Advisor-Advisee Research Overlap and Its Implications for Scientists' Early-Career Performance in the U.S.

ABSTRACT

A genealogical training process, in which senior (advisor) scientists engage in cross-generational transfer of skills and knowledge to junior (advisee) scientists is one of the core organizational features of modern science. In this paper, we examine the consequences of the tension faced by all junior scientists: to build upon an advisor's skills or to strike out on one's own? We study the implications of advisor-advisee research overlap for emerging scientists' performance by constructing a novel, bibliometric-record-based dataset on 15,271 U.S. biomedical scientists (advisees) who were trained in 7,924 PI advisors' labs between 1972 and 2009. We assessed the junior scientists' performance in the first ten years of their careers as independent PIs. Tests across multiple research-overlap measures and model specifications reveal a consistently positive relationship between maintaining a higher degree of proximity to advisor's research areas and the junior scientist's early-career funding and publication performance. However, evidence is weak regarding scientific impact and non-existent regarding research disruptiveness. We further test how advisor status moderates the research overlap-performance relationship using both a large-sample analysis comparing the performance of academic siblings, and a more stringent difference-in-difference analysis leveraging the exogenous timing of the status elevation events experienced by the advisor scientists when they receive major scientific awards. Both tests yield consistent evidence that the positive relationship between advisor-advisee research overlap and advisee's early-career performance is reduced as the advisor's status increases. Taken together, these findings provide a more complete understanding of how advisor-advisee relationships shape new scientists' performance during early careers.

Advisor-Advisee Research Overlap and Its Implications for Scientists' Early-Career Performance in the U.S.

"The social ties between scientific masters and apprentices are, as we shall see, enduring and consequential, for it is in the course of apprenticeships that young scientists learn the scientific role."
(The Nobel Elite; H. Zuckerman; pp. 96)

INTRODUCTION

A quintessential element in the organization of modern science is a genealogical training process whereby junior scientists (advisees) are apprenticed with senior scientists (advisors) to acquire knowledge and tacit skills on the road to independent scholarship. Over the years, a significant amount of research has been devoted to the examination of scientific training and the production of knowledge, resulting in a broad array of generalized insights, including social stratification (Cole and Cole 1973), the bureaucratization of work (Blau 1994), and the emergence of academic entrepreneurship (Azoulay et al. 2017, Stuart and Ding 2006), to name a few topics among many.

Broadly, two avenues of inquiry have sustained research interest in the scientific training process. First, a core mandate of scientific training rests upon the cross-generational transfer of complex knowledge and skills from the advisor to the advisee (Partha and David 1994, Roche 2023, Sauermann and Roach 2012). Implicit in this transmission process is the notion of replication, whereby the tacit skills of the scientific trade are propagated from one generation to the next. Concomitantly, a second mandate for the junior scientist is the deployment of acquired knowledge and skills to prospect across uncharted territory, mapping out novel patterns beyond the existing body of knowledge passed down from the advisor (Uzzi et al. 2013). Consequently, one of the hallmarks of a mature, successful scientist is intellectual independence.

In this paper, we argue that this dual mandate—the requirement both to build upon an advisor's skills or to strike out on one's own—represents a fundamental tension shaping the research agendas of junior scientists. We instantiate this tension in our core research question: *whether and when an advisee's research agenda should overlap with their advisor's*. Theoretically, this question of advisor-advisee research overlap is situated at the nexus of a number of literatures ranging from those that foster research overlap (e.g., knowledge transfer and acquisition, as in Blackburn et al. 1981, Liénard et al. 2018) to those that preclude it (e.g., niche-based competition, as in Baum and Singh 1994, Podolny et al. 1996, Popielarz and McPherson 1995). Topically, a

better understanding of advisor-advisee research overlap and its consequences is of central importance to our understanding of the organization of science, as well as the rate and direction of innovative research (Rosenberg 1974). Lastly, we would be remiss to ignore the consequences of managing this tension, which comes at a particularly critical juncture in the careers of junior scientists as they transition into the role of new principal investigators and are under pressure to obtain independent research funding for their own labs.

Despite the salience of these issues, existing theory and evidence on the topic of advisor-advisee research overlap remains both thin and mixed. In support of high overlap, one older study, Blackburn et al. (1981), reported the success of advisees who were similar to the advisors based on qualitative evidence from a small sample. A newer study, Patsali et al. (2024) based on a sample of French STEM PhDs yielded mixed results, indicating both advantages and disadvantages to low overlap (i.e., advisee independence). This mixed evidence is mirrored in two large-scale studies that have included advisor-advisee research overlap as a control variable. Liénard et al.'s (2018) study of 18,856 biomedical researchers drawing on a crowd-sourced, voluntary-report-based dataset found that greater publication *similarity* between a trainee and his or her advisor (mentor) is associated with a higher probability of the trainee's obtaining academic placement. On the other hand, in Ma et al.'s (2020) models, research topic *dissimilarity* between a protégé and his or her advisor is significantly and positively related to the odds of the protégé later becoming an academic superstar (as measured by citations, NAS membership, and winning scientific prizes and awards). As advisor-advisee research overlap was not the central focus in either study, neither article delved into unpacking the mechanisms driving either the positive or negative consequences of advisor-advisee research overlap.

We undertake a three-step approach to examine the issue of advisor-advisee research overlap. First, we constructed a novel panel dataset of 15,271 new biomedical faculty members (advisees) at U.S. research universities who were trained in the labs of 7,924 PI advisors between 1972 and 2009. We then utilize this dataset to ask, broadly, whether there are positive (or negative) consequences of advisor-advisee research overlap. Lastly, we turn to an established stream of literature on status-based competition (e.g., Bothner et al. 2007, Piezunka et al. 2018, Reschke et al. 2018) to elucidate one mechanism underlying the conditions when research overlap is (or is not) particularly beneficial to junior scientists. Specifically, we hypothesize that changes

in an advisor's status can shift the competitive dynamic between the advisor and the advisee, and that increases in competitive pressure are more salient under conditions of high research overlap.

Our dependent variable of interest is the grant-funding performance of advisees, as grant funding is both an essential component for the survival of any academic lab as well as a key metric for promotion. We focus on U.S. grant-funding in the first ten years after an advisee departs from his or her advisor's lab, as this is the period when research imprinting on the part of the advisor remains most salient (Marquis and Tilcsik 2013). To complement grant-funding, we also examine an array of alternative performance indicators, including publication counts, citations, and the disruptiveness of research of the advisee scientists.

To begin to unpack the conditions under which research overlap is beneficial, we explore status-based competition between advisor and advisee through a series of relatively exogenous shifts in advisor status upon their receipt of important awards. Using a full-sample analysis with two discrete measures of status, as well as a difference-in-difference analysis with a smaller, more restricted sample, we demonstrate that while advisor-advisee research overlap is largely beneficial to advisees with modal advisors, it has different implications for advisees with high-status advisors. For these advisees, there is a weaker link between research overlap and early-career performance, implying that advisees with high-status advisors having the freedom to tackle a broader array of scientific topics.

RESEARCH QUESTION AND HYPOTHESIS

Advisor-advisee Research Overlap and Performance

Genealogical relationships—apprenticeships between an advisee and his or her advisor—remain the cornerstone of training in academia. For advisors, the reward system in science confers esteem upon advisors whose advisees develop into eminent scientists themselves. However, a preoccupation with academic phylogenies is not simply narcissistic introspection on the part of scientists (e.g., academictree.org). Genealogical relationships yield manifold consequences, particularly for advisees. Mentorship ties have been shown to shape perceptions and norms regarding commercial and entrepreneurial science, valuable social networks, as well as status and prestige (Azoulay et al. 2017, Long and McGinnis 1985, Roche 2023, Stuart and Ding 2006). Advisor-advisee mentorship relationships have even been suggested to present a more accurate

reflection of a scientist's true impact, as they capture both direct (knowledge creation) as well as indirect (scientist training) effects of scientific research (Barres 2013, Liénard et al., 2018, Ma et al., 2020, Malmgren et al. 2010). Referencing the difficulties of quality assessment by non-specialist outsiders, Partha and David (1994) point to the twin pillars of scientist training and scientist evaluation (by both the advisor and the broader scientific community) as central, irreplaceable functions of academia.

While a plethora of reasons may underpin the two-sided match between an advisor and an advisee, there is little doubt that a central, motivating element is a quest for training: the junior scientist hopes to acquire useful skills from his/her advisor. However, the genealogical training process encompasses far more than the transfer of directly applicable knowledge and skills. For instance, Zuckerman (1977) emphasized that "one point on which the [Nobel] laureates are largely agreed is that the least important aspect of their apprenticeship was the acquiring of substantive knowledge from their master… … [training] involves acquiring the norms and standards, the values and attitudes, as well as the knowledge, skills, and behavior patterns associated with particular statuses and roles." [pp 123]. Put more simply, Markus Meister (1998) stated that "…their [Howard Berg and Denis Baylor] thinking and their attitude towards science and also towards people really framed my view of things…., … in practice I'm following a lot of the same attitudes…" In short, even more important than the acquisition of codified knowledge lies the assimilation of tacit knowledge and skills that serve to blossom the advisee into an independent scholar.

Polanyi (1983, p. 4) defines tacit knowledge as the type of knowledge that "we can know more than we can tell." Tacit knowledge cannot be transmitted via a set of codified instructions or manuals. Accordingly, tacit knowledge is often only developed gradually over time, through extensive trial and error. Within the context of scientific inquiry, examples of such tacit skills include the ability to identify good research questions, the facility to solve problems by recombining or redeploying existing methods or tools in new and creative ways, and the communication of complex discoveries to others (including reviewers of manuscripts and grant proposals) in a clear and convincing manner.

In contrast to codified knowledge, tacit knowledge is intrinsically difficult to transfer between individuals. To ease the transmission of tacit skills, it is often necessary to be situated in an environment where

direct, firsthand interactions occur between the pair. This notion of inducing situations that foster training is reflected by an advisee stating "When I came to his laboratory, he [advisor] said. 'What would you like to do? Do you have a problem that you'd like to work on?' I said: 'There are things I had in mind that I could do but I would rather work with something you are interested in.' This is advice I'd give to anyone. You get far more out of the experience of training… …. if you work on something he is interested in" [Zuckerman, 1977, pp 139]. In the context of training within scientific labs, we expect that the degree of research overlap plays a similarly important role in determining transfer effectiveness, as the senior (advisor) scientist is presumably more capable of helping the junior (advisee) scientist when he or she has strong expertise in the domain where the advisee is developing his or her own research projects. Meaningful interactions between the advisor and the advisee are more likely to occur when the pair works on similar questions, easing the transmission of tacit skills between the two. Consistent with this theorizing, past research on organizations has suggested a mentor's effect on subsequent advisee performance is contingent on collaboration (Long and McGinnis 1985), and that domain similarity between advisor and advisee is a critical scope condition in facilitating the efficient transmission of tacit knowledge (Ribeiro 2013). Similarly, early career physicists who switch topics away from their graduate work experience detrimental effects (Zeng et al. 2019).

While there is little doubt that benefits exist to advisor-advisee research overlap, (at least) two reasons lead us to suspect that a high level of advisor-advisee research overlap may also have negative implications for the advisee's future performance. First, there is an imperative to demonstrate intellectual independence on the part of junior scientists. Novelty is a key indicator of scientific creativity and intellectual dexterity (Simonton 2004). This notion of the need for intellectual independence is reflected in the positive association between research independence and the odds of starting an academic career among French STEM PhDs (Patsali et al. 2024). This mandate for independence is reflected by Carolyn Bertozzi, a Nobel chemistry laureate, who simply stated that "… if you apply for a job with proposals that look like what your old advisor did, and it's too close, you won't get a job" (Bertozzi, 2003). Research independence is not only important for job placements, but also for subsequent funding. Linking the nature of advisor-advisee research overlap to grant resources, Tom Alber, a renowned structural biologist at UC Berkeley, stated that "There's no way that

somebody would fund my lab to work on point mutants of lysozyme… … I'm not going to be viewed as an original scientist or a creative scientist by continuing to do exactly what's happening in [post-doc advisor] Brian Matthews' lab" (Alber, 1993). Lastly, scientists that fail to evolve their research trajectories beyond their mentors' have been shown to be less likely to achieve academic stardom (Ma et al. 2020). To summarize, the lack of independent thinking can have significant consequences for both job prospects as well as subsequent funding opportunities.

A second negative consequence of high advisor-advisee research overlap is the possible competitive tension between advisors and advisees when they occupy similar niches in a research domain. In research spanning multiple organizational contexts, scholars have found that the degree of niche overlap (i.e., the extent to which actors or organizations require resources from the same, or similar, environment) often has negative implications for organizational survival and performance. For example, Baum and Singh's (1994) study of day care centers reveals that the density of competing firms occupying similar niches threatens the survival chances of the focal population of day care centers. Similar patterns of competitive pressure are also observed in studies of technological firms (Podolny et al. 1996) and voluntary associations (Popielarz and Mcpherson 1995). These studies generally found that when organizations occupy similar (overlapping) niches, they tend to experience more intense competition against each other for common resources, and thereby face a tougher environment for survival and performance. In the domain of science, actively managing and sidestepping potential advisor-advisee competition is often implicit, if not outright explicitly discussed. For example, Charles Zuker, a biochemistry and neuroscience professor at Columbia University and a Howard Hughes Medical Institute (HHMI) Investigator, stated that "[Postdoc advisor] Gerry [Rubin] and I had very different interests. I mean, he wanted to study development of the eye. I have no interest whatsoever in any aspect of developmental biology but rather the function of the cells… … So there was a very clean break, and we made a strong commitment that I won't work on development and he won't work on function. So things worked out very well" (Zuker, 1992).

Both positive and negative consequences are reflected in inconsistent empirical findings regarding the relationship between advisor-advisee research overlap and advisee future performance in the literature on

scientists' careers. For example, Blackburn et al. (1981) reported that when a sample of academic advisors (mentors) were asked to name their successful advisees (protégés), the mentors in the study overwhelmingly nominated as their most successful protégés those whose careers were essentially identical to their own—i.e., their "clones." A more recent study by Liénard et al. (2018) of a larger sample of 18,856 biomedical researchers in an online, crowd-sourced database also reports a similar pattern—that greater publication similarity between an advisee and his or her advisor is associated with a higher probability of the advisee obtaining academic placement. By contrast, Ma et al. (2020) reported that research topic dissimilarity— the degree to which a protégé studies topics that differ from the topics studied by his or her advisor—is significantly and positively related to the odds of that protégé later becoming a superstar scientist, as measured by citations to publications, NAS membership, and important scientific prizes and awards. Similarly, another study of French STEM Ph.D. students reports a positive association between research independence of the advisee (i.e., semantic dissimilarity as revealed in the abstracts of theses) and the likelihood that the advisee enters an academic career and his or her publication performance, though a negative association with citation performance (Patsali et al. 2024).

Given mixed theory and evidence regarding how the degree of advisor-advisee research activity overlap might impact advisee future performance, our main research question (RQ) asks: i*s research overlap between advisee and advisor positively (or negatively) associated with advisee performance?*

Status Hierarchy and Competition

As discussed above, empirical investigations of advisor-advisee research overlap are few and the findings in this small set of studies remain mixed. Varied evidence is often an indicator of the contingent consequences of advisor-advisee research overlap, pointing towards a need to unpack the underlying mechanisms. Yet, to our knowledge, no study has attempted to delve into the mechanisms underlying the beneficial implications (or lack of them) of advisor-advisee research overlap for advisee performance.

To deepen our understanding of the relationship between advisor-advisee research overlap and performance, we turn to the literature on status. We anchor our theorizing on two key insights from organizational research on status: a) that status is a coveted resource that is oftentimes the target of competition among actors (Gould 2002, Bothner et al. 2007, Piezunka et al. 2018) and b) that status can serve as an

endorsement signal, spilling over across relationships (Podolny 1993, Stuart et al. 1999). Specifically, we propose a two-fold effect when an advisor occupies a higher position in the status hierarchy. First, a high-status advisor exhibits greater competitive pressure on the advisee, and this pressure is likely to be more salient if the advisor and advisee are pursuing similar research agendas (i.e., have high advisor-advisee research overlap). Second, a high-status advisor acts as a patron to endorse individuals lower on the status hierarchy (i.e., advisee), and this endorsement is particularly salient if the advisee is probing scientific topics farther afield from his or her intellectual base (i.e., have low advisor-advisee research overlap). Taken together, these arguments lay out the theoretical basis explaining how shifts in an advisor's status can be leveraged as a tool to gain insight into the mechanisms underlying advisor-advisee research overlap's performance implications.

