Research Policy 50 (2021) 104332

Contents lists available at ScienceDirect

Research Policy

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

Long-term effects from early exposure to research: Evidence from the NIH "Yellow Berets"☆

Pierre Azoulay [a,b,*], Wesley H. Greenblatt [a], Misty L. Heggeness [c]

[a] MIT Sloan School of Management, 100 Main Street, Cambridge, MA, 02142, United States
[b] National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA, 02138, United States
[c] U.S. Census Bureau, Research and Methodology Directorate, 4600 Silver Hill Road, Suitland, MD, 20746, United States

ARTICLE INFO

JEL classification:
J24
O31
O33

Keywords:
Biomedical workforce
Scientific and technical human capital
Career imprinting
Mentorship
Translational medicine

ABSTRACT

Can a relatively short but intense exposure to frontier research alter the career trajectories of potential innovators? To answer this question, we study the careers and productivity of 3075 medical school graduates who applied to the Associate Training Programs (ATP) of the National Institutes of Health (NIH) during the turbulent period of the Vietnam War, 1965–1975. Carefully selecting on observables, we compare physicians who attended the program to those who passed a first admission screen but were ultimately not selected. We find that program participants were twice as likely to choose a research-focused position after training, and considerably less likely to switch to purely clinical endeavors as their careers unfolded. Over the life cycle, NIH trainees also garnered publications, citations, and grant funding at a much higher rate than synthetic controls, and went on to mentor more trainees who themselves became successful researchers. The direction of their research efforts was durably imprinted by their training experience. In particular, NIH trainees appear to have acquired a distinct "translational" style of biomedical research which became an implicit training model for physician-scientists as ATP alumni came to occupy the commanding heights of academic medicine throughout the United States.

> "[The ATP] did not help [my career], it made it…I followed a pathway that was a combination of hard work, some talent and being in the right place at the right time…None of that would have happened had I not come down here as a Clinical Associate…[I would have] gone to Vietnam for a few years in the Navy, [and then] I would have probably returned to New York Hospital. I would probably be practicing medicine right now on 69th Street and First Avenue. The Clinical Associate program put me on a career track that I am still on".
>
> Anthony Fauci, Director, NIAID
>
> *Oral History (1998)*

1. Introduction

It has become a truism among policy-makers that innovation and technological advances are a key determinant of economic growth (Aghion and Howitt, 1992; Romer, 1990; Solow, 1957). But innovation is fundamentally constrained by the supply of innovators—those individuals whose skills and knowledge put them at risk of bringing forth a useful "new-to-the-world" idea. Innovators are made, rather than simply born, and growth possibilities are shaped by the institutions, incentives, and norms that nudge would-be innovators to receive the training necessary to bring themselves to the frontier. Indeed, over the past century, macro evidence suggests that only by steadily increasing

☆ This project was supported by the National Institute on Aging, USA (grant number R24-AG048059) to the National Bureau of Economic Research. We thank Dr. Michael Gottesman, Dr. Michael Lauer, and Dr. Kay Lund for their steadfast support and encouragement. We are indebted to Dr. Anthony Fauci, Dr. Michael Gottesman, Dr. Richard G. Wyatt, and Dr. John Gallin for reflecting on their experiences in the ATP and encouraging us to investigate the impact of the NIH Intramural Program during the Vietnam War. We would also like to thank Daniel Boehlert, Andrew Breazeale, Delaney Cruickshank, Daphne Hines, Maria Isabel Larenas, Kyle Myers, Ryan Pfirrmann-Powell, Lindsey Raymond, and Ying Zeng for their diligent efforts in sorting, entering, matching, and curating the data. Lastly, this project would not be feasible without the help of Barbara Harkins and Christopher Wanjek from the NIH Office of History. We also thank Scott Stern as well as seminar audiences at Georgia Tech, Wisconsin, and Northwestern for constructive feedback. Any opinions (and all errors) are solely those of the authors and do not represent any official position of the U.S. Census Bureau or NBER.

* Corresponding author at: MIT Sloan School of Management, 100 Main Street, Cambridge, MA, 02142, United States.
E-mail addresses: pazoulay@mit.edu (P. Azoulay), wesley.greenblatt@mit.edu (W.H. Greenblatt), misty.l.heggeness@census.gov (M.L. Heggeness).

https://doi.org/10.1016/j.respol.2021.104332
Received 31 January 2021; Received in revised form 19 July 2021; Accepted 21 July 2021
Available online 3 August 2021

the number of workers engaged in formal R&D activities has a steady growth rate in income per capita been sustained (Jones, 1995).

In the medium run at least, designing institutions that might increase the supply of potential innovators is therefore of crucial policy importance. Yet, severe headwinds frustrate efforts to broaden the innovator pipeline. First, because scientific and engineering training is protracted, individual career choices are often shrouded in uncertainty, both with respect to the monetary payoffs and the direction of human capital investments likely to earn the best labor market returns. Witness, for example, the dismal track record of "manpower analysis" and the perennially flawed predictions of "innovator shortage" (Freeman, 1975; Teitelbaum, 2014). Second, innovative careers are fragile (Milojevic et al., 2018) both because of the winner-take-most aspect of the scientific reward system, and because skills at the frontier depreciate rapidly, leading many initial entrants to abandon the idea sector and reenter the production sector (Deming and Noray, 2020). Third, especially for countries with domestic training capabilities, restrictions on high-skilled immigration can act as a brake on plugging leaks in the innovator pipeline (Kerr, 2018). As a result of these headwinds and the elimination of mandatory retirement in academia in the mid-1990s, the scientific workforce is aging rapidly (Blau and Weinberg, 2017).

Despite the paucity of research examining the allocation of talent to innovative activities, some recent evidence points to an important friction, that of exposure to research during an individual's formative years. In a telling anecdote, pro-footballer turned Math Ph.D. student John Urschel recounts how his athletic prowess was identified and nurtured from a young age, whereas his mathematical talents were left undeveloped until a chance encounter with an inquisitive college instructor (Urschel and Thomas, 2019). More systematically, Bell et al. (2019), using IRS tax records linked to U.S. patent data, provide evidence of a strong association between fathers and sons' propensity to patent in the exact same narrow patent class, a finding most easily explained by early socialization opportunities regarding the feasibility and desirability of a research career.

The existence of exposure effects might at first blush appear surprising, but their potential importance is better appreciated if one remembers that early research careers exhibit both brittleness—in the sense that small negative shocks can shift individuals back to the production sector of the economy (Hill, 2018)—and malleability—in the sense that the flexibility to alter one's research trajectory declines over the life cycle (Higgins, 2005). Together, brittleness and malleability suggest that transient but intense formative experiences in the early career may significantly influence potential innovators' decision to enter the "ideas sector" of the economy, as well as their choice of research trajectory, domain, or methodology.

Despite the empirical plausibility of exposure effects, providing convincing evidence of their existence and magnitude presents seemingly insurmountable challenges. Three necessary ingredients are required. First, one needs to identify a population of "naïve to research" individuals who nonetheless possess much of the human capital required to propel themselves to the research frontier. Second, one requires an intervention consisting of a short but intense exposure to research in a rarefied intellectual environment to a (preferably random) subset of this population. A final requirement is the opportunity to observe these individuals for a long period with minimal loss to follow-up, and see their career unfold.

In this paper, we study an intervention in physician training that comes close to bringing together these three ingredients: The Associate Training Program (ATP) of the National Institutes of Health (NIH). The ATP brought recent MD graduates to the intramural campus of the NIH in Bethesda, Maryland for two to three years to participate in research under the supervision of NIH investigators. A unique aspect of the program is that participation fulfilled a draftee's military service requirement (Berry, 1976). After the war ended, trainees began to refer to themselves ironically as "Yellow Berets", a derogatory term used to contrast draft dodgers with the elite Green Berets—the U.S. Army Special Forces (Baskir and Strauss, 1978; Klein, 1998). Though quite small when the program was founded in 1953, its scale steadily grew with applications dramatically increasing during the years of the Vietnam War. The ATP can be considered a large human capital intervention not because it selected a particularly large cohort (even at its 1973 peak, the program drafted only 229 associates, or approximately 2.5% of graduating male students) but because it induced a very high proportion of eligible participants to actually apply, from around 20% in 1963 to close to 80% in 1971.[1] Though some applicants had prior exposure to biomedical research in medical school or during their undergraduate studies, the unpopularity of the war drove many physicians who otherwise would not have been interested in a research career to apply for one of those coveted positions (Varmus, 2009). This unique confluence of events provides us with a quasi-experimental lever to disentangle the role of sorting from that of training and mentorship, always a vexing challenge in empirical studies of the scientific labor market.

We study the careers and productivity of all 3075 male medical school graduates who applied to the ATP and were interviewed on campus between 1965 and 1975. We build a rich hand-collected dataset containing the complete training and career histories for these individuals, including all publications, patents, NIH grants, and citations. Carefully selecting on observables, we compare physicians who attended the program to those who passed a first admission screen but were ultimately not selected. Despite lasting only two to three years, we find that the ATP had a large and sustained impact on the careers of those who attended. Relative to synthetic control applicants, program participants were twice as likely to sort into research-focused positions, and dramatically less prone to switch to purely clinical endeavors as their careers unfolded. Over the life cycle, NIH trainees also garnered publications, citations, and grant funding at a much higher rate than synthetic controls, with over a 75% higher odds of joining the biomedical research elite.[2] They also mentored more trainees who themselves became successful researchers, providing a way their impact could persist through the training of the next generation. Moreover, the direction of their research efforts was durably imprinted by their training experience. In particular, ATP attendees appear to have acquired a distinct "translational" style of biomedical research which became an implicit training model for physician-scientists as ATP alumni came to occupy the commanding heights of academic medicine throughout the United States (Khot et al., 2011).

In addition to the unique historical importance of the NIH ATP (Klein, 1998), our study sheds light on the forces that shape skill acquisition in medicine, and how medical training influences the rate and direction of medical progress. Much of the training physicians receive in medical school, internship, and residency is fungible between medical care and medical research. Early in their career, physicians invest heavily in human capital, but then typically go on to apply their skills narrowly, for the benefits of their (private) patients. These same skills, however, can be redeployed in research activities, where physician effort also generates social returns. In fact, it has been a long-standing policy goal of the medical elite to steer a larger number of physicians towards research careers (Wyngaarden, 1979). As a result, studying the NIH training programs in the Vietnam War era provides a unique window on the long-term consequences of exogenously shifting a well-defined population from the "production sector" of the economy (i.e., clinical care) to its "ideas sector" (i.e., biomedical research, including bench, clinical, and translational research).

[1] Since records on the total number of applicants in each year have not survived, the first figure comes from a back of the envelope calculation (see footnote 4), whereas the second stems from anecdotal accounts that are plausible, but hard to substantiate empirically.

[2] Defined as receiving the Nobel Prize, being appointed Howard Hughes Medical Institute investigator, being elected to the National Academy of Science/Medicine, or winning an NIH R37 MERIT award.

Our study also speaks to how the institutional environment of scientific training programs shapes their participants' research careers. A limited number of studies have examined how mentorship during or after training (Ginther et al., 2020; Shibayama, 2019), funding level and source (Blume-Kohout and Adhikari, 2016; Broström, 2019; Ginther and Heggeness, 2020; Jacob and Lefgren, 2011), and their interaction with trainee background (Graddy-Reed et al., 2019) may impact the outcomes of pre- and post-doctoral scientific training programs. By studying medical doctors who were pushed to seek research training by a unique confluence of historical events but many of whom ultimately received training in other settings, the NIH ATP provides a lens on how the content of training programs shapes the type and quality of the scientific talent it nurtures.

The rest of the manuscript proceeds as follows. Section 2 provides institutional background on the NIH ATP program, including the procedures used to select the trainees. Section 3 describes our sample construction, provides descriptive statistics, and discusses our econometric approach. Section 4 presents our main results. Section 5 puts the results in context, and discusses their implications for the design of scientific training programs in the twenty-first century.

2. Institutional setting

Relative to other professional or creative endeavors, the scientific labor market is notable for the extent to which, at any given point of time, a handful of research institutions are responsible for training a disproportionate share of the future elite in a field while simultaneously providing an extraordinary environment for breakthrough discoveries. Examples abound from a wide variety of scientific fields. In physics, the Cavendish laboratory was the prime breeding ground of atomic physicists in the first half of the twentieth century (Rhodes, 1986); the Laboratory of Molecular Biology, also located at the University of Cambridge, played a similar role for biomedical research after the second world war (Bynum, 2012; Rubin, 2006). This phenomenon is not limited to the physical sciences. For example, the MIT economics department stands out from those located at other universities in the extent to which it spawned a community of academics who went on to exert a profound influence on the discipline (Svorencík, 2014).

During the period of our study, the intramural campus of the NIH, located in Bethesda, Maryland, was widely recognized as one of the preeminent biomedical research institutions. One aspect setting it apart from other elite institutions, however, was its unique ability to attract recently minted physicians eager to pursue a research career. Due to the confluence of multiple factors—the Doctor Draft, plentiful federal funding, and the opening of a massive clinical research center in 1953—the NIH had probably no equal in the world with respect to the training of "physician-scientists" (Park, 2003). We draw on historical evidence, including a large archive of oral histories curated by the NIH Office of History to describe this setting in more detail, review the genesis and development of the Associate Training Program (ATP), and describe how trainees were selected and trained during this period (see Appendix E for additional details).

2.1. The Associate Training Program

The NIH ATP started in 1953 with about 15 medical graduates to provide research training to physicians (Klein, 1998). Associates would come to Bethesda and do research under the supervision of NIH investigators, usually after completing a portion of their residency training. Two years were typically spent in the program, with the option to extend training an additional year. From the start, the program was focused on turning physicians into independent medical investigators well-grounded in scientific knowledge and methods. The goal was on learning how to do research more than simply doing research itself and on bringing the physicians into close contact with accomplished scientists. In addition to the research, the NIH also hosted a set of after-hours basic sciences courses for program participants that could rival the offerings of major universities. Christian Anfinsen, a Nobel Laureate and NIH investigator during the early years of the program, describes its key features as *"... the importance of having the [associates] work on problems of [their] own choice rather than be 'servants' in the research problems of the preceptor, and the importance of providing the student[s] with some integrated and organized basic knowledge as a foundation that would permit them to do their own integrating of knowledge later"* (Anfinsen, 1963). While the focus was on research, some participants were able to get credit for their time at the NIH towards their required clinical training for board certification.

