Does Grant Peer Review Penalize Scientific Risk Taking? Evidence from the NIH*

Wesley H. Greenblatt
MIT
Sloan School of Management
100 Main Street
E62-485
Cambridge, MA 02142

Suman K. Maity
MIT
Department of Brain and Cognitive Sciences
77 Massachusetts Ave.
Building 46-3027D
Cambridge, MA 02139

Roger P. Levy
MIT
Department of Brain and Cognitive Sciences
77 Massachusetts Ave.
Building 46-3033
Cambridge, MA 02139

Pierre Azoulay
MIT and NBER
Sloan School of Management
100 Main Street
E62-487
Cambridge, MA 02142

June 21, 2023

Abstract

Scientific projects that carry a high degree of risk may be more likely to lead to breakthroughs yet also face challenges in winning the support necessary to be carried out. We analyze the determinants of renewal for more than 100,000 R01 grants from the National Institutes of Health between 1980 and 2015. We use four distinct proxies to measure risk taking: extreme tail outcomes, disruptiveness, pivoting from an investigator's prior work, and standing out from the crowd in one's field. After carefully controlling for investigator, grant, and institution characteristics, we measure the association between risk taking and grant renewal. Across each of these measures, we find that risky grants are renewed at markedly lower rates than less risky ones. This penalty appears relatively homogeneous across variation in investigator demographic characteristics, prominence, and grantsmanship experience. We also provide evidence that the magnitude of the risk penalty is magnified for more novel grants, consistent with the academic community's perception that current scientific institutions do not motivate exploratory research adequately.

Keywords: risk, scientific productivity, breakthrough innovation, government funding, biomedical innovation.

JEL Classification: O32, O38, H51

* We thank Scott Stern and Danielle Li, as well as seminar participants at the NBER for useful feedback.

1. Introduction

No funder of science has ever prided itself on supporting "low risk, low reward" projects, yet few ideas elicit as much consensus as the claim that peer review punishes risk taking. These laments appear in the popular press, in U.S. Congressional testimony, errant tweets, or the pages of *Nature*, and are authored by successful investigators and leading science policy makers (Kent 2018, Kolata 2009, Kornberg 2007, Nielson 2022, Sutter 2022, Woolston 2014). Prominent scientists have penned editorials satirizing the behavior of risk-averse grant funders, and ponder whether Nobel Prize winners could receive funding for their seminal discoveries in the modern funding climate (Fields 2014, Petsko 2012). While economists would caution that a benevolent social planner might want to tame individual scientists' risk appetites, the proposition that current institutions do not adequately reward risky research strikes practicing scientists as self-evidently true.

Yet, for all the colorful anecdotes, blue-ribbon panels, and editorials, there is surprisingly little evidence to support the claim that peer review punishes risk taking, and less evidence still that society as a whole would be better served if funders could clamp down on their conservatism. Azoulay et al. (2011) compare two grant funding mechanisms—the Howard Hughes Medical Institute Investigatorships and National Institutes of Health (NIH) R01 grants—and provide evidence that the former encourages more risk taking than the latter, but this does not necessarily imply that peer review at NIH punishes risk taking. Other scholars have examined how novelty shapes grant funding outcomes (Ayoubi et al. 2021, Packalen and Bhattacharya 2020, Veugelers et al. 2022). However, novelty is distinct from risk taking: investigators can pursue novel research which isn't risky, and risky research which isn't novel. What is widely agreed upon is that grant funders' tolerance for risk has direct implications for what research is undertaken in the first place. In a recent survey of investigators supported by Fast Grants, 78% stated they would make significant changes to their research program if they were not constrained by the demands of grant funding agencies, including pursuing more ambitious research programs, pivoting to new topics, and testing hypotheses others see as unlikely to succeed (Collison et al. 2021).

More generally, the choice of what scientists select to study is fundamental to determining the rate and direction of innovation. An assumption typically made by models of the innovation process is that exploratory projects are less likely to bear fruit than projects that merely seek incremental advances, but with more upside if they are indeed successful (March 1991; Manso 2011). This is the sense in which science funders always claim to crave "high risk, high reward"

projects even though the trade-off between risk and reward in the market for scientific ideas is less obvious than in financial markets (Nielsen and Qiu 2022). While both risky exploratory research and incremental lines of inquiry are valuable from a societal point of view, a first-order concern in contemporary science policy is that funding programs tilt the mix of projects they support too far in the direction of incrementalism. Moreover, the societal costs of undersupplying risky research may be especially prominent for early-stage, basic research, which may lead to substantial positive spillovers (Azoulay and Li 2022).

What causes individual scientists to select risky projects? Rather than being solely determined by an individual's preference for risk taking, this choice is shaped in important ways by the incentives, organization, and supporting institutions that surround a researcher. Tolerance for failure (Tian and Wang 2014), long time horizons (Lerner and Wulf 2007), team structure (Lee et al. 2015, Wu et al. 2017), and the structure of financial incentives (Graff Zivin and Lyons 2020) shape how much risk innovators choose to bear when selecting projects. The choice of research approach is made more challenging by the difficulty in predicting the ultimate benefits if the project is successful; this is especially the case for early stage and highly uncertain projects (Rosenberg 1995).

In this paper, we study how a specific institution, grant peer review at NIH, shapes scientific risk taking. In grant peer review, scientific proposals are evaluated by a panel of experts, with the resulting scores having a major influence on the likelihood of funding. Scientists who are awarded funding use these resources to conduct the investigations they committed to in their proposal. At the end of the grant cycle, a funded scientist has the option of extending the research program by applying for an additional cycle of support. To study the relationship between peer review and risk taking, one would ideally have access to all proposals (regardless of funding status), reviewer scores, and grant outcomes in the form of publications and citations that can be traced to a particular funding stream.[1] With access limited to publicly available data, we focus on funded applications only, and we ask whether risk taking is associated with the likelihood that a grant is renewed competitively for an additional cycle. Since we are studying the impact of risk taking conditional on a successful initial grant application, the estimates we present should be construed as a lower bound on the risk taking penalty that would be observed if we had access to the full set of grant applications, included those that were not supported. We further link each grant cycle

[1] These data are typically not made available to researchers. NIH employees have access to funded and unfunded grant applications and average scores; even these insiders do not have access to the full distribution of scores at the proposal level, as these scores are destroyed in order to protect reviewer confidentiality.

with the publications they engendered, capitalizing on the requirement that authors acknowledge their NIH grant support in their published work.