Status is predicated on the deference an actor can garner from others (Stuart et al. 1999, Azoulay et al. 2014). Deference, in turn, stems from the allocation of attention. Given time constraints to attention, the amount of deference an actor can allocate is ultimately a zero-sum resource and as with other valuable, limited resources, status as a deferential asset is the target of intense competition. Consistent with the nature of zero-sum resources, the elevation in status of any one actor must necessarily result in a fall in status of others in the hierarchy (Bothner et al. 2010). In concordance, the sociological perspective views status-based social structures as intrinsically rank ordered and hierarchical (Gould 2002). When competitors have an intrinsically asymmetric relationship—such as that between an advisor and advisee—competitive advantage in the competition for status naturally accrues to the more senior partner. Thus, we put forward the suggestion that status is an under-appreciated element of genealogical relationships in science as advisors and advisees often occupy related but rank-ordered (i.e., filial) positions in the status hierarchy.

Moreover, there is reason to suspect that these effects are compounding. Occupants of higher positions in a status hierarchy enjoy cumulative advantages through a process known as the Matthew Effect (Merton 1968), in which actors who have already attained higher status reap disproportionate rewards over others who occupy (even just marginally) lower status. If two actors—such as the advisor-advisee pairs that are the subject of this paper—are active in similar domains, competition for status, as well as its consequences, will be particularly intense. Two recent studies reinforce our argument on the (potential) negative consequences that

come with association to high-status (or enhanced status) actors. First, Reschke et al. (2018) examined the reallocation of attention to observe that (the lower-status) "neighbors" (papers) of an HHMI-appointed paper attracted less attention in the post-appointment period. In a disparate setting, Piazza et al. (2020) analyzed observations of 1,385 bands and found that a new music band's risk of financial failure increased when the band has performed gigs at the concerts of higher-status, more established bands. Both findings, in different empirical (i.e., knowledge generation and creative) contexts, suggest that close association with high-status actors can hurt the performance of the lower-status actor.

On the other hand, a second body of literature also points to positive consequences of affiliation with high-status actors: an actor's set of relationships often serves as a signaling device within the marketplace. Over twenty years ago, Podolny (2001) stated that "the presence (or absence) of a tie between two market actors is an informational cue on which others rely to make inferences about the underlying quality." Spurred on by this notion that relationships can serve as indirect "prisms", there has been a groundswell of theory and evidence suggesting the presence of perceptual cues inferred from the presence (or absence) of connections to prominent actors (Hochberg et al. 2007, Shipilov and Li 2008, Sorenson and Stuart 2001). The core insight in this line of work has been the "transferability of status", that status can be attained through one's affiliation with high-status actors whose quality has been previously socially determined (Podolny and Phillips 1996).

The salience of status transfers may be particularly beneficial when a scientist enters a disparate field. In pivoting their research trajectory, the itinerant scientist enters a domain where he or she accrues less benefits from the advisor's social and knowledge capital. As a result, these pivots—while often necessary due to the imperative for intellectual independence described above—incur intrinsic risks to the focal advisee, often requiring additional time and resources for the pivot to yield benefits. Endorsements by higher status actors (e.g., advisors) may serve as a buffer in mitigating this risk. The higher the status of the endorsing actor (i.e., advisor), the greater the potential benefits to the endorsee (i.e., advisee) may be. Thus, the salience of status transfers in the form of a positive status spillover may be particularly potent for a junior (advisee) scientist entering a disparate field. For example, James D Watson, discoverer of the triple helix DNA structure, stated that "I think it is extraordinarily important that you have a scientific patron because there'll

be times when you are bound to strike it bad and you'll need somebody to convince people that you are not irresponsible" (Bradbury 1970). In comparison, for a junior (advisee) scientist whose research is adjacent to the advisor's, his evaluators (e.g., grant proposal and publication reviewers) are likely in the same field and already have in-depth knowledge about the topic. From the perspective of these evaluators, the impact of an advisor's endorsement is likely limited, as evaluators have sufficient internal capabilities to ascertain the quality of the advisee (and advisor) in their chosen domain. Thus, our assumption is that when advisees are in more distant scientific fields (i.e., have low advisor-advisee research overlap), an increase in the advisors' status may lead to a greater flow of endorsement benefits to the advisee. This argument, taken in conjunction with the status-induced competition argument above, lead us to expect that:

Hypothesis 1 (H1): *A (potentially) positive relationship between advisor-advisee research overlap and advisee performance will decrease with increasing advisor status.*

DATA, MEASURES & METHODS

Data

We study in this paper biomedical researchers who were trained in their PI advisors' labs and later have transitioned into the role of independent scientists. To do so, we obtained and merged data from several different sources. In summary, we first started with the "Author-ity" dataset, constructed by Torvik and Smalheiser (2009), that applied similarity-based metrics to disambiguate author names in historical Medline publications. Second, we retrieved each scientist's publications from the National Library of Medicine's Medline database, from which we identified the advisor-advisee dyads as well as each scientist's research field and productivity (https://www.nlm.nih.gov/medline/medline_home.html). Third, we tracked citation records of their publications from Microsoft Academic Graph Analytics (https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/), which has been utilized in previous research to track publication and citation records (Gomez et al. 2022). Fourth, we collected NIH (National Institutes of Health) grant information for each scientist using NIH RePORTER tool (https://reporter.nih.gov) to track funding performance of the advisee scientists *after* they have departed from their advisors' labs. Lastly, we used the

receipt of a set of prestigious awards in the biomedical sciences to identify advisors' status changes following Azoulay et al. (2010).

The "Author-ity" project is an NSF-sponsored data curation effort by Torvik and Smalheiser (2009) that tracks the community of biomedical researchers as reflected in Medline publications. The computational scientists used an array of machine-learning tools to assign sets of publication-authorship positions to a unique, disambiguated author. In this manner, the careers and activities of life scientists can be traced across time through their bibliometric records. While proprietary, the Author-ity dataset has also been used in other social studies of science (e.g., Foster et al. 2015, Mohnen 2022). The Author-ity dataset covers scientists that have published papers between 1972 and 2009. We restricted our dataset to U.S.-based university scientists because we measured scientists' funding performance through NIH grants, a funding source that is largely closed to non-U.S.-based scientists.

To identify advisor-advisee pairs, we leveraged a long-standing norm in the life sciences: an advisor (Principal Investigator or PI) of a lab is almost always listed as the last author and a member (usually a graduate student or postdoc) of that lab is almost always listed as the first author. Using the bibliographic information that the Author-ity dataset provided us, we identified advisees as individuals that a) have one or more first-author publications, b) have one or more last-author publications (i.e., they have become PIs after leaving parent's lab), and c) whose median year of first-author publications is earlier than the median year of last-author publications. We then identified the modal last-author on each advisee's first-author publications, bracketing the 10 years around which the advisee transitioned into an independent PI role, and assigned this last-author as the advisee's advisor.

We validated this approach of identifying advisor-advisee pairs in three complementary ways. First, we conducted qualitative interviews with established biomedical scientists about authorship norms in the field who all reassured us that we were not misinterpreting bibliometric norms. Second, we verified our data collection strategy using a randomized sample. Specifically, we drew a random sample of 100 advisees from our data and asked research assistants to verify the scientists' identity as well as their advisors based on publicly available information. The validation rate was on average 80% across two RAs. Third, we used detailed personnel

information of labs at the MIT Department of Biology as a "ground-truth" dataset (Conti and Liu 2015). We again found high consistency between our inferences and this data source.

Our final sample consists of 144,566 advisee-year observations with 15,271 advisee scientists mentored by 7,924 unique advisors, including only those advisors who mentored at least two advisees. This criterion was applied as we need to use advisor fixed effect in our models to control for unobservable, advisor-level, time-invariant characteristics such as quality or mentoring style. We explain our analytical strategy in greater detail in the subsequent section. The dataset focuses on advisee scientists at U.S. academic institutions between 1972 and 2009, structured as an unbalanced panel. We monitored each advisee scientist over the first 10 years of the career after his or her departure from the advisor's lab, creating time-varying variables to document the scientist's activities annually during this period.

Variables

Dependent Variables

Funding Performance. To comprehensively assess the performance of advisee scientists across various dimensions, we developed several dependent variables. First, we constructed five dependent variables for scientists' grant funding performance in a given year: (i) the *Grant Support Index (GSI)*, (ii) a dichotomous indicator for *having National Institutes of Health (NIH) grant* (or not), (iii) the *number of NIH grants* received, (iv) the *number of R01 grants* received (R01 is one of the most significant types of NIH research grant), and (v) the total *project costs* associated with NIH-funded projects (i.e., amount of NIH grant awards). The GSI, following the methodology suggested by the NIH Deputy Directory of Extramural Research (Lauer et al. 2017), eschews traditional financial metrics in favor of a mechanism-focused approach, thereby mitigating the inherent variability in funding amounts across different research fields. The methodology underpinning the calculation of the Grant Support Index (GSI) score was developed through a collaborative and objective process by the NIH program. Initially crafted by NIH staff independently of research outcomes, the scoring scheme underwent further refinement under the auspices of the trans-NIH Extramural Activities Working Group (EAWG) subcommittee. This effort was led by Jon Lorsch, Director of the National Institute of General Medical Sciences (NIGMS) and involved extensive deliberations among dozens of NIH program and grant

management staff. The concerted effort aimed to ensure the GSI's robustness and applicability across NIH research project grants, with the definitive point values detailed in Appendix 1. Despite its innovative focus, the GSI's weighting system—such as assigning different points to solo versus collaborative R01 projects—may introduce biases, which prompted us to include the other four alternative funding-related measures to ensure the robustness of our assessment.

Research Performance. In addition to funding performance, we also constructed dependent variables characterizing scientists' performance in research. For a scientist in each of the first ten years of his or her career, we constructed three distinct measures: (i) the *number of last-authored publications* in a given year, capturing a scientist's productivity as an independent PI; (ii) the log of *number of forward citations* as of 2018 (the last year of our observation window) to the papers published by the scientist in the year, reflecting the impact of the scientist's research products; and (iii) the *disruption score* of his or her research publications following the method described in Funk and Owen-Smith (2017) and Park et al. (2023). These two groups of measures on funding and research performance together provide a comprehensive assessment of a scientist's multi-dimensional performance, extending our understanding of the relationship between advisor-advisee research overlap and early-career performance of the advisee.

Independent Variables

Research Overlap. Our main independent variable is the *research overlap* between an advisee and his or her advisor, specifically measuring the similarity in research topics during the advisee's training in the advisor's lab. To operationalize this variable, we utilized Medical Subject Headings (MeSH) associated with each article indexed in Medline. These MeSH terms, assigned by the National Library of Medicine, objectively categorize the content of articles, thereby providing an unbiased representation of research topics.

To construct the measure of research overlap, we aggregated MeSH keywords from journal articles published by both the advisee and the advisor within a five-year window preceding the advisee's departure (inclusive of the departure year). The overlap metric is then calculated as the proportion of MeSH terms shared between the pair of scientists to the total number of MeSH terms associated with the advisee's publications during this period. For instance, if an advisee left an advisor's lab in 1990, we assessed the overlap in research

topics by examining publications from 1986 to 1990. This involved counting shared MeSH keywords between the pair and dividing by the total MeSH keywords attributed to the advisee's work. Importantly, this measure is based on both independent publications by the advisee and co-authored publications with the advisor (we have also calculated an alternative overlap measure based solely on independent publications; see below). It is worth noting that this is a time-invariant variable capturing the degree of research similarity between advisees and their advisors during the mentorship period. Therefore, this research overlap helps us assess the imprinted effect of maintaining a certain degree of similarity to advisor's research on the advisee's early-career performance.

For our robustness analyses, we constructed several alternative research overlap measures, each designed to capture different facets of the concept and facilitate robust testing of the research-overlap and advisee performance relationship. These alternative measures are featured in our robustness tests in Appendix 4.

First, we constructed an alternative to the above (main) research overlap measure by excluding the coauthored papers with the advisor from the calculation. This alternative measure thus addresses the independent works by the advisee.

Second, while our main research-overlap measure is based on the proportion of overlapped MeSH keywords, we constructed an alternative measure that examines, in each scientist-year, the proportion of an advisee's publications that significantly mirror the advisor's research. The determination of the advisee's significantly similar publications to the advisor's is based on MeSH keyword overlap in their publications. Specifically, we assessed each of the advisee's articles against the advisor's during the training period, identifying papers with over 50% keyword overlap. The measure is then calculated as the ratio of these significantly similar papers (to the advisor's) to the total number of publications by the advisee during training.

Third, we constructed an alternative research overlap measure that is derived from the PubMed's Relatedness Algorithm. This algorithm identifies articles with thematic similarities through an analysis of titles, abstracts, and MeSH terms. For each of the advisee's articles published during the training period, PubMed's algorithmic API identifies a list of similar papers. We counted from this list number of articles by the focal

advisee's advisor and computed the proportion of the advisee's five-year pre-departure publications that the algorithm considered as similar to the advisor's works.

Fourth, we assessed the extent of research overlap between an advisee's *post-lab* publications (i.e., *after* their departure from the advisor's lab) and those of the advisor's during the training period. This alternative measure differs from our main research overlap measure in that it is based on the advisee's publications after he or she has left the advisor's lab, and it is an annually updated metric in the first ten years of the junior scientist's career. This alternative measure captures the dynamic changes in an advisee's research similarity/distance with his or her advisor's during an early career stage. We refer to this variable as "dynamic research overlap" in our interpretation of the results.

Advisor's Status. Our second major independent variable is an advisor's status. We measure an advisor's status in two different ways. First, we followed previous research (e.g., Azoulay et al. 2014) that used awards as a status-boosting indicator and constructed an indicator variable "*advisor has received an award",* which takes on a value of one in years t+1 and onward if an advisor received one or more of the major prizes and awards in year t: For example, if an advisor received any of the major scientific awards in 2000, this measure will be valued as 1 from 2001 onward, as status attainment tends to be sticky (Betancourt and Wezel 2016, Podolny and Phillips 1996). Following previous research, we considered the following major awards and prizes in biomedical sciences: Pew Searle Scholars, Packard Award, Beckman Award, Rita Allen Scholarship, MacArthur Fellowship, March of Dimes, HHMI appointments, National Institutes of Health Director's Pioneer Award, Nobel Prize, Lasker Clinical Awards, Lasker Basic Awards, IOM Fellowship, and National Academy of Sciences induction.

Since the odds of attaining any one of these prizes and awards is very small and any one of these prizes and awards can significantly elevate a scientist's status in biomedical fields, we used the dichotomous rather than a count variable of the number of prizes and awards received by an individual. In our data, we have identified a total of 1,045 advisors who have received one or more of the above-listed prizes and awards during the time window of our study, accounting for 13.2% of our sample. At the advisee-year level, 15.4% of observations are in years when the advisee's advisor has received an award.

In our second advisor status measure, we employed institutional (or employer) status as an alternative proxy for the advisor's status. The *advisor's employer status* is based on the Carnegie classification of institutions of higher education. The classification system rates institutions based on numerous criteria, including the types of degrees or institution's focus on specific fields of study and the ratings range from 0 to 4, with 4 being the most prestigious. We used this alternative advisor status measure in the robustness test reported in Appendix 5.

Control Variables

Lastly, we constructed the following research productivity and collaboration network control variables for advisees in each of the first ten years of his or her career: (i) *number of publications* in year *t*, (ii) *number of coauthors*, (iii) *average forward citations* to papers published in year t (i.e., total number of forward citations, as of 2018, to the scientist's publications in year *t* divided the number of publications by the scientist in year *t*), (iv) *career age* as an independent researcher (i.e., number of years since the advisee departed from advisor's lab) and (v) *average field maturity* of his or her papers published in year t.

For (v) *average field maturity*, we used MeSH keywords as the building blocks. First, for every MeSH keyword in the PubMed data, we calculated its maturity in a given year *t* by taking the ratio between the number of authors who published papers on the MeSH in the year to the maximum number of authors publishing on the MeSH across all years from 1971 to 2009. For example, a MeSH word's maturity is measured as 0.5 in year *t* if the number of authors who published on the MeSH in that year was 300 while the maximum number of authors who published on the MeSH was 600 in any year between 1971 to 2009. This maturity variable ranges from 0 to 1, where 1 indicates that the number of authors working on the MeSH keyword was at a peak. Next, for each scientist-year, we took a weighted average of maturity level of all the MeSH words on which he or she has published papers in that year, with the weights based on the proportion of publications devoted to each MeSH.

