Private R&D and NIH Science, 1985-2000

	(1)	(2)	(3)	(4)	(5)	(6)
Private R&D$_{(t-11,t-1)}$	0.596***		0.223	1.069***		0.763***
	(0.163)		(0.265)	(0.246)		(0.222)
NIH$_{(t-20,t-15)}$		0.949***	0.828*		0.812**	0.759
		(0.300)	(0.476)		(0.395)	(0.477)
Time$_{(1989-1992)}$				0.0305	0.114	-0.0258
				(0.134)	(0.150)	(0.142)
Time$_{(1993-1996)}$				-0.171	0.167	-0.256
				(0.234)	(0.214)	(0.228)
Time$_{(1997-2000)}$				-0.514	0.0941	-0.564
				(0.403)	(0.320)	(0.381)
N	224	224	224	224	224	224

Note: Standard errors in parentheses, clustered at the drug-category level. Fixed effect (drug category) conditional Poisson specifications. Time dummies are relative to the earliest time period, 1985-1988. All continuous independent variables are log-transformed.* p<0.10, ** p<0.05, *** p<0.01.

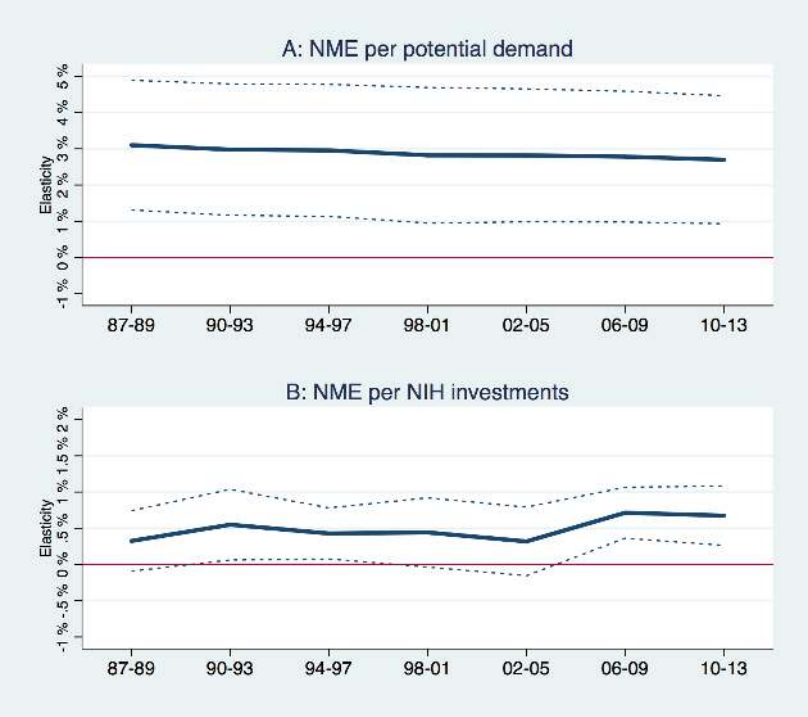
This first pass at the cost to the NIH per NME identifies an estimate modestly larger than

prior work, such as Toole (2012) who identified an average marginal cost of $741 million using an unconditional (no category fixed effects) Poisson model that treated both private and NIH spending as exogenous in the same specification. What is perhaps noteworthy from a policy viewpoint is how large this public investment cost per new drug is relative to private investment, and how large it makes the total cost of bringing a new drug to market, with NIH spending swamping the magnitude of private spending over which there has been so much controversy.

As indicated in Section 2, we assume the two major exogenous inputs to the NME production function to be the potential for demand driven by demographic trends and investments in science by the NIH. If these inputs do indeed stimulate private investment in R&D, or directly result in any NME discoveries (i.e. NIH only), then they should be significantly correlated with NME approvals. Furthermore, we are interested specifically in whether or not these correlations changed significantly at all over the past two and a half decades. If the pharmaceutical industry experienced a significant decline in their productivity during the cost increases post-2000, then this would manifest in significantly lower NME output per changes in demand or NIH investments as it implies the industry was no longer able to convert valuable ideas into new drugs. Figure 4 plots the coefficients of NME counts regressed on these two inputs and uses interactions with the time dummy variables in order to present the trends in NME productivity per potential demand and NIH investments. It is immediately obvious that there is no significant change in the responsiveness of the industry to opportunities created by changes in either consumer demand or the supply of basic science. The estimated magnitudes of the elasticity of NME approvals with respect to changes in potential demand are very similar to the coefficients estimated in Acemoglu and Linn (2004) initial development of that measure. A number of related studies have investigated the elasticity of pharmaceutical innovations with respect to changes in demand using a variety of methods to identify exogenous changes in demand and measures of innovation. Our results and estimated magnitudes fit squarely in line with this previous work[11].