R01 grants are the "bread and butter" of academic biomedical research, and a requirement for scientists who aspire to establish and make viable their laboratory over the long run. We study how risk taking is penalized or rewarded in 103,164 NIH-funded R01 and R01-equivalent grant cycles between 1980 and 2015. After controlling for a robust set of covariates for investigator demographics, institution quality, and grant characteristics, we estimate the relationship between risk taking and whether the grant was competitively renewed. Risk taking, much like creativity, is a nebulous concept. There is no consensus on what constitutes risk in science nor how it should be measured (Althaus 2005, Franzoni and Stephan 2023). Since it is unlikely that a single measure can fully represent the phenomenon, we develop four separate measures of risk taking, each of which captures a different aspect of the risks taken or borne by scientists, and examine if our findings depend on the particular measure used.

First, risk taking may be associated with more extreme tail outcomes, with greater variance in outcomes and both more exceptionally good and exceptionally bad results (Azoulay et al. 2011). Second, risky research may be more disruptive, seeking to overturn the status quo, rather than consolidative, reinforcing earlier results (Funk and Owen-Smith 2017). Third, researchers may pivot from what they have done in the past to follow new scientific opportunities, capturing the idea that risk is not inherent to the project itself but also depends on whom is undertaking the project. Finally, researchers may "stand out from the crowd," selecting projects that are intellectually distant from those of other investigators, thus making the success (or failure) more salient in the eyes of the scientific community. Across each of these measures, we examine effects across the full distribution of risk taking, eschewing strong functional form assumptions in our econometric analysis.

Across each of these measures, we find that risk taking is penalized. When comparing grant cycles in the top and bottom decile of risk taking, grants with greater risk taking have a 9.5% lower renewal rate (20.5% decline) when measuring risk taking using extreme tail outcomes, an 11.1% lower renewal rate (24.4% decline) when measuring risk taking by its disruptiveness, a 7.7% lower renewal rate (16.9% decline) when measuring risk taking by an investigator pivoting from her prior research, and a 6.1% lower renewal rate (12.4% decline) when measuring risk taking by standing out from what other investigators are studying. In contrast to our measures of risk taking, novelty is associated with higher grant renewal rates in our data. At the very least, this finding

buttresses our claim that novelty and risk taking are different concepts that it is best not to conflate.

We also examine heterogeneity in the penalty for risk taking across career stages, gender, prior status, and "grantsmanship." Risk taking is penalized across all subgroups examined, with weak evidence of an additional penalty faced by early career investigators.

Of course, it may well be that the observed penalty applied to riskier projects is appropriate, given the real-world constraints faced by funders. For public funding at least, the committees and officials doling out funding to individual proposals are accountable to political paymasters, and may feel pressure to present evidence of accumulating and steady success, rather than a series of failed projects occasionally punctuated by an extraordinary outcome (Clancy 2023). We present one piece of evidence consistent with the claim that funders penalize risk even beyond what a benevolent social planner might wish: the gradient in the renewal/risk taking relationship appears steeper for novel proposals, relative to less novel ones.

This study's primary contribution is to provide direct evidence that grant peer review may punish scientific risk taking. In addition to the potential policy importance for public and philanthropic scientific grant programs, this emphasizes how the type of innovations a system produces is shaped—deliberately or not—by the institutions and managerial processes that support it. While we focus on peer review for scientific grants, our study has implications for the design of organizational processes to select innovative projects in other settings that value risk taking.

The rest of the paper proceeds as follows. In Section II we develop how peer review might shape scientific risk taking. Section III presents the setting, data, empirical strategy, and descriptive statistics. We report results in Section IV. Section V concludes by discussing broader implications of our findings.

2. Peer review and risk taking

Peer review plays a prominent role in the selection of scientific projects (Chubin and Hacket 1990, Stephan 2012). While some degree of risk is inherent in nearly all innovation projects, a central challenge facing projects with high levels of risk is both that they are less likely to succeed, and that conditional on success, it may be more difficult to predict their ultimate impact and benefits (Rosenberg 1995). A number of distinct processes may potentially be at play in shaping how peer review may punish risk taking (Azoulay and Li 2022, Franzoni et al. 2022). These may occur at the level of the investigator, individual peer reviewer, peer review committee, and funding institution. While some of these factors may be idiosyncratic to

particular grant funding institutions, others may be present in how innovation projects are selected across a range of different settings.

Peer review may shape the degree of risk taking investigators pursue through several indirect mechanisms. Many academic labs rely on external grants to fund their operations (Stephan 2012). An inability to win external funding may result in closure of the lab, laying off dependent lab personnel, and loss of the investigator's job and potentially career (Ruben 2017). To the extent investigators *perceive* grant peer review to punish risk taking, this will lead them to instead propose safer, more incremental projects (Kent 2018, Langer 2012, Nielson 2022). Similarly, scientific trainees may become imprinted with a less risky approach to research after observing cautious, incremental projects proposed by their mentors in response to a perceived penalty for risk taking (Azoulay et al. 2021, Higgins 2005). In this way, peer review may shape the supply of risk taking in research outside of its direct role in evaluating particular projects.

Penalization of risk taking may also occur at the level of individual peer reviewers. Individuals may have a psychological bias against the uncertainty that characterizes risky research compared to more incremental projects (Carson et al. 2022, Fox and Tversky 1995, Mueller et al. 2012). In line with this, evaluators of research proposals have been found to give lower scores to proposals with greater novelty (Boudreau et al. 2016) and to emphasize feasibility (Kreiger et al. 2022, Lane et al. 2022a). The bias against novelty may also lead to delayed recognition of the underlying potential of a scientific work (Wang et al. 2017).

Peer review is conventionally undertaken in a committee. Bias against risk taking can occur due to both committee processes and committee membership. Proposals with greater risk taking may have greater variance in reviewer assessment of their quality (Langer 2012). Committee processes that emphasize consensus and uniformly high evaluations across its members to be funded may penalize risk taking (Azoulay and Li 2022, Franzoni et al. 2022).[2] Similarly, given the wider variation in evaluations of risky projects, processes where each member independently fully evaluates a proposal prior to the committee meeting may be more likely to result in identifying a judge who enthusiastically supports the project compared to processes where one or two committee members have primary responsibility for evaluating each proposal and then present their recommendation to the committee. Exposure to negative assessments of other committee members prior to the final judgment my lead evaluators to revise their own assessment downwards more so than exposure to positive assessments leads them to revise their assessments upwards, biasing against projects with more disagreement (Lane et al. 2022b). Who serves on the committee is also important. The penalty for novelty may be larger from those investigators in the same field as the novel proposal (Wang et al. 2017). This suggests intellectual diversity in committee membership may reduce the bias against risk taking. Additionally, a scientist's prior risk taking may shape their likelihood of penalizing

[2] For instance, a process where a proposal's overall score is simply the average of each judge's evaluation will emphasize consensus more than having scores aggregated after dropping the lowest judge score(s) or a process which emphasizes having a few judges with very high levels of support.

risk taking when evaluating project proposals. To the extent this is true, considering prior risk taking when determining committee membership may reduce the committee's risk penalty (Nicholas and Ioannidis 2012).