By the early 1960s, the Associate Training Program had been expanded to include three separate tracks. Clinical associates would divide their time between clinical care at the NIH Clinical Center and laboratory research. Research associates would spend most of their time on research and had limited clinical responsibilities. Staff associates also had training in research administration as well as undertaking clinical or laboratory research.

Oral histories from NIH staff are replete with claims attesting to the cutting edge research, breadth of expertise, and concentration of talent in biomedical research within the confines of the intramural campus that resulted in a rarefied environment (Appendix E). In addition, many ATP fellows came to view the focus on what would later be called translational research as a distinctive element of the approach to research at the NIH. This was no accident. James Shannon, one of the early leaders of the NIH, carefully structured the intramural program to facilitate close cooperation between basic and clinical research (Goldstein and Brown, 1997; Park, 2003). Anthony Fauci, an ATP alumni and prominent HIV/AIDS researcher, recalls, *"What the Clinical Associate Program does is it gives you a very interesting perspective on the relationship between disease and the basic science that you have to study to be able to approach disease... Also the link, as we used to say, between 'the bed and the bench,' you see something at the bedside, you bring it back and ask the question at the bench or you make a discovery at the bench and you go back and apply it to the bedside, that bedside to bench phenomena was really what the Clinical Associates program was all about"* (Fauci, 1998).

Since the NIH, through historical accident, grew out of a laboratory within one of the U.S. Navy Marine Hospitals, ATP applicants applied to the program under the auspices of the U.S. Public Health Service and those selected became commissioned officers. This allowed service with the U.S. Public Health Service to fulfill any military service obligation a physician may have if drafted.[3] The interest in and level of competition for spots in the program increased in proportion to the perceived hardship of military service. The program, however, was highly competitive even before the increased interest during the Vietnam War. Unfortunately, there is no reliable information on the total number of applicants to the program, except in a single year before the start of our information period: 1963. That year 53 of 1464 physician applicants were selected (NIH Office of Research Information, 1963).[4] At its peak, in 1973, the program included 229 associates (Klein, 1998). In contrast, in the year following the 1973 Paris Peace Accords which effectively led to an end to the military draft, the NIH was not able to fill its associateship quota for the year, and by 1976 included only 108 physicians, down over 50% from its peak (Klein, 1998).

[3] Of note, in addition to the NIH, the U.S. Public Health Service had other programs through which physicians could apply to spend two years of service, including at the Center for Disease Control, the Food and Drug Administration, and the Indian Health Service.

[4] In 1963 there were 7265 graduates from US Medical Schools (Association of American Medical Colleges, 2016), an estimated 5.6% of which were female (Snyder, 1993). Using this, we can conclude approximately 21% of eligible male medical students actually applied to the NIH ATP in 1963.

While certainly some of the physicians would have applied to and attended the program regardless of the war, avoiding the draft was a significant motivation. Donald Fredrickson, a former director of the NIH and one of the first clinical associates in the program in 1953, later played a role in determining who to admit to the program during the 1960s and 1970s. He recalled, *"The NIH Associates program would never have been as popular or as competitive as it was without the draft"* (Fredrickson, 1998). Anthony Fauci, a program alumni and Director of the National Institute of Allergy and Infectious Disease, echoed these sentiments *"... every single physician went into military service... essentially, I came down to the NIH because I didn't have any choice"* (Fauci, 1989).

2.2. The application process

Applications to the NIH ATP were typically submitted two years in advance, during the final year of medical school with a planned program start date after completing internship and the first year of residency training. Applications included academic transcripts, references, publications, and planned post-graduate training institutions. After a first screen based on these documents, a small number of applicants were invited to interview on campus at the NIH in order to match with a particular laboratory and mentor. Unfortunately, much of this written documentation was destroyed, leaving only the application index cards of the subset of candidates who cleared the first admission hurdle and attempted to match with a laboratory. There is also no official record of the labs with which each participant attempted to match or offers made. The data can only tell us that out of these second round applicants, roughly 63% accepted an ATP position and attended the program. According to the NIH's official documentation, these final appointments were made based upon intellectual attainment and demonstrated research interest and ability (NIH, 1968).

Applicants were undoubtedly positively selected from the eligible population—male medical school graduates. In Appendix C, Table C1, we can see that compared to a random sample of non-applicants drawn from the American Medical Association (AMA) Physician Master File, applicants graduated from more selective medical schools (as measured by NIH grants) and published at significantly higher rates than non-applicants before application (0.9 vs. 0.3 publications on average). However, it would be wrong to conclude from this evidence that applicants displayed a preternatural disposition for research career prior to application. For instance, the median number of publications for applicants is zero; the overwhelming majority of applicants do not hold a PhD degree; and applicants do not appear particularly precocious, relative to the eligible population (kernel densities corresponding to the age distribution at the time of application for applicants and non applicants is depicted in Figure C1; the two curves are nearly identical).[5]

The oral and written historical records also speak to the difficulty in evaluating research potential and making decisions between candidates. Donald Fredrickson, an ATP alumnus who later served on the selection committee for the program in the 1960s and 1970s, recalls that *"... the main objective was getting people who would use this environment to turn into scientists"*, but also notes selecting participants was *"extremely difficult because all we really had was the scholastic record of most people. Very few had done any research... so the art of picking out of a whole group of qualified people those who might become successful scientists was extremely difficult... We would have to pick them with a certain amount of variety because our programs needed people of diverse interests"* (Fredrickson, 1998). Harry Kimball, another alumnus of the program who was also later involved in applicant selection remembers *"It was truly astonishing how qualified these people were and the kind of close decisions you had to make as to who to offer a spot in the program"* (Kimball, 1997). Harold Varmus describes how the decisive factor in his own selection into the program likely did not hinge on his promise as a budding scientist. Rather, he writes that during his interview with Ira Pastan *"My schooling in literature turned out to be more important than my interest in endocrinology, Ira's field, because Ira's wife Linda, a poet, had often complained that Ira's colleagues seldom talked about books. Ira, himself an enthusiastic reader, thought it might be helpful to have someone with my background in his lab"* (Varmus, 2009).

2.3. Prior evaluations

A handful of prior studies have examined the program. Klein (1998) provides a thorough description of the ATP and the NIH during the Vietnam era grounded primarily in the conduct and review of historical documents and interviews. We have drawn on her analysis to provide much of the necessary institutional background required to guide our empirical analysis. Khot et al. (2011) analyze the careers of NIH ATP attendees from 1955 to 1973, comparing them to a random sample of medical school faculty that graduated in the same years selected from the Association of American Medical Colleges Faculty Roster. The authors show that relative to these controls, ATP participants were 150% more likely to achieve the rank of full professor, twice as likely to become a department chair, and three times as likely to become a medical school dean. Matching the population of attendees with a series of prestige markers appropriate for biomedical researchers, they found in their sample nine winners of the Nobel Prize in Physiology or Medicine, ten recipients of the National Medal of Science, 44 members of the National Academy of Sciences, and 125 members of the Institute of Medicine. Our study improves on their design with a more appropriate control group, that of unsuccessful applicants to the ATP, which helps shed light not simply on the effect of ATP attendance on the intensive margin—articles, citations, grants, patents—but also on the extensive margin: how did selection shape applicants choice of career, in particular participation in research activities as opposed to purely clinical endeavors?

3. Empirical design, data, and descriptive statistics

3.1. Data

The application index cards for the NIH Associate Training Programs form the raw material for the creation of our dataset. While the cards for successful applicants had been previously digitized and used in prior research efforts (e.g., (Khot et al., 2011)), the index card for applicants who did not attend the program were previously thought to have been destroyed. In 2015, carton boxes containing a subset of these index cards—those corresponding to applicants who interviewed on campus but were ultimately not offered a position—were discovered at the National Archives by the NIH archivist, Barbara Harkins. Fig. 1 displays the number of index cards in our dataset in each year belonging to our observation window, 1965 and 1975. While the ratio of successful to unsuccessful applicants is approximately 2:1 over the entire period, this average masks large swings, with the years 1970, 1971, and 1972 exhibiting a greater proportion of unsuccessful applicants. These years correspond to the height of the Vietnam War mobilization effort.

We limited our analysis to those who applied to the program between 1965 and 1975. To arrive at the final list of 3075 applicants, we eliminated 22 applicants who did not hold an MD degree, three unsuccessful applicants who applied at the very start of medical school (and did not reapply), and eight who died while in training, or soon

[5] An additional piece of evidence argues against viewing the applicant population as being dominated by science "geniuses": matching carefully the applicant roster with the Directory of Rhodes Scholars, we found only seven matches (four treated physicians and three control physicians). Note that comparisons with "non-applicants" are subject to an important caveat: since we do not know the identity of the first-round applicants, our sample of non-applicants could in fact include individuals who did not pass the first application screen.

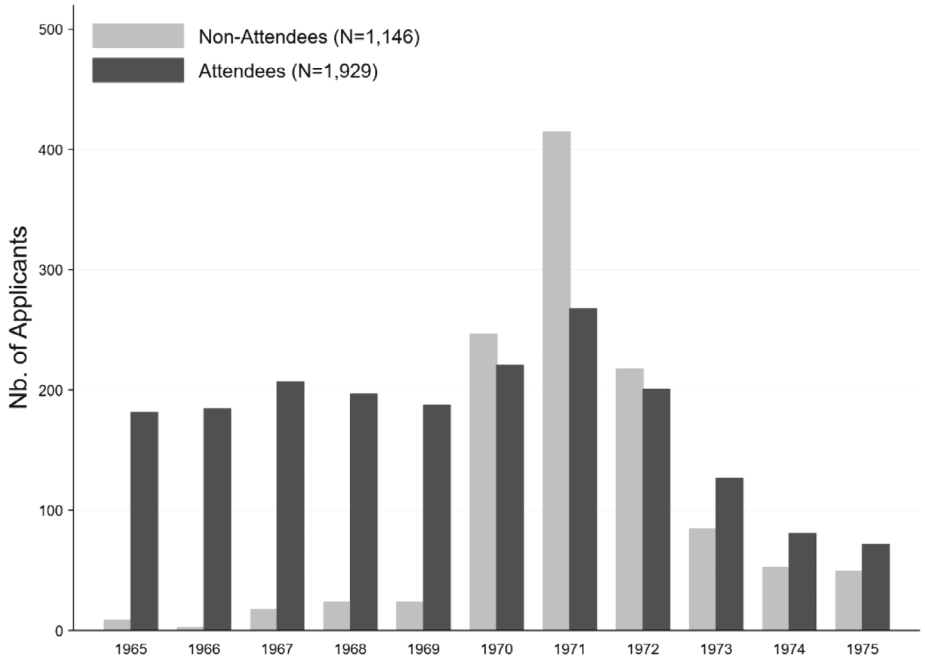

Fig. 1. NIH ATP interviewed candidates by year. *Note*: Number of second-round applicants, by year and treatment status. N = 3075 applicants (1929 attendees; 1146 non-attendees). *Sources:* ATP Index Cards.

thereafter. We also excluded 33 female and 22 foreign medical school graduates as their motivations to apply may have been very different from applicants subject to the draft. Despite our best effort, we also lost 13 applicants to follow-up (less than 0.42% of the total). In the case of repeated applications for the same applicant, we retained only the latest one.

For each of these physicians, we manually collected their training and career history using a mix of Google, Doximity, and LinkedIn searches; medical licensure records; professional profiles and CVs; Who's Who profiles; and other publicly available internet sources. These were supplemented with physician biographical information contained in the AMA Physician Masterfile. To ascertain treatment status, participation in the ATP was verified with the biographical resources above as well as NIH telephone directories and internal human resource records (additional details on dataset construction are available in Appendix A).[6] Applicants who were appointed to the Public Health Service Commissioned Corps but served at the Center for Disease Control (CDC) or the Indian Health Service (IHS) were assigned to the control group. Of course, many members of the control group received research training in traditional academic medical settings, some of them after a period of military service, though only one applicant in the sample appears to have served in the Vietnam military theater. The final sample contains the records of 3075 physicians (1929 program attendees and 1146 non-attendee controls).

We distinguish between three career phases for all applicants. First, the education, or pre-application phase, which ends at the end of medical school. Second, the training phase, which covers internship, residency, post-residency fellowships, as well as national service regardless of the setting where it was served (Army/Navy, NIH, CDC, IHS). Finally, the independent phase of the career begins immediately after the end of the training phase, and ends with retirement or death. When referring to career choice in the rest of the paper, we refer to the choice of employment in this last career phase. 277 (9.01%) applicants pass away prior to their retirement; 762 (24.78%) retire prior to 2017, the end of our observation period; and for 2036 applicants (66.21%), the career is still ongoing as of 2017. Though these observations are technically censored, it is important to acknowledge that the youngest applicant in our sample was 65 years old in 2017 and in his thirty-first career year. To a first order of approximation, these physicians are therefore at the twilight of the active phase of their research or clinical careers.

Publications, citations, patents, and NIH grants were collected for each individual from PubMed, the Web of Science, the U.S. Patent and Trademark Office (USPTO), and the NIH's Consolidated Grant Applicant File, respectively, and carefully name-disambiguated. For publications we include only original research articles, excluding other types of publications such as letters, editorials, and review articles. Importantly for our analysis, we use the richness of the individual profiles collected to measure participation in research independently of the applicants' employers. For instance, the career of many of our applicants unfolds within academic medical centers in purely clinical positions where there is no expectation of publication. In contrast, other applicants work in industry or other non-academic institutions and yet amass a respectable publication record in the context of non-traditional research careers. Since our motivation is to understand how early career interventions might influence long-run engagement with the idea sector of the economy, distinguishing between career *locus* (academic versus non-academic jobs) and career *focus* (research jobs versus clinical jobs) is important.