[11] Finkelstein (2004) explores a policy change related to the expected value of certain vaccines and finds

Figure 18: Trends in NME output per Potential Demand and NIH Basic Science, 1987-2013



Note: Potential demand is per the exogenous market size measure and NIH investments are a sum of lagged investments in basic science; both measures are described in Section 4. Coefficients are plotted from the same conditional Poisson model of the count of NMEs regressed on drug-class specific demand and NIH measures interacted with each time dummy variable. Non-interacted time dummies are also included in the model to capture variation across all drug classes over time.

The elasticities for NIH investments are at times not significantly different from zero; however, taken altogether the trend indicates a relatively stable relationship that, if anything, has improved in the past decade. This finding in the aggregate that NIH investments gen-

a 1% increase in expected market size stimulates a 2.5% increase in the number of clinical trials for affected diseases. Duggan and Morton (2010) and Blume-Kohout and Sood (2013) utilize the enactment of Medicare Part D as a plausibly exogenous shock to consumers' willingness to pay for pharmaceuticals and also find corresponding increases in clinical trials for drug categories expected to grow the largest following Part D's enactment. In closely related work, Dubois et al. (2015) utilize detailed global revenue data and instrumental variables approach to estimate this elasticity. Directly comparing their estimates with ours is not straightforward given the different data sources and measures of demand; however, we can get a sense of consistency by comparing the estimating change in demand necessary to induce a new NME. At the sample means, our estimates imply that roughly $1.2 billion is necessary, with Dubois et al. (2015) finding that $2.5 billion is necessary, despite their elasticity estimate being an order of magnitude less than ours (0.25 versus 3.0). We consider these estimates to be relatively consistent given the differences in sample periods and the fact that we lack data on worldwide revenues.

erate valuable downstream outcomes has received support from Azoulay et al. (2015c) who use more fine-grain data and careful econometric techniques.

Table 16: Private R&D Investments per NIH and Potential Demand, 1985-2000

	(1)	(2)	(3)
Potential Demand$_t$	1.556***		1.366***
	(0.221)		(0.210)
NIH$_{(t-20,t-15)}$		0.361***	0.311***
		(0.0901)	(0.0914)
Time$_{(1989-1992)}$	0.201***	0.245***	0.157***
	(0.0230)	(0.0347)	(0.0351)
Time$_{(1993-1996)}$	0.454***	0.563***	0.360***
	(0.0306)	(0.0556)	(0.0565)
Time$_{(1997-2000)}$	0.487***	0.836***	0.401***
	(0.0523)	(0.0825)	(0.0806)
N	224	224	224
adj. R^2	0.883	0.881	0.899

Note: Standard errors in parentheses, clustered at the drug-category level. Fixed effect (drug category) OLS specifications. Time dummies are relative to the earliest time period, 1985-1988. All continuous variables are log-transformed. * p<0.10, ** p<0.05, *** p<0.01.

These results also provide support for the first stage of our model - that changes in these inputs are the major forces spurring industry investment. To explore this relationship specifically, Table 16 presents the results of estimating equation (2) using OLS with drug-class fixed effects and time dummies. Because this disaggregated data on private R&D is only available pre-2001 we can only estimate the equation for this portion of our timeframe. Within this period, we find evidence that our measures of demand and NIH investments are both significantly correlated with investments by the pharmaceutical industry.

The magnitudes of the coefficients suggest that for each 10% increase in the potential market size of a drug-class for a given year, the industry will have invested 14% more into R&D. Likewise, for each 10% increase in NIH investments in the years prior, the industry will invest 3% more into R&D. Relatively speaking, it appears that the pharmaceutical industry is significantly more responsive to changes in demand compared to changes in the supply of

science. However, as noted earlier, the magnitude of NIH investments relative to the market sizes is much larger - directly comparing these relative elasticities should be done so with this in mind.