Finally, some aspects of the potential for peer review to penalize risk taking emerge at the level of the funder and innovation ecosystem. Of course, funders may shape any potential risk-taking penalty by how they structure committee procedures and membership, set expectations for project stage and any required preliminary supporting evidence, and determine explicit evaluation criteria. Political pressure to show constituents funders are effective stewards of money may shift towards the selection of reliable, incremental research projects (Lorsch 2015). Risk taking may also be facilitated by shifting the focus of selection from individual projects to a *portfolio* of projects (Azoulay and Li 2022, Franzoni et al. 2022). Finally, the vigorous competition that emerges from the low overall rate of funding may push investigators and reviewers to prioritize short-term advances and inhibit creativity (Alberts et al. 2014, Fang and Casadevall 2015).

Prior literature has studied the association between aspects of research that are related to risk taking and the evaluation and funding of scientific proposals and publications. This includes novelty (Ayoubi et al. 2021, Boudraeu et al. 2016, Packalen and Bhattacharya 2020, Teplitskiy et al. 2022, Veugelers et al. 2022), interdisciplinary research (Banal-Estanol et al 2019, Bromham et al 2016), and the combination of novelty and feasibility (Krieger et al. 2022, Lane et al. 2022a). Our work builds on these studies in several notable ways. First, we develop a set of eclectic measures of risk taking, each of which captures a different aspect of risk, and empirically emphasize effects across the full distribution of different levels of risk taking. Prior research has often conflated novelty with risk taking; we discuss the distinction between the two in more detail in section III and empirically distinguish between risk taking and novelty in our results. Second, we focus on measures of risk taking in the grant proposal itself and publications emerging directly from the grant. In contrast, prior studies have frequently focused on risk-related measures of an investigator's research in the years preceding a grant application and used this to predict grant award. Yet, investigators not only evolve over time in their approach to risk taking, they frequently manage a portfolio of research projects each of which may have different levels of risk taking. An investigator's past research may not accurately reflect risk taking in a focal grant proposal, and indeed, investigators may deliberately avoid proposing projects with high risk taking if they perceive the funding agency to punish this, instead either forgoing these projects or funding them through other sources. Boudraeu et al. (2016) experimentally assigns grant proposals and focuses on individual judge evaluations, while Carson et al. (2022) experimentally adjusts research project evaluations and focuses on individual funder project selection. These approaches provide important causal insights, but do not capture selection committee dynamics, realized investigator behavior after a grant is funded, as well as faces the usual challenges of generalizing from experimental to real-world conditions. Finally, by studying the universe R01-equivalent NIH grants, we focus on a population of substantial policy interest.

3. Empirical design, data and descriptive statistics

Setting: The National Institutes of Health

The NIH plays a prominent role in funding biomedical research. In 2012, it comprised 27% of all biomedical research funding in the United States and nearly two-thirds of all public and philanthropic research funding (Moses et al. 2015). In 2021, the NIH spent $32 billion to support extramural research projects (Lauer 2022). While the NIH funds clinical research (including clinical trials, health services research, and behavioral studies among others), the bulk of their funding supports basic research aimed at advancing the frontier of scientific understanding. With research project grants having an average size of over $580,000 in 2021, NIH grants are a major source of funding for the researchers they support (Stephan 2012). Not only does this funding play an important role in funding lab materials and the salaries of lab employees, many academic researchers are in "soft money" positions, funding their own salary out of awarded grants. The ability to successfully raise grant funding is an important metric upon which academic employment and promotion decisions are made.

Understanding the process the NIH uses to evaluate and fund grant applications can provide useful context for our empirical analysis (Azoulay and Li 2022, Gerin et al. 2018). Briefly, the NIH is comprised of 27 separate institutes or centers (IC) which are typically organized around a set of diseases (e.g., National Cancer Institute) or organ systems (e.g., National Heart, Lung and Blood Institute). Each IC is responsible for funding research that is potentially relevant to its mission. Scientific evaluation of grant proposals occurs primarily in standing review committees known as study sections (e.g., "Cellular and Molecular Immunology" or "Atherosclerosis and Vascular Inflammation"). Each study section may review and score grants from multiple ICs, which are then funded by the ICs in order of their study section score until the IC exhausts its funds.[3] The amount of resources devoted to reviewing grant proposals—both from the NIH with staff to manage the process and from study section members—is substantial. In 2021, the NIH reviewed nearly thirty-eight thousand applications for R01-equivalent grants (Lauer 2022).

The NIH is an attractive setting for studying how peer review might respond to risk taking. First, professional norms lead most academic research to be published in peer-reviewed journals (Dasgupta and David 1994). When combined with rich bibliometrics in biomedicine, this provides a robust "paper trail" we can follow and characterize to measure the degree of grant risk taking. Second, as NIH research is a source of significant improvements in health and positive economic spillovers, understanding how the NIH may more effectively utilize its funds to support breakthrough innovation is of considerable policy interest (Azoulay et al. 2019, Fleming et al. 2019).

Data

[3] Institute directors do have discretion to fund applicants out of order considering factors such as fit with the institute's mission, the overall portfolio of projects being supported, and their own evaluation of application quality.

As we do not have observe grant applications, we focus empirically on whether an awarded grant was renewed. We rely on the NIH's Consolidated Grant Applicant File to identify all those researchers who received an NIH grant, 1980-2015. We start our analysis in 1980 as we are not able to reliably link publications to the NIH grants which supported them before this date, and end in 2015 to allow sufficient time for follow up to observe research outcomes and renewal status of grants awarded in this year. We limit our analysis to R01-equivalent grants.[4] These investigator-initiated grants are a major funding mechanism to support the work of principal investigators who have achieved career independence and run their own laboratory. We also exclude grant awards from special emphasis study sections, which may not have an option for grant renewal, and grants without associated publications for which many of our measures of risk taking would be undefined. We focus our analysis on the grant cycle, which corresponds to a single competitive review of a funding proposal. A cycle starts the year a grant proposal is funded and ends either at the conclusion of grant funding or when a proposal for grant renewal is competitively reviewed and funded. As grants may have multiple principal investigators and investigator characteristics are explicitly considered during renewal decisions, we analyze our sample at the investigator-grant-cycle level.[5] This gives a final sample of 103,164 investigator-grant-cycles from 63,101 grants and 37,222 investigators.