3.2. Descriptive statistics

Pre-application characteristics. Table 1a presents descriptive statistics regarding ATP applicants at the time of application. Applicants with stronger academic credentials, or with evidence of involvement in research activities are also more likely to attend the program. For instance, applicants holding a PhD degree, those with a publication record, those inducted in an elite medical school honors society

[6] Our set of treated applicants include fellows who completed their training outside of the confines of the NIH intramural campus in Bethesda, such as the Baltimore Cancer Research Center or the Food and Drug Administration (FDA). Other NIH locations were even more far-flung such as the Rocky Mountain Laboratory (located in Hamilton, Montana) or the Panama Control Zone. As a robustness check, we repeated our analysis excluding the 267 ATP attendees not located on the main NIH campus in Bethesda with similar results obtained.

Table 1a
Descriptive statistics: Pre-application data.
Sources: ATP Index Cards, PubMed, CGAF.

	Unweighted sample			Lasso IPTW Reweighting		
	Non-attendees	Attendees	*t*-stat	Non-attendees	Attendees	*t*-stat
Ph.D.	0.013	0.036	3.738	0.030	0.027	0.216
	(0.003)	(0.004)		(0.011)	(0.003)	
Age in the Year of Last Application	25.931	26.016	1.596	26.021	26.145	0.923
	(0.042)	(0.033)		(0.056)	(0.139)	
Applies more than once	0.027	0.028	0.154	0.043	0.028	1.434
	(0.005)	(0.004)		(0.010)	(0.004)	
Number of Applications	1.028	1.029	0.093	1.044	1.029	1.426
	(0.005)	(0.004)		(0.010)	(0.004)	
Number of Institutes Applied For	2.948	3.933	11.789	3.501	3.596	0.730
	(0.061)	(0.053)		(0.115)	(0.057)	
Number of Associate Tracks Applied For	1.828	2.068	7.811	1.962	1.991	0.729
	(0.025)	(0.019)		(0.034)	(0.021)	
AΩA Honor Medical Society	0.257	0.383	7.162	0.305	0.328	0.894
	(0.013)	(0.011)		(0.019)	(0.016)	
Pre-ATP Nb. of Publications	0.582	1.005	7.367	0.841	0.846	0.063
	(0.037)	(0.039)		(0.080)	(0.045)	
Pre-ATP JIF-weighted Nb. of Publications	3.288	6.595	6.835	6.107	5.289	0.854
	(0.292)	(0.330)		(0.967)	(0.330)	
NIH Grants for Applicant's Medical School	170.323	207.006	6.879	185.481	189.629	0.428
	(3.792)	(3.438)		(6.033)	(7.441)	
NIH Grants for Applicant's Internship Hospital	90.915	97.153	1.978	90.758	92.300	0.294
	(2.590)	(1.882)		(3.505)	(3.863)	
Attended Harvard Medical School	0.075	0.142	5.614	0.101	0.121	1.238
	(0.008)	(0.008)		(0.014)	(0.008)	
Attended Johns Hopkins School of Medicine	0.047	0.059	1.356	0.043	0.056	1.576
	(0.006)	(0.005)		(0.007)	(0.006)	
Attended Columbia University	0.050	0.044	0.725	0.051	0.042	0.965
	(0.006)	(0.005)		(0.008)	(0.005)	

Note: N = 3075 applicants (1929 attendees; 1146 non-attendees). Means, standard errors, and *t*-statistics are reported; reweighting is performed using average treatment effect inverse probability of treatment weights. T-statistics are calculated using IPTW-weighted OLS regression of the variable of interest on an indicator variable for ATP attendance. Harvard, Johns Hopkins, and Columbia are the three most common medical schools attended in the sample. For NIH grants, original amounts were deflated using the Biomedical R&D Producer Price Index (2015 dollars) and presented in units of millions of dollars. JIF—journal impact factor.

Table 1b
Descriptive statistics: Career choice.
Sources: ATP Index Cards, AMA Physician Masterfile, doximity.com, state licensure records, NIH telephone directories.

	Non-Attendees		Attendees	
	Mean	Std. Dev.	Mean	Std. Dev.
Deceased	0.075	0.264	0.100	0.299
Years of Post-graduate Training	5.864	1.688	6.425	1.556
Nb. of Career Years (censored in 2017)	37.651	5.805	38.149	6.389
First Job in Academia	0.572	0.495	0.757	0.429
Ends Career in Academia	0.381	0.486	0.546	0.498
Researcher First Job	0.460	0.499	0.694	0.461
Ends Career as Researcher	0.300	0.459	0.519	0.500
First Job in Clinical Practice	0.535	0.499	0.296	0.457
Ends Career in Clinical Practice	0.657	0.475	0.441	0.497

Note: Academia includes both universities/medical schools and research settings such as the NIH or private non-profit institutes (e.g., The Salk Research Institute). Researcher jobs is different from academia in that it includes for-profit industry research positions but excludes clinical university faculty. Clinical practice includes both those in community practice as well as medical school clinical faculty. All variables except years post-graduate training and number of career years are indicator variables.

($A\Omega A$),[7] and those having graduated from elite medical schools (as proxied by the NIH funding received by its affiliated faculty members) are more likely to be selected.[8] Recall that these applicants all survived a first screen, so one might have expected that covariates observable before this initial screen would not influence the selection decision at the interview stage. The fact that observable markers of "research preparedness" do in fact predict selection imply that interviewing "skills" are correlated with these markers, or alternatively, that the ultimate decision makers place positive weights on them even at the second stage of the process. However, one must remember that due to the young age of the applicants, the signals of research potential upon which the selection decision relies are necessarily noisy. For instance, 59.4% of applicants have no publication to their name within two years of their ATP application (67.4% for attendees; 54.7% for non-attendees). ATP attendees also applied to more NIH institutes (3.9 vs.

[7] Criteria for selection into $A\Omega A$ varies by school, but typically weighs academic and clinical excellence most heavily.

[8] Appendix Table B1 lists the 10 most frequent medical schools from which physicians in the sample graduated, separately for attendees and non-attendee controls. Appendix Figure B1 provides a histogram for the distribution of the number of original publications published up to the year of ATP application, weighted by the journal impact factor of the publication outlet in which they appeared.

Table 1c
Descriptive statistics: Research outcomes.
Sources: ATP Index Cards, PubMed, CGAF, USPTO, Marx and Fuegi (2020) "reliance on science" publication-to-patent linkages.

	Non-Attendees		Attendees	
	Mean	Std. Dev.	Mean	Std. Dev.
Nb. of Pubs, Training Period	2.400	4.079	6.050	6.389
Career Nb. of Pubs	37.313	80.078	77.773	109.584
Career Citations	1988	5345	5131	10,391
Nb. of Patents	0.657	3.729	1.738	6.569
Career Citations to Patents in Patents	7.506	53.651	20.227	106.080
Career Citations to Pubs in Patents	80.095	347.028	252.029	914.263
NIH Grant Recipient	0.206	0.405	0.442	0.497
Career NIH Grants ($ 2015)	4,511,372	35,192,232	12,436,209	42,898,984
Career NIH R01 Grants ($ 2015)	1,193,642	5,035,673	3,149,951	8,197,320
Nb. NIH-R01-funded Trainees	0.214	0.885	0.758	1.914
Trainee Career NIH R01 Grants ($2015)	1,167,519	6,091,129	4,722,876	14,203,229

Note: Except in the first row, all outcomes should be understood to be restricted to output in the post-training (i.e., independent) phase of the career. NIH grant recipient is an indicator variable equal to 1 if an individual ever received an NIH grant.

Table 1d
Notable achievements.
Sources: ATP Index Cards, CGAF, Nobel Prize, HHMI, and NAS web sites.

	Nobel Prize	Natl. Academies Member	Howard Hughes Med. Investigator	NIH MERIT [R37] Awardee
Non-Attendees	0 (0.00%)	14 (1.12%)	0 (0.00%)	14 (1.22%)
Attendees	7 (0.36%)	90 (4.67%)	32 (1.66%)	79 (4.10%)
Total	7 (0.23%)	104 (3.34%)	32 (1.04%)	93 (3.02%)

Table 1e
Descriptive statistics: Research style.
Sources: ATP Index Cards, PubMed.

	Non-Attendees		Attendees	
	Mean	Std. Dev.	Mean	Std. Dev.
Basic Science Articles	0.107	0.200	0.199	0.248
Translational Medicine Articles	0.209	0.234	0.273	0.232
Clinical Trial Articles	0.097	0.161	0.107	0.162
Other Clinical Articles	0.467	0.324	0.338	0.292
Articles Appearing in "Translational" Journals	0.012	0.065	0.016	0.039
Inspires Translational Research	0.088	0.135	0.118	0.137
Builds on Translational Research	0.068	0.131	0.078	0.130
Articles Cited in Patents	0.109	0.149	0.162	0.162

Note: N = 2584 scientists (491 scientists with zero publications cited at least once in the independent phase of the career are excluded). Statistics correspond to the fraction of each scientist's work with the corresponding characteristic.

2.9), perhaps signaling greater interest in or motivation for research undertakings.[9]

Career choice. Table 1b provides basic statistics regarding career outcomes, with a particular focus on the first job following the end of the training phase and the last job held by each applicant before the earliest of 2017, retirement, or death (Appendix Tables B2 and B3 provide a finer-grained occupational breakdown). It is immediately apparent that ATP attendees choose academic (76% vs. 57%) and research (69% vs. 46%) careers at a more pronounced rate, relative to non-attendees, following the end of their training. These differences reflect in part time spent in training, though this contrast is not especially stark: On average, ATP attendees spend an additional 6.7 months in post-graduate training prior to achieving career independence, relative to non-attendees. The gap does not seem to narrow as their career unfolds, though one can observe attrition in the subsample of attendees. The proportion of fellows in research positions falls from 69% to 52% between the beginning and the end of the career. Overall, these univariate comparisons corroborate the claims made by ATP alumni regarding the effect of their training on career orientation. For instance, Harry Keiser, an ATP alumnus and later clinical director of the National Heart, Lung and Blood Institute, mentions that *"if I had gone back to Northwestern. . . I would have almost certainly gone out into private practice. . . I certainly would not have continued to devote the rest of my life to research"* (Keiser, 1998).

Research outcomes. Table 1c reports descriptive statistics on a variety of research outcomes. ATP attendees garner over twice the number of career publications on average (77.8 vs. 37.3). Similar differences can be observed for patents (1.7 vs. 0.7), NIH extramural grant funding ($12.4 vs. $4.5 million), and citation impact (5,131 vs. 1988 for article-to-article citations; 20.2 vs. 7.5 for patent-to-patent citations). ATP attendees' publications are also more heavily cited in patents (252 vs. 80). Attendees receive greater NIH R01 funding as well, with $3.1 million compared to $1.2 million over their career.

We also examine the "fecundity" of ATP applicants by identifying the set of individuals they train over their career who go on to be awarded NIH R01 funding, a key marker of research independence in U.S. academic medicine. In this way, the impact of training institutions

[9] The right-most columns of Table 1a provide a comparison of means in the reweighted sample which reflects the methodology presented in Section 3.3 and Appendix F. In the pseudo-population of trainees created by this procedure, the differences in baseline covariates are no longer statistically significant.

can ripple through a much larger community of scholars as yesterday's trainees become the trainers of today. In the context of our data, a trainee is an individual who, in a window centered on the time of her highest degree, appears as first author on a publication jointly with the ATP applicant in last authorship position. We then match the names of these individuals with the NIH Consolidated Grant Applicant File, allowing us to identify the subset of trainees who go on to be awarded NIH funding (more details are provided in Appendix H). This is a relatively sparse outcome, but there again, successful applicants appear more prolific than unsuccessful ones (0.76 vs. 0.21 R01-funded trainees and $4.7 vs. $1.2 million in trainee career R01 grants on average).

Panels A, B, and C of Appendix Figure B2 display histograms for the distribution of career publications, citations, and NIH funding by treatment status. The differences in achievement between attendees and non-attendees are even more pronounced in the right-tails of these distributions. This is also reflected in the rate at which attendees accrue markers of research excellence over the career, relative to non-attendees (Table 1d). In the control group, no physician ever receives a Nobel Prize or a Howard Hughes Medical Institute (HHMI) Investigatorship (the corresponding numbers in the treatment group are 7 and 32, respectively). The differences in the rate at which treatment and control physicians become Members of the National Academies or NIH MERIT awardees are less stark, but still large in magnitude.

Research style. We develop a battery of measures to capture differences in research style across physicians in the sample. In particular, we take a first stab at measuring "translational" biomedical research. Translational research does not have an agreed-upon definition (Butler, 2008; Woolf, 2008). For the purposes of this paper, we will build upon the view of David Nathan, an NIH ATP alumni and former president of the Dana-Farber Cancer Institute (Nathan, 2005):

> *"Translational clinical investigators come in at least two flavors... One class includes physician-scientists interested in disease mechanisms... But these almost never interact in their research with an intact patient/subject. Such disease-oriented researchers are content to study tissue samples, cell lines, and model systems such as mice, fish, and yeast and do so with great benefit... Their career paths are only slightly distinguishable from those of basic scientists... The other class of physician-scientists include patient oriented researchers. They actively search for patients who may enable them to uncover the secrets of complex diseases, care for those patients, and with their permission, undertake to explore new diagnostic and therapeutic approaches to treating their diseases".*

As a concrete (and famous) example of translational research of the first type, consider the work of NIH ATP alumni Joseph Goldstein and Michael Brown, recipients of the 1985 Nobel Prize for Medicine and Physiology. Their initial investigations were inspired by observations of patients with familial hypercholesterolemia they saw at the NIH Clinical Center (Goldstein and Brown, 1997). Through patient-inspired basic investigations performed at the laboratory bench, they identified the underlying root cause of this disease as a lack of low-density lipoprotein receptors. These discoveries in turn informed drug development efforts, ultimately leading to the market introduction of statins. The work of Goldstein and Brown illustrates well the importance of both the "bench to bedside" and "bedside to bench" transitions which are a recurring theme in the oral histories of ATP alumni.