For robustness, we also constructed an alternative measure of average field maturity, in which the maturity for each MeSH is based on the age of the MeSH term (i.e., number of years since the MeSH first

appeared in PubMed database). All other steps in the construction of average maturity for each scientist-year are the same as the above measure.

Analytical Strategy

We used a three-step analytical approach in our empirical estimations. First, to answer our main research question (RQ) regarding the relationship between advisor-advisee research overlap on the advisee's early-career performance, we employed models that incorporate advisor fixed effects in the estimations, thereby comparing "academic siblings" who are trained by the same advisor. The inclusion of advisor fixed effects help mitigate endogeneity arising from unobserved, time-invariant advisor characteristics, such as mentorship quality and style. The absence of these effects could lead to omitted variable bias, given that unobserved advisor traits might concurrently influence research overlap and advisee performance. For instance, an intensive mentorship style could positively correlate with both outcomes, introducing a positive bias in our estimates. In addition, the research overlap measure in our main analysis is fixed at the time the advisee departed from his or her lab of training, rendering it less vulnerable to potential biases caused by reverse causality (e.g., early-career funding pattern drives the advisee scientists to adopt certain level of research proximity to the advisor). Nonetheless, the correlation between the dynamic research-overlap with advisor (i.e., research overlap in the years when the junior scientist has become an independent PI) and advisee performance may still be an interesting question to examine. We report correlational results using dynamic research-overlap in our "Robustness" section (see Appendix 4).

Second, in testing our Hypothesis 1 regarding the role of advisor status in moderating the main relationship between advisor-advisee research overlap and advisee performance, we employed a difference-in-difference research design. This is prompted by our concern that the advisor fixed effects cannot eliminate the potential omitted variable bias arising from advisor-level time-variant factors that may occur concurrently with advisor's status change. For this reason, in the second step of our analytical approach, we focused on a smaller subset of scientists. Specifically, we followed prior researchers (e.g., Azoulay et al. 2014, Reschke et al. 2018) who have used major scientific awards as indicators for a status-elevation to identify how status may influence scientific outcomes. In this strategy, the central part of the identification of advisors' status effect comes from

within-advisor comparison of advisees' performance (i.e., the comparison of performance of sibling advisees trained by the same advisor) before and after the advisor received a scientific award. This identification strategy relies on the assumption that the timing of prizes and awards is often exogenous. Many (future) laureates are on a short-list of awardees (e.g., we all knew Doudna and Charpentier would receive the Nobel and it was just a question of when rather than whether), but few scientists can accurately *time* when they receive a prize, let alone their advisees. After the (relatively exogenous timing of) award of a scientific prize, the advisor's status is boosted, thereby increasing the strength of any competitive mechanism or lack-of-independence discount underlying the (main) overlap-performance relationship. Therefore, such a status change is expected to moderate (reduce) any positive relationship between overlap and performance.

Following Azoulay et al.'s (2017) study of career imprinting effect of advisors on advisees' academic entrepreneurship, we restricted the analytical sample in this step to *sibling* advisees who left advisors' lab *before* their advisors have received the prize or award. By restricting our observations to advisees who departed from (future) award-winner advisors' labs *prior to* their advisors' status-elevation event (due to receiving an award), we can avoid potential advisors' award-induced quality sorting between advisees and advisors. In addition, because the advisees were all trained before the advisors had become stars in their fields, the mentorship quality should also be stable, less affected by advisors' post-award distractions.

Third, to bolster confidence that the changes we observed in advisee performance are linked to the advisors' status change, rather than to increases in the capabilities of advisors during the natural course of their careers, we constructed a propensity-score matched sample of advisors, who are comparable to the award-winning advisors on key dimensions including publication count, grants, employer university ranking and research distance to their advisees (to proxy for the advisors' style of training students). In this third step of the analysis, the non-award-winning advisors in this matched sample only observably differ from award-winning advisors in that they have not experienced a status-elevation event (see comparative statistics for the award-winning and matched-sample advisors in Appendix 3). We then ran a triple difference-in-difference analysis that compares the performance gap between sibling advisees with different levels of research overlap to their

advisors *before* and *after* advisor's status change, benchmarked against the before-and-after-status-change comparison of outcomes of the advisees trained by comparable non-award-winning advisors.

RESULTS

Summary Statistics

Table 1 provides summary statistics of variables used in our analysis, with a correlation matrix provided in Appendix 2. Table 1 shows that the mean GSI score among advisee scientists is 3.3, reflecting a moderate level of success in securing funding. About 35% of scientist-year observations are with NIH funding. At the scientist level, about 71% of scientists have received at least one year of NIH funding. The mean for our main research overlap measure (MeSH keyword-based, measured with publications during training period) is about 0.71 with a standard deviation of 0.29. If we exclude coauthored publications, the mean for this measure is at 0.19, with a standard deviation of 0.27.

Figure 1 shows binscatter plots of the distribution of the (main) research overlap levels across the years in our dataset. Overall, the peak of research overlap level appears to be between 1990 and 1995. Advisees of award-winning (higher-status) advisors have a higher level of research overlap than advisees of non-award-winning (lower-status) advisors. Data show a similar pattern when we separated advisors' status by their employer institutions' Carnegie ranking. In terms of field maturity, research overlap appears to be at a higher level for advisees in mature (rather than fledgling) fields.

Tests for Main RQ: Relationship between Overlap and Performance

In Table 2, we sought to answer our main research question (RQ) regarding the relationship between advisor-advisee research overlap and advisee's performance in the first ten years of the advisee's independent PI career (after he or she has departed from the advisor's lab). The advisor-advisee research overlap is calculated based on the ratio of MeSH keyword overlap in all of the advisee's publications during the period of training (including coauthored papers with the advisor). Other advisor-advisee research overlap measures are used in robustness tests (see the "Robustness" section below and Appendix 4). Except for model 8, Poisson Pseudo-Maximum Likelihood (PML) estimators are used as our dependent variables are non-negative. Poisson PML

has been recommended for analyzing dependent variables of non-negative values (Gourieroux et al. 1984). While the negative binomial estimator is often suggested for addressing over-dispersion in count data (Allison and Waterman 2002), the Poisson PML estimator was preferred for its capability to produce consistent, robust standard errors. Model 8 uses an ordinary least squares (OLS) estimator as the dependent variable contains negative values. All models have controlled for advisor fixed effects—thus basing inference on comparison across academic siblings of the same advisor. Our models also controlled for calendar-year temporal influences through year fixed effects, as well as characteristics of the advisees including whether the advisee has an award-winning advisor, the employer status of the advisor, career age, number of publications, average citations, number of coauthors, and average research field maturity.

We estimated the relationship between research overlap and five funding performance indicators and three research performance indicators. The results are illustrated in Figure 2. Across the indicators, the effect of research overlap is consistently positive and statistically significant on funding performance indicators. One standard deviation (29%) increase of advisee's research overlap with the advisor from the mean level is associated with an increase of 26.7% in the advisee's GSI score, 26.2% in the likelihood of being funded by NIH, 21.8% in the number of NIH grants received, 25.8% in the number of R01 grants received, and 16.9% in the funding amount (project costs) received during the first ten years of his or her career as an independent PI. The effects on productivity indicators are also positive, though at smaller magnitude: one standard deviation increase of research overlap is associated with a 13.9% increase in his or her last-authored publication count, and a 1.3% increase in his average (per-paper) citations during the early career[1]. It is worth noting that we have included the total publication count of the advisee in his early-career years as a control variable to ensure that the observed effect specifically attributes to the influence of research overlap on the outcome of last-authored publications in which the junior scientist takes on the PI role.

The only exception is with the advisee's research disruption score, for which we didn't see support for a positive relationship in the data. There are two possible explanations for this finding. One is that highly

[1] Right-truncation of citation-count measures is not an issue in this estimation as we have controlled for year fixed effects in the models, which allow inference from comparison of papers published in the same year. Thus, right-truncation of citation count does not introduce systematic biases into our estimation.

disruptive research requires the scientist to be more pathbreaking. By staying under the wing of the advisor, the junior scientist may not have as strong an impetus to generate disruptive research as his or her academic siblings who have opted to break away from the advisor's research before the onset of his or her career. The other explanation is timing—it may be possible that truly disruptive research emerges at mid-to-late career stage, outside of our observation window of the first ten years during a junior scientist's career.

With the exception of research disruption, the findings from Table 2 present a relatively consistent answer to our main research question (RQ): a high degree of research overlap between advisees and advisors is positively associated with various performance metrics of the advisee scientist during early career, including funding, research productivity, and scientific impact. However, this relationship does not extend to the generation of disruptive scientific contributions.

Tests for Hypothesis 1: Moderating Effect of Advisor Status

To test Hypothesis 1 with regard to the moderating effect of advisor status on the overlap-performance relationship studied in our assessment of the main RQ, we first ran, in Table 3, models with research overlap-advisor status interaction in the models, controlling for advisor fixed effects. The model setup in Table 3 is the same as in Table 2, with the exception of the addition of the interaction term. The advisor status change in the table is measured as a change in the advisor's award status. After a year when the junior scientist's advisor was awarded a prize, this indicator variable turned from 0 to 1 and stayed as 1 onward. We also have an alternative measure for advisor status based on advisor's employer status, which we used in a robustness test (see "Robustness" section and Appendix 5).

The first five models in Table 3 examine the moderating effect of advisor status shift on the main relationship between research overlap and advisee funding performance. The findings across all five models reveal a consistently negative interaction between research overlap and advisor status, lending support to our Hypothesis 1 which posits that any positive research overlap-performance relationship may be reduced as the advisor gains more status. We used model 1 (estimation on advisee GSI score) to illustrate the moderating effect of advisor status in Figure 3A. First, the figure reveals that research overlap has an overall positive relationship with performance, regardless of the advisor's status. However, it also reveals that for the advisees

of higher-status (award-winning) advisors, the slope of increase is flatter than the slope for the advisees of lower-status advisors, which suggests to us a negative mechanism is at work dampening any positive relationship between overlap and funding performance. Notably, at the level of one standard deviation above mean in research overlap, the advisees of lower status advisors even match the advisees of higher-status advisors in funding performance. Among the non-funding performance indicators, H1 also appears to be supported in the estimation with the advisee's last-authored publication count as the dependent variable, which we visualized in Figure 3B.

In our second step of analysis, we tested H1 with the difference-in-difference setup as described in the second step of our analytical strategy. This step relies on (i) the relatively exogenous *timing* of the advisor's status change (through receiving an award) and (ii) requiring advisees to depart from advisors' labs before the status-elevation event took place to help avoid status-based sorting of advisors and advisees and also stabilize the mentorship quality of advisors.

Model 1 of Table 4 conducts a test of H1 with the restricted sample of advisees of future award-winner advisors. With this much smaller sample and more stringent test setup, the hypothesized moderating effect of advisor status still holds for advisee funding performance (GSI), as reflected in the negative interaction of the "research overlap" and "after advisor has received an award". We rendered this interaction effect in the left panel of Figure 4 ("Advisees of the Treated Sample"). The figure displays a pattern consistent with that observed in Figure 3A based on the full-sample estimates, indicating that the marginal benefits of research overlap have decreased subsequent to an advisor's status elevation event.

Notably, in this more stringent estimation step, the moderating effect of an advisor's status on overlap-performance relationship is even more salient than that revealed in Table 3 (and in Figure 3A). Specifically, the positive relationship between research overlap and (funding) performance exists only in the years *before* the junior scientist's advisor has become an academic star through the receipt of an award (revealed in the blue dashed line in the left panel of Figure 4). For these advisees, the optimal strategy appears to be closer to their advisors' research domains. However, the almost-flat, red-solid line in the same figure for the years *after* the advisor has become award-winning academic stars suggests that the junior (advisee) scientists of the very-high-

status, award-winner advisors actually have more freedom as they can choose to be either close to or distant from their advisors.

In models 2 and 3, we executed the third step of our empirical strategy by incorporating a matched sample to further demonstrate that the observed advisor-status effect (that moderates the overlap-performance relationship) results truly from a status change, rather than from expected improvement of the advisors' skills and capabilities. As described in our analytical strategy, we constructed a matched sample of advisors to the star-scientist advisors featured in model 1. These matched advisors follow similar productivity trajectories as the star-scientist advisors, but they will fail to receive an award in the future. If the status-related effect in model 1 is spurious in the sense that it results from the advisors' skill improvement (as opposed to award-related status elevation), then we should see a similar negative interaction in this matched sample analysis in model 2. We did not observe such a negative interaction in model 2 of the table. The right panel of Figure 4 further reveals that the relationship between research overlap and funding performance does not differentiate for the advisees of these matched-sample advisors before and after their pseudo-treatment year (the time point when the advisor was matched to one of the star-scientist advisors in model 1). Model 3 runs a pooled-sample analysis with three-way interactions, thus a triple difference-in-difference analysis. The significant three-way interaction estimate gives us additional assurance that the difference in the overlap-performance slopes observed in model 1 (and left panel of Figure 4) is driven by the advisor's status change rather than the advisor's normal productivity improvement.

In Table 5, we report results from the similar analytical structure as in Table 4, but for the dependent variable of advisee's last-authored publications. The interaction between "research overlap" and advisors' status shift ("after advisor has received an award") in model 1 is negative but only marginally significant. The three-way interaction effect in model 3 is no longer significant. Our conclusion is that although there is a similar pattern of advisor's status change moderating the relationship between research overlap and advisee research (in particular, publication) performance, this pattern appears to be weaker than that for the relationship between overlap and funding performance.

Multiple explanations may underlie the weaker moderating effect of advisor status on publication than on funding. One possibility that could dampen the positive overlap-performance relationship is the potential competition between the pair of advisor and advisee who conduct research in proximate domains. Funding in biomedical sciences is more resemblant of a zero-sum game than publications. Researchers, particularly those working in similar research areas, compete in the same NIH funding pool. By comparison, there is a wider pool of potential venues for publication. Even though competitive pressure in publication certainly exists among researchers in similar areas, we speculate that publication competitive intensity is less than that for funding, leading to our observation of ameliorated advisor status effects on the overlap-performance relationship in Table 5.

Robustness Analysis

Robustness Test with Alternative Research Overlap Measures. We conducted additional analyses using four alternative measures for calculating research overlap between advisors and advisees, as outlined in our previous section on "Variables". We reported in Appendix 4 (Table A4a) coefficient estimates and standard errors of these additional research overlap measures on advisee performance following the structure in Table 2. This table shows that our findings on the main RQ are not sensitive to the choice of research overlap measures in the assessment of funding and research productivity (last-authored publication count) performance. Citation-performance, however, is not consistently replicated across these alternative measures.

In Table A4b, we reported coefficient estimates and standard errors for (i) the alternative research overlap measure featured in the row, (ii) the variable "*Advisor has received an award*", and (iii) their interaction term on the dependent variable featured in the column heading. All cells in this table are based on models using the same regression setup as in Table 3. The results with these alternative measures remain consistent with those in Table 3 for all funding performance variables, and largely for the last-authored publication count outcome variable. The citation performance again can no longer be replicated when using these alternative measures.

Robustness Test with Alternative Advisor Status Measure. We used an award-based advisor status measure in our main analysis. For robustness, we re-ran our Table 3 by leveraging advisor's employer status (based on

Carnegie classification) as the advisor status measure. Note that this advisor's employer status measure is already controlled for in all main tables. This robustness test treats having an award-winning advisor as the control and instead interacts research overlap with advisor's employer status. We reported the test results in Appendix 5 (Table A5), which are consistent with results for the interaction effects reported in Table 3. While this variable of advisor status confirms our findings in Table 3, we prefer the award-winning based status measure to this one given the relative exogeneity of the timing of prize awards. With this measure, advisor's status gain when moving to a higher-ranked employer can be endogenous with advisee's choice of research streams and their performance outcomes.

Using Research Field Maturity to Test Competitive Mechanism. Our main analyses use changes in advisor status to help unpack mechanisms underlying the overlap-performance relationship. Advisor status is by no means the only factor shaping the relationship between the advisee and his or her advisor. Rather, many other factors can drive the competitive intensity in the advisor-advisee relationship as relating to their research overlap. In Appendix 6 (Table A6a and A6b), we tested one candidate for such an alternative factor, the maturity of a research field. Growing fields on an expanding trajectory are likely to have an influx of resources and hence, be more accommodating to advisors and advisees, regardless of their research overlap with each. In contrast, in relatively mature fields where growth has stalled, there will be a stable level of resources, leaving research-overlapped advisees and advisors to compete for the same zero-sum funding pool. Therefore, we anticipate that the negative (competitive) pressure from being close to advisor's research will be stronger in mature fields than in fledgling fields. Hence, the average maturity of an advisee's research fields should negatively moderate any positive overlap-performance relationship. In Tables A6a (using researcher-count-based maturity measure) and A6b (using MeSH-age-based maturity measure), we observed empirical support for our argument across all models for funding outcomes, and partially (in A6b) for the research impact (log of number of forward citations) outcome.