Our publication data is from *PubMed*, an online resource from the National Library of Medicine to support access to the biomedical research literature that indexes over 30,000 journals and over 34 million articles. We limit our analysis to original research publications; this excludes other types of publications such as review articles, letters, and erratum. For each grantee, we identify their career publications in *PubMed* using Author-ity (Torvik and Smalheiser 2021, Torvik and Smalheiser 2009, Torvik et al. 2005). Author-ity disambiguates authors using the insight that authors of the same name sharing coauthors, affiliations, keywords, and other such characteristics are likely, in fact, to be the same person. Author-ity has been shown to be highly accurate (Lerchenmueller and Sorenson 2016).

We use NIH RePORTER to link publications attributable to each NIH grant. This data relies on the grant having been listed in the publication's acknowledgement section as a funding source, which is required as a condition of accepting NIH funding. We consider publications attributable to a grant cycle if they are published up to one year after the last year of the grant cycle to allow for publication delays.[6] As investigators may have different degrees of risk taking across different projects and funding sources, in all

[4] We consider R01-equivalent grants to be those with an NIH activity code of R01 or R37. We exclude several grant mechanisms considered by the NIH to be R01-equivalent but which may have distinct dynamics around risk taking. For instance, we exclude DP1, DP2 and DP5 which are part of the NIH's effort to fund high risk, high reward projects; R56 and RL1 which may not be applied for by the investigator; U01 in which NIH staff actively support the research; and R35 and RF1 which provide longer-term support to investigators.

[5] One hundred thirty eight grant cycles appear in our sample more than once as they have multiple principal investigators.

[6] Among publications acknowledging a focal NIH grant in our sample, 20.8% are published more than one year after the end of grant funding. These publications and the results they are based on will typically not have been available to study sections when they review a grant for competitive renewal and so do not inform our outcome or measures of interest.

cases, our measures of risk taking are calculated solely using publications during a grant cycle that acknowledge funding from the grant rather than also including publications during the grant cycle which do not acknowledge funding from the grant.

To measure research topic and methodology, we capitalize on Medical Subject Headings (MeSH). MeSH is a hierarchical controlled vocabulary created and maintained by the National Library of Medicine used for indexing and searching the biomedical literature. In the 2019 edition of MeSH, there are over 29,000 terms, with an average of about 12 MeSH keywords associated with each PubMed publication. Crucially, MeSH are assigned by professional indexers. This excludes any concern authors selecting MeSH might do so strategically (Bachrach and Charen 1978). While it is difficult to assess accuracy as there is no single "correct" way to index articles, MeSH have generally been found to be consistent across indexers (Coletti and Bleich 2001, Funk and Reid 1983). In addition to MeSH associated with grant publications, we use the National Library of Medicine Medical Text Indexer to assign MeSH to grant abstracts. Grant abstract text is available from the NIH RePORTER.

We supplement publication data with citations from Thomas Reuter's *Web of Science*. All grants are adjusted to 2020 dollars using the Biomedical Research and Development Price Index.

Measuring risk taking

There is not consensus on what constitutes risk in science nor how it should be measured (Althaus 2005, Franzoni and Stephan 2023). For instance, within science, there are different types of risk, including technical risk stemming from what nature's answer is to the question asked, execution risk in being able to successfully carry out the planned experiments, competitive risk from other investigators studying the same or similar question, and risk in how a successfully completed project will be perceived by peers or advance the investigator's career. Some aspects of scientific risk are not inherent to the science itself, but also depend on who is undertaking the study and the environment that supports it. Our measures of risk taking do not correspond one-to-one with these different types of risk, but instead emphasize different combinations of them.

To illustrate how projects may have different aspects of risk, consider the challenges of a single measure in fully capturing risk taking dynamics across disparate projects. The Human Genome Project, once well underway, had little technical risk or danger of being viewed as a marginal contribution, but substantial competitive risk in the race to be the first to sequence the genome. In contrast, early efforts to develop cancer immunotherapy had very substantial technical risk, as it may be very difficult or impossible to develop immunotherapy specifically targeting cancer but not normal tissue, but much lower levels of competitive risk. Similarly, studying new systems to encourage handwashing may have little technical or competitive risk, but may entail career risk from being viewed as a marginal contribution.

Additionally, connecting scientific novelty with risk taking is particularly challenging (Ayoubi et al. 2021, Foster et al. 2015, Uzzi et al. 2013). While pursuing novel research may be linked in some cases with

risk taking, it conflates technical risk, competitive risk, and career risk in ways that may vary across different projects. More importantly, investigators can take risks without being novel, and can be novel without taking large risks. Consider efforts to develop non-mRNA COVID-19 vaccines. While this research was novel, targeting a new disease, there was arguably only modest levels of technical risk. In line with this, many research groups around the world were independently successful in developing such a vaccine, and some experts viewed it is a matter of when, not if, such efforts would be successful despite its novelty. In contrast, consider Judah Folkman's initial findings on the role of angiogenesis in cancer. Other scientists viewed building on this work as so technically risky few were willing to do so until over a decade later, by which time arguably such follow on work may no longer be considered novel.

As an additional complicating factor in taking novelty as a proxy for risk taking, scientists are judged for employment and promotion on their perceived scientific contribution. If a scientist persistently undertakes work which isn't novel or cutting edge, the resulting career risk of being viewed as having only low-impact contributions may be greater than the technical risk of studying cutting edge science (Franzoni and Stephan 2023). Additionally, even if a funding agency is risk averse, it still has a mandate to pursue new-to-the-world frontier science, and thus may look for "incremental or safe novelty" in projects. We see the study of factors which promote scientific novelty as essential research and novelty as a valuable metric in and of itself. Yet, we believe care should be taken in considering novelty as a proxy for risk taking.

To address the challenges of measuring risk taking, we employ four separate approaches. We argue each of these measures captures a distinct aspect of scientific risk taking. As each measure has different strengths and weaknesses in how they correspond to risk taking, by integrating across them we can develop a more complete understanding than examining each separately. In each case, we are non-parametric in our approach and examine the full distribution of different levels of risk taking. This lets the data speak for itself for whether any risk penalty is due primarily to differences in more conventional levels of risk taking or is instead driven by risk taking in the tails of the distribution.