Conversely, Philip Pizzo personifies an approach to translational research closely connected with patient care. After his clinical associateship, Pizzo stayed on at NIH, becoming Chief of Pediatrics and Scientific Director of the Division of Clinical Sciences at the National Cancer Institute before being named Physician-in-Chief of Boston Children's Hospital and later Dean of Stanford Medical School. An expert in infectious disease and cancer, examples of his contributions include the first use of antiretroviral medication in children with HIV, a phase I trial of a solubilized receptor used by HIV for cell attachment, assessing the effectiveness in cancer patients of a diagnostic test for invasive fungal infection previously studied only in animal models, and in vitro testing of approaches to rescue neutrophil dysfunction using HIV patient samples.

The MeSH thesaurus from the National Library of Medicine provides the raw material necessary to create our measures of research style. MeSH consists of terms arranged in a hierarchical structure that permit searching at various levels of specificity (there are over 29,000 descriptors in the 2019 edition of MeSH). Almost every publication in *PubMed* is tagged with a set of MeSH terms (between 1 and 68 in the current edition of *PubMed*, with both the mean and median approximately equal to 10). For each article published by a scientist in the sample, we measure disease orientation by the presence of a disease MeSH term. To capture bench research, we take note of the presence of MeSH terms for molecular biology techniques—such as *nucleic acid amplification techniques* or *cell migration assays*, MeSH terms corresponding to model organisms—such as the nematode *caenorhabditis elegans* or the fruit fly *drosophila melanogaster*, MeSH terms related to cellular structures and macromolecules—e.g., *DNA topoisomerase IV*, or MeSH terms denoting biochemical and cellular processes—e.g., *oxidative phosphorylation* (See Appendix G for further details).

In a second step, we partition the bibliome into four mutually exclusive styles: (i) *Basic science* articles are not disease-oriented, are tagged by at least one bench science keyword, and are not clinical trials; (ii) *translational* articles are disease-oriented, tagged by at least one bench science keyword, and not clinical trials; (iii) clinical trials (identified using MeSH terms and the publication type field in *PubMed*); and (iv) *"other" clinical* articles, which are disease-oriented, not clinical trials, and not tagged by any bench MeSH keywords.[10]

We create four additional approaches to uncover the empirical signature of a translational research style. First, a natural way for the transition from bench to bedside to take place is for clinical researchers to further develop translational work, for example by performing a clinical trial. We designate an article as "inspiring translational research" whenever it is translational according to the above criteria *and* is cited by a clinical trial publication. Second, in the same spirit, we identify work that "builds on translational research": articles that report the results of a clinical trial *and* also list a translational publication in their references. Third, we identify papers published in six high-impact journals that prominently advertise their translational focus (the *Journal of Clinical Investigation*, the *Journal of Translational Medicine*, *Science Translational Medicine*, *Nature Medicine*, *Translational Research: The Journal of Laboratory and Clinical Medicine*, and the *Journal of Experimental Medicine*). Finally, a different way to facilitate the bench-to-bedside transition is to enable biopharmaceutical firms to build on the applicant's published research, since many health-related innovations cannot reach patients unless firms invest in bringing them to market (Azoulay et al., 2009). To capture this, we tag each article that garners at least one citation in the header of a patent subsequently granted by the USPTO (Marx and Fuegi, 2020). This provides a crude way to capture the extent to which biopharmaceutical firms build on the work of the scientists in the sample to inform their applied R&D efforts.

Table 1e reports descriptive statistics for the research style measures. Because these measures are only meaningfully defined for publishing researchers, we create a subsample that only includes the 2584 scientists (1730 treated and 854 controls) who publish at least one article after the end of their training. Rather than focusing on the levels of these variables, we normalize them by the total number of articles published by each scientist in the independent career phase.

Non-attendees and attendees differ markedly in the style composition of their published work. The proportion of basic science articles is almost twice as high for successful applicants (19.9% vs. 10.7%);

[10] Jointly, these styles comprise 93% of the applicant's published output. For the style analysis, we ignore the residual unclassifiable publications.

the proportion of translational articles is approximately 30% higher; and the proportion of clinical trials is approximately 10% higher. This means that a higher fraction of the non-attendees' output falls into the "other" clinical category. Similarly, univariate comparisons point to higher translational orientation for attendees, relative to non-attendees, using additional measures of research style. For instance, a higher fraction of attendees' articles appear in a small set of explicitly translational journals, are referenced in patents, or inspire follow-on translational research. Below, we explore whether these differences subsist when comparing treated and control physicians with similar observable characteristics.

3.3. Econometric considerations

The univariate comparisons point to large differences in outcomes between attendees and non-attendees of the NIH ATP. It would be hazardous to interpret these differences as reflecting the causal effect of the ATP "treatment", since it is obviously a goal of NIH laboratory heads to admit applicants with the most research promise. Recall that all applicants in our sample already passed a first selection screen. Yet residual sources of selection might remain at the interview stage, e.g., the admissions committee might extract relevant information regarding an applicant's suitability for a research career in a series of relatively short interviews. To address this fundamental identification challenge, we adopt a propensity score weighting methodology which belongs to a broad class of "selection-on-observables" techniques (additional details are provided in Appendix F).

Inverse probability of treatment weighted estimation. Let us assume that the NIH principal investigators recruiting fellows at the interview stage are unable to select applicants on the basis of covariates unobserved by the econometrician and correlated with research career success—the "unconfoundedness" assumption. This assumption is not refutable and it places strong demands on the data generating process. In addition, we must assume that, for all included values of the covariates predicting treatment, the likelihood of being selected to attend is positive—the "common support" assumption. Under these assumptions, Hirano and Imbens (2001) show that various treatment effects of attending the NIH ATP, conditional on exogenous applicant characteristics, can be recovered by weighted least squares or weighted maximum likelihood estimation where the weights correspond to the inverse probability that each observation is treated. Our weighting procedure effectively creates a pseudo-population of applicants in which observable covariates no longer predict assignment to treatment and the causal association between treatment and the outcome variable is unchanged from the original population. We refer to this as the Inverse Probability of Treatment Weighted (IPTW) estimation (Austin and Stuart, 2015; Xu et al., 2010).

Informative censoring. Although we focused on the problem of non-random selection into treatment, a second problem arises because some applicants might fail to engage in research activities for the sole reason that their chosen position does not afford them the possibility to publish, seek external grants, or train the next generation of scientists. This problem is distinct from informative loss to follow-up. These physicians' careers are observed in full and yet it does not seem meaningful to compare the research productivity of a full-time, tenure-track academic researcher with that of a clinician who very occasionally dabbles in research. We deal with this problem by treating early exit from research as another treatment. As Robins et al. (2000) note, adjusting for this type of informative censoring is tantamount to estimating the causal effect of ATP attendance on an outcome if, contrary to the fact, all applicants had remained engaged in research rather than followed their censoring history. We model the exit decision as a function of the same pre-application covariates used to model selection into treatment, and compute weights corresponding to the probability of exit given these observables. The final weight, obtained by multiplying the weights corresponding to the inverse probability of treatment and inverse probability of censoring, is the probability an applicant would have followed his own treatment *and* censoring history, conditional on observables. We label this methodology Inverse Probability of Treatment and Censoring-Weighted (IPTCW) estimation in what follows.

Selection on unobservables. Despite a long list of observable covariates to predict selection into the ATP, IPTW estimation does little to address the threat to identification due to factors unobservable to the econometrician. The time period of the study suggests an instrumental variable approach based on draft eligibility, as in Angrist (1990). However, medical school graduates, having already deferred their service for educational purposes, were not, in effect, eligible to participate in the lottery (Crowell, 1971; Rousselot, 1971). Table D1 in Appendix D verifies that having one's number called in the lottery does not help predict ATP attendance.

Estimation procedure. Many of the outcomes we study, including publication counts and NIH grants awarded, are skewed and non-negative with a large mass point at zero (see Figures B2a, B2b, and B2c). For example, 426 (13.9%) of the applicants do not publish after their training; approximately two thirds of the sample never receive any NIH grant funding over the career. Following a long-standing tradition in the study of scientific and technical change, for these skewed outcomes we present Poisson quasi-maximum likelihood (hereafter QML) estimates (Santos Silva and Tenreyro, 2006). Because the Poisson model is in the linear exponential family, the coefficient estimates remain consistent as long as the mean of the dependent variable is correctly specified (Gouriéroux et al., 1984). QML (i.e., "robust") standard errors are computed using the outer product of the gradient vector (and therefore does not rely on the Poisson variance assumption).

4. Results

The exposition of the econometric results proceeds in stages. We first explore empirically the determinants of selection into the ATP. Using the predicted probabilities from these models as regression weights, we then report estimates of the effect of ATP attendance on (i) career choice outcomes; (ii) research productivity (including trainee mentorship outcomes); and (iii) research style outcomes. Finally, we perform a battery of robustness tests to probe the plausibility of the unconfoundedness assumption in our context.

4.1. Selection into the NIH ATP

We model the likelihood of selection in a logit framework using an extensive list of covariates observed at the time of selection (Table 2).[11] We capture the research orientation of the medical school and intended internship hospital for each applicant with the NIH funding that accrue to principal investigators in these institutions. We also include an indicator variable for applicants who received a PhD before they applied, and an indicator variable for election to the $A\Omega A$ Honor Medical Society. The most informative indicator of research promise is probably demonstrated engagement in research activities, as ascertained by an applicant's list of scientific works published, or soon-to-be-published at the time of application. We weight each of these student publications by the impact factor of the journal in which they appeared as a crude quality adjustment (raw counts produce similar results).

Columns 1a and 1b report logit coefficients and find the signs for most of the covariates are in the expected direction. Relative to

[11] In fact, most of these factors might have been observed at the initial selection stage (e.g., medical school attended) while for others the timing is more ambiguous as they might become known to the applicant between the first and second stage of the ATP selection process (e.g., intended internship hospital, accepted or forthcoming journal publications).

Table 2
Modeling selection into the NIH ATP.
Sources: ATP Index Cards, PubMed, CGAF.

	Program selection			*Informative censoring*		
	Parsimonious Model [Logit]		Saturated Model [Lasso]	Parsimonious Model [Logit]		Saturated Model [Lasso]
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
Log(Pre-ATP Nb. of Publications)		0.307**	0.329**		−0.192**	−0.210**
		(0.071)	(0.071)		(0.064)	(0.066)
Ln(NIH Grants for Applicant's Medical School)	0.357**	0.317**		−0.193**	−0.158*	
	(0.090)	(0.091)		(0.067)	(0.066)	
Ln(NIH Grants for Applicant's Internship Hospital)	0.019*	0.017†		−0.031**	−0.029**	
	(0.009)	(0.010)		(0.008)	(0.009)	
Ph.D.	0.932**	0.577†	0.794*	−1.347**	−1.036**	−1.141**
	(0.334)	(0.342)	(0.311)	(0.354)	(0.358)	(0.359)
No Internship	1.962*	1.763*	1.117	−2.716*	−2.583*	−3.936**
	(0.839)	(0.849)	(0.981)	(1.056)	(1.068)	(0.879)
Applies more than once	−0.039	−0.091	0.071	0.076	0.115	−0.028
	(0.300)	(0.295)	(0.273)	(0.246)	(0.249)	(0.242)
AΩA Honor Medical Society	0.688**	0.701**	0.661**	−0.346**	−0.346**	−0.347**
	(0.105)	(0.106)	(0.101)	(0.087)	(0.088)	(0.087)
Constant	−3.278†	−2.649		3.049*	2.329†	
	(1.748)	(1.775)		(1.311)	(1.310)	
Medical School Fixed Effects	No	No	Yes	No	No	Yes
Internship Hospitals Fixed Effects	No	No	Yes	No	No	Yes
Nb. of Non-zero Predictors			151			168
Nb. of Potential Predictors			372			372
χ^2 Test Statistic			67.96			51.23
Pseudo-R^2	0.251	0.265		0.056	0.073	
Log-likelihood	−1521	−1493		−1945	−1910	
Nb. of Applicants	3075	3075	3075	3075	3075	3073

Note: The dependent variable is an indicator variable equal to one for attendees, zero for non-attendees (first three columns) or an indicator variable equal to one for attendees who exit research immediately after training (last three columns). Estimates are displayed as coefficients from logit specifications. All models incorporate a full suite of medical school graduation year effects; a set of indicator variables for the applicant's age at the time of application; indicator variables for the number of distinct NIH component institutes that received the application; indicator variables for the number of tracks applied to within the Associate Training Program; indicator variables for the number of years between the application and the medical school graduation year; and a series of indicator variables capturing if the applicant (1) intended to postpone his internship until after training, (2) intends to perform his internship abroad, (3) intends to intern in a hospital affiliated with the Veterans Affairs Administration, or (4) has missing information regarding his intended internship hospital. All models except (1a) and (2a) also include an indicator variable for applicants without any publication before application. Estimates in columns [1c] and [2c] correspond to the results of a cross-fit partialling-out lasso logit procedure with ten folds, as described in Chernozhukov et al. (2018). The specification includes all the covariates mentioned above, plus a full suite of medical school indicator variables and a full suite of internship hospitals indicator variables, but only a subset of this list is selected for inclusion (151 out of 372 in model [1c]; 168 out of 372 in model [2c]. In both models [1c] and [2c], a Wald test rejects the hypothesis that the "coefficients of interest" (i.e., those that are constrained to appear in the model, and for which inference is performed) are jointly equal to zero. Robust errors in parentheses ($^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$).

applicants without publications and at the mean of all other covariates, computed marginal effects suggest applicants with one publication are 7% more likely to be selected; those with two publications or more, 20% more likely.