The time dummies in these models also indicate that aside from our two measures, there also appears to have been a significant increase in R&D investments over this timeframe with roughly 40% larger investments observed in the late 1990's compared to the late 1980's, conditional on our inputs and fixed differences across the drug classes. And notably, these models explain upwards of 90% of the variation in R&D investments within each drug class, supporting our assumption that these are likely the major drivers of R&D investment decisions in this industry.

Given the stability and significance of our model of private R&D determinants, we use the estimates from this first stage in order to project out predicting spending post-2000 under the assumption that the relationships identified in Table 16 column (3) persist. The result of these linear projections are aggregated to industry total annual investments and compared to the observed data (see Figure 15). Our model appears to predict R&D investments reasonably well in the post-2000 period where we observe only inputs and not actual category-specific investments. But notably, our model regularly underpredicts investments in this period on the order of 20-30%. In a later section, we speculate that this may be due to entry drawing down above normal economic profits.

Table 17 presents the results from re-estimating the models presented in Table 15 where private R&D was assumed endogenous and the data was limited to pre-2001. Comparing the coefficients on our predicted R&D measure to the coefficients on the time dummies allows us to examine whether any changes in productivity appeared to occur for marginal or average investments, respectively. Our predicted R&D measure is significantly correlated with NME output, and per column (3) this relationship did not appear to change significantly during the so-called crisis of the 2000's.

Table 17: NME Production with Endogenous Investments, 1985-2013

	(1)	(2)	(3)	(4)
Pred. R&D$_{(t-11,t-1)}$	1.796***		1.799***	1.963***
	(0.549)		(0.539)	(0.747)
NIH Basic$_{(t-20,t-15)}$		0.459**		-0.0869
		(0.223)		(0.331)
Pred. R&D$_{(t-11,t-1)}$			-0.0191	
$\times$ 2001-on			(0.232)	
Time$_{89-92}$	-0.0962	0.183	-0.0970	-0.111
	(0.164)	(0.131)	(0.163)	(0.162)
Time$_{93-96}$	-0.237	0.304	-0.239	-0.268
	(0.276)	(0.189)	(0.272)	(0.274)
Time$_{97-00}$	-0.701	0.263	-0.704	-0.768
	(0.479)	(0.298)	(0.473)	(0.478)
Time$_{01-04}$	-1.183**	-0.00462	-0.742	-1.264**
	(0.520)	(0.278)	(5.523)	(0.523)
Time$_{05-08}$	-1.699***	-0.312	-1.255	-1.793***
	(0.517)	(0.268)	(5.553)	(0.528)
Time$_{09-13}$	-1.550***	-0.0418	-1.105	-1.652***
	(0.587)	(0.338)	(5.550)	(0.595)
N	464	464	464	464
Period	Full	Full	Full	Full

Note: Standard errors in parentheses, clustered at the drug-category level. Fixed effect (drug category) conditional Poisson specifications. Time dummies are relative to the earliest time period, 1985-1988. All independent continuous variables are log-transformed. Predicted R&D is the linear projection of private R&D investments based on the correlations from Table 16, Column (3), without time-period dummies since they are not identified in the post-2000 timeframe. * p<0.10, ** p<0.05, *** p<0.01.

Rather, it appears that the significant declines in NME output occurred for the average idea as the time dummies for NME outputs post-2000 are significantly less than zero. In other words, the estimates indicate that after 2000, across all therapeutic categories, the average rate of NME approvals was roughly 150% lower than before.

Certainly we are not the first to identify this decline in approvals conditional on observed investments. However, the important point and major contribution of this Chapter is that these results imply the declines were amongst the average drug discovery project and not the marginal project. As argued in Section 2, this distinction has important policy implications. Our results do not point to any significant difference in this industries' ability to turn the next (marginal) idea into a valuable drug, relative to what was observed in decades prior. Rather, it appears that the average expected costs of operating in this industry are what increased.

Referring back to Figure 2, this indicates that the distribution of ideas is relatively linear and that as forces of supply (NIH science) and demand (demographic trends) increase the value of the marginal drug idea, firms are willing to incur higher average costs in order to continue to produce profits. While we cannot directly connect our results to changes in social welfare, this result implies that, should policymakers seek to increase the rate at which new drugs are introduced, policies should focus on decreasing the fixed costs common to all new drug ideas. These costs might include efforts such as establishing clinical trial networks across the country (or globe), or the ability to access new results from publicly funded biomedical research.