Robustness Test with Junior Awards. In addressing potential variations in the impact of awards on the competitive relationship between advisors and advisees, we distinguished between major lifetime achievement awards (e.g., NAS induction, Nobel Prize) and contemporary research stream awards (e.g., HHMI, MacArthur).

Lifetime achievement awards typically are given post-discovery at a scientists' later career stage. By contrast, contemporary research stream awards are forward-looking and meant to stimulate future achievements (e.g., HHMI), and are typically given at earlier career stages. As our empirical strategy rests upon a shift in advisor status (and the consequences of that status boost) post-award, status shifts will likely be larger in the second type (young-scientist-oriented) of awards. As a result, we categorized awards into junior awards, typically bestowed upon scientists under the age of 50, and senior awards, generally awarded to those over 50.[2] The age distribution among the prizes and awards we used is provided in Figure A1. We re-ran the analysis with only the advisees of junior award recipients in Appendix 7. In Table A7a, which replicates Table 3, we observed findings in a consistent direction as in Table 3, with larger magnitude for the interaction effects though weaker statistical significance. In Table A7b, we used this junior-awards-based advisor status measure to re-trest Table 4 and found consistent findings mirroring those in Table 4.

Robustness Test with Advisee Fixed Effects Included in Models. The models in our main analysis have included advisor fixed effects. Another way to robust test our models is to include advisee fixed effects, which can help tease apart time-invariant unobserved heterogeneity across advisee scientists. In regression specifications with advisee fixed effects, the main effect of research overlap cannot be estimated due to the time-invariant nature of the measure. However, we can still test the interaction effect of research overlap and advisor status change. We reported tests with advisee fixed effects in Appendix 8. Table A8a replicates Table 3 with advisee fixed effects, in which the estimates for the interaction effects remain consistent with those in Table 3 for all the funding variables. The estimate for the publication count outcome remains in the same direction, although the standards errors are larger. Table A8b reports the replication of Table 4. The direction and significance of the three-way interaction term in model 3 is consistent with that in Table 4.

Robustness Test with Additional Control Variables of Advisor Productivity. Our theoretical framework, grounded in status competition theory, suggests that high-status advisors may impose greater competitive pressure on their advisees. An alternative hypothesis could attribute such pressures to the advisors' productivity

[2] Junior awards include Packard, Searle, Beckman, Rita Allen, Mod, Pew, HHMI, MacArthur and NIH. Senior awards include Lasker, NAS, IOM, and Nobel.

levels. To differentiate between these mechanisms, we incorporated three control variables into our analysis: the size of the advisors' lab at the time of advisees' departure, the advisors' publication count, and advisors' funding performance. These variables serve as proxies for advisor productivity. The inclusion of these variables aims to tease apart the influence of advisor productivity on our estimates for advisor status effect. The findings, as reported in Appendix 9 (Table A9a-c, replicating Tables 2-4), affirm the consistency of our original results, even with these additional controls.

Robustness Test with Standard Errors Clustered at Advisor Level. The set of tests in Appendix 10 clusters standard errors at the advisor level (rather than at the advisee level in our main models). Table A10a replicates Table 2 consistently. In Table A10b (replication of Table 3), we lost some statistical significance for the "research overlap – advisor has received an award" interaction term in the models for project costs (model 5) and for last-authored publications (model 6) though the direction of the estimates remains consistent to that in Table 3. Replication of Table 4 is confirmed in Table A10c.

CONCLUSION AND DISCUSSION

This paper examines the relationship between a junior (advisee) scientist's research overlap with his or her advisor and the junior scientist's early-career performance. We posit that while working in an area close to that of the advisor can lead to the more effective transfer of knowledge and skills, a high degree of research overlap may also induce stronger competitive pressure from the advisor as well as triggering lack-of-independence concerns. We examined this relationship by building a novel panel dataset of 15,271 new biomedical faculty members (advisees) at U.S. research universities who were trained in the labs of 7,924 PI advisors between 1972 and 2009. Drawn from bibliometric records, our dataset is thus less prone to reporting biases that exist in crowd-sourced academic genealogical datasets.

Overall, we observed a positive relationship between advisor-advisee research overlap and the advisee's early-career (grant-funding and publication) performance. This pattern is consistent and robust to several alternative measures for research overlap as, advisor status well as alternative model specifications. Taken together, these results suggest that *the modal advisee in our dataset benefits from asking scientific research questions in close proximity to his or her advisor.*

This generalized conclusion, however, needs to be conditioned, as our theory suggests. Towards this direction, we hypothesized a negative moderating effect of advisor's status on the (main) relationship between research overlap and advisee early-career performance (i.e., any positive overlap-performance relationship will decrease as the advisor's status increases). We tested this hypothesis both in a large-sample analysis that incorporated advisor fixed effects (thus comparing the outcomes of academic siblings trained by the same advisor), and in a restricted sample with a difference-in-difference test that leveraged the (random) timing when some of the advisors received major scientific awards. Empirical tests with both approaches confirmed our hypothesis and enabled us to fine-tune the scope conditions delimiting our main RQ. *Junior scientists with higher-status advisors, but not those with lower-status advisors, appear to have greater latitude in terms of research proximity (or distance) to their advisors, as the overlap-performance relationship slope is flatter for them.*

Our findings have important implications for the scientific community. Which research areas to focus on and how much proximity (or distance) to maintain vis-à-vis one's advisor have always been critical questions for the broader research community in general, and junior scientists specifically, during the formative period of training and early career advancement. Yet, answers to these questions based on large-scale empirical data remain thin and inconclusive, with underlying mechanisms and scope conditions underexplored. The key finding from our data—that on balance there is an overall positive relationship between research overlap with advisor and the advisee's future performance, particularly for junior scientists with lower-status advisors—conditions the normative advice that is often passed down to junior scientists to craft an independent identity.

At first glance, our findings also appear to diverge from the Ma et al. (2020), in which the authors reported a negative relationship between research overlap and the likelihood that the junior (protégé) scientists will later become an academic superstar (defined as being elected to National Academy of Sciences membership, winning major scientific awards or prizes, or being in top quartile of citation impact). However, our studies differ in three respects. First, the outcomes we study, such as securing funding support for their labs and successfully publishing papers in journals, occur far more frequently than the outcomes listed above (e.g., election to NAS). Second, we study the (existential) events earlier in the career of scientists, rather than

(approbatory) outcomes that typically occur at a mid-to-late career stage. Third, our estimation is based on a wider range of junior scientists across the quality and status spectrums. Nonetheless, the disparity in these results suggest that an examination of outlier outcomes—and we especially note the absence of a correlation with research disruption scores (see Table 2)—as well as differences across the career life cycle may be particularly fruitful avenues of future research.

While our paper touches upon issues that are of clear relevance to the science and knowledge production community, our findings may also provide guidance to the broader literature on organizations, particularly the literature on organizational genealogies (Phillips, 2002). Academic laboratories are managed by PIs (CEOs) who are charged with securing resources, setting the strategy, hiring employees, and organizing the production of scientific research and promotion of research products (Conti and Liu 2015). Successful advisees may depart from their parents' (advisors') labs to form their own independent labs (Freeman et al., 2001), just as employees may depart from their prior employers to found their own spinoff firms (e.g., Agarwal et al. 2004, Chatterji 2009, Burton et al. 2002, Gaonkar and Moeen 2022, Klepper 2009). Importantly, both the newly formed academic labs and organizational spinoffs must manage a complex relationship with their parent organizations: the parents are both a source of potential benefits (particularly for resource-starved new organizations) as well as a potential threat (Sakakibara and Balasubramanian 2020, Starr et al. 2018, Vaznyte et al. 2020, Walter et al. 2014). As such, our study points to the possible intermingling of these two research streams—the study of scientific and organizational genealogies may serve to inform each other in complementary ways.

This paper also makes contributions to the burgeoning literature that illustrates a negative status spillover between actors and organizations (e.g., Piazza et al. 2020, Reschke et al. 2018). Our paper contributes to this body of work by bringing in the role of organizational genealogies and negative spillovers across familial ties. Taken together, these findings shed greater light on the many faces under which proximity in activities between actors at different positions in the status hierarchy may have beneficial or deleterious effects.

Lastly, this paper also serves to draw attention to organizational genealogies and familial ties. By mapping advisor-advisee pairs, we draw attention to the complex relationships between parent and progeny. For example, advisor-advisee dyads may serve as the foundational elements to partition community structures into "family trees". We speculate that relations within families, with shared norms and trust may be different from relations between families. Given Padgett and Ansell's (1993) emphasis on multivocal identities, our mapping of genealogical ties may enable the study of this topic in a more modern setting comprised of scientists.

This paper is not without its limitations. A first concern is the specialized nature of the setting. While an examination of scientists and their advisors has afforded us a rich setting with traction on causal inference and sample self-selection, these empirical advantages are offset by scope conditions. Unlike more traditional business settings that examine the departure of employees, often under conditions of disagreement, employee (i.e., trainee) departures in this educational setting are structural and by design. As a result, our findings may be specific to the scientist parent-progeny pairs and may not extend to the modal observation of interest in the management field, or even into the for-profit biotechnology sector (e.g., Bikard et al. 2019). Even when the natural sciences are contrasted with proximate educational fields such as the social sciences, differences abound. For example, the social sciences are dominated by direct one-on-one mentorship ties, while the natural sciences are dominated by differentiated groups of individuals embedded in laboratories. In other words, the norms and behaviors endemic to these respective fields are significantly different.

A second concern is a question of the empirical validity of our findings. Any study that focuses on the issue of choice, particularly consequential strategic choices such as those faced by the advisees in this study, faces empirical hurdles. One method in which we tackled this obstacle was by first constructing a large-scale dataset, within which we could identify advisors with multiple advisees. Given the longitudinal nature of our data, we were then able to exploit subsequent advisor awards to examine how the timing of these (exogenous to the advisee) events shifted the dynamics between advisor and advisee. Moreover, given the modern advent of large, publicly accessible datasets, closely matched cases drawn from large sample datasets have become increasingly feasible (as one recent example, see Marx and Hsu 2021). One of the advantages of

examining our hypotheses in the academic life sciences is the ability to construct detailed productivity variables to guide our matching strategy. Nonetheless, retesting our hypotheses in other settings is likely to be a worthwhile exercise.

Lastly, in this paper we have used established theories on status and organizations to inform the behavior of scientists. It is our hope that in turn, the study of scientists may ultimately lead to new theories that serve to enrich the study of organizations more generally. Not only is the study of scientists important in and of itself—and the utility of new vaccine technologies in stemming the tide of COVID infections is just one demonstration of this—it is our belief that the social studies of science may also serve as a model system for examining organizations and their structures.

REFERENCES

Agarwal R, Echambadi R, Franco AM, Sarkar MB (2004) Knowledge transfer through inheritance: Spin-out generation, development, and survival. *Academy of Management journal* 47(4):501–522.

Alber TC (1993) Pew Charitable Trusts. Chemical Heritage Foundation Oral History Program.

Allison PD, Waterman RP (2002) Fixed–effects negative binomial regression models. *Sociological methodology* 32(1):247-265.

Azoulay P, Graff Zivin JS, Wang J (2010) Superstar Extinction. *The Quarterly Journal of Economics* 125(2): 549-589.

Azoulay P, Liu CC, Stuart TE (2017) Social influence given (partially) deliberate matching: Career imprints in the creation of academic entrepreneurs. *American Journal of Sociology 122*(4):1223-1271.

Azoulay P, Stuart T, Wang Y (2014) Matthew: Effect or fable? *Management Science* 60(1):92–109.

Barres BA (2013) How to pick a graduate advisor. *Neuron* 80(2):275-279.

Baum JA, Singh JV (1994). Organizational niches and the dynamics of organizational mortality. *American journal of sociology* 100(2):346-380.

Bertozzi CR (2003) Pew Charitable Trusts. Chemical Heritage Foundation Oral History Program.

Betancourt N, Wezel FC (2016) The credibility of social climbing: When does inter-firm mobility [Not] influence organizational status? *Organization Science* 27(6):1435-1452.

Bikard M, Vakili K, Teodoridis F (2019) When collaboration bridges institutions: The impact of university-industry collaboration on academic productivity *Organization Science* 30(2): 426-445.

Blackburn RT, Chapman DW, Cameron SM (1981) "Cloning" in academe: Mentorship and academic careers. *Research in Higher Education* 15:315-327.

Blau PM (1994) *The Organization of Academic Work*, 2nd ed. (Routledge, New York).

Bothner MS, Godart FC, Lee W (2010) What is Social Status? Comparisons and Contrasts with Cognate Concepts. Working paper, University of Chicago Booth School of Business.

Bothner, MS, Kang J, Stuart TE (2007) Competitive crowding and risk taking in a tournament: Evidence from NASCAR racing. *Administrative Science Quarterly* 52(2): 208-247.

Bradbury W (1970) Genius on the prowl. Life. 69:57-67.

Burton M, Sørensen JB, Beckman CM (2002) Coming from good stock: Career histories and new venture formation. *Research in the Sociology of Organizations* 19:229–262.

Chatterji AK (2009) Spawned with a silver spoon? Entrepreneurial performance and innovation in the medical device industry. *Strategic Management Journal* 30(2):185–206.

Cole JR, Cole S (1974) *Social stratification in science.* (University of Chicago Press, Chicago).

Conti A, Liu CC (2015) Bringing the lab back in: Personnel composition and scientific output at the MIT Department of Biology. *Research Policy 44*(9):1633-1644.

Dehejia RH, Wahba S (2002) Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics* 84:151-161.

Foster JG, Rzhetsky A, Evans JA (2015) Tradition and innovation in scientists' research strategies. *American Sociological Review 80*(5):875-908.

Freeman J (1986) Entrepreneurs as organizational products: Semiconductor firms and venture capital firms. *Advances in the study of entrepreneurship, innovation, and economic growth* 1:33-52.

Funk RJ, Owen-Smith J (2017) A dynamic network measure of technological change. *Management science* 63(3):791-817.

Gaonkar S, Moeen M (2022) Standing on the Parent's Shoulder or in its Shadow? Alliance Partner Overlap Between Employee Spinouts and Their Parents. Available at SSRN: https://ssrn.com/abstract=4124267.

Gomez CJ, Herman AC, Parigi P (2022) Leading countries in global science increasingly receive more citations than other countries doing similar research. *Nature Human Behaviour* 6(7):919-929.

Gould RV (2002) The Origins of Status Hierarchies: A Formal Theory and Empirical Test. *American Journal of Sociology* 107(5):1143–1178.

Gourieroux C, Monfort A, Trognon A (1984) Pseudo maximum-likelihood methods - applications to poisson models. *Econometrica* 52(3):701-720.

Hochberg YV, Ljungqvist A, Lu Y (2007) Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance* 62(1):251-301.

Klepper S (2009) Spinoffs: A review and synthesis. *European Management Review* 6:159-171.

Lauer M, Roychowdhury D, Patel K, Walsh R, Pearson K (2017) Marginal returns and levels of research grant support among scientists supported by the National Institutes of Health. *BioRxiv* 142554.

Liénard JF, Achakulvisut T, Acuna DE, David SV (2018) Intellectual synthesis in mentorship determines success in academic careers. *Nature communications* 9(1):4840.

Long JS, McGinnis R (1985) The effects of the mentor on the academic career. *Scientometrics* 7:255-280.

Ma Y, Mukherjee S, Uzzi B (2020) Mentorship and protege success in STEM fields. *Proceedings of the National Academy of Sciences* 117(25):14077–14083.goul

Malmgren RD, Ottino JM, Nunes Amaral LA (2010) The role of mentorship in protégé performance. *Nature* 465(7298):622-626.

Marquis C, Tilcsik A (2013) Imprinting: Toward a multilevel theory. *The Academy of Management Annals* 7(1):195-245.

Marx M, Hsu DH (2021) Revisiting the Entrepreneurial Commercialization of Academic Science: Evidence from "Twin" Discoveries. *Management Science.*

Meister MD (1998) Pew Charitable Trusts. Chemical Heritage Foundation Oral History Program.