Estimates in column 1c correspond to the results of a cross-fit partialling-out lasso logit procedure with ten folds, as described in Chernozhukov et al. (2018). The specification includes all the covariates mentioned above, plus a full suite of medical school indicator variables and a full suite of internship hospitals indicator variables, for a total of 372 covariates, 151 of which the procedure selects for inclusion as control variables. This procedure allows for statistical inference to be performed on five covariates of interest also included in the specification in column 1b, enabling the coefficients and standard errors to be compared across columns.[12]

Columns 2a, 2b, and 2c perform a similar exercise, but the response variable is not selection in this case, but rather exit from research at the end of training. The signs of the coefficient estimates for the predictive covariates are flipped, relative to the specifications in columns 1a, 1b, and 1c.

The specifications used to compute selection probabilities and regression weights for each applicant depart ever so slightly from those in columns 1c (for the selection weights) and 2c (for the informative censoring weights). Since the estimation of the propensity score is solely a prediction exercise, we favor an abundance of explanatory variables in these models. Our least restrictive specification includes 94 fixed effects for medical schools and 238 indicator variables for intended internship hospitals. We constrain the model to include the same variables as the specification in column 1c and 2c as well as the inverse hyperbolic sine of medical school and internship hospital NIH grant funding. The other variables are selected via a logit procedure with a lasso penalty term, using ten-fold cross-validation to prevent overfitting the data. The predicted probabilities from this model are used to generate the benchmark set of lasso weights used below to estimate treatment effects.[13,14]

Table 1a confirms that pre-application covariates appear balanced across treated and control observations in the sample appropriately

[12] Note that medical school and internship hospital funding variables are not separately identified from the fixed effects and drop out of the specification. The χ^2 test statistic (*i.e.*, the Wald test of the hypothesis that the coefficients of these five covariates are jointly equal to zero) is equal to 78.85 ($p < 0.01$).

[13] We test the quality of our predictions by splitting the sample into a prediction subsample (2460 or 80% of the observations) and a hold-out sample (615 or 20% of the observations). The out-of-sample deviance ratio (a measure of goodness of fit for logit models) is equal to 0.70 of the corresponding in-sample value, which is acceptable.

[14] As a robustness exercise, we repeat our analysis using logit weights computed using the model in Table 2, columns 1b and 2b (Figure B3, Tables B4a and B4b). Note that the correlation between the predicted selection probabilities from column 1b and that of the model with lasso regularization is 0.919. As a result, the magnitudes and precision of the IPTCW estimates are not very sensitive to the choice of weights.

Table 3
Career choice outcomes.
Sources: ATP Index Cards, AMA Physician Masterfile, doximity.com, state licensure records, NIH telephone directories.

	X-Sect.	Lasso weights	
	Naive	ATE	ATET
Poisson estimates			
Years of Post-graduate training	1.091**	1.082**	1.069**
	(0.012)	(0.015)	(0.019)
Nb. of career years	0.988†	0.988†	0.989
	(0.006)	(0.007)	(0.008)
Logit estimates			
First Job in Academia	0.160**	0.113**	0.085**
	(0.018)	(0.021)	(0.025)
Ends Career in Academia	0.146**	0.132**	0.111**
	(0.020)	(0.032)	(0.027)
Researcher First Job	0.212**	0.170**	0.153**
	(0.018)	(0.022)	(0.026)
Ends Career as Researcher	0.216**	0.192**	0.175**
	(0.019)	(0.030)	(0.025)
First Job in Clinical Practice	−0.212**	−0.171**	−0.155**
	(0.018)	(0.022)	(0.026)
Ends Career in Clinical Practice	−0.215**	−0.189**	−0.174**
	(0.019)	(0.029)	(0.025)
Joins the Research Elite	0.056**	0.025*	0.032*
	(0.013)	(0.012)	(0.013)
Number of applicants	3075	3075	3075

Note: Each cell contains an estimate for the treatment effect in a separate regression. The dependent variables are listed in the left-most column. All models incorporate a full suite of medical school graduation year effects as well as an indicator variable for holding a Ph.D. degree at the time of application. Column 2 and 3 perform inverse probability of treatment weighted estimation for first career position and training length outcomes (rows 1, 2, 3, 5, and 7) and inverse probability of treatment and censoring weighted estimation for all other outcomes; the corresponding estimates can be interpreted as the ATE/ATET of NIH training, under the assumption of unconfoundedness. On the first two rows, the estimates stem from Poisson regressions. Exponentiated coefficients are presented; subtracting 1 yields magnitudes interpretable as elasticities. For example, the estimate in the top cell of the first column imply that attendees stay 100×(1.091]-1) = 9.1% longer in training, relative to non-attendees; the effect is highly statistically significant. On the next six rows, the estimates stem from logistic regressions. The marginal effects for the treatment indicator are reported. For instance, the coefficient in the third row of the first column implies that attendees are 16.0% more likely than non-attendees to be initially placed in academia after completing their training. Robust errors in parentheses ($^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$).

weighted using the fitted selection probabilities to construct the selection weights according to the method described in Section 3.3 and Appendix F.

4.2. Career choice

Table 3 reports estimates of the treatment effect of ATP attendance on career outcomes. For each outcome (which differ across rows), the first column reports the naïve cross-sectional estimate. The remaining columns report the average treatment effect (ATE) and the average treatment effect on the treated (ATET) using inverse probability of treatment and censoring lasso weights (computed using the model in Table 2 columns 1c and 2c).

The first two rows of Table 3 report the ATP effect on the length of the training period as well as the length of the career overall. Each estimate in the table corresponds to the coefficient on a treatment indicator variable (and its associated robust standard error) from a Poisson model where the outcome of interest is regressed on an indicator variable for holding a PhD degree at the time of application and a full suite of medical school graduation year effects in addition the treatment variable.

Exponentiated coefficients are presented; subtracting one yields a magnitude interpretable as an elasticity. For example, the estimates in the first cell of Table 3 imply that ATP attendees spend $100\times(1.091-1) =$ 9.1% longer in training than non-attendees—an additional six months on average. This is a meaningful yet rather small increase relative to the time of commitment of the ATP (two years). It underscores the extent to which our results pertain to the effect of the *content* of training, rather than to the mere fact that training was received. We also find that NIH training reduces slightly the length of the overall post-independence career, but the effect is small (between 1 and 2%, or seven months on average), and imprecisely estimated in some specifications.

The next six rows of Table 3 pertain to the effect of the program on the choice of career. We report the marginal effects from logistic regressions of these career choice indicators on the treatment indicator and our usual set of controls. Across columns, we observe that attending the ATP greatly increases the likelihood of embarking on an academic or research career. For instance, using the average treatment effect estimated using lasso weights, the marginal effect of starting in academia is 0.11, which corresponds to an odds ratio of 1.77. The program increases the probability of a research-focused initial job even more (the marginal effect is 0.17, which translates into an odds ratio of 2.17) for treated physicians, relative to controls. The effects are also persistent, with similar magnitudes observed when analyzing the program's impact on end-of-career positions. Conversely, attending NIH ATP appears to make it markedly less likely to choose a clinical career (an odds ratio of 0.46).

We also create a composite outcome for joining the biomedical research elite over the course of one's career, which we define as either

Table 4
Research outcomes.
Sources: ATP Index Cards, PubMed, Web of Science, CGAF, USPTO, Marx and Fuegi (2020) "reliance on science" publication-to-patent linkages.

	X-Sect.	Lasso Weights	
	Naïve	ATE	ATET
Career Nb. of Pubs	1.922**	1.639**	1.668**
	(0.146)	(0.128)	(0.152)
Career Nb. of Pubs, First/Last Authorship Position	1.970**	1.673**	1.701**
	(0.146)	(0.131)	(0.157)
Career Citations	2.316**	1.775**	1.884**
	(0.227)	(0.190)	(0.220)
Nb. of Patents	2.515**	1.635*	1.559†
	(0.521)	(0.368)	(0.406)
Career Citations to Pubs in Patents	3.015**	1.892**	1.938**
	(0.511)	(0.386)	(0.454)
Career NIH Grants	2.558**	1.846*	1.807*
	(0.591)	(0.489)	(0.517)
Career NIH R01 Grants	2.285**	1.668**	1.774**
	(0.352)	(0.296)	(0.340)
Nb. NIH-R01-Funded Trainees	2.410**	1.621*	1.689*
	(0.365)	(0.349)	(0.429)
Trainee Career NIH R01 Grants	2.677**	1.890**	2.039**
	(0.501)	(0.423)	(0.541)
Number of Applicants	3075	3075	3075

Note: Each cell contains an estimate for the treatment effect in a separate regression. All estimates stem from Poisson regressions. The dependent variables are listed in the left-most column. All models incorporate a full suite of medical school graduation year effects as well as an indicator variable for holding a Ph.D. degree at the time of application. Exponentiated coefficients are presented; subtracting 1 yields magnitudes interpretable as elasticities. For example, the estimate in the first cell imply that attendees publish 100×(1.922-1) = 92.2% more original articles during the independent phase of their career, relative to non-attendees; the effect is highly statistically significant. Columns 2 and 3 perform inverse probability of treatment and censoring weighted; the corresponding estimates can be interpreted as the ATE/ATET of NIH training, under the assumption of unconfoundedness. Robust errors in parentheses ($^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$).

(i) receiving the Nobel Prize; (ii) being elected to the National Academy of Sciences or the National Academy of Medicine; (iii) being appointed Investigator of the Howard Hughes Medical Institute; or (iv) getting a MERIT designation from the NIH in at least one R01 grant cycle. Only 173 (5.6%) of the applicants belong to this select group by career's end (7.7% of the attendees; 2.2% of the non-attendee controls). Adjusting for selection and censoring based on observable covariates dampens somewhat this difference: the average treatment effect corresponds to an odds ratio of 1.77.

4.3. Research outcomes

Whereas Table 3 focused on the effect of NIH training at the extensive margin (i.e., the choice to begin a research career or to stay in one), Tables 4 and 5 hone in on the effect of the program at the intensive margin (the intensity of research effort over the career, as it is being converted into publications, patents, and grants).

Table 4 reports estimates of the treatment effect of ATP attendance on various metrics of research output over the career. Each outcome variable has been constructed to exclude output that results from research undertaken as a student or a trainee: they correspond to research output for the entire post-training (i.e., "independent") career. We consider nine different outcomes: publication count; publication count excluding those where the applicant is in the middle of the authorship list;[15] cumulative citation count accrued by 2015; USPTO patent count (by 2016); count of references to the scientist's publications appearing on the front page or within the body of patents (Marx and Fuegi, 2020); cumulative NIH grant funding received as a principal investigator; cumulative NIH R01 grant funding received as a principal investigator; count of trainees who go on to receive NIH R01 funding during their own independent careers; and the amount of NIH R01 funding accrued by these trainees.

Synthesizing the results across rows and columns of Table 4, a number of patterns emerge. First, the magnitude of the treatment effects are large, even when they filter out the effect of selection and censoring under the maintained assumption of unconfoundedness. Using the lasso weights, for example, the ATE for publications corresponds to an increase of 63.9%, and the ATET to an increase of 66.8%. Second, modeling selection based on observable covariates does shrink the magnitude of the estimated effects by 25 to 50%, depending on the outcome. Third, the ATE and ATET typically have similar magnitudes, which is logical since control scientists are drawn from the same underlying population. All estimates are precisely estimated, although the ATET specification for patents is only significant at the 10% level.[16]

Citation analysis. The estimates for the effect on overall citations in Table 4 conflate the effect of treatment on the quantity of output with

[15] A robust social norm in the life sciences systematically assigns last authorship to the principal investigator, first authorship to the junior author who was responsible for the conduct of the investigation, and apportions the remaining credit to authors in the middle of the authorship list, generally as a decreasing function of the distance from the extremities (Dance, 2012; Sauermann and Haeussler, 2017). Therefore, the first- and last-authored publications correspond to those associated most closely with each applicant.

[16] This is not entirely surprising since applicants in clinical research careers are at very low risk of patenting (only 20% of the physicians on the sample are awarded at least one patent over the course of their career). In contrast, all applicants in the sample are at risk of publishing.

Table 5
Publication outcomes, by citation quantiles.
Sources: ATP Index Cards, PubMed, Web of Science.

	X-Sect.	Lasso weights	
	Naïve	ATE	ATET
Career Nb. of Pubs, Total (with citation data available)	1.951**	1.637**	1.680**
	(0.151)	(0.130)	(0.156)
Career Nb. of Pubs Top 50% of the Citation Distribution	2.065**	1.696**	1.735**
	(0.169)	(0.146)	(0.175)
Career Nb. of Pubs Top 25% of the Citation Distribution	2.157**	1.717**	1.771**
	(0.189)	(0.162)	(0.194)
Career Nb. of Pubs Top 5% of the Citation Distribution	2.348**	1.800**	1.884**
	(0.247)	(0.202)	(0.240)
Career Nb. of Pubs Top 1% of the Citation Distribution	2.653**	1.963**	2.176**
	(0.347)	(0.284)	(0.322)
Career Nb. of Pubs Top 0.1‰ of the Citation Distribution	2.814**	1.968**	2.195**
	(0.533)	(0.401)	(0.424)
Number of Applicants	3075	3075	3075

Note: Each cell contains an estimate for the treatment effect in a separate regression. All estimates stem from Poisson regressions. The dependent variables are listed in the left-most column. All models incorporate a full suite of medical school graduation year effects as well as an indicator variable for holding a Ph.D. degree at the time of application.
Exponentiated coefficients are presented; subtracting 1 yields magnitudes interpretable as elasticities. For example, the estimate in the bottom cell of the first column imply that attendees publish 100×(2.814-1) = 181.4% more articles in the top 0.1‰ of the citation distribution during the independent phase of their career, relative to non-attendees; the effect is highly statistically significant. Columns 2 and 3 perform inverse probability of treatment and censoring weighted estimation; the corresponding estimates can be interpreted as the ATE/ATET of NIH training, under the assumption of unconfoundedness. Robust errors in parentheses ($^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$).

the effect of treatment on the quality of output. Table 5 sheds light on the effect of NIH training on citation impact (a reasonable proxy for publication quality) specifically. For each publication, we use the *Web of Science* to ascertain its percentile in the vintage-specific article-level citation distribution.[17] This makes it possible to meaningfully aggregate, for each applicant, the number of his post-training publications whose eventual impact falls above the jth-percentile of the citation distribution, even though these publications might have appeared at different times. The structure of Table 5 is otherwise identical to that of Table 3.