4.5.2. Explaining the Increase in Investments beyond the NIH and Demand

Recall from Figure 15 that, while our model of R&D was able to relatively accurately predict the trend in investments, it appears we under-predict investments for the majority of years post-2000, typically on the order of 20-30%. As argued previously, this "overinvestment" (relative to the expected response to NIH investments and shifts in demand) stands in

contrast to a growth supply-side frictions that prior literature has frequently suggested[12]. If firms were truly faced with new problems in their governance, or regulations became suddenly more burdensome, this should have been met with a slower growth in R&D costs (relative to pre-2000) as markets grew. In contrast, our regressions indicate that the marginal idea (as proxied by marginal R&D spending) resulted in the same number of NMEs in expectation after 2000 compared to before (See Table 16, Column 3). Likewise, Figure 4 indicates a very stable relationship between changes in the demand of new drugs and the rate of new drug approvals.

Instead we propose two alternatives[13]: (1) the growing value of health, and (2) competing away of supranormal profits. To the first point, while the Acemoglu and Linn (2004) measure of potential demand is empirically useful in that it identifies plausibly exogenous shifts in demand, it assumes that the elasticity of health spending with respect to income is fixed at one. Thus, the value of health to each individual is held constant. Hall and Jones (2007) present and solve a detailed model of health investment and production which, upon solving numerically, implies an elasticity significantly greater than 1[14]. These magnitudes imply that our measure underestimates changes in demand at times of income growth, hence the underpredictions of our full model. However, micro-level evidence has yet to provide robust support for elasticities much greater than one (Acemoglu et al. 2013), and so we are reluctant to propose this mechanism as being fully responsible for our finding.

Our second proposed mechanism is based on the notion that, in industries such as this, the enormous barriers to entry will prevent new firms from immediately competing away excess

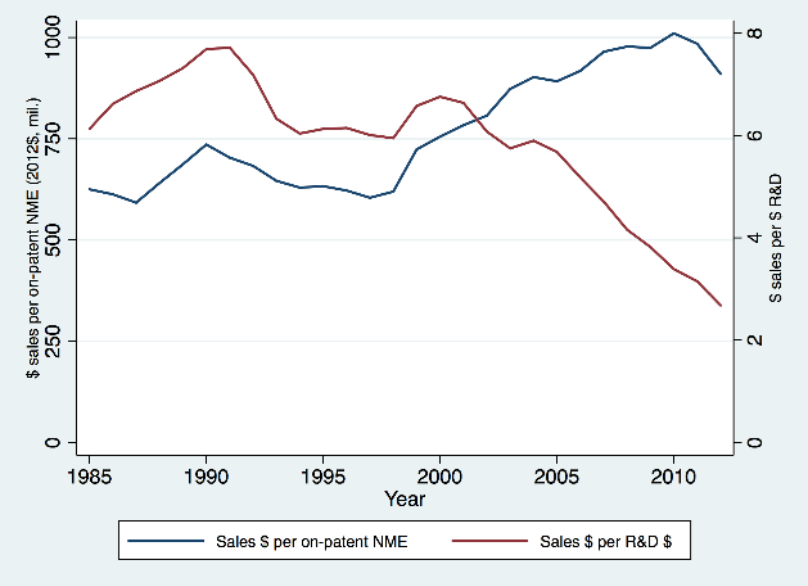
[12]We are agnostic to stable frictions that may exist on the supply-side in this market, such as the "short-termism" of managers discussed by Budish et al. (2015) or the effect of regulatory requirements at the FDA. We are only concerned with whether *changes* in the effects of these frictions might be responsible for the growth in R&D costs observed.

[13]The introduction of Medicare Part D in 2003 should also be noted. The demand measure does not perfectly capture the portion of growth in potential market size driven by this major policy. In order to maintain exogeneity, it holds age-category by drug-category expenditure shares (S_{jt} in eq. 1) fixed based on the sample average which includes seven years prior to Part D's introduction. Thus, therapeutic categories that experience the largest growth in utilization by Part D eligible individuals likely underestimate true growth in demand post-2003.