First, risk taking may be associated with both more exceptionally good and exceptionally bad outcomes (Azoulay et al. 2011). Projects with this type of risk reflect a higher reward if successful, but also a higher chance of failure. In line with this, work with atypical combinations of ideas has been found to have a wider dispersion in outcomes (Schilling and Green 2011, Wang et al. 2017). We term this measure "extreme tail outcomes" and operationalize it as the difference between the highest and lowest vintage-adjusted citation percentiles among grant cycle publications. We focus on long-run vintage-adjusted citation percentiles here for several reasons. Early citations can be noisy and there may be a delay in recognizing the contributions of highly novel articles (Wang et al. 2017). This lets us focus on the long-run view of the publication's underlying quality and contribution. Additionally, this allows us to readily compare articles across different publication years.

Research may either be consolidative, reinforcing earlier results, or disruptive, overturning the status quo. Another way of taking risks is to undertake this type of disruptive research. To provide evidence of

this, we employ the disruption index proposed by Funk and Owen-Smith (2017) and previously used in the study of science (Bornmann et al. 2020, Wu et al., 2019). In this index, a highly disruptive article is one in which future research cites the focal paper but not the focal paper's references, while highly consolidative research has future research citing both the focal paper and the focal paper's references. Our measure of "disruption" for a grant cycle is the maximum of the disruption index percentile, calculated relative to all publications in that year in PubMed, among grant cycle publications.

As a third measure, we focus on intellectual distance between grant cycle publications and an investigator's prior work. This takes seriously the idea that some aspects of risk are not inherent to the project itself but are also a function of who is undertaking the research. Scientists recombining their existing expertise to address new topics, employ new techniques, or explore newly identified opportunities may lead to promising previously unexplored findings. On the other hand, such exploration may devalue their prior expertise and decrease the quality of their output after pivoting (Arts and Fleming 2018, Hill et al. 2021).[7] We term this type of risk taking "pivoting." To capture this, we identify the fraction of MeSH term pairs among grant cycle publications that are new compared to the MeSH pairs from an investigator's publications in the 5 years preceding the grant cycle.

As a final approach to measuring risk taking, we examine intellectual distance from what other NIH-funded investigators are studying. This reflects scientists studying unique recombinations of ideas and techniques (Fleming 2001, Uzzi et al. 2013). This speaks to the risks of being contrarian by choosing projects that do not fall within the mainstream of what their peers are studying and emphasizes career and competitive risks (Fang and Casadevall 2015). We refer to this measure as "standing out" from the crowd and measure it by identifying the fraction of MeSH term pairs from grant cycle publications in which the focal investigator is the only researcher using this term pair among all funded grant proposal abstracts in the same IC-study section-year.[8,9]

We also explore novelty to empirically distinguish it from risk taking. Prior literature has used a number of different approaches to measuring novelty, including examining unusual or intellectually distant combinations of references (Shibayama et al. 2021, Uzzi et al. 2013, Wang et al. 2017), unusual combinations

[7] Franzoni and Stephan (2023) point out that persisting in a researcher's current trajectory rather than pivoting isn't always associated with lower risk. If the current research line shows little promise, there may be less risk in pivoting.
[8] The relevant sample size varies slightly across each of these measures. Extreme tail outcomes is only defined for those grant cycles resulting in at least two publications indexed in Web of Science. Disruption is only defined for those grant cycles resulting in at least one publication that is cited at least once. Pivoting is defined only if there is at least one publication with MeSH terms in both the current grant cycle and the preceding 5 years. Standing out is defined only if there is one publication with MeSH terms in the current cycle and funded grant proposal abstracts in the same IC-study section-year are available.
[9] We compare an investigator's grant cycle publication MeSH to that for all available awarded grant proposal abstracts, including those that are not R01-equivalent grants, in the same IC-study section-year. Among the 103,164 investigator-grant cycles in our sample, we have abstracts for other grants in the same IC-study section-year for 83,451. The mean number of abstracts in the comparison group for each investigator-grant-cycle is 99. We prefer this approach to alternatives, as while data is not available for all possible investigator-grant-cycles, the comparison group is thicker for those that are available, facilitating our focus on the tail-measures of intellectual distance from other investigators.

of patent classes (Fleming 2001, Verhoeven et al. 2016), citation network centrality (Shibayama and Wang 2020), keyword and keyword combination age (Misha and Torvik 2016, Packalen and Bhattacharya 2020), and phenomena-specific measures such those based on chemical structure (Krieger et al. 2022, Rzhetsky et al. 2015). We follow Boudreau et al. (2016) in focusing on new combinations of MeSH terms. Our measure of novelty is the fraction of MeSH pairs among grant cycle publications that were first used by any publication within PubMed within the past three years.[10]

Empirical strategy

We observe the universe of NIH grants actually awarded, but unfortunately, do not observe grant applications which are not ultimately funded by the NIH. Because of this, we empirically focus on NIH R01-equivalent grants that were awarded and measure whether or not the grant was renewed. This approach lets us reliably observe all investigators who were at risk of having their grant renewed. Among this population, a grant may not be renewed either because the investigator applied for but did not receive a competitive renewal, or because they never applied for a renewal. Given the career importance of grant funding in maintaining a lab, we expect most investigators will apply for renewal.[11] Furthermore, if an investigator with risky research is dissuaded from applying for grant renewal due to their perception of a low likelihood of renewal and the time involved in applying, the implications of this for our research question are very similar to having applied and not been selected. As our sample evaluates the degree of risk taking among scientists all of whom have an R01-equivalent grant, this suggests these scientists will have similarly benefited from the act of applying for and receiving an R01-equivalent grant (Ayoubi et al 2019, Jacob and Lefgren 2011).

The largest threat to unbiased estimation of the impact of risk taking on grant renewal is potential omitted variable bias: that there is a factor correlated with both risk taking and grant renewal that is not included in our model. To address this, we use a rich set of controls for investigator, institution, and grant cycle observable characteristics.[12]

We also consider controls related to grant research outputs. Study sections reviewing applications for grant renewal will consider an investigator's progress and achievement during the current cycle as reflected in its publications. Omitting measures of research productivity and impact during a grant cycle might lead to omitted variable bias. For instance, there may be some element of investigator quality that is

[10] 7.9% of MeSH pairs in our sample have three or fewer years between when they were first used in PubMed and their use in a grant cycle publication. Our findings are robust to using thresholds other than three years, including using MeSH pairs new to PubMed in the past year.

[11] Applications for competitive grant renewal, using the NIH's more inclusive definition of R01-equivalent (see footnote 3), were funded at a rate of 34.5% in 2015 (Lauer 2016). This is similar to the 35.9% renewal rate in our sample in the same year, consistent with most grantees applying for renewal.