Merton RK (1968) The Matthew effect in science: The reward and communication systems of science are considered. *Science* 159(3810):56–63.

Mohnen, M (2022) Stars and Brokers: Knowledge Spillovers Among Medical Scientists. *Management Science* 68(4):2513–2532.

Padgett JF, Ansell CK (1993) Robust action and the Rise of the Medici, 1400-1434. *American Journal of Sociology* 98(6):1259-1319.

Park M, Leahey E, Funk RJ (2023). Papers and patents are becoming less disruptive over time. *Nature* 613(7942):138-144.

Partha D, David PA (1994) Toward a new economics of science. *Research policy* 23(5):487-521.

Patsali S, Pezzoni M, Visentin F (2024) Research independence: drivers and impact on PhD students' careers. *Studies in Higher Education* 1-24.

Phillips DJ (2002) A Genealogical Approach to Organizational Life Chances: The Parent-Progeny Transfer among Silicon Valley Law Firms, 1946-1996. *Administrative Science Quarterly* 47(3):474–506.

Piazza A, Phillips DJ, Castellucci F (2020) High-status affiliations and the success of entrants: New bands and the market for live music performances, 2000–2012. *Organization Science* 31(5):1272–1291.

Piezunka H, Lee W, Haynes R, Bothner MS (2018) Escalation of competition into conflict in competitive networks of Formula One drivers. *Proceedings of the National Academy of Sciences* 115(15):E3361-E3367.

Podolny JM (1993) A Status-Based Model of Market Competition. *American Journal of Sociology* 98(4):829–872.

Podolny JM (2001) Networks as the pipes and prisms of the market. *American journal of sociology* 107(1):33-60.

Podolny JM, Phillips DJ (1996) The dynamics of organizational status. *Industrial and Corporate Change* 5(2):453-471.

Podolny JM, Stuart TE, Hannan MT (1996) Networks, knowledge, and niches: Competition in the worldwide semiconductor industry, 1984-1991. *American Journal of Sociology* 102(3):659-689.

Polanyi M (1983) *The Tacit Dimension* (Peter Smith, Gloucester, MA).

Popielarz PA, McPherson JM (1995) On the edge or in between: Niche position, niche overlap, and the duration of voluntary association memberships. *American Journal of Sociology* 101(3):698-720.

Reschke BP, Azoulay P, Stuart TE (2018) Status spillovers: The effect of status-conferring prizes on the allocation of attention. *Administrative Science Quarterly* 63(4):819–847.

Ribeiro R (2013) Tacit knowledge management. *Phenomenology and the cognitive sciences* 12:337-366.

Roche M (2023) Academic entrepreneurship: Entrepreneurial advisors and their advisees' outcomes. Organization Science 34(2): 959-986.

Rosenberg N (1974) Science, invention and economic growth. *The Economic Journal* 84(333):90-108.

Sakakibara M, Balasubramanian N (2020) Human capital, parent size, and the destination industry of spinouts. *Strategic Management Journal* 41(5):815–840.

Sauermann H, Roach M (2012) Science PhD Career Preferences: Levels, Changes, and Advisor Encouragement. *PLOS One* 7(5): e36307.

Shipilov AV, Li SX (2008) Can you have your cake and eat it too? Structural holes' influence on status accumulation and market performance in collaborative networks. *Administrative Science Quarterly* 53(1):73-108.

Simonton DK (2004) *Creativity in science: Chance, logic, genius, and zeitgeist* (Cambridge University Press).

Sorenson O, Stuart TE (2001) Syndication networks and the spatial distribution of venture capital investments. *American journal of sociology* 106(6):1546-1588.

Starr E, Balasubramanian N, Sakakibara M (2018) Screening spinouts? How noncompete enforceability affects the creation, growth, and survival of new firms. *Management Science* 64(2):552–572.

Stuart TE, Ding WW (2006) When Do Scientists Become Entrepreneurs? — The Social Structural Antecedents of Commercial Activity in the Academic Life Sciences. *American Journal of Sociology* 112(1): 97-144.

Stuart TE, Hoang H, Hybels RC (1999) Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative science quarterly* 44(2):315–349.

Torvik VI, Smalheiser NR (2009) Author name disambiguation in MEDLINE. *ACM transactions on knowledge discovery from data* 3(3):11.

Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. *Science* 342(6157):468-472.

Vaznyte E, Andries P, Demeulemeester S (2020) "Don't leave me this way!" Drivers of parental hostility and employee spin-offs' performance. *Small Business Economics* 1–29.

Walter SG, Heinrichs S, Walter A (2014) Parent hostility and spin-out performance. *Strategic Management Journal* 35(13):2031– 2042.

Zeng A, Shen Z, Zhou J, Fan Y, Di Z, Wang Y, Stanley HE, Havlin S (2019) Increasing trend of scientists to switch between topics. *Nature communications* 10(1):3439.

Zuckerman H (1977) *Scientific elite: Nobel laureates in the United States*. Transaction Publishers.

Zuker CS (1992) Pew Charitable Trusts. Chemical Heritage Foundation Oral History Program.

Table 1 Summary Statistics

VARIABLES	Mean	Std. Dev.	Min	Max
Grant Support Index (GSI)	3.304	5.376	0	116
Having NIH grant (Indicator)	0.355	0.479	0	1
Num. of NIH grants	0.766	1.220	0	133
Num. of R01 grants	0.383	0.655	0	8
Project costs (logged US dollars)	4.137	5.823	0	16.004
Num. of last-authored publications	0.784	1.264	0	24
Num. of forward citations (logged)	1.673	1.964	0	10.403
Disruption score	0.143	0.321	-0.667	1
Research overlap (Main measure, MeSH overlap with advisor's based on pubs within 5 years before lab departure)	0.713	0.292	0	1
Research overlap (excl. coauthored pubs with advisor)	0.194	0.268	0	1
Research overlap (proportion of similar papers to advisor's based on 50% MeSH overlap)	0.508	0.414	0	1
Research overlap (proportion of similar papers to advisor's based on PubMed algorithm)	0.236	0.262	0	1
Post-departure "dynamic" research overlap (MeSH overlap in pubs, yearly updated)	0.466	0.221	0	1
Advisor has received an award	0.154	0.361	0	1
Advisor's employer status	2.932	1.639	0	4
Career age	5.471	2.836	1	10
Num. of publications	2.293	2.327	0	35
Avg. forward citations	1.079	4.836	0	1,176.6
Num. of co-authors	8.366	11.052	0	67
Avg. field maturity (based on cumulative author count)	0.481	0.153	0.053	0.975
Avg. field maturity (alternative, based on MeSH age)	46.375	8.448	15.706	85.6
Year	1992.2	8.809	1972	2009

Notes: Unless otherwise noted, all variables are measured for the (focal) advisee scientists. Number of scientist-year observations = 144,566; number of advisee scientists = 15,271; number of advisor scientists =7,924)

Table 2 Estimates of Advisor-advisee Research Overlap on Advisee Performance (Tests for Main Research Question)

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption score
Regression Model	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.651**	0.640**	0.531**	0.628**	0.412**	0.338**	0.033*	-0.0004
	(0.026)	(0.022)	(0.022)	(0.026)	(0.022)	(0.020)	(0.016)	(0.005)
Advisor has received an award	0.053*	0.041+	0.042+	0.07`**	-0.032	0.059*	0.004	0.010
	(0.027)	(0.023)	(0.023)	(0.027)	(0.024)	(0.023)	(0.019)	(0.006)
Advisor's employer status	0.040**	0.028**	0.013	0.030**	0.028**	-0.0005	0.017**	0.002
	(0.008)	(0.007)	(0.009)	(0.009)	(0.008)	(0.006)	(0.006)	(0.002)
Career age	0.163**	0.138**	0.108**	0.164**	0.080**	0.137**	0.00001	-0.0003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.0004)
Number of publications	0.077**	0.065**	0.071**	0.079**	0.059**	0.309**	0.120**	0.040**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.001)
Average forward citations	0.002**	0.001**	0.001**	0.002**	0.001**	0.002**		0.004**
	(0.001)	(0.0004)	(0.0004)	(0.001)	(0.0004)	(0.0007)		(0.001)
Number of co-authors	-0.001	-0.003**	0.0003	-0.002**	-0.002**	-0.024**	0.012**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.0002)
Average field maturity	0.540**	0.555**	-0.009	0.669**	0.657**	0.165**	0.518**	0.038**
	(0.077)	(0.068)	(0.079)	(0.082)	(0.065)	(0.064)	(0.055)	(0.015)
Constant	-0.610**	-2.647**	-1.293**	-2.699**	0.512**	-2.098**	-0.197**	0.041**
	(0.052)	(0.046)	(0.062)	(0.055)	(0.048)	(0.043)	(0.038)	(0.010)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453964	-89922	-145852	-90696	-465034	-134658	-234801	
R-squared								0.209
Observations (scientist-year)	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes: (1) Unless otherwise noted, all variables in the models are measured for advisee scientists. (2) Research overlap is calculated based on MeSH overlap in publications of the advisee and the advisor during training period (within five years before the advisee departed from his or her advisor's lab). (3) The "Advisor has received an award" variable is assigned a value of one in years after the advisee's advisor has received an award. (4) The numbers of observations in models 4, 5 and 7 are reduced due to the inclusion of advisor fixed effects. (5) Project costs are natural-log transformed after adding one to the value to account for the skewed distribution. Project costs are in US dollars. The amount is adjusted to the year 2000 values to account for inflation. (6) FE OLS regression model was employed for model 8, as opposed to a Poisson PML, because the disruption scores include negative values. (7) In model 7, where the DV is the total forward citations assigned to advisee scientists, the average citation control was omitted due to multicollinearity. Citations are natural-log transformed after adding one to the value to account for the skewed distribution. (8) Robust standard errors in parentheses ** $p<0.01$, * $p<0.05$, + $p<0.1$.

Table 3 Moderating Effect of Advisor Status (Tests for Hypothesis 1)

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored pubs	(7) Log of Num. of forward citations	(8) Disruption score
Regression Model	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.744**	0.724**	0.575**	0.729**	0.462**	0.369**	0.030+	0.004
	(0.029)	(0.025)	(0.025)	(0.030)	(0.025)	(0.021)	(0.017)	(0.005)
Advisor has received an award	0.373**	0.352**	0.211**	0.407**	0.126**	0.168**	0.013	0.032**
	(0.051)	(0.044)	(0.044)	(0.053)	(0.043)	(0.044)	(0.036)	(0.012)
Research overlap	-0.393**	-0.384**	-0.207**	-0.415**	-0.194**	-0.135**	-0.014	-0.028*
× Advisor has received an award	(0.055)	(0.047)	(0.047)	(0.056)	(0.045)	(0.047)	(0.038)	(0.012)
Advisor's employer status	0.039**	0.028**	0.013	0.030**	0.028**	-0.001	0.015**	0.002
	(0.008)	(0.007)	(0.009)	(0.009)	(0.008)	(0.006)	(0.006)	(0.002)
Career age	0.164**	0.138**	0.108**	0.164**	0.080**	0.138**	0.0004	-0.0003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.0004)
Number of publications	0.077**	0.065**	0.071**	0.079**	0.059**	0.309**	0.126**	0.040**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.001)
Average forward citations	0.002**	0.001**	0.001**	0.002**	0.001**	0.003**		0.004**
	(0.0005)	(0.0004)	(0.0005)	(0.0006)	(0.0004)	(0.001)		(0.001)
Number of co-authors	-0.001	-0.003**	0.0003	-0.003**	-0.002**	-0.024**	0.009**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.0002)
Average field maturity	0.561**	0.575**	0.002	0.692**	0.669**	0.174**	0.616**	0.040**
	(0.077)	(0.068)	(0.079)	(0.082)	(0.066)	(0.064)	(0.054)	(0.015)
Constant	-0.689**	-2.718**	-1.330**	-2.786**	0.468**	-2.124**	-0.303**	0.038**
	(0.054)	(0.047)	(0.063)	(0.057)	(0.049)	(0.043)	(0.038)	(0.010)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453797	-89905	-145842	-90674	-464984	-134653	-230296	
R squared								0.209
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of Advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of Advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes for Table 2 apply to this table.

Table 4 Difference-in-difference Analysis of the Moderating Effect of Advisor Status (Tests for Hypothesis 1 on Funding Performance)

DV: Advisee funding performance (GSI score)	(1) Treated Sample: Advisees of award-winner advisors	(2) Matched Sample: Advisees of matched-sample advisors	(3) Full
Research overlap	0.898**	0.991**	0.981**
	(0.101)	(0.090)	(0.089)
After advisor has received an award (Yes=1)	0.688**	-0.152	-0.119
	(0.150)	(0.151)	(0.148)
Research overlap × After advisor has received an award	-0.751**	0.186	0.208
	(0.169)	(0.172)	(0.170)
After advisor has received an award × Treated sample			0.737**
			(0.204)
Research overlap × Treated sample			-0.077
			(0.133)
Research overlap × After advisor has received an award × Treated sample			-0.942**
			(0.240)
Advisor's employer status	0.062	0.097*	0.074*
	(0.048)	(0.040)	(0.030)
Career age	0.167**	0.169**	0.168**
	(0.007)	(0.007)	(0.005)
Number of publications	0.083**	0.128**	0.097**
	(0.011)	(0.012)	(0.008)
Average forward citations	4.706	15.67**	10.59**
	(5.067)	(3.578)	(3.264)
Number of co-authors	-2.170	-2.535	-1.149
	(2.619)	(2.718)	(1.938)
Average field maturity	0.650*	1.135**	0.855**
	(0.313)	(0.285)	(0.212)
Constant	-1.097**	-1.779**	-1.395**
	(0.265)	(0.221)	(0.169)
Year fixed effects	Yes	Yes	Yes
Parent (advisor) fixed effects	Yes	Yes	Yes
Log-Likelihood	-27780	-32820	-60902
Observations	7,771	11,546	19,317
Number of Advisees	785	1,086	1,871
Number of Advisors	218	323	541

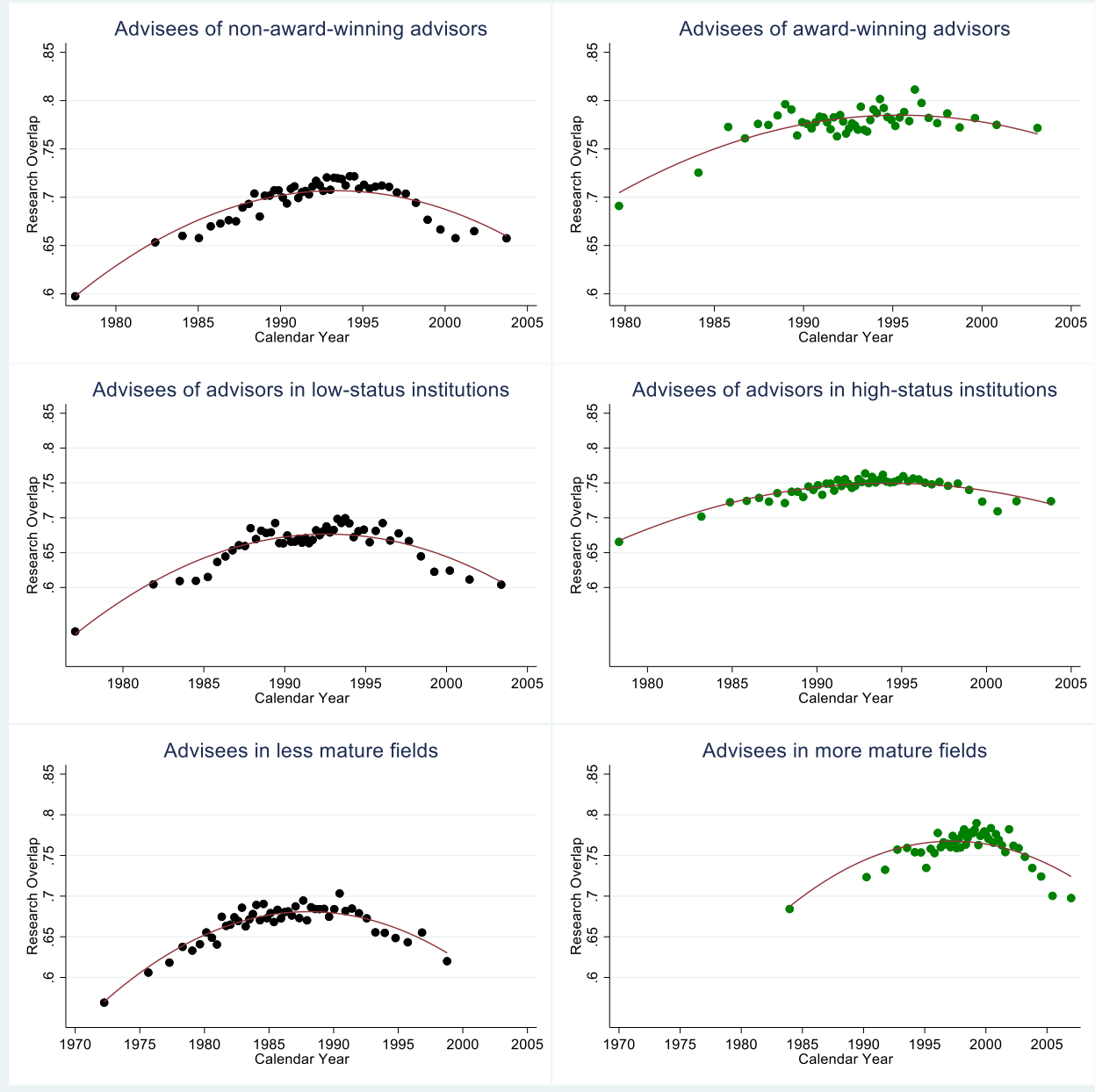
Notes: (1) In model 1, the sample includes advisees of star-scientist advisors who will later receive an award. These advisees departed from the advisors' labs *before* the advisor had received the status-elevation award. (2) In model 2, we used PSM to generate a matched sample of equal-quality advisors. PSM balance statistics reported in Appendix 3. Similar to the subsample in model 1, we required all advisees of these matched-sample advisors to depart before the pseudo-treatment year (the year in which an advisor in this PSM sample was matched to a comparable star-scientist advisor featured in model 1). (3) Model 3 pools the subsamples in models 1 and 2. (4) The variable "After advisor has received an award" is the same as the variable "Advisor has received an award" in Tables 2-3, with the exception that it also changes from 0 to 1 for matched advisors after their pseudo-treatment year. (5) Robust standard errors in parentheses; ** $p<0.01$, * $p<0.05$, + $p<0.1$.