[14]Depending on their method of solving for the elasticity and the assumptions involved, Hall and Jones' (2007) point estimates range between 1.5 and 2.

returns (relative to other industries). The appearance of supranormal profits being a key feature of this industry prior to 2000 has been highlighted previously, and it was hypothesized that new entry should eventually bid these returns down to levels seen in other industries (Scherer 2001; Danzon 2011). In order to explore trends in industry profits, Figure 5 plots two measures of profitability using PhRMA reported sales and R&D investment data: (1) annual revenues per "on-patent" NMEs, and (2) annual revenues per prior R&D investments. The first measure speaks to the value of new drugs. We define "on-patent" NMEs as those

Figure 19: Trends in Revenues per Drugs and R&D Expenditures, 1985-2012



Note: Sales are US-based revenues. On-patent NME refers to New Molecular Entities approved in the 10 years prior. R&D investments are lagged 10 years relative to the reported year, and are capitalized using a 10% annual interest rate. Based on data in PhRMA Annual Reports and FDA new drug approvals.

approved by the FDA in the ten years prior to the focal year, given the average length of patent protection post FDA approval[15]. Prior to 2000, the average revenues per on-patent NME was relatively stable at roughly $625 million per year per drug. A noticeable increase in the value of these drugs is obvious, beginning around 2000 and continuing to our most

[15]This simplifying definition is useful because connecting individual patents to individual molecules is not a straightforward exercise as multiple patents may cover a single "drug", which may also hold different market exclusivities (either via a patent of FDA policy) for different indications.

recent data, reaching nearly $1 billion per year per drug.

Conversely, the trend in revenues per R&D investments - a measure indicative of overall industry profitability - suggests a significant decline in overall returns. Prior to 2000 the data indicates that each dollar of R&D was generating upwards of six dollars in future revenues, even after accounting for the capitalization of costs during the long development times (assumed here to be 10 years on average). Certainly our measure of costs is not perfectly comprehensive: only "R&D" costs are reported in the PhRMA data. Still, this magnitude of returns to capital, even under generous assumptions about unmeasured costs[16], is objectively large.

But after 2000, the return to R&D has essentially halved while the average returns for each new drug has increased by roughly 30%. This pattern of lower overall returns even amidst increasingly valuable (based on revenue per product) products is in line with the notion competitive pressures decreased profit margins even as the demand for new drugs continued to grow. A full-fledged analysis of competition in this industry is beyond the scope of this Chapter, so we only offer some speculations as to why these returns were so excessive in earlier periods. Many observers have noted the growth in entrepreneurial firms engaging in biomedical research, which is likely a function of decreasing fixed costs of R&D prior to clinical trials. Others have argued that the financial arrangements unique to the healthcare sector have given rise to a "medical innovation premium" (Koijen et al. 2016), which may be declining over time as the uncertainties surrounding new drug production become resolved in the long-run.

[16]Data on industry-wide non-R&D expenditures is not readily available. Arguably, aside from R&D the industry's largest expenditures may be in the form of physician-based payments via detailing or other interactions. Reports from the Open Payments function of the Center for Medicare and Medicaid services indicate that in between 2013 and 2015, roughly $3 to $7 billion was paid to physicians in some form (e.g. meals, travel, education) annually (Source: `openpaymentsdata.cms.gov`). Relative to approximately $50 billion in annual R&D investments, these additional costs would not substantially alter the magnitude of returns presented in Figure 5.

4.6. Conclusion

This Chapter introduces an alternative perspective on the productivity changes observed in the pharmaceutical industry over the past twenty years. The traditional view has taken firm R&D investments as given, and speculated about what supply-side frictions have arisen to cause the declines in productivity - drugs per R&D dollar - observed. In contrast, we take the amount invested (and expectations of success and profits) as endogenous, and suggest why firms would pursue less productive investments in a world of fixed resources (i.e. new drug ideas). We attribute this pursuit in large part to a documented growth in the demand for new products and fluctuations in NIH spending that increased the returns to the marginally profitable drug idea. Much like the farmers studied by Ricardo, pharmaceutical firms appeared willing to take on increased costs given expectations of increased demand - productivity declines were a profitable choice, not a worrisome problem.