[12] The NIH instructs reviewers to scores each application based on five criteria: significance, innovation, approach, investigator, and environment. While the first three assess the importance, novelty, and feasibility of the proposed project itself, the last two evaluate investigator qualifications and ability of the institution to support the work, respectively.

correlated with both research productivity and risk taking during a grant cycle. As the association between risk taking and research quantity and quality may vary in magnitude and direction across different levels of risk taking, it is difficult to predict the direction of any bias if these controls are excluded. On the other hand, publication quantity and quality is a *result* of the mechanism through which risk taking may impact grant renewal (Angrist and Pischke 2008). Due to this, their inclusion may also result in biased estimates of the effect of risk taking on grant renewal. For this reason, we present each core result both with and without controls for publication quantity and quality during the grant cycle and view our estimates as an association rather than causal effect.

Econometric considerations

Our primary estimating equation relates the dependent variable of interest, $renewal_{ij}$, of investigator i and grant cycle j. Formally, we estimate variations of

$$E\left[renewal_{ij} = 1 \mid \delta_i, \gamma_j\right] = \beta_0 + \beta_1 risktaking_{ij} + \delta_i + \gamma_j$$

where $renewal_{ij}$ is an indicator variable equal to 1 if the grant cycle was renewed and 0 otherwise. We include high dimensional vectors of controls for investigator and grant cycle characteristics, δ_i and γ_j, respectively. This includes sets of indicator variables for investigator age, degree and gender; log investigator career citations, log prior investigator NIH funding, an indicator variable for no prior investigator NIH funding, and log institutional NIH total funding as measures of investigator and institutional quality; and sets of indicator variables for year of evaluation, grant length, and the interaction between the IC and study section. Select specifications also include controls based on measures of research quality and quantity during the grant cycle, including log of mean journal impact factor, sets of indicator variables for the number of publications during the grant cycle separately both for those publications acknowledging and not acknowledging funding from the grant, and indicator variables for having publications in highly selective journals.[13,14] We use robust standard errors with two-way clustering at the investigator and grant throughout. The maintained assumption is that, after conditioning on observables, measures of risk taking are uncorrelated with the error term.

Our primary specification is a linear probability model (LPM). Since LPM may generate fitted values outside of the interval [0,1], its use would not be appropriate for predictive analysis. However, our object here is the estimation of the marginal effects from the conditional expectation function. For this purpose, the LPM is typically more appropriate (Angrist and Pischke 2008: pp. 94-106). In particular, while logit and probit models may have the benefit of bounding estimates of the dependent variable between 0

[13] We use journal impact factor as a proxy for publication quality for grant cycle publications as this metric is visible to peer review committees at the time of evaluation and early citations are a noisy signal of quality. Our results are robust to instead controlling for citations at the time of grant cycle evaluation and long-run citations.
[14] For this measure we include those publications which were published in *Nature*, *Cell*, *Science*, and the *New England Journal of Medicine*.

and 1, the resulting estimates may be biased by the incidental parameter problem (Neyman and Scott 1948). This may be particularly problematic when including many fixed effects as is necessary in our case.

Descriptive Statistics

Table 1a and 1b present descriptive statistics of investigators and grant cycles, respectively. In fitting with the very high levels of competition largely from established investigators to win an R01-equiavelent grant, the sample is comprised of accomplished scientists. Investigators started the grant cycle at a mean of 18.7 years after completing their terminal degree with 56 prior publications and 2,234 career citations. A full 5% of the investigators achieve elite status during their careers, which we define as having ever won a Nobel Prize, Lasker Award, Howard Hughes Medical Institute Investigatorship, or membership in the Institute of Medicine or National Academy of Sciences. The majority of an investigator's research output is not directly supported by the focal R01-equivalent grant of interest. Among publications during the cycle period, 39.9% directly acknowledge the focal R01-equivalent NIH grant; 59.6% of publications which do not draw support from the focal NIH grant still had NIH support, whether from a non-R01-equivalent NIH grant to the investigator or a grant to one of their coauthors. Overall 45.2% of grant cycles were renewed; the renewal rate was lower for the first cycle of a grant (40.4%) than for subsequent cycles (51.5%).[15] Grant cycles that were renewed had investigators that were younger (17.6 vs. 19.6 years), greater grant research productivity (8.1 vs. 7.0 publications and 663 vs. 422 citations among papers directly acknowledging the focal grant), and a greater percentage of the investigator's research activity during the cycle attributable to the focal grant (46% vs. 38.2% of all publications during the cycle had funding from the focal grant). Figure 1 provides the full distribution of the number of publications per grant cycle by renewal status.

Table 1c provides descriptive statistics for each of the four measures of risk taking and for novelty. While for most measures the mean level of risk taking was higher among grant cycles that were not renewed than those that were, the absolute differences are small. Figure 2 shows the full distribution of each of these measures by grant cycle renewal status. This provides initial descriptive evidence that the differences in renewal rate are not evenly distributed across different levels of risk taking for each of these measures. Table 1d presents the correlation for measures of risk taking and novelty. The relatively low correlation across these measures emphasizes how each is capturing a distinct approach to risk taking and how novelty contrasts with risk taking.

4. Results

Risk taking and grant renewal

[15] In contrast, the success rate of applicants for a new R01-equivelent grant is lower. In 2015, 16.1% of applications for a new R01-equivalent grant were funded (Lauer 2016).

Table 2 reports estimates for the impact of risk taking by each of our four measures on grant cycle renewal. Overall, we observe that high levels of risk are associated with lower renewal rates, and conversely, that low levels of risk are associated with higher renewal rates across each of these measures. Not only is the effect highly statistically significant, but the magnitude is substantial as well. For instance, when compared to grant cycles in the 10-90%ile for extreme tail outcomes, cycles in the bottom decile, reflecting low levels of risk taking, have a 5.1% higher renewal rate (an 11.0% increase) while those over the top decile have a 4.4% lower renewal rate (a 9.5% decrease) (Table 2A, column 2). Similar patterns are seen across the other measures of risk taking for the top and bottom deciles, including disruption (8.9% higher [19.6% increase] vs. 2.2% lower renewal rate [4.8% decrease]), pivoting (5.3% higher [11.6% increase] vs. 2.4% lower renewal rate [5.3% decrease]), and standing out (3.8% higher [7.7% increase] vs. 2.3% lower renewal rate [4.7% decrease]).[16] These findings are robust to inclusion of controls for publication quantity and quality during a grant cycle, although in some cases this reduces the magnitude of the risk penalty. We do not find evidence that the risk-taking penalty decreases in subsequent grant cycles after a grant is initially renewed.