Table 5 Difference-in-difference Analysis of the Moderating Effect of Advisor Status (Tests for Hypothesis 1 on Research Productivity Performance)

DV: Number of Last-authored pubs	(1) Treated Sample: Advisees of award-winner advisors	(2) Matched Sample: Advisees of matched-sample advisors	(3) Full
Research overlap	0.392**	0.592**	0.539**
	(0.082)	(0.071)	(0.065)
After advisor has received an award (Yes=1)	0.210+	-0.115	-0.074
	(0.127)	(0.120)	(0.115)
Research overlap × After advisor has received an award	-0.282+	0.016	-0.013
	(0.145)	(0.134)	(0.129)
After advisor has received an award × Treated sample			0.214
			(0.171)
Research overlap × Treated sample			-0.129
			(0.105)
Research overlap × After advisor has received an award × Treated sample			-0.220
			(0.204)
Advisor's employer status	0.049	0.051+	0.038+
	(0.039)	(0.027)	(0.022)
Career age	0.145**	0.152**	0.149**
	(0.006)	(0.006)	(0.004)
Number of publications	0.204**	0.353**	0.270**
	(0.014)	(0.012)	(0.011)
Average forward citations	13.62**	18.63**	16.75**
	(3.534)	(4.653)	(2.784)
Number of co-authors	-6.272*	-30.72**	-17.67**
	(2.956)	(2.670)	(2.227)
Average field maturity	0.994**	1.240**	1.118**
	(0.256)	(0.204)	(0.160)
Constant	-2.711**	-3.296**	-2.936**
	(0.209)	(0.167)	(0.128)
Year fixed effects	Yes	Yes	Yes
Parent (advisor) fixed effects	Yes	Yes	Yes
Log-Likelihood	-7751	-10374	-18296
Observations	7,771	11,546	19,317
Number of Advisees	785	1,086	1,871
Number of Advisors	218	323	541

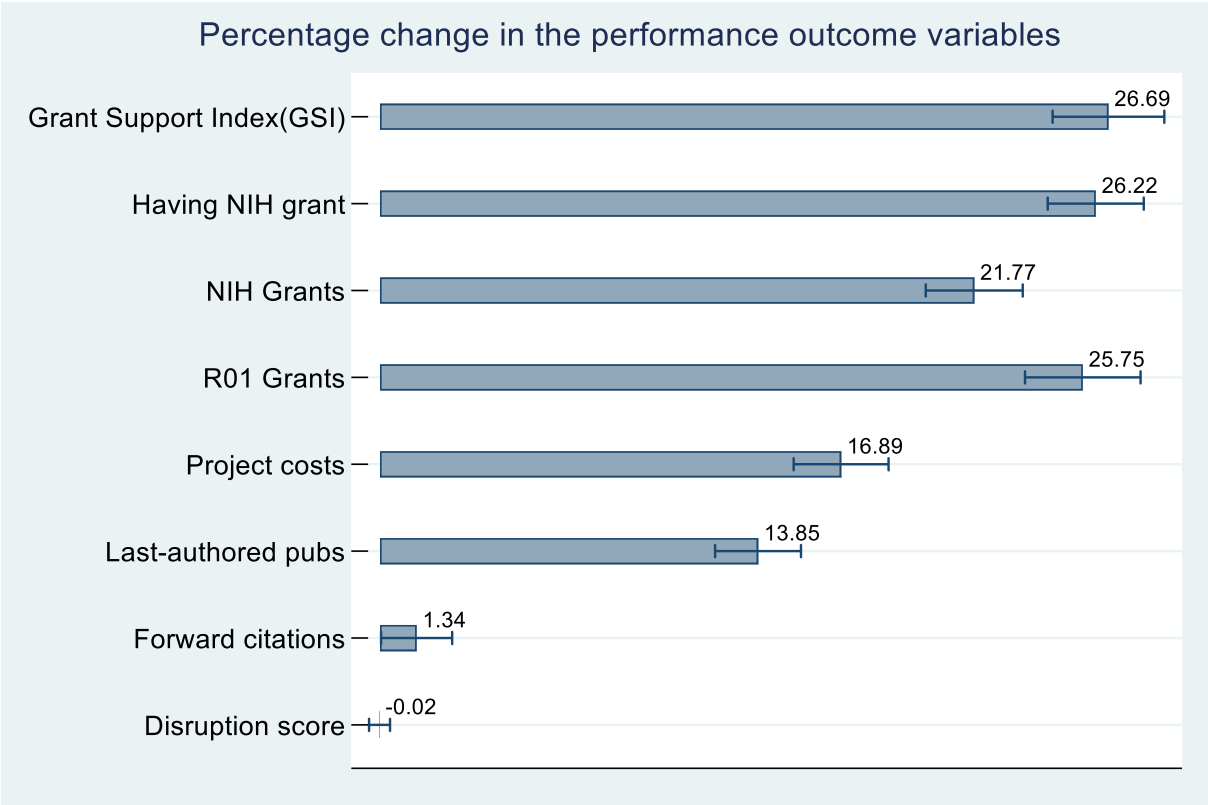
Notes to Table 4 apply to this table.

Figure 1 Trends in Research Overlap Over Time,
Split by Advisor Awards, Advisor Employer Status, and Research Field Maturity



Notes: Figure 1 displays binscatter plots complemented by quadratic fit lines. The same set of control variables used in the regression models is incorporated in the creation of this plot, with the number of bins specified as 50. Research overlap is the proportion of overlapped MeSH keywords during training period (main measure). "Low-status institutions" are defined as those institutions with a Carnegie classification ranking below the median. "Less mature fields" refer to research fields that have a maturity score below the median.

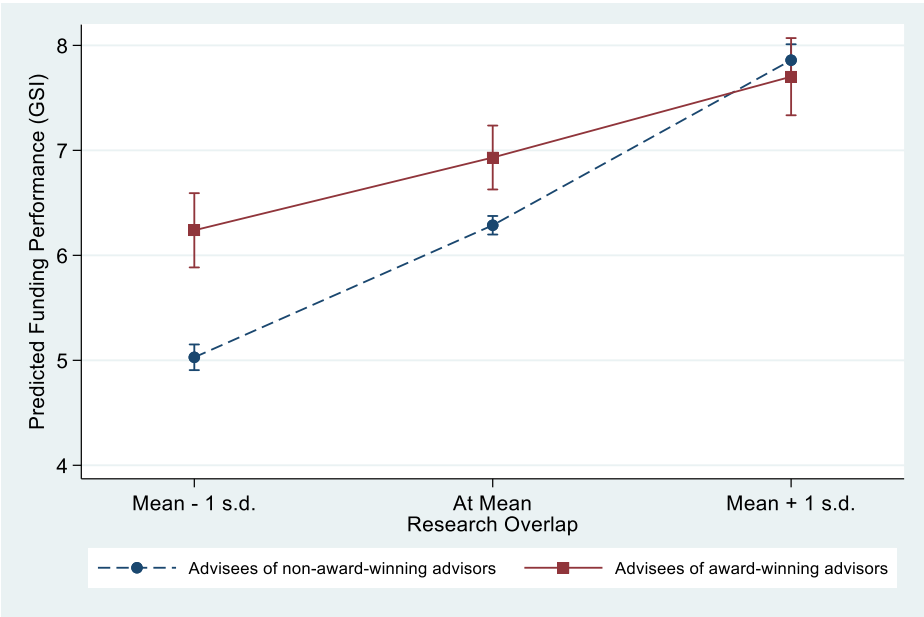
Figure 2 Marginal Effects of Research Overlap on Performance



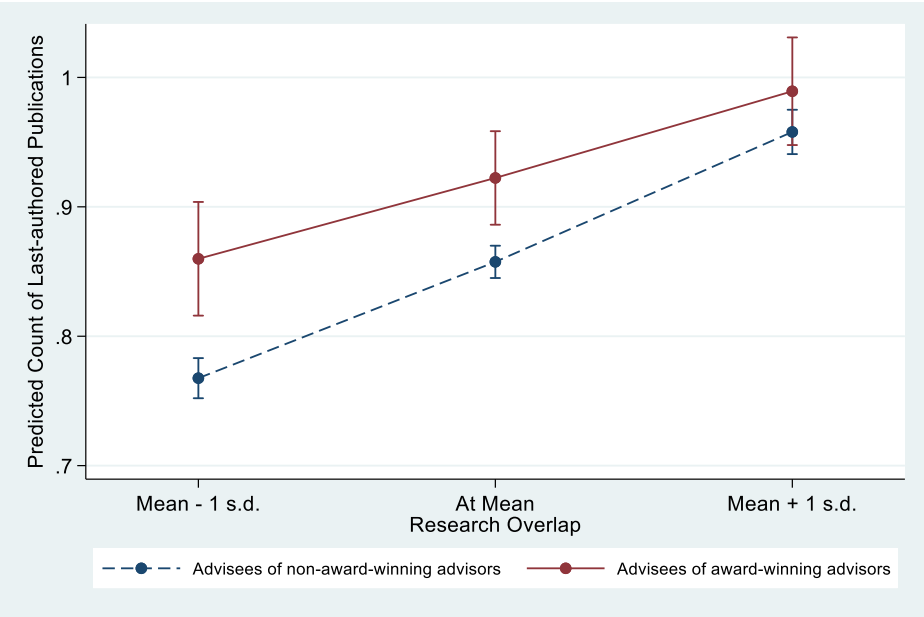
Notes: Figure 2 summarizes marginal effects of one standard deviation of change in research overlap (29% change) from the mean on the eight performance outcome variables assessed in Table 2. Numeric values next to each bar indicates % change in the dependent variable as the advisee's research overlap with his or her advisor moves 29% away from the mean level.

Figure 3 Effect of Research Overlap on Advisee's Predicted Performance Based on Full-Sample Analysis

3A

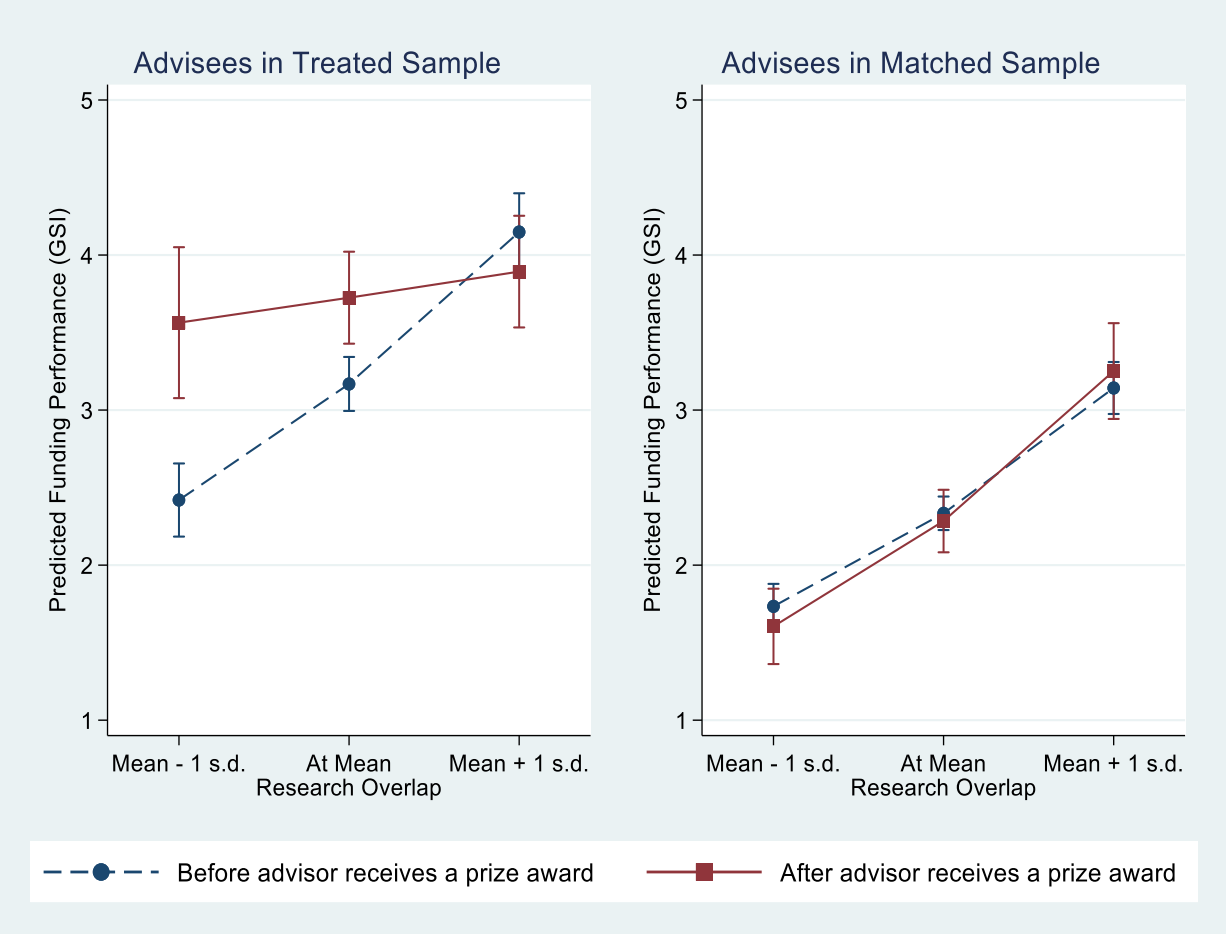


3B



Notes: This figure illustrates the effect of advisor-advisee research overlap on advisee's predicted performance (funding GSI in 3A and last-authored publications in 3B), differentiating between advisees of higher-status advisors (who have received status-elevating awards) represented by solid red lines, and advisees of lower-status advisors (who have not received awards), represented by dashed blue lines. Results are based on Models 1 and 6 of Table 3.

Figure 4 Effect of Research Overlap on Predicted Advisee Funding Performance from Matched-Sample Analysis



Note: Figure 4 illustrates the difference in the moderating effect of "Advisor has received an award" on the research overlap-funding performance relationship between the treated sample (left panel) and the matched sample (right panel), based on models 1 and 2 of Table 4. The definitions of treated and matched sample are in the notes to Table 4 and in text. Advisees mentored by advisors who have *already* received an award are depicted by a solid line with red squares, while those with advisors who have *not yet* received any award are shown with a dashed line and blue circles.