The first row of Table 5 replicates the first row of Table 4, with the caveat that we exclude from the publication count variable those for which citations are not available because they appear in a journal indexed by *PubMed* but not the *Web of Science*.[18] The next five rows progressively restrict the count to those whose citations put them above an impact percentile threshold: above the 50th, above the 75th, above the 95th, above the 99th, and above the 99.9th percentile. Looking across rows, the magnitude of the treatment effects increases slightly as one moves up the tail of the impact distribution (except when focusing on the one in a thousand "citation hits"). The more important conclusion is that ATP attendance increases dramatically the number of low-impact as well as the number of high-impact publications over the career.

4.4. Research style

Table 6 examines the impact of NIH training on the style of the research published by applicants to the ATP. Since the style measures cannot be computed absent publications, we limit the analysis in this section to the 2584 applicants (1730 attendees and 854 non-attendees) who publish at least once in the post-training phase of the career.[19] The effect on the overall number of publications for the restricted sample of publishers appears in the first row of Table 6 as a benchmark.

A hallmark of the training received at NIH was exposure to laboratory research for young physicians that might have had only limited exposure to the bench as undergraduates or medical school students (and might be unable to receive that style of training in postgraduate fellowships outside of NIH), with an emphasis placed in the oral history on facilitating the "bench to bedside" transition of translational research. Recall that we partition the bibliome into four mutually exclusive styles—basic science, translational medicine, clinical trials, and "other" clinical. The results imply that the program increases output regardless of style, but not evenly. The effect on the number of basic science publications is unambiguously the largest in magnitude, followed by translational and clinical trial publications, with the "other clinical" experiencing only modest and imprecisely estimated increases.[20] We also find that relative to controls, treated physicians publish much more in six high-impact journals prominently advertising

[17] When referring to the vintage-specific, article-level distribution of citations, the relevant universe to compute quantiles is not limited to the articles authored by scientists who belong to our applicant sample. Rather, the relevant universe includes the entire set of 17,312,059 articles that can be cross-linked between *PubMed* and the *Web of Science*.

[18] These account for 13,853 of 192,785 (7.2%) of all post-independence original research publications for the sample of applicants.

[19] The inverse probability of treatment and censoring weights are recomputed on the restricted sample to take into account the fact that the publication constraint disproportionately drops unsuccessful applicants from the data.

[20] Estimating these four specifications jointly enables us to compare the magnitudes explicitly. χ^2 tests strongly reject the hypothesis that the coefficient for basic science is equal to any of the other three categories ($p < 0.01$). Similarly,

Table 6
Research style.
Sources: ATP Index Cards, PubMed.

	X-Sect.	Lasso Weights	
	Naive	ATE	ATET
Career Nb. of Pubs	1.609**	1.478**	1.505**
	(0.117)	(0.112)	(0.131)
Basic Science Articles	2.787**	2.197**	2.184**
	(0.320)	(0.291)	(0.348)
Translational Medicine Articles	1.830**	1.542**	1.570**
	(0.196)	(0.179)	(0.207)
Clinical Trial Articles	1.584**	1.544**	1.674**
	(0.188)	(0.178)	(0.205)
Other Clinical Articles	1.056	1.121	1.132
	(0.092)	(0.108)	(0.121)
Articles Appearing in Translational Journals	2.545**	2.048**	2.188**
	(0.424)	(0.357)	(0.457)
Inspires Translational Research	1.799**	1.583**	1.623**
	(0.211)	(0.186)	(0.207)
Builds on Translational Research	1.692**	1.595**	1.728**
	(0.213)	(0.197)	(0.224)
Articles Cited in Patents	2.138**	1.767**	1.800**
	(0.226)	(0.215)	(0.262)
Number of applicants	2584	2584	2584

Note: Each cell contains an estimate for the treatment effect in a separate regression. All estimates stem from Poisson regressions. The dependent variables are listed in the left-most column. All models also include a full suite of medical school graduation year effects as well as an indicator variable for holding a Ph.D. degree at the time of application. Exponentiated coefficients are presented; subtracting 1 yields magnitudes interpretable as elasticities. For example, the estimate in the cell at the bottom left imply that attendees publish 100×(2.138-1) = 113.8% more articles cite by patents, relative to non-attendees; the effect is highly statistically significant. Columns 2 and 3 perform inverse probability of treatment and censoring weighted; the corresponding estimates can be interpreted as the ATE/ATET of NIH training, under the assumption of unconfoundedness. Robust errors in parentheses ($^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$).

Table 7a
Robustness analyses.

	IPTC Lasso weights				Zero Pre-ATP Pubs	
	No Winsoring	Winsoring, 95th pctl.	Winsoring, 90th pctl.	CEM	Top 10 Med Schools	Other Med Schools
Career Nb. of Pubs	1.639**	1.516**	1.457**	1.921**	2.695**	1.675**
	(0.128)	(0.124)	(0.123)	(0.219)	(0.426)	(0.235)
Log Pseudo-Likelihood	−152,478	−129,712	−113,235	−53,656	−47,084	−49,882
Number of applicants	3075	2,69	2461	1036	849	988

Note: Each cell contains an estimate for the treatment effect in a separate regression. All estimates stem from Poisson regressions. All models incorporate a full suite of medical school graduation year effects as well as an indicator variable for holding a Ph.D. degree at the time of application. Exponentiated coefficients are presented; subtracting 1 yields magnitudes interpretable as elasticities. For example, the estimate in the first column imply that attendees publish 100×(1.639-1) = 63.9% more articles during the independent phase of their career, relative to non-attendees. The first three columns vary the sample to reflect the winsorization of the inverse probability of treatment and censoring (IPTC) regression weights. In the fourth column, CEM refers to coarsened exact matching, a blocking technique to guarantee balance on a small set of covariates. The last two columns restrict sample to the set of applicants with no research experience prior to application, separately for those having graduated from elite and non-elite medical schools. Robust errors in parentheses ($^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$).

a translational focus. In addition, attendees both greatly "inspire" clinical researchers to further develop their translational work, and "stand on translational shoulders" by publishing clinical trials that backward-reference translational articles. Finally, we find that the NIH ATP increases published output that will eventually be cited in one or more USPTO patents.

Considered as a whole, these results points to a durable intellectual imprint associated with the training received at NIH. Some of the trainees became bench scientists, indistinguishable in their output from PhD-holding scientists trained in biology or other basic science departments. Harold Varmus, who went on to win the Nobel Prize in 1989 for his discovery of oncogenes with J. Michael Bishop, is an exemplar of the subset of trainees who leveraged their training to embark on a career at the laboratory bench. Many others, however, did not forsake clinical work completely, but rather acquired in Bethesda an approach to clinical research that was informed by basic research advances, seeding academia with a new generation of who saw themselves as "physician-scientists" rather than "clinician-researchers".

we can reject the hypothesis that the coefficient for translational medicine and clinical trials are equal to the coefficient for "other clinical" articles. However, we fail to reject the hypothesis that the translational medicine and clinical trial coefficients are in fact equal.

Table 7b
Robustness analyses.

	Nb. of Pubs		$\sinh^{-1}$(Nb. of Pubs)	
	IPTC Lasso Weights	Double Lasso	IPTC Lasso weights	Double Lasso
ATE	27.776** (3.981)	25.776** (4.000)	0.901** (0.114)	0.868** (0.085)
Number of Applicants	3075	3075	3075	3075

Note: Each cell contains an estimate for the average treatment effect in a separate regression. All estimates stem from OLS regressions. The dependent variable is either the number of post-training publications in levels (first pair of columns) or the inverse hyperbolic sine of the number of post-training publications (second pair of columns). The first and third columns perform inverse probability of treatment and censoring (IPTC) weighted estimation as in Table 4. The second and fourth column report an estimate of the average treatment effect using the "post-double-selection" lasso estimator due to Belloni et al. (2014). Robust errors in parentheses ($^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$).

4.5. Mechanisms

It is likely that attending the NIH ATP may impact career and research trajectories through multiple mechanisms, including skill building, signaling, status, peer, and network effects, or instilling values and aspirations (Argote and Fahrenkopf, 2016). Distinguishing between these mechanisms is difficult with the data available, and indeed more than a single mechanism might be responsible for the treatment effects we estimate.

It is notable that many physicians in the control group had exposure to research opportunities outside of the NIH; there was only a small difference in total training time compared to ATP attendees relative to the length of the program. This suggests that the NIH treatment entails more than mere exposure to research. In line with this, we repeat our main analysis, but exclude the 218 applicants who did not attend the ATP and started their independent career immediately upon finish residency training, those individuals in the control group least likely to have had substantive research exposure during postgraduate training. The treatment effect magnitudes using this control subsample are, if anything, higher than those observed when using the entire sample (Appendix B, Table B5). This reinforces our contention that the treatment effect should be interpreted as the effect of receiving NIH training relative to "traditional training" and emphasizes the importance of the content rather than the quantity of training received.

Dose–response relationship. Table B6 in Appendix B reports the results of an analysis contrasting the effect of different levels in the intensity of treatment, as proxied by the number of years spent in the ATP. Within the set of 1929 attendees, 12 (0.6%) spent a year or less at NIH, which we interpret as reflecting the decision to quit the program and receive training elsewhere; 1321 (68.5%) spent exactly two years as trainees; and 596 (30.9%) three years or more.[21] In these analyses, we model ATP attendance as a multi-valued treatment (Imbens, 2000), and use an ordered logit specification to generate inverse probability of treatment weights. The results uncover a strong dose–response relationship. Across several outcomes, "quitters" and non-attendees exhibit similar outcomes (with the caveat that the effect of quitting is very imprecisely estimated). The effect of spending an additional year within the program is large, and precisely estimated. For example, relative to non-attendees, those staying 3 years publish more over their careers (106% vs. 49%), gather more citations (153% vs. 42%) and are more likely to enter a research job after training (26% vs. 17%) than those staying only the two years necessary to fulfill their service obligation. Once again, we must interpret these results with a great deal of caution, since exposure length is endogenous, and after two years, preceptors are presumably better able to ascertain correctly the research potential of a trainee. While not rejecting selection as a plausible mechanism, this dose–response relationship appears inconsistent with an interpretation of the results based on signaling or status, since it is unlikely that additional years spent in the program would shift future employers' perceptions, or elevate one's status even more in the minds of collaborators, funders, editors, and referees. In addition, the research style evidence seems hard to reconcile with a simple status or signaling story.

Research independence. Another potential mechanism is research independence during training (Shibayama, 2019). This emphasis was reflected in the oral histories of the ATP (see Section 2.1). A particular lens on research independence is to examine the extent to which NIH trainees "outperform" their mentors once they leave the nest and become responsible for their own agenda. To provide evidence on this point within the constraints of our data, we first identify the peak vintage-adjusted citation percentile achieved by articles published during training for each trainee. Then, looking only at last-authored publications during career independence, we record the number of articles the former trainee published that exceeded this citation benchmark. Despite having higher peak citation percentiles during training, ATP attendees more frequently exceed their training peaks, relative to non-attendees (Tables B7a and B7b).

Coauthorship-driven peer effects. ATP Fellows could avail themselves to a much broader peer community than in the typical laboratory or clinical fellowship where non-attendees might have completed training, and the oral histories are replete with mentions of the concentration in talent at the NIH brought about by the draft (see Appendix E). ATP attendees may have pushed their colleagues to work harder, to hold higher internal standards of scientific excellence, or helped instill values such as the inherent worth of translational investigations. Unfortunately, while we can observe the cohort of each ATP trainee, we cannot do the same for the non-attendee controls. But we can shed light on the importance of a particular variety of peer effects, those driven by coauthorship.

In Appendix B, Table B8b, we estimate the total number of coauthored papers using a Poisson model with an offset for the total number of publications after career independence. The coefficients are small in magnitude and not statistically significant when comparing all ATP attendees to non-attendees (except for the subset of ATP attendees who stay at NIH after the completion of their training). While we do not find evidence supporting a large role for coauthorship-driven peer effects, this does not preclude an important role for other types of peer effects.

In summary, it is not feasible, within the limitations of our data, to unambiguously pin down a set of mechanisms for the treatment effect magnitudes we estimate. However, the collage of evidence above, together with the results on research style (Section 4.4) imply that skill-based explanations must have played a meaningful role in driving the outcomes we observe. In the conclusion, we argue that in spite of the tentative nature of the evidence regarding mechanisms, some of the ATP's distinctive features provide clues to policy makers as they design training programs adapted to the challenges faced by the 21st century scientific ecosystem.

[21] This last category includes a small set of about sixty attendees who transitioned from the ATP to another postdoctoral fellowship within NIH, before securing a permanent position.

4.6. Robustness analyses

We perform a number of robustness checks to probe the sensitivity of our estimates to alternative modeling assumptions and subsamples. Recall that in addition to unconfoundedness, the validity of IPTW estimates requires common support. Fig. 2 displays the histogram corresponding to the predicted probabilities generated by the selection model in column 1c of Table 2. One can readily observe that the common support assumption is violated in the tails: our model predicts a high probability of selection for very few controls, and low probability of selection for very few treated applicants. The first three columns of Table 7a vary the extent of winsorization for the regression weights: no winsorization (as in Table 4), winsorization at the 5th and 95th percentiles of the distribution of lasso weights; and winsorization at the 10th and 90th percentile of the distribution of lasso weights. The magnitudes of the average treatment effect (corresponding to a single outcome, the number of post-training publications) increases slightly. The violation of the common support assumption is therefore not a first-order concern to assess the robustness of our results.

Rather than weighting by the inverse probability of treatment, the next set of estimates uses coarsened exact matching (Iacus et al., 2011) to match attendees and non-attendees on a handful of covariates: year of medical school graduation, medical school attended, and quintile of the distribution of the pre-application publication count, weighted by journal impact factor. Any treated applicant for whom we cannot find a matched control based on this list of pre-application covariates is simply dropped from the estimation sample. We find that the estimated treatment effect is similar in magnitude to that reported earlier (Table 4).