We also explore the full distribution of different levels of risk taking for its association with grant renewal (Figures 3, 4, 5, and 6). We estimate specifications in which there is a separate indicator variable for each ventile of the risk measure. Similar to Table 2, we find higher renewal rates for lower levels of risk taking and lower renewal rates for higher levels of risk taking. Across several of these measures, we find the striking pattern of a nearly monotonic decrease in renewal rates with increasing levels of risk taking.[17]

Novelty and risk taking

We next explore empirically the connection between risk taking and novelty. Figure 7 presents these results. We estimate specifications that are similar to Figures 2-5, but in addition to an indicator variable for the ventile of the measure of risk taking of interest we also include an indicator variable for each ventile of novelty. Two findings emerge from this analysis. First, in contradistinction to our measures of risk taking, we find novelty is rewarded. Additionally, the magnitude of the reward for novelty is similar to the magnitude of the risk-taking penalty. Second, inclusion of novelty does not significantly reduce the observed penalty for risk taking. This provides empirical evidence that novelty and measures of risk taking may have distinct dynamics, consistent with the idea that in some settings novelty is not a valid proxy for risk taking. Appendix Figure A2 repeats this analysis, but measures the effect of novelty on grant renewal without the inclusion of measures of risk taking. Similar to Figure 7, we find that novelty is associated with a higher probability of grant renewal in this analysis.

[16] For extreme tail outcomes, pivoting, and standing out, percentiles are calculated relative to other grant cycles in the same evaluation year. For disruption, we emphasize work that is disruptive in an absolute sense and so use the percentile of the most disruptive paper, calculated relative to all publications in PubMed in the same year, for the grant cycle.

[17] An exception to this is with disruption, where there is a non-linearity in the association between disruption and grant renewal: grant cycles with the very highest degree of disruption have higher renewal rates than those with elevated but less extreme levels of disruption.

We also look at if the magnitude of the penalty for risk taking is different for novel research. Across all four measures of risk taking, grant cycles with above median novelty have a larger penalty for risk taking than those that are less novel (Table 3). At higher thresholds of novelty, the penalty for risk taking increases for pivoting and standing out, but is not statistically different from less novel research for the extreme tail outcomes and disruption measures (Appendix Table A1).

Who is penalized for risk taking?

We also explore heterogeneity in who is penalized for risk taking, with a focus on informing risk-taking dynamics as investigators progress through their career. We first focus on career age, measured as the time from terminal degree until grant cycle proposal. We then repeat the analysis in Figures 3-6, but interact indicator variables for the decile of each risk measure with an indicator variable for either being below or above the median career age. We similarly examine heterogeneity by gender, elite status as proxied by having won a prominent biomedical research award prior to grant cycle renewal, and grantsmanship as measured by the level of career NIH funding prior to the current grant cycle.

Figures 8 present these heterogeneity results for risk taking by extreme tail outcomes.[18] Integrating across them, several trends emerge. First, risk taking is penalized across each subgroup examined. There are no subgroups explored for which risk taking wasn't penalized across each of the four measures. Second, early career investigators seem to face a larger penalty than their older colleagues for the same level of risk taking. Finally, some of these results are imprecisely estimated. This is especially the case for heterogeneity by status given the small number of elite investigators.

Robustness analysis and ancillary results

We also perform a number of robustness checks to assess the sensitivity of our estimates to alternative assumptions and subsamples. First, we extensively test alternative ways of measuring each of our approaches to risk taking and novelty (Appendix Table A2). In all cases, our results are robust to these alternative measures. This includes the difference between the maximum and median, median and minimum, winsorized maximum and minimum, and variance in vintage-adjusted citation percentiles for extreme tail outcomes (Appendix Figures A3 and A4); the mean rather than the maximum disruption index and a variation of the index following Bornmann et al. (2020) for disruption (Appendix Figures A5 and A6); fraction new single MeSH rather than MeSH pairs and cosine dissimilarity score for single MeSH terms after term frequency-inverse document frequency weighting for pivoting (Appendix Figures A7 and A8); fraction unique single MeSH rather than MeSH pairs and cosine dissimilarity score for single MeSH terms after term frequency-inverse document frequency weighting for standing out (Appendix Figures A9 and A10), and

[18] Similar patterns are seen for heterogeneity measuring risk taking by disruption, pivoting and standing out.

mean MeSH and MeSH pair age for novelty (Appendix Figures A1 and A2).[19,20] Emphasizing how pivoting and standing out are distinct from novelty, the findings with these measures are robust to exclusion of MeSH terms and MeSH term pairs which were first used in PubMed within the past decade.

The penalty for pivoting might be smaller for those grants for which the investigator proposed—and peer review approved and funded—to pivot in the initially submitted planned research studies. To examine this, we identify the fraction of MeSH terms shared between the grant proposal abstract and an investigator's prior work and split the sample at the median for this measure. We find overall a similar magnitude penalty for pivoting across both groups (Appendix Table A3). In other words, grants which proposed a pivot and then investigators actually pivoted in the work carried out are still renewed at lower rates.

Author-ity has been shown to be highly accurate (Lerchenmueller and Sorenson 2016). It is likely the case, however, that its accuracy of publication disambiguation is higher for rare names within PubMed. To ensure our results are not driven by investigators with poorly disambiguated publications in Author-ity, we split the sample at the median name frequency based on frequency within the entire corpus of PubMed and find similar results across both groups (Appendix Table A4).

One reason investigators may choose not to renew their grant is that their primary research focus lies outside of the scope of what the NIH typically supports. For instance, some organic chemists, electrical engineers, or sociologist investigators may have an isolated project at the interface of biomedicine within the context of a broader research agenda elsewhere. To ensure our results are not driven by investigators with only a weak attachment to NIH funding, we calculated the fraction of career publications with any NIH funding and split the sample at the median of this measure (Appendix Table A5). We find a similar penalty for risk taking across both those with above and below the median career attachment to NIH funding.

One might be concerned, especially for articles published well before the digital era, that not all articles that in fact received NIH funding are recorded as having acknowledged it. Care must be taken when examining heterogeneity over time, as the degree of competition for R01-equivalent grants has been increasing and NIH policy has evolved over time. To reassure that our results are not driven by this potential measurement error, we split our sample at the median grant cycle evaluation year and find similar results across both groups (Appendix Table A6).