We motivate our analyses with a simplified production function that accounts for this endogeneity of R&D investments. Although, the data requires that we analyze relatively aggregate categories of investments across therapeutic classes. In particular, our estimates imply that the marginal value of new drug ideas has not changed significantly over the past thirty years, while the average value of these ideas did decline substantially post-2000. These results suggest that policies intent on stimulating further innovation in this industry may be most effective when designed to lower the fixed costs of R&D shared by all active R&D projects, such as establishing clinical trial networks, as opposed to subsidizing marginal expenditures with tax credits.

Examining trends in revenues and costs more generally, we also speculate that industry-wide profitability was above normal prior to 2000 - a feature documented previously (Danzon 2011) - so new investment at lower (but still positive) marginal profit was the result of entry (eventually) bidding down returns. We think this view places less emphasis on the idea that there is something on the supply side that needs to be "fixed" (by changes in firm strategies, trial design, or new institutes of health) and more emphasis on a natural economic model

of investment where transitions between equilibria, during which profits are supranormal, takes time. Looking forward, if the demand for new drugs continues to grow, and there is cause to think it will (a la Hall and Jones 2007), and there is not a major boost to NIH investments, we expect continued declines in the number of new drugs per R&D dollar, but do not see this a case for major policy interventions or concern.

BIBLIOGRAPHY

D. Acemoglu and J. Linn. Market size in innovation: Theory and evidence from the pharmaceutical industry. *The Quarterly journal of economics*, 119(3):1049–1090, 2004.

D. Acemoglu, A. Finkelstein, and M. J. Notowidigdo. Income and health spending: Evidence from oil price shocks. *Review of Economics and Statistics*, 95(4):1079–1095, 2013.

P. Aghion and J. Tirole. Formal and real authority in organizations. *Journal of political economy*, pages 1–29, 1997.

P. Aghion, M. Dewatripont, and J. C. Stein. Academic freedom, private-sector focus, and the process of innovation. *The RAND Journal of Economics*, 39(3):617–635, 2008.

K. Arrow. Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, pages 609–626. Princeton University Press, 1962.

P. Azoulay, J. S. Graff Zivin, and G. Manso. Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3):527–554, 2011.

P. Azoulay, C. Fons-Rosen, and J. S. G. Zivin. Does science advance one funeral at a time? Technical report, National Bureau of Economic Research, 2015a.

P. Azoulay, J. L. Furman, J. L. Krieger, and F. Murray. Retractions. *Review of Economics and Statistics*, 97(5):1118–1136, 2015b.

P. Azoulay, J. S. G. Zivin, D. Li, and B. N. Sampat. Public r&d investments and private-sector patenting: evidence from nih funding rules. Technical report, National Bureau of Economic Research, 2015c.

P. Azoulay, I. Ganguli, and J. S. G. Zivin. The mobility of elite life scientists: Professional and personal determinants. Technical report, National Bureau of Economic Research, 2016.

P. Bajari, H. Hong, J. Krainer, and D. Nekipelov. Estimating static models of strategic interactions. *Journal of Business & Economic Statistics*, 28(4):469–482, 2010.

G. S. Becker. Investment in human capital: A theoretical analysis. *The journal of political economy*, pages 9–49, 1962.

S. Berry and P. Reiss. Empirical models of entry and market structure. *Handbook of industrial organization*, 3:1845–1886, 2007.

A. Blasco, O. S. Jung, K. R. Lakhani, and M. Menietti. Motivating effort in contributing to public goods inside organizations: Field experimental evidence. Technical report, National Bureau of Economic Research, 2016.

M. E. Blume-Kohout and N. Sood. Market size and innovation: Effects of medicare part d on pharmaceutical research and development. *Journal of public economics*, 97:327–336, 2013.

R. Blundell, R. Griffith, and F. Windmeijer. Individual effects and dynamics in count data models. *Journal of econometrics*, 108(1):113–131, 2002.

G. Borjas. Self-selection and the earnings of immigrants. *American Economic Review*, 77 (4):531–53, 1987.

K. J. Boudreau, N. Lacetera, and K. R. Lakhani. Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science*, 57(5):843–863, 2011.

K. J. Boudreau, E. C. Guinan, K. R. Lakhani, and C. Riedl. Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science*, 2016a.

K. J. Boudreau, K. R. Lakhani, and M. Menietti. Performance responses to competition across skill levels in rank-order tournaments: field evidence and implications for tournament design. *The RAND Journal of Economics*, 47(1):140–165, 2016b.