The last set of two columns in Table 7a focuses on the subset of 1837 applicants (59.7% of the sample) who had little—if any—research preparation at the time they applied for the program, as ascertained by a lack of any published output. It is of course possible that interviewers were able to divine research potential at the second stage of the selection process, but they would not have had a strong evidentiary record to back up their intuition. The results show that the magnitude of the average treatment effect is just as high, if not higher, in this subpopulation.

Table 7b reports estimates using the "post-double-selection" lasso (hereafter pds-lasso) estimator due to Belloni et al. (2014). This estimator uses the lasso to select covariates to predict both the treatment and the outcome variable, and then estimates the treatment effect of interest by the linear regression of the outcome on the treatment variable and the union of the set of variables selected in the two variable selection steps. The resulting estimator is "doubly robust" in that it allows for imperfect variable selection in either (but not both) of the covariate selection steps. Since the theoretical properties of the pds-lasso estimator have been demonstrated for a linear model, we apply it to our data using ordinary least squares to model the impact of the NIH ATP on the count of post-training publications.[22] The estimates yielded by this procedure are once again large in magnitude, very similar to those associated with IPTW estimation using OLS, and precisely estimated. The point estimate of 26 extra publications, corresponds to 64% of the raw mean difference in the number of publications between attendees and non-attendees.

We also use the bounding technique recently proposed by Oster (2019) to gauge the sensitivity of our results to a failure of the unconfoundedness assumption. The intuition behind this approach is that the stability of the coefficient for the treatment effect when varying the set of control variables included in the model, scaled by movement in R^2, provides information about the potential impact of unobserved covariates. To generate these bounds, the analyst must assume proportionality between the covariances of the outcome with observed and unobserved covariates, and posit a maximum value for R^2 if the regression could include all observed and unobserved covariates. Oster's technique generates a bound δ, the covariance ratio that would be required to reduce the magnitude of the treatment effect to zero. Table 7c reports the results of this exercise for a number of research outcomes. In all cases, δ is far above one, the threshold value recommended by Oster to suggest robustness to the influence of unobservable covariates.

Table 7c
Robustness analyses.

	Oster's δ
Career Nb. of Pubs	1.743
Career Nb. of Pubs, Top 5% of the Cit. Distrib.	1.789
Career Nb. of Pubs, Top 1% of the Cit. Distrib.	1.767
Career Citations	1.748
Nb. of Patents	2.282
Career NIH Grants ($ 2015)	1.516

Note: The score reported corresponds to the δ parameter from Oster (2019), the ratio between the covariances of the outcome with observed and unobserved covariates, respectively. All outcomes are transformed using the inverse hyperbolic sine function, and δ is computed using OLS regression and the list of covariates selected by the pds-lasso estimator of Belloni et al. (2014), and chosen to produce an estimate of the treatment effect equal to zero. We follow Oster's recommendation of setting $R^{max} = 1.3 \times R^2$ from the fully saturated specification.

Appendix B includes a number of other robustness checks and ancillary analyses, including: isolating the effect of informative censoring from selection into treatment (Table B10), dynamics of treatment effect over time (Figure B5a and B5b), evidence of imprinting during training (Table B11), heterogeneity in treatment effect by year of program attendance (Table B12), and heterogeneity in treatment effect by program track within the ATP (Table B13).

5. Conclusion

We examine the role of early career exposure to research on sorting into the "ideas sector" of the economy, as well as research trajectory and productivity within this domain. The NIH ATP had a large impact on attendees' careers on both the intensive and extensive margins. Attendees entered research positions at higher rates after training and remained in them for longer. They not only published more and earned more grant funding, their influence persists through training more second generation scientists and their work was more impactful as measured by citations. More specifically, ATP attendees acquired at NIH a more "translational" style of research, with a greater focus on the bench-to-bedside transition. Remarkably, these changes were sustained throughout their subsequent careers. It is notable that, while there are more "superstars" among ATP attendees than in the set of non-attendee controls, the average physician showed a substantial treatment effect as well. All in all, it is a remarkable impact for a two- to three-year training experience.

Our conclusions depend on the maintained assumption that, conditional on an extensive list of covariates observable at the time of application, selection into the program was essentially random. At first blush, this would appear to be an untenable assumption. While we have adopted a variety of econometric strategies to minimize omitted variable bias, we recognize that at least some of our results could be explained by factors observed by the scientists in charge of selecting the trainees, but not by the econometric analyst. Yet, the institutional setting and the details of the selection process suggest that decision-makers were equally unaware of whom, among the applicants, was decidedly poised for research greatness.

Our control group includes only those who have also applied to the program, which eliminates interest in the program as a potential omitted variable (Jones et al., 2019). In addition, the set of non-attendee

[22] We also use the inverse hyperbolic sine function to transform the publication count. This generates estimates that can approximately be interpreted as elasticities, and therefore compared to those presented in Table 4.

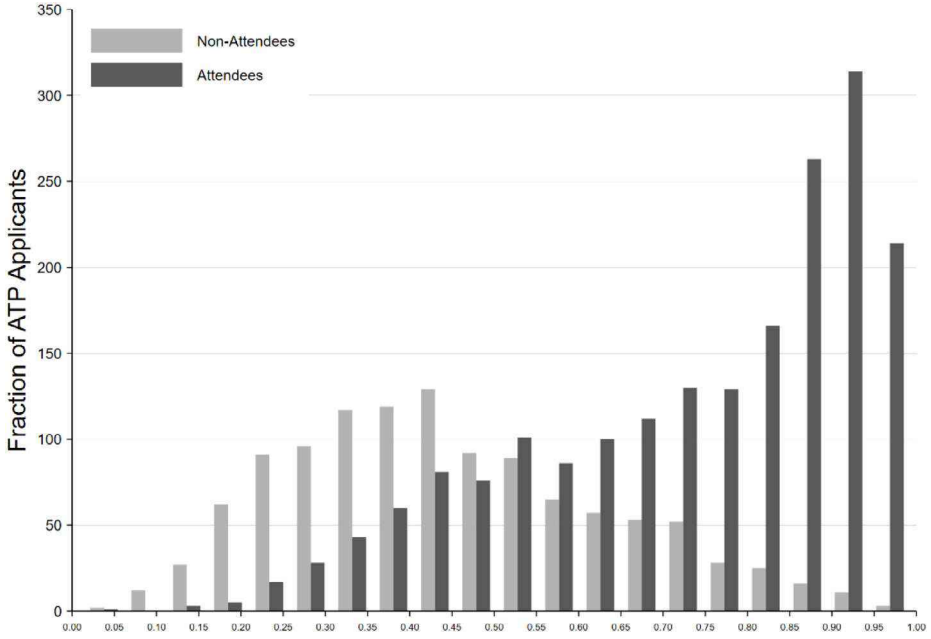

Fig. 2. Predicted probability of selection. *Note*: Predicted probabilities from the lasso penalized logit procedure described in the last paragraph of Section 4.1 of the manuscript.

controls consists exclusively of those who reached the final interview stage for program admittance and are therefore already highly selected. While we would of course prefer to have interview notes to model the influence of unobservable covariates directly, a large literature suggests that unstructured interviews provide only limited additional information, relative to what is observable on a curriculum vitæ (Dana et al., 2013; Huffcutt et al., 1996; McDaniel et al., 1994; Wiesner and Cronshaw, 1988; Wright et al., 1989). In fact, psychological research has shown that the addition of noisy signals may in fact impair the quality of decision making (Hall et al., 2007; Nisbett et al., 1981). Our reading of this literature leads us to doubt that the unstructured NIH ATP interviews enabled the selection of individuals poised for research greatness. Indeed, medical education is one of a handful of settings where the limited usefulness of interviews has been documented in the field (Milstein et al., 1981).[23] In line with this literature, the oral histories corroborate the difficulty faced by the interviewers in discerning the scientific potential of applicants at such an early career stage. Finally, the evidence on research style does not appear to be consistent with the view that selection alone accounts for the results. It strains credulity that the demand side of this labor market might have been able to evaluate aptitude for translational research specifically, in addition to more general research abilities.

Many of the ATP alumni's oral histories evoke the feeling of "being in the right place at the right time". In light of these accounts, the sociological concept of imprinting offers a powerful lens to interpret our results. This stream of research finds that organizations and individuals often exhibit a sensitive period, during which they are susceptible to external influences and come to reflect aspects of this environment, and these aspects can persist despite subsequent environmental changes (Marquis and Tilcsik, 2013; Stinchcombe, 1965). While much of the work on imprinting has focused on firms, there is evidence that imprinting also occurs in the context of individual careers (Baron et al., 1999; Boeker, 1988; Burton and Beckman, 2007; Hannan et al., 1996; Higgins, 2005). During career imprinting, individuals absorb a set of capabilities, connections, and cognitive models from one employer which persist as they change employers later on. Careers are more likely to exhibit the characteristics of an early imprint when their current environment allows them to be surrounded by colleagues with the same imprint, offers them considerable freedom in how they might express an imprint, and if they believe the imprint contributed to prior success (Higgins, 2005). The NIH ATP and the academic medicine context would appear particularly conducive to career imprinting: not only was the ATP an intense experience early in the career, when an imprint is more likely to be absorbed, but the program also had many alumni who seeded the expansion of U.S. Medical Schools in the period immediately following the end of the Vietnam War.[24] Finally, academic research offers a considerable degree of leeway to investigators in structuring the direction and style of their research, and the senior NIH investigators who had acted as mentors to the ATP trainees during the program exemplified the creative use of this autonomy.

The pool of ATP applicants is not diverse by today's standards.[25] This may pose a challenge to external validity if members of different socioeconomic groups respond differently to the mechanisms of the ATP treatment effect. In particular, interventions that provide a status boost to young scientists can have very different impacts along gender lines (Graddy-Reed et al., 2019). While status is one potential mechanism for the NIH ATP treatment effect, the "dose–response" relationship between the length of training at the ATP and career outcomes argues strongly against an interpretation of the effect mostly, or solely reflecting status considerations. Rather, this evidence is consistent with the idea that trainees are durably imprinted with

[23] For instance, the University of Texas Medical School at Houston was forced to admit an additional 50 students, all of whom were initially rejected for admission post-interview, due to a legislative decree in 1979; these students had no meaningful difference in clinical performance, academic performance and honors, or attrition at either the end of medical school or the first year of postgraduate training (Devaul et al., 1987).

[24] Between 1975 and 2005, the number of faculty members at US Medical Schools increased by a factor of more than two (AAMC Data Book, various editions; (Jolly, 1988)).

[25] We came across two African–American physicians in the entire sample. Similarly, we found 36 female applicants (include 2 foreign medical graduates and 1 lost to follow up) in the NIH index cards (15 attendees, 21 non-attendees), who may have been discriminated against in the application process because any spot occupied by a woman entailed that a male physician would serve in the armed forces, possibly in the Vietnam theater (although in fact few among our control appear to have served in South Asia, if they did serve at all).

specific research skills during their stay at NIH. If this view is correct, then there is less reason to fear the findings would not be observed in a more demographically-balanced cohort of trainees.

In light of the unique historical circumstances within which physician research training took place at NIH during the period of our study, we must exercise caution to suggest wider policy implications.[26] Certainly, part of the effectiveness of the ATP in turning physicians into researchers owes much to the extreme concentration of talent in one institution that was facilitated by the Vietnam War. The effects of the ATP may have been large and long-lasting precisely because the exposure received was intense. Yet, this program provides an existence proof for the proposition that it is possible to design interventions to turn individuals who in the main would not have had scientific careers into frontier researchers. This stands in contrast with many other active labor market policies often studied by economists. The effects of these programs are typically modest in magnitude, and their effects relatively transitory (Heckman et al., 1999). Conversely, the labor market effects of military service appear to mostly correspond to loss of experience, as the earnings profiles of veterans and non-veterans converge relatively quickly (Angrist, 1990; Angrist et al., 2011).

There have been attempts to recreate the "hot house" environment that characterized the intramural campus of the NIH in the 1960s and 1970s (Rubin, 2006). But which characteristics of the NIH ATP were instrumental in its ability to push attendees towards the heights of the biomedical research elite? The unique set of circumstances is unlikely to occur again, and it would be depressing to suggest that the exigencies of wartime are a necessary condition for the design of effective scientific training programs. Such pessimism is not warranted. We emphasize three features of the ATP relevant for the design of training programs today, within and beyond the setting of biomedicine. First is the timing of training receipt, which for many ATP attendees was their first serious engagement with scientific investigation. In many respects, the ATP was more akin to a "pre-doc" than a graduate school or postdoctoral experience. Second is the size of each cohort. The ATP cohorts were much larger than those in the typical Medical Scientist Training Program or other research fellowships. This may intersect in important ways with the role of peer effects, facilitated by the concentration of talent in one location during the NIH ATP. Third, the ATP stressed building independence. This is very different from the modern setting, where typically all papers are automatically coauthored with the principal investigator, there is often little scope to deviate from the principal investigator's research agenda, and many budding scientists linger in training, typically in a sequence of post-doctoral positions (Kahn and Ginther, 2017).

Yet, despite these distinctive features, it is difficult to offer firm guidance for scientific training programs. Our evidence unfortunately does not allow us to empirically isolate the individual mechanisms explaining the effect of ATP attendance. This is an opportunity. The very success of the ATP suggests policy makers should experiment with design features that were its hallmarks. We conclude with a call for more systematic and rigorous evaluation of training programs. It is an unfortunate paradox that Randomized Control Trials (RCTs) are a staple of the biomedical research enterprise, and yet seem to be viewed as out of place in the context of funding and training policies. In our view, the lowest hanging fruit available to designers of training programs—especially those with more applicants than available seats—is to build in evaluation in the design phase, instead of treating it as an afterthought.

[26] Appendix E contains a discussion of the estimated program costs and return on investment.