APPENDIX

Appendix 1 Information on Grant Support Index (GSI)

Grant Support Index (GSI), a grant weighting formula, aims to measure input provided by NIH to each project of a principal investigator. Instead of focusing on money and the number of grants each PI holds, GSI considers differences between the level of costs and grant types. With the lead of a trans-NIH Extramural Activities Working Group (EAWG) subcommittee, a group of NIH program and grants management staff developed the GSI (originally designed the "Research Commitment Index" or RCI) (Lauer et al., 2017). They started with assigning 7 points to the most common type of award, the R01 (Lauer et al., 2017). Based on that benchmark and consultation within the group, they assigned more or fewer points to the other grants. The table below shows the GSI scores for a PI per activity code depending on having one or multiple PIs per project.

Table A1 Grant Support Index (GSI) point values for NIH Research Project Grants

Activity Code	Single PI point assignment	Multiple PI point assignment
UM1, UM2	11	10
Subprojects under multi-component awards	6	6
R01, R33, R35, R37, R56, RC4, RF1, RL1, P01, P42, UC4, UF1, UH3, U01, U19, DP1, DP2, DP3, DP4	7	6
R00, R21, R34, R55, RC1, RC2, RL2, RL9, UG3, UH2, U34, DP5	5	4
R03, R24, P30, UC7	4	3

Source: Lauer et al., 2017.

Appendix 2

Table A2 Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Grant Support Index (GSI)	1.00															
(2) Having NIH grant (Yes=1)	0.83	1.00														
(3) Number of NIH grants	0.68	0.55	1.00													
(4) Number of R01 grants	0.88	0.79	0.57	1.00												
(5) Project costs (logged)	0.57	0.65	0.50	0.58	1.00											
(6) Num. of last-authored pubs	0.31	0.26	0.26	0.29	0.24	1.00										
(7) Num. of forward citations (ln)	0.17	0.14	0.15	0.15	0.19	0.34	1.00									
(8) Disruption score	0.02	0.01	0.02	0.02	0.00	0.11	0.33	1.00								
(9) Research overlap	0.05	0.05	0.04	0.05	0.08	0.02	0.05	-0.02	1.00							
(10) Advisor has received an award	0.06	0.05	0.04	0.06	0.08	0.07	0.08	0.01	0.12	1.00						
(11) Advisor's employer status	0.01	-0.01	0.01	0.01	0.03	0.03	0.07	0.02	0.19	0.21	1.00					
(12) Career age	0.28	0.29	0.20	0.27	0.24	0.31	0.17	-0.01	-0.01	0.05	0.00	1.00				
(13) Num. of pubs	0.21	0.16	0.20	0.17	0.16	0.65	0.52	0.20	-0.01	0.02	0.01	0.19	1.00			
(14) Average forward citations	0.06	0.05	0.05	0.05	0.08	0.08	0.38	0.04	0.02	0.05	0.03	0.05	0.12	1.00		
(15) Num. of co-authors	0.18	0.13	0.19	0.13	0.17	0.44	0.54	0.12	0.03	0.03	0.02	0.18	0.84	0.18	1.00	
(16) Average field maturity	0.09	0.09	0.06	0.08	0.28	0.11	0.30	-0.12	0.18	0.13	0.08	0.27	0.08	0.14	0.21	1.00

Note: Number of observations is 144,566.

Appendix 3

Table A3s Balance Between Award-winning Advisors and Non-award-winning Advisors to Prepare for the Final Subsample Used in Table 4 Difference-in-Difference Analysis

	Matched Sample Non-award-winning Advisors Mean (s.d.)	Treated Sample Award-winning Advisors Mean (s.d.)	p-value
Advisor's GSI score	13.307 (12.666)	13.692 (12.250)	0.51
Publication stock	7.328 (6.948)	7.515 (6.239)	0.55
Employer university status	3.727 (.766)	3.735 (1.073)	0.85
Advisor's career age	25.964 (9.679)	25.537 (9.283)	0.34
Pre-award research overlap with advisee	.570 (.160)	.5642 (.148)	0.41
N	1,356	689	

Notes: This table compares the characteristics of award-winning advisors (those who have received an award at any point in their career) with that of the non-award-winning advisors (those who are yet to receive an award at the point of matching) whom we matched to the award-winning set using propensity-score-matching (the nearest two neighbors, psmatch2 package in Stata). In line with Stuart and Rubin (2008), we set the caliper at one-quarter of the standard deviation of the propensity score to refine our matching process.

Table A3b Balance Between Award-winning Advisors and Non-award-winning Advisors of the Advisees Used in Table 4 Difference-in-Difference Analysis

	Matched Sample Non-award-winning Advisors Mean (s.d.)	Treated Sample Award-winning Advisors Mean (s.d.)	p-value
Advisor's GSI score	12.552 (15.494)	14.954 (12.845)	0.06
Publication stock	7.975 (7.156)	8.963 (8.248)	0.14
Employer university status	25.238 (18.314)	31.534 (16.950)	<0.001
Advisor's career age	24.932 (9.315)	25.064 (8.387)	0.87
Pre-award research overlap with advisee	.560 (.121)	.553 (.112)	0.50
N	323	218	

Notes: This table compares the characteristics of award-winning advisors of the advisees featured in the Treated sample in Model 1 of Table 4 of main text, with that of the non-award-winning advisors of the advisees featured in the Matched sample in Model 2 of Table 4. These two sets of advisors are based on the corresponding sets in Table A3a but become smaller in the final analysis in Table 4 due to sample attrition caused by the inclusion of advisor fixed effects—when advisor fixed effects are included, those advisors (along with their advisees) who do not have >1 advisees in their observation window are dropped from analysis. Consequentially, the matched-sample advisees in Table 4 have advisors whose employer statuses are higher. We included control for advisor employer status in Table 4 estimation. As part of our robustness checks, we incorporated control for the advisor's GSI score in the estimation presented in Table 9c.

Appendix 4

Table A4a Coefficient Estimates and Standard Errors from Replication of Table 2 with Alternative Research Overlap Measures

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption score
Regression Model	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap (original used in main tables)	0.651**	0.640**	0.531**	0.628**	0.412**	0.338**	0.033*	-0.0004
	(0.026)	(0.022)	(0.022)	(0.026)	(0.022)	(0.020)	(0.016)	(0.005)
Research overlap (different from the original by excluding coauthored pubs with advisors)	0.440**	0.439**	0.390**	0.437**	0.310**	0.158**	-0.034*	-0.006
	(0.021)	(0.018)	(0.019)	(0.022)	(0.017)	(0.018)	(0.014)	(0.004)
Research overlap (proportion of similar papers to advisor's based on 50% MeSH overlap)	0.198**	0.210**	0.158**	0.235**	0.146**	0.081**	0.019+	-0.003
	(0.015)	(0.013)	(0.014)	(0.016)	(0.013)	(0.012)	(0.010)	(0.003)
Research overlap (proportion of similar papers to advisor's based on PubMed algorithm)	0.342**	0.331**	0.296**	0.367**	0.255**	0.082**	0.140**	0.024**
	(0.022)	(0.020)	(0.021)	(0.023)	(0.019)	(0.020)	(0.015)	(0.005)
Dynamic Research Overlap (based on publications after advisee has departed from advisor lab, yearly updated)	0.227**	0.275**	0.204**	0.270**	0.209**	0.024	-0.054**	-0.006
	(0.0251)	(0.022)	(0.023)	(0.026)	(0.022)	(0.021)	(0.018)	(0.005)

Notes: each cell reports the coefficient estimate and standard error based on a regression using the "research overlap" in that row on the outcome variable in that column. Except for the "research overlap" measure, all other aspects of model specifications are as in Table 2.

Table A4b Coefficient Estimates and Standard Errors from Replication of Table 3 with Alternative Research Overlap Measures

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption score
Regression Model	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap (original used in main tables)	0.744**	0.724**	0.575**	0.729**	0.462**	0.369**	0.029	0.004
	(0.029)	(0.025)	(0.025)	(0.030)	(0.025)	(0.021)	(0.018)	(0.005)
Advisor has received an award	0.373**	0.352**	0.211**	0.407**	0.126**	0.168**	-0.010	0.032**
	(0.051)	(0.044)	(0.044)	(0.053)	(0.043)	(0.044)	(0.037)	(0.012)
Research overlap (original used in main tables)	-0.393**	-0.384**	-0.207**	-0.415**	-0.194**	-0.135**	0.017	-0.028*
× Advisor has received an award	(0.055)	(0.047)	(0.047)	(0.056)	(0.045)	(0.047)	(0.039)	(0.012)
Research overlap (different from the original by excluding coauthored pubs with advisors)	0.501**	0.496**	0.421**	0.512**	0.337**	0.227**	-0.044**	-0.010+
	(0.025)	(0.022)	(0.024)	(0.027)	(0.021)	(0.019)	(0.017)	(0.005)
Advisor has received an award	0.092**	0.0797**	0.059*	0.120**	-0.018	0.109**	-0.005	0.006
	(0.029)	(0.025)	(0.025)	(0.030)	(0.026)	(0.025)	(0.021)	(0.006)
Research overlap	-0.179**	-0.176**	-0.106**	-0.213**	-0.079*	-0.213**	0.033	0.016+
× Advisor has received an award	(0.042)	(0.036)	(0.038)	(0.044)	(0.034)	(0.040)	(0.031)	(0.010)
Research overlap (proportion of similar papers to advisor's based on 50% MeSH overlap)	0.236**	0.233**	0.177**	0.278**	0.178**	0.0861*	0.015	-0.003
	(0.018)	(0.015)	(0.016)	(0.019)	(0.015)	(0.013)	(0.011)	(0.003)
Advisor has received an award	0.130**	0.087**	0.081**	0.159**	0.036	0.066*	-0.007	0.009
	(0.034)	(0.029)	(0.030)	(0.035)	(0.030)	(0.029)	(0.025)	(0.007)
Research overlap	-0.139**	-0.089**	-0.081**	-0.151**	-0.109**	-0.021	0.015	0.001
× Advisor has received an award	(0.032)	(0.027)	(0.028)	(0.033)	(0.026)	(0.027)	(0.023)	(0.007)
Research overlap (proportion of similar papers to advisor's based on PubMed algorithm)	0.458**	0.449**	0.371**	0.499**	0.374**	0.132**	0.174**	0.027**
	(0.026)	(0.023)	(0.025)	(0.028)	(0.023)	(0.021)	(0.017)	(0.005)
Advisor has received an award	0.155**	0.151**	0.112**	0.187**	0.075**	0.102**	0.040+	0.014*
	(0.030)	(0.026)	(0.027)	(0.031)	(0.027)	(0.026)	(0.022)	(0.007)
Research overlap	-0.375**	-0.407**	-0.276**	-0.424**	-0.386**	-0.168**	-0.121**	-0.014
× Advisor has received an award	(0.047)	(0.040)	(0.042)	(0.049)	(0.039)	(0.046)	(0.035)	(0.011)
Dynamic Research Overlap (based on pubs after advisee has departed from advisor lab, yearly updated)	0.295**	0.307**	0.255**	0.340**	0.247**	0.021	-0.055**	-0.004
	(0.028)	(0.025)	(0.026)	(0.030)	(0.024)	(0.022)	(0.019)	(0.005)
Advisor has received an award	0.205**	0.109**	0.165**	0.225**	0.047	0.048	0.001	0.017+
	(0.039)	(0.034)	(0.035)	(0.041)	(0.034)	(0.034)	(0.030)	(0.009)
Research overlap	-0.319**	-0.154**	-0.260**	-0.322**	-0.165**	0.013	0.003	-0.013
× Advisor has received an award	(0.058)	(0.051)	(0.050)	(0.060)	(0.047)	(0.050)	(0.043)	(0.013)

Notes: each cell reports the coefficient estimates and standard errors based on a regression using the "research overlap" in that row, "advisor has received an award", and the two-way interaction term on the outcome variable in that column. Except for the "research overlap" measure, all other aspects of model specifications are as in Table 3.

Appendix 5

Table A5 Replication of Table 3 with an Alternative Advisor's Status Measure (Employer Status Based on Carnegie Classification)

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Model	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.879**	0.825**	0.728**	0.909**	0.601**	0.424**	0.0857*	0.0120
	(0.055)	(0.049)	(0.046)	(0.060)	(0.050)	(0.041)	(0.036)	(0.010)
Advisor's employer status	0.088**	0.068**	0.055**	0.089**	0.068**	0.018+	0.028**	0.005+
	(0.013)	(0.012)	(0.012)	(0.014)	(0.013)	(0.010)	(0.009)	(0.0024)
Research overlap × advisor's employer status	-0.073**	-0.059**	-0.063**	-0.089**	-0.059**	-0.028*	-0.017+	-0.004
	(0.016)	(0.014)	(0.014)	(0.017)	(0.014)	(0.012)	(0.010)	(0.003)
Advisor has received an award	0.057*	0.044+	0.046*	0.074**	-0.030	0.061**	0.004	0.001+
	(0.027)	(0.023)	(0.023)	(0.027)	(0.024)	(0.023)	(0.020)	(0.006)
Constant	-0.761**	-2.769**	-1.423**	-2.884**	0.385**	-2.153**	-0.231**	0.034**
	(0.062)	(0.055)	(0.064)	(0.066)	(0.057)	(0.049)	(0.043)	(0.011)
Controls included as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent (advisor) fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453892	-89916	-145838	-90684	-464981	-134655	-234798	-23862
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of Advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of Advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes to Table 3 apply. Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1

Appendix 6

Table A6a Moderating Effect of Average Research Field Maturity (based on researcher count) on Overlap-Performance Relationship

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Model	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	1.029**	1.091**	0.930**	1.042**	0.963**	0.239**	0.050	0.013
	(0.068)	(0.058)	(0.057)	(0.071)	(0.066)	(0.055)	(0.046)	(0.013)
Average field maturity (researcher-count-based)	1.145**	1.277**	0.634**	1.335**	1.458**	0.009	0.544**	0.059**
	(0.127)	(0.109)	(0.110)	(0.134)	(0.111)	(0.102)	(0.082)	(0.023)
Research overlap × Average field maturity	-0.811**	-0.975**	-0.863**	-0.890**	-1.076**	0.211+	-0.036	-0.029
	(0.134)	(0.114)	(0.116)	(0.141)	(0.120)	(0.113)	(0.087)	(0.024)
Advisor has received an award	0.057*	0.045*	0.046*	0.074**	-0.027	0.059*	0.0038	0.010+
	(0.027)	(0.023)	(0.023)	(0.027)	(0.024)	(0.023)	(0.019)	(0.006)
Advisor's employer status	0.037**	0.025**	0.010	0.027**	0.025**	0.0003	0.017**	0.002
	(0.008)	(0.007)	(0.010)	(0.009)	(0.008)	(0.006)	(0.006)	(0.002)
Constant	-0.875**	-2.960**	-1.572**	-2.989**	0.120+	-2.029**	-0.209**	0.033**
	(0.069)	(0.059)	(0.068)	(0.072)	(0.066)	(0.055)	(0.047)	(0.012)
Controls included as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453842	-89903	-145820	-90679	-464809	-134656	-234800	-23863
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of Advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of Advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes to Table 3 apply. Average research field maturity is measured based on the ratio of count of researchers publishing on a MeSH area in a given year to the maximum researcher count in the MeSH area throughout the observation window. Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1

Table A6b Moderating Effect of Average Research Field Maturity (based on MeSH Term Age) on Overlap-Performance Relationship

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Model	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	1.400**	1.312**	1.113**	1.371**	1.011**	0.412**	0.347**	0.015*
	(0.075)	(0.069)	(0.059)	(0.080)	(0.070)	(0.067)	(0.054)	(0.007)
Average field maturity (MeSH term age-based)	0.016**	0.016**	0.013**	0.016**	0.012**	0.025**	0.033**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0009)	(0.0001)
Research overlap × Average field maturity	-0.017**	-0.016**	-0.013**	-0.017**	-0.013**	-0.002	-0.007**	-0.0005**
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.0001)
Advisor has received an award	0.063*	0.051*	0.049*	0.081**	-0.020	0.066**	0.019	0.012*
	(0.027)	(0.023)	(0.023)	(0.027)	(0.024)	(0.023)	(0.019)	(0.006)
Advisor's employer status	0.037**	0.026**	0.010	0.028**	0.027**	-0.001	0.0156**	0.002
	(0.008)	(0.007)	(0.009)	(0.009)	(0.008)	(0.007)	(0.006)	(0.002)
Constant	-1.019**	-3.019**	-1.856**	-3.053**	0.316**	-3.133**	-1.358**	-0.030**
	(0.067)	(0.061)	(0.058)	(0.071)	(0.063)	(0.056)	(0.047)	(0.007)
Controls included as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453615	-89880	-145752	-90663	-464890	-133955	-231631	-23504
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of Advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of Advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes to Table 3 apply. Average research field maturity is measured based on the age of each MeSH term (number of years since a MeSH term first appeared in PubMed). Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1