J. L. Bower. Managing the resource allocation process: A study of corporate planning and investment. 1970.

L. Brunt, J. Lerner, and T. Nicholas. Inducement prizes and innovation. *The Journal of Industrial Economics*, 60(4):657–696, 2012.

K. A. Bryan and J. Lemus. The direction of innovation. *Mimeo*, 2016.

E. Budish, B. N. Roin, and H. Williams. Do firms underinvest in long-term research? evidence from cancer clinical trials. *The American economic review*, 105(7):2044–2085, 2015.

C. Catalini. Microgeography and the direction of innovative activity. Technical report, Working Paper, 2015.

Y.-K. Che and I. Gale. Optimal design of research contests. *The American Economic Review*, 93(3):646–671, 2003.

F. S. Collins. Reengineering translational science: the time is right. *Science translational medicine*, 3(90):90cm17–90cm17, 2011.

P. M. Danzon. The economics of the biopharmaceutical industry. 2011.

P. Dasgupta and P. A. David. Toward a new economics of science. *Research policy*, 23(5): 487–521, 1994.

R. S. Decker, L. Wimsatt, A. G. Trice, and J. A. Konstan. A profile of federal-grant administrative burden among federal demonstration partnership faculty. Technical report, Federal Demonstration Partnership, 2007.

W. Dessein. Authority and communication in organizations. *The Review of Economic Studies*, 69(4):811–838, 2002.

J. A. DiMasi, R. W. Hansen, H. G. Grabowski, and L. Lasagna. Cost of innovation in the pharmaceutical industry. *Journal of health economics*, 10(2):107–142, 1991.

J. A. DiMasi, R. W. Hansen, and H. G. Grabowski. The price of innovation: new estimates of drug development costs. *Journal of health economics*, 22(2):151–185, 2003.

G. Dosi. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3):147–162, 1982.

P. Dubois, O. De Mouzon, F. Scott-Morton, and P. Seabright. Market size and pharmaceutical innovation. *The RAND Journal of Economics*, 46(4):844–871, 2015.

M. Duggan and F. S. Morton. The effect of medicare part d on pharmaceutical prices and utilization. *The American Economic Review*, 100(1):590–607, 2010.

A. Finkelstein. Static and dynamic effects of health policy: Evidence from the vaccine industry. *The Quarterly Journal of Economics*, 119(2):527–564, 2004.

L. Fleming and O. Sorenson. Technology as a complex adaptive system: evidence from patent data. *Research policy*, 30(7):1019–1039, 2001.

R. Freeman and J. Van Reenen. What if congress doubled r&d spending on the physical sciences? In *Innovation Policy and the Economy, Volume 9*, pages 1–38. University of Chicago Press, 2009.

J. L. Furman and S. Stern. Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *The American Economic Review*, 101(5):1933–1963, 2011.

G. Gavetti and D. Levinthal. Looking forward and looking backward: Cognitive and experiential search. *Administrative science quarterly*, 45(1):113–137, 2000.

T. F. Gieryn and R. F. Hirsh. Marginality and innovation in science. *Social Studies of Science*, 13(1):87–106, 1983.

E. L. Glaeser, A. Hillis, S. D. Kominers, and M. Luca. Predictive cities crowdsourcing city government: Using tournaments to improve inspection accuracy. *The American Economic Review*, 106(5):114–118, 2016.

A. Goolsbee. Does government r&d policy mainly benefit scientists and engineers? *The American Economic Review*, 88(2):298, 1998.

R. E. Hall and C. I. Jones. The value of life and the rise in health spending. *The Quarterly Journal of Economics*, 122(1):39–72, 2007.

D. Hegde and B. Sampat. Can private money buy public science? disease group lobbying and federal funding for biomedical research. *Management Science*, 61(10):2281–2298, 2015.

B. Holmstrom and P. Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7: 24–52, 1991.

S. T. Howell. Financing constraints as barriers to innovation: Evidence from r&d grants to energy startups. *Available at SSRN 2687457*, 2015.

B. A. Jacob and L. Lefgren. The impact of research grant funding on scientific productivity. *Journal of public economics*, 95(9):1168–1177, 2011.

L. B. Jeppesen and K. R. Lakhani. Marginality and problem-solving effectiveness in broadcast search. *Organization science*, 21(5):1016–1033, 2010.