CRediT authorship contribution statement

Pierre Azoulay: Conceptualization, Methodology, Investigation, Formal analysis, Writing – review & editing. **Wesley H. Greenblatt:** Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Misty L. Heggeness:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Data: Appendices A through H

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.respol.2021.104332.

References

Aghion, Philippe, Howitt, Peter, 1992. A model of growth through creative destruction. Econometrica 60 (2), 323–351.

Anfinsen, Christian B., 1963. History of the Research Associate Program. Office of NIH History.

Angrist, Joshua D., 1990. Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. Am. Econ. Rev. 80 (3), 313–336.

Angrist, Joshua D., Chen, Stacey H., Song, Jae, 2011. Long-term consequences of Vietnam-era conscription: New estimates using social security data. American Economic Review 101 (3), 334–338.

Argote, Linda, Fahrenkopf, Erin, 2016. Knowledge transfer in organizations: The roles of members, tasks, tools and networks. Organ. Behav. Hum. Decis. Processes 136, 146–159.

Association of American Medical Colleges, 2016. Diversity in Medical Education: Facts & Figures 2016. AAMC, Washington, D.C.

Austin, Peter C., Stuart, Elizabeth, 2015. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat. Med. 34 (28), 3661–3679.

Azoulay, Pierre, Ding, Waverly, Stuart, Toby, 2009. The effect of academic patenting on the rate, quality, and direction of (public) research output. J. Ind. Econ. 57 (4), 637–676.

Baron, James N., Diane Burton, M., Hannan, Michael T., 1999. Engineering bureaucracy: The genesis of formal policies, positions, and structures in high-technology firms. J. Law Econ. Organ. 15 (1), 1–41.

Baskir, Lawrence N., Strauss, William A., 1978. Chance and Circumstance: The Draft, the War, and the Vietnam Generation. Vintage Books, New York, NY.

Bell, Alexander M., Chetty, Raj, Jaravel, Xavier, Petkova, Neviana, Reenen, John Van, 2019. Who becomes an inventor in America? The importance of exposure to innovation. Q. J. Econ. 134 (2), 647–713.

Belloni, Alexandre, Chernozhukov, Victor, Hansen, Christian, 2014. Inference on treatment effects after selection among high-dimensional controls. Rev. Econom. Stud. 81 (2), 608–650.

Berry, Frank B., 1976. The Story of 'The Berry Plan'. Bull. New York Acad. Med. 52 (3), 278–282.

Blau, David M., Weinberg, Bruce A., 2017. Why the US science and engineering workforce is aging rapidly. Proc. Natl. Acad. Sci. 114 (15), 3879–3884.

Blume-Kohout, Margaret E., Adhikari, Dadhi, 2016. Training the scientific workforce: Does funding mechanism matter? Res. Policy 45 (6), 1291–1303.

Boeker, Warren, 1988. Organizational origins: Entrepreneurial and environmental imprinting at time of founding. In: Carroll, Glenn R. (Ed.), Ecological Models of Organizations. Ballinger, Cambridge, MA, pp. 33–51.

Broström, Anders, 2019. Academic breeding grounds: Home department conditions and early career performance of academic researchers. Res. Policy 48 (7), 1647–1665.

Burton, M. Diane, Beckman, Christine M., 2007. Leaving a legacy: Position imprints and successor turnover in young firms. Am. Sociol. Rev. 72 (2), 239–266.

Butler, Declan, 2008. Translational research: Crossing the valley of death. Nature 453 (7197), 840–842.

Bynum, William, 2012. What makes a great lab? Nature 490 (7418), 31–32.

Chernozhukov, Victor, Chetverikov, Denis, Demirer, Mert, Duflo, Esther, Hansen, Christian, Newey, Whitney, Robins, James, 2018. Double/debiased machine learning for treatment and structural parameters. Econom. J. 21 (1), C1–C68.

Crowell, James A., 1971. The Procurement of Physicians for the United States Armed Forces. Army War College Working Paper AD-769 594.

Dana, Jason, Dawes, Robyn, Peterson, Nathanial, 2013. Belief in the unstructured interview: The persistence of an illusion. Judgm. Decis. Mak. 8 (5), 512–520.
Dance, Amber, 2012. Who's on first? Nature 489 (7417), 591–593.
Deming, David J., Noray, Kadeem, 2020. Earnings dynamics, changing job skills, and STEM careers. Q. J. Econ. 135 (3), 1965–2005.
Devaul, Richard, Jevey, Faith, Chappell, James, Caver, Patricia, Short, Barbara, O'Keefe, Stephen, 1987. Medical school performance of initially rejected students. JAMA 257 (1), 47–51.
Fauci, Anthony S., 1989. In: Harden, Victoria (Ed.), In Their Own Words AIDS Oral History Series, Office of NIH History, National Institutes of Health.
Fauci, Anthony S., 1998. In: Klein, Melissa (Ed.), Clinical Associates Program Oral History Series, Office of NIH History, National Institutes of Health.
Fredrickson, Donald S., 1998. In: Klein, Melissa (Ed.), Clinical Associates Program Oral History Series, Office of NIH History, National Institutes of Health.
Freeman, Richard B., 1975. Supply and salary adjustments to the changing science manpower market: Physics, 1948-1973. Amer. Econ. Rev. 65 (1), 27–39.
Ginther, Donna K., Currie, Janet M., Blau, Francine D., Croson, Rachel T.A., 2020. Can mentoring help female assistant professors? An evaluation by randomized trial. Am. Econ. Assoc. Pap. Proc. 110 (5), 205–209.
Ginther, Donna K., Heggeness, Misty L., 2020. Administrative discretion in scientific funding: Evidence from a prestigious postdoctoral training program. Res. Policy 49 (4), 103953.
Goldstein, Joseph L., Brown, Michael S., 1997. The clinical investigator: Bewitched, bothered, and bewildered—But still beloved. J. Clin. Invest. 99 (12), 2803–2812.
Gouriéroux, Christian, Montfort, Alain, Trognon, Alain, 1984. Pseudo maximum likelihood methods: Applications to Poisson models. Econometrica 53 (3), 701–720.
Graddy-Reed, Alexandra, Lanahan, Lauren, Eyer, Jonathan, 2019. Gender discrepancies in publication productivity of high-performing life science graduate students. Res. Policy 48 (9), 103838.
Hall, Crystal C., Ariss, Lynn, Todorov, Alexander, 2007. The illusion of knowledge: When more information reduces accuracy and increases confidence. Organ. Behav. Hum. Decis. Processes 103 (2), 277–290.
Hannan, Michael T., Diane Burton, M., Baron, James N., 1996. Inertia and change in the early years: Employment relations in Young, high technology firms. Ind. Corp. Change 5 (2), 503–536.
Heckman, James J., Lalonde, Robert J., Smith, Jeffrey A., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, Orley C., Card, David (Eds.), Handbook of Labor Economics, vol. 3A. Elsevier North-Holland, Amsterdam, pp. 1865–2097 (Chapter 31).
Higgins, Monica C., 2005. Career Imprints: Creating Leaders Across an Industry. Jossey-Bass/Wiley, San Francisco.
Hill, Ryan Reed, 2018. Searching for Superstars: Research Risk and Talent Discovery in Astronomy. Working Paper, MIT.
Hirano, Keisuke, Imbens, Guido W., 2001. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Serv. Outcomes Res. Methodol. 2, 259–278.
Huffcutt, Allen I., Roth, Philip L., McDaniel, Michael A., 1996. A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. J. Appl. Psychol. 81 (5), 459–473.
Iacus, Stefano M., King, Gary, Porro, Giuseppe, 2011. Multivariate matching methods that are monotonic imbalance bounding. J. Amer. Statist. Assoc. 106 (493), 345–361.
Imbens, Guido W., 2000. The role of the propensity score in estimating dose–response functions. Biometrika 87 (3), 706–710.
Jacob, Brian A., Lefgren, Lars, 2011. The impact of NIH postdoctoral training grants on scientific productivity. Res. Policy 40 (6), 864–874.
Jolly, Paul, 1988. Medical education in the United States, 1960-1987. Health Affairs 7 (Suppl. 2), 144–157.
Jones, Charles I., 1995. R&D-based models of economic growth. J. Polit. Econ. 103 (4), 759–784.
Jones, Damon, David, Molitor, Reif, Julian, 2019. What do workplace wellness programs do? Evidence from the Illinois workplace wellness study. Q. J. Econ. 143 (4), 1747–1791.
Kahn, Shulamit, Ginther, Donna K., 2017. The impact of postdocs on early careers in biomedicine. Nature Biotechnol. 35 (1), 90–94.
Keiser, Harry, 1998. In: Klein, Melissa (Ed.), Clinical Associates Program Oral History Series, Office of NIH History, National Institutes of Health.
Kerr, William R., 2018. The Gift of Global Talent: How Migration Shapes Business, Economy and Society. Stanford University Press.
Khot, Sandeep, Park, Buhm Soon, Longstreth Jr., W.T., 2011. The Vietnam war and medical research: Untold legacy of the U.S. doctor draft and the NIH yellow berets. Acad. Med. 86 (4), 502–508.
Kimball, Harry R., 1997. In: Klein, Melissa (Ed.), Clinical Associates Program Oral History Series, Office of NIH History, National Institutes of Health.
Klein, Melissa K., 1998. The Legacy of the 'Yellow Berets': The Vietnam War, the Doctor Draft, and the NIH Associate Training Program. Manuscript. NIH History Office, National Institutes of Health.
Marquis, Christopher, Tilcsik, Andras, 2013. Imprinting: Toward a multilevel theory. Acad. Manag. Ann. 7 (1), 193–243.
Marx, Matt, Fuegi, Aaron, 2020. Reliance on science: Worldwide front-page patent citations to scientific articles. Strateg. Manag. J. 41 (9), 1572–1594.
McDaniel, Michael A., Whetzel, Deborah L., Schmidt, Frank L., Maurer, Steven D., 1994. The validity of employment interviews: A comprehensive review and meta-analysis. J. Appl. Psychol. 79 (4), 599–616.
Milojevic, Staša, Radicchi, Filippo, Walsh, John P., 2018. Changing demographics of scientific careers: The rise of the temporary workforce. Proc. Natl. Acad. Sci. 115 (50), 12616–12623.
Milstein, Robert M., Wilkinson, Leland, Burrow, Gerard N., Kessen, William, 1981. Admission decisions and performance during medical school. J. Med. Educ. 56 (2), 77–82.
Nathan, David G., 2005. The several Cs of translational clinical research. J. Clin. Invest. 115 (4), 795–797.
NIH, 1968. Associate Training Program in the Medical and Biological Sciences at the National Institutes of Health. Department of Health, Education, and Welfare.
NIH Office of Research Information, 1963. New Class of 101 New Physicians Join NIH Research Training Programs. NIH Office of Research Information, National Institutes of Health.
Nisbett, Richard E., Zukier, Henry, Lemley, Ronald E., 1981. The dilution effect: Nondiagnositc information weakens the implications of diagnostic information. Cogn. Psychol. 13 (2), 248–277.
Oster, Emily, 2019. Unobservable selection and coefficient stability: Theory and evidence. J. Bus. Econom. Statist. 37 (2), 187–204.
Park, Buhm Soon, 2003. The development of the intramural research program at the national institutes of health after world war II. Perspect. Biol. Med. 46 (3), 383–402.
Rhodes, Richard, 1986. The Making of the Atomic Bomb. Touchstone, New York.
Robins, James M., Hernan, Miguel A., Brumback, Babette, 2000. Marginal structural models and causal inference in epidemiology. Epidemiology 11 (5), 550–560.
Romer, Paul M., 1990. Endogenous technological change. J. Polit. Econ. 98 (5), S71–S102.
Rousselot, Louis M., 1971. Doctor draft. Archives of Surgery 102 (1), 88–89.
Rubin, Gerald M., 2006. Janelia farm: An experiment in scientific culture. Cell 125 (2), 209–212.
Santos Silva, João M.C., Tenreyro, Silvana, 2006. The log of gravity. Rev. Econ. Stat. 88 (4), 641–658.
Sauermann, Henry, Haeussler, Carolin, 2017. Authorship and contribution disclosures. Sci. Adv. 3 (11), e1700404.
Shibayama, Sotaro, 2019. Sustainable development of science and scientists: Academic training in life science labs. Res. Policy 48 (3), 676–692.
Snyder, Thomas D. (Ed.), 1993. 120 Years of American Education: A Statistical Portrait. U.S. Department of Education, Washington, D.C.
Solow, Robert M., 1957. Technical change and the aggregate production function. Rev. Econ. Stat. 39 (3), 312–320.
Stinchcombe, Arthur L., 1965. Social structure and organizations. In: March, James G. (Ed.), Handbook of Organizations. Rand McNally, Chicago, IL, pp. 142–193.
Svorencík, Andrej, 2014. MIT's rise to prominence: Outline of a collective biography. Hist. Political Econ. 46 (Suppl. 1), 109–133.
Teitelbaum, Michael S., 2014. Falling Behind? Boom, Bust and the Global Race for Scientific Talent. Princeton University Press, Princeton, NJ.
Urschel, John, Thomas, Louisa, 2019. Mind and Matter: A Life in Math and Football. Penguin Group, New York.
Varmus, Harold, 2009. The Art and Politics of Science. W. W. Norton & Company.
Wiesner, Willi H., Cronshaw, Steven F., 1988. A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. J. Occup. Psychol. 61 (4), 275–290.
Woolf, Steven H., 2008. The meaning of translational research and why it matters. JAMA 299 (2), 211–213.
Wright, Patrick M., Lichtenfels, Philip A., Pursell, Elliot D., 1989. The structured interview: Additional studies and a meta-analysis. J. Occup. Psychol. 62 (3), 191–199.
Wyngaarden, James B., 1979. The clinical investigator as an endangered species. New England J. Med. 301 (23), 1254–1259.
Xu, Stanley, Ross, Colleen, Raebel, Marsha A., Shetterly, Susan, Blanchette, Christopher, Smith, David, 2010. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health 13 (2), 273–277.