Appendix 7

Table A7a Replication of Table 3 (with Advisor's Award only Counting Junior, <50 Age Bracket, Prize Awards)

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Models	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.734**	0.720**	0.536**	0.711**	0.422**	0.374**	0.051**	0.006
	(0.032)	(0.028)	(0.028)	(0.034)	(0.028)	(0.023)	(0.019)	(0.006)
Advisor has received an award	0.945**	0.897**	0.647*	0.944**	0.788*	0.458	0.229	0.105+
	(0.334)	(0.328)	(0.267)	(0.354)	(0.327)	(0.319)	(0.253)	(0.060)
Research overlap × Advisor has received an award	-1.162*	-1.149*	-0.869*	-1.261*	-0.794+	-0.658+	-0.127	-0.105
	(0.481)	(0.481)	(0.373)	(0.527)	(0.444)	(0.393)	(0.326)	(0.071)
Controls included as in Table 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent(advisor) fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-338383	-68212	-110258	-67152	-350430	-100867	-175188	-16036
Observations	111,421	111,421	111,497	98,706	81,778	111,522	109,580	111,596
Number of Advisees	11,804	11,804	11,825	10,449	9,752	11,825	11,614	11,854
Number of Advisors	6,790	6,790	6,803	5,835	5,658	6,804	6,604	6,830

Notes to Table 3 apply. In computing "Advisor has received an award" variable, we used only the subset of awards targeting the below-50 age bracket. See awards' age distribution in Figure A1. All advisee-year observations associated with advisors who have received senior (above-50 age bracket) awards are excluded from analysis. Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1

Table A7b Replicating Table 4 (with Advisor's Award only Counting Junior, <50 Age Bracket, Prize Awards)

DV: Advisee funding performance (GSI score)	(1)	(2)	(3)
Sample:	Treated Sample: Advisees of award-winner advisors	Matched Sample: Advisees of matched-sample advisors	Full
Research overlap	4.331**	0.114	0.172
	(0.789)	(0.172)	(0.172)
After advisor has received an award (Yes=1)	2.252**	-0.537+	-0.396
	(0.814)	(0.311)	(0.314)
Research overlap × After advisor has received an award	-3.030**	0.364	0.217
	(0.929)	(0.354)	(0.357)
After advisor has received an award× Treated sample			2.035**
			(0.764)
Research overlap × Treated sample			3.142**
			(0.657)
Research overlap × After advisor has received an award × Treated sample			-2.176*
			(0.896)
Controls as in Table 5	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes
Log-Likelihood	-1133	-7333	-8705
Advisee-year observations	406	2,289	2,695

Notes to Table 4 apply.

Appendix 8

Table A8a Replicating Table 3 with Advisee Fixed Effects (replacing Advisor FE)

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Models	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Advisor has received an award	0.612**	0.670**	0.492**	0.645**	0.856**	0.171*	0.065	0.036
	(0.106)	(0.100)	(0.087)	(0.111)	(0.111)	(0.087)	(0.080)	(0.026)
Research overlap × Advisor has received an award	-0.688**	-0.822**	-0.499**	-0.722**	-1.065**	-0.132	-0.102	-0.047
	(0.127)	(0.119)	(0.106)	(0.133)	(0.130)	(0.105)	(0.099)	(0.032)
Controls as in Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisee fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-326244	-77130	-122801	-74718	-344815	-126818	-221442	-17375
Observations	105,753	105,753	119,397	91,374	84,532	144,530	139,282	144,566

Notes to Table 3 apply. The numbers of observations in the models are reduced, relative our main Table 3, due to the inclusion of advisee fixed effects.

Table A8b Replication of Table 4 with Advisee Fixed Effects (replacing Advisor FE)

DV: Advisee funding performance (GSI score)	(1)	(2)	(3)
Sample:	Treated Sample: Advisees of award-winner advisors	Matched Sample: Advisees of matched-sample advisors	Full
After advisor has received an award) (Yes=1)	0.829**	0.176	0.193
	(0.175)	(0.159)	(0.157)
Research overlap × After advisor has received an award	-1.024**	-0.368*	-0.355+
	(0.202)	(0.187)	(0.186)
After advisor has received an award × Treated sample			0.581*
			(0.231)
Research overlap × After advisor has received an award × Treated sample			-0.639*
			(0.272)
Other controls in Table 5	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Advisee fixed effects	Yes	Yes	Yes
Log-Likelihood	-36317	-16584	-19615
Observations	5,243	6,406	11,649
Number of advisees	529	623	1,152
Number of advisors	218	323	541

Notes to Table 4 apply. The numbers of observations in the models are reduced, relative our main Table 4, due to the inclusion of advisee fixed effects. The coefficients for research overlap and the interaction term (Research overlap × Treated sample) are omitted due to the inclusion of advisee fixed effects. Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1.

Appendix 9

Table A9a Replication of Table 2 with Additional Advisor Productivity Controls Included

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Models	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.646**	0.638**	0.524**	0.627**	0.415**	0.337**	0.033*	0.000004
	(0.026)	(0.022)	(0.022)	(0.026)	(0.022)	(0.020)	(0.016)	(0.005)
Advisor has received an award	0.051+	0.041+	0.041+	0.071*	-0.032	0.066**	0.004	0.010
	(0.027)	(0.023)	(0.023)	(0.028)	(0.024)	(0.023)	(0.019)	(0.006)
Advisor lab size	0.006	-0.018+	0.026*	-0.013	-0.033**	-0.013	-0.005	-0.002
	(0.012)	(0.010)	(0.010)	(0.012)	(0.009)	(0.010)	(0.008)	(0.002)
Advisor's num. of publications	-0.004**	-0.002+	-0.004**	-0.005**	-0.003**	-0.013**	0.0006	0.0005+
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.0003)
Advisor's funding performance (GSI)	0.005**	0.005**	0.004**	0.004**	0.002**	0.001*	0.0004	-0.0002
	(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0004)	(0.0001)
Other Controls in Table 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453492	-89885	-145782	-90673	-464947	-134554	-234798	-23860
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes to Table 2 apply. This table has three added advisor productivity control—advisor's lab size (measured by number of other members in the advisor's lab), advisor's number of publications, and advisor's funding performance (measured by advisor's yearly Grant Support Index). Robust standard errors in parentheses ** $p<0.01$, * $p<0.05$, + $p<0.1$.

Table A9b Replication of Table 3 with Additional Advisor Productivity Controls Included

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Models	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.737**	0.721**	0.566**	0.727**	0.465**	0.369**	0.029+	0.005
	(0.029)	(0.025)	(0.025)	(0.030)	(0.025)	(0.021)	(0.018)	(0.005)
Advisor has received an award	0.364**	0.347**	0.204**	0.404**	0.124**	0.178**	-0.011	0.032**
	(0.051)	(0.044)	(0.044)	(0.053)	(0.043)	(0.044)	(0.037)	(0.012)
Research overlap × Advisor has received an award	-0.385**	-0.377**	-0.201**	-0.411**	-0.192**	-0.139**	0.018	-0.028*
	(0.055)	(0.047)	(0.047)	(0.056)	(0.045)	(0.047)	(0.039)	(0.012)
Advisor lab size	0.005	-0.018+	0.026*	-0.013	-0.033**	-0.013	-0.005	-0.002
	(0.012)	(0.010)	(0.010)	(0.012)	(0.009)	(0.010)	(0.008)	(0.002)
Advisor's num. of publications	-0.004**	-0.002+	-0.004**	-0.005**	-0.003**	-0.013**	0.0006	0.0005+
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.0003)
Advisor's funding performance (GSI)	0.005**	0.005**	0.004**	0.004**	0.002**	0.0009*	0.0004	-0.0002
	(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0004)	(0.0001)
Other Controls in Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453332	-89870	-145772	-90652	-464897	-134549	-234798	-23857
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes to Table 3 apply. This table has three added advisor productivity control—advisor's lab size (measured by number of other members in the advisor's lab), advisor's number of publications, and advisor's funding performance (measured by advisor's yearly Grant Support Index). Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1.

Table A9c Replication of Table 4 with Additional Advisor Productivity Controls Included

DV: Advisee funding performance (GSI score)	(1)	(2)	(3)
Sample:	Treated Sample: Advisees of award-winner advisors	Matched Sample: Advisees of matched-sample advisors	Full
Research overlap	0.890**	1.105**	1.116**
	(0.206)	(0.211)	(0.222)
After advisor has received an award) (Yes=1)	0.700**	-0.004	0.066
	(0.242)	(0.284)	(0.282)
Research overlap × After advisor has received an award	-0.737**	-0.054	-0.060
	(0.280)	(0.322)	(0.326)
After advisor has received an award × Treated sample			0.543
			(0.379)
Research overlap × Treated sample			-0.247
			(0.313)
Research overlap × After advisor has received an award × Treated sample			-0.633
			(0.439)
Advisor lab size	0.032	0.051	0.044
	(0.103)	(0.084)	(0.066)
Advisor's num. of publications	0.004	-0.004	-0.0003
	(0.006)	(0.009)	(0.005)
Advisor's funding performance (GSI)	0.001	0.004	0.002
	(0.003)	(0.003)	(0.002)
Other controls in Table 5	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes
Log-Likelihood	-27772	-33791	-61962
Observations	7,771	11,546	19,317
Number of advisees	785	1,086	1,871
Number of advisors	218	323	541

Notes to Table 4 apply. This table has three added advisor productivity control—advisor's lab size (measured by number of other members in the advisor's lab), advisor's number of publications, and advisor's funding performance (measured by advisor's yearly Grant Support Index). Robust standard errors in parentheses ** $p<0.01$, * $p<0.05$, + $p<0.1$.

Appendix 10

Table A10a Replication of Table 2 with Standard Errors Clustered at advisor level

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Models	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.651**	0.640**	0.531**	0.628**	0.412**	0.338**	0.033	-0.0004
	(0.059)	(0.051)	(0.052)	(0.061)	(0.051)	(0.034)	(0.023)	(0.006)
Advisor has received an award	0.053	0.041	0.042	0.071	-0.032	0.060+	0.004	0.010
	(0.053)	(0.044)	(0.047)	(0.054)	(0.044)	(0.033)	(0.031)	(0.008)
Controls in Table 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453964	-89922	-145852	-90696	-465034	-134658	-234801	-23864
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

Notes to Table 2 apply. Standard errors are clustered at advisor level. ** p<0.01, * p<0.05, + p<0.1

Table A10b Replication of Table 3 with Standard Errors Clustered at advisor level

Dependent Variables	(1) Grant Support Index (GSI)	(2) Having NIH grant	(3) Num. of NIH grants	(4) Num. of R01 grants	(5) Project costs	(6) Num. of last-authored publications	(7) Log of Num. of forward citations	(8) Disruption Score
Regression Models	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	PML Poisson	OLS
Research overlap	0.744**	0.724**	0.575**	0.729**	0.462**	0.369**	0.029	0.004
	(0.067)	(0.057)	(0.058)	(0.069)	(0.058)	(0.034)	(0.025)	(0.006)
Advisor has received an award	0.373**	0.352**	0.211*	0.407**	0.126	0.168*	-0.010	0.032*
	(0.104)	(0.092)	(0.090)	(0.107)	(0.090)	(0.073)	(0.051)	(0.014)
Research overlap × Advisor has received an award	-0.393**	-0.384**	-0.207*	-0.415**	-0.194+	-0.135	0.017	-0.028+
	(0.114)	(0.101)	(0.099)	(0.118)	(0.099)	(0.083)	(0.055)	(0.015)
Controls in Table 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log-Likelihood	-453797	-89905	-145842	-90674	-464984	-134653	-234800	-23860
Observations	144,566	144,566	144,566	130,581	107,361	144,566	142,553	144,566
Number of advisees	15,271	15,271	15,271	13,783	12,766	15,271	15,044	15,271
Number of advisors	7,924	7,924	7,924	6,898	6,685	7,924	7,708	7,924

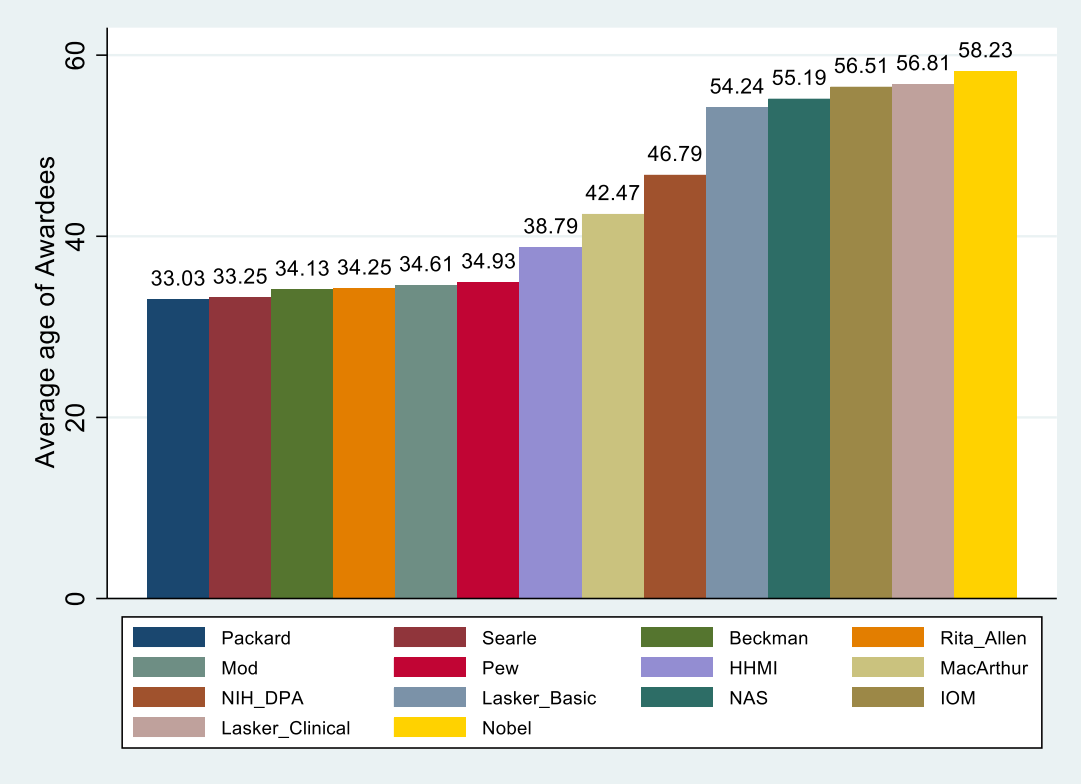
Notes to Table 3 apply. Standard errors are clustered at advisor level. ** p<0.01, * p<0.05, + p<0.1

Table A10c Replication of Table 4 with Standard Errors Clustered at advisor level

DV: Advisee funding performance (GSI score)	(1)	(2)	(3)
Sample:	Treated Sample: Advisees of award-winner advisors	Matched Sample: Advisees of matched-sample advisors	Full
Research overlap	0.898**	1.033**	1.018**
	(0.206)	(0.194)	(0.193)
After advisor has received an award) (Yes=1)	0.688**	-0.199	-0.156
	(0.235)	(0.297)	(0.291)
Research overlap × After advisor has received an award	-0.751**	0.158	0.178
	(0.282)	(0.332)	(0.328)
After advisor has received an award × Treated sample			0.777*
			(0.370)
Research overlap × Treated sample			-0.089
			(0.281)
Research overlap × After advisor has received an award × Treated sample			-0.924*
			(0.436)
Controls in Table 5	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Advisor fixed effects	Yes	Yes	Yes
Log-Likelihood	-61567	-27780	-33360
Observations	7,771	11,546	19,317
Number of advisees	785	1,086	1,871
Number of advisors	218	323	541

Notes to Table 4 apply. Standard errors are clustered at advisor level. ** p<0.01, * p<0.05, + p<0.1

Figure A1 Distribution of Age Cutoff in the Prizes and Awards Used in Analysis



Note: This figure presents the average age of awardees (award-winning-advisors) for each award.