Finally, we also use a bounding technique proposed by Oster (2019) to gauge the sensitivity of our results to a failure of the unconfoundedness assumption. The intuition behind this approach is that the stability of the coefficient for risk taking when varying the set of control variables included in the model,

[19] Inverse document frequency was performed relative to all original research publications in PubMed in the year the grant cycle came up for renewal.

[20] Since for standing out we are interested in identifying investigators whose research substantially departs from other investigators, as a robustness check, we repeat the analysis limiting it only to investigators with at least 10 other funded researchers with abstracts available in the same IC-study section-year with a larger risk penalty observed.

scaled by the movement in R^2, provides information about the potential impact of unobserved covariates. To generate these bounds, the analyst must assume proportionality between the covariances of the outcomes with observed and unobserved covariates and posit a maximum value for R^2 if the regression could include all observed and unobserved covariates. Oster's technique generates δ, which can be interpreted as the degree of selection on unobservables relative to observables necessary to reduce the magnitude of the effect of the regressor of interest to zero. Appendix Table A7 reports the results of this exercise. For extreme tail outcomes, disruption, and pivoting, δ is far above one, the threshold recommended by Oster to suggest robustness to the influence of unobservable covariates, with standing out having a δ of -0.70 and -0.59.[21]

5. Discussion

Using the universe of awarded NIH R01-equivalent grants and across a range of different measures of risk taking, we find that investigators with high levels of risk taking are renewed at lower rates than those with low levels of risk taking. In this respect, our findings provide systematic evidence that peer review at the NIH punishes risk taking. We also empirically demonstrate that novelty has distinct dynamics from our measures of risk taking, emphasizing that in some settings it is not a valid proxy for risk taking. In addition to informing peer review and innovation project selection processes more generally, this is a population of substantial policy interest given the prominence of the NIH and the R01 grant mechanism in funding biomedical research in the United States.

A strength of our approach is its focus on measuring the association between risk taking and grant renewal across the full distribution of observed levels of risk taking. This emphasis on the tail-values of risk taking reflects that not all scientific innovation impact scientific progress equally, and it may be those with the largest degree of risk are most likely to be associated with radical innovation (Rzhetsky et al. 2015, Veugelers and Wang 2019, Wang et al. 2017). It is also striking that peer review punished risk taking in light of the mixed evidence on the ability of peer review scores to predict scientific productivity and impact (Cole et al. 1981, Danthi et al. 2014, Li and Agha 2015, Pier et al. 2018, Rothwell and Martyn 2000).[22]

The validity of our results relies upon the maintained assumption there is no significant omitted variable bias: that is, that there are no covariates associated with both risk taking and grant renewal which are not included in our empirical strategy. We are reassured this is not a large concern given the rich set of controls we employ as well as how pervasive and robust our findings are that risk taking is penalized. Additionally, there are also several reasons to think our estimates may reflect a lower bound of the possible

[21] Some values for δ are negative. This reflects that the correlation between unobservables and risk taking would need to be in the opposite direction as the correlation between observables and risk taking to explain away the result.
[22] This parallels the dynamics seen for innovation in other settings, such as entrepreneurship. For instance, even among VC-backed startups, a handful companies that are outliers in their success account for the majority of VC profits and impact on the larger economy. Likewise, while investors may be able to identify those companies with little growth potential, they have only a very limited ability to identify among VC-funded companies which ones are likely to have outlier success (Kerr et al. 2014).

risk penalty. First, several of our approaches to measuring risk rely on the presence of publications. To the extent high risk taking is more likely to result in the absence of publications and that the absence of publications leads to a low renewal rate, measuring risk taking conditional on having publications may underestimate the total penalty for risk taking. Second, as we are unable to observe unfunded grant applications, entry into our sample is conditional upon having first applied for and received an NIH R01-equivalent grant. Any risk-taking punishment in winning an initial NIH grant would lead to our sample having lower overall levels or less extreme distributions of risk taking than the population of initial NIH R01-equivalent applicants. This suggests the magnitude of risk-taking penalization in our study is lower than we might expect if we could repeat our study using initial applications for NIH R01-equivalent grants. Similarly, investigators who prioritize high levels of risk taking in their research, aware of the NIH's reputation for penalizing risk, may not seek NIH funding given the time and opportunity costs involved in applying (Langer 2012). This would represent another indirect way grant peer review penalizes risk taking not captured in our study. It is also possible that researchers in training—many of whom work under the mentorship of NIH-funded investigators—may imprint with a less risky approach to research after observing their mentors avoid proposing risky projects for NIH funding (Azoulay et al. 2021, Higgins 2005).

We argue each measure of risk taking captures of a different aspect of what it means to perform risky research, each with its accompanying strengths and weaknesses. For instance, an individual scientist exploring a new area relative to their prior research may reflect promising recombination of ideas, or alternatively it may reflect devaluing of the scientist's prior experience with a decreased likelihood of success and lower ultimate impact (Arts and Fleming 2018). Similarly, scientists who pursue research few other scientists are interested in may reflect potential breakthrough innovation, or it may reflect simply a less promising research question. The consistency of our results across these different measures is reassuring against the particular weaknesses of each individual measure that what we are capturing, in fact, reflects scientific risk taking. As it is generally very difficult to publish failed experiments and null results, that our findings are exclusively based on risk taking measured from successfully published work is reassuring that our results are not driven by grant non-renewal after a risky project resulted in a failed experiment or null outcome.

There are many potential mechanisms through which peer review may act to penalize risk. Our study does not differentiate between the relative contributions of these various mechanisms towards the observed risk penalty. Similarly, while we observe which grants are renewed, we do not observe if they were not renewed because an investigator chose not to apply or applied but was not selected. We expect exploring these issues to be fruitful areas for future research.

Both high-risk, high-reward and low-risk, incremental projects are valuable and play a role in advancing the scientific frontier. Yet there are reasons to think the NIH playing a larger role supporting high-risk research could be valuable. In particular, one would expect that established, mature lines of scientific inquiry should contain more incremental research—and less risk taking—than novel, cutting-edge

lines of scientific inquiry which would be best served by greater exploration and risk taking. Instead, we find the opposite: the penalty for risk taking is larger for research that is more novel. This argues that the current degree of penalization for risk taking at the NIH results in a welfare loss. In addition, if the NIH penalizes risk taking, this may have a disproportionate impact on research for which other parties have limited incentives to provide. Public investments in early-stage basic research are often justified with the positive knowledge spillovers and difficulty for private industry to appropriate its value (Aghion et al. 2008). Indeed, the government is precisely the agent best positioned to bear risks that are too large for other individuals or organizations to bear. Corporations have also