B. F. Jones. The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317, 2009.

R. S. Koijen, T. J. Philipson, and H. Uhlig. Financial health economics. *Econometrica*, 84 (1):195–242, 2016.

I. Kola and J. Landis. Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, 3(8):711–716, 2004.

N. Lacetera. Different missions and commitment power in r&d organizations: Theory and evidence on industry-university alliances. *Organization Science*, 20(3):565–582, 2009.

E. P. Lazear and S. Rosen. Rank-order tournaments as optimum labor contracts. *The Journal of Political Economy*, 89(5):841–864, 1981.

J. Lerner. *The architecture of innovation: The economics of creative organizations*. Harvard Business Press, 2012.

S. G. Levin and P. E. Stephan. Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, pages 114–132, 1991.

D. A. Levinthal. Adaptation on rugged landscapes. *Management science*, 43(7):934–950, 1997.

D. A. Levinthal. The slow pace of rapid technological change: gradualism and punctuation in technological change. *Industrial and corporate change*, 7(2):217–247, 1998.

D. Li. Expertise versus bias in evaluation: Evidence from the nih. *American Economic Journal: Applied Economics*, 2017.

D. Li and L. Agha. Big names or big ideas: Do peer-review panels select the best science proposals? *Science*, 348(6233):434–438, 2015.

F. R. Lichtenberg. The allocation of publicly funded biomedical research. In *Medical care output and productivity*, pages 565–590. University of Chicago Press, 2001.

J. Lin and W. J. Wilbur. Pubmed related articles: a probabilistic topic-based model for content similarity. *BMC bioinformatics*, 8(1):1, 2007.

B. Moldovanu and A. Sela. The optimal allocation of prizes in contests. *American Economic Review*, pages 542–558, 2001.

J. G. Mork, A. Jimeno-Yepes, and A. R. Aronson. The nlm medical text indexer system for indexing biomedical literature. In *BioASQ@CLEF*, 2013.

H. Moses, D. H. Matheson, S. Cairns-Smith, B. P. George, C. Palisch, and E. R. Dorsey. The anatomy of medical research: Us and international comparisons. *Jama*, 313(2):174–189, 2015.

R. R. Nelson. The simple economics of basic scientific research. *The Journal of Political Economy*, pages 297–306, 1959.

R. R. Nelson and S. G. Winter. An evolutionary theory of economic change, 1982.

A. D. Roy. Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2): 135–146, 1951.

H. Sauermann and W. M. Cohen. What makes them tick? employee motives and firm innovation. *Management Science*, 56(12):2134–2153, 2010.

J. W. Scannell, A. Blanckley, H. Boldon, and B. Warrington. Diagnosing the decline in pharmaceutical r&d efficiency. *Nature reviews Drug discovery*, 11(3):191–200, 2012.

F. M. Scherer. The link between gross profitability and pharmaceutical r&d spending. *Health Affairs*, 20(5):216–220, 2001.

J. A. Schumpeter. *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*, volume 55. Transaction publishers, 1934.

J. C. Stein. Internal capital markets and the competition for corporate resources. *The Journal of Finance*, 52(1):111–133, 1997.

S. Stern. Do scientists pay to be scientists? *Management science*, 50(6):835–853, 2004.

T. E. Stuart and W. W. Ding. When do scientists become entrepreneurs? the social structural antecedents of commercial activity in the academic life sciences. *American Journal of Sociology*, 112(1):97–144, 2006.

C. Terwiesch and K. T. Ulrich. *Innovation tournaments: Creating and selecting exceptional opportunities*. Harvard Business Press, 2009.

A. A. Toole. The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry. *Research Policy*, 41(1):1–12, 2012.

V. I. Torvik and N. R. Smalheiser. Author name disambiguation in medline. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(3):11, 2009.

G. Tullock. Efficient rent-seeking. In R. T. J.M. Buchanan and G. Tullock, editors, *Toward a theory of the rent-seeking society*, pages 97–112. College Station: Texas A&M University Press, 1980.

H. Williams. Innovation inducement prizes: Connecting research to policy. *Journal of Policy Analysis and Management*, 31(3):752–776, 2012.

H. L. Williams. Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy*, 121(1):1–27, 2013.

O. E. Williamson. *Markets and hierarchies*. New York: Free Press, 1975.

B. D. Wright. The economics of invention incentives: Patents, prizes, and research contracts. *The American Economic Review*, 73(4):691–707, 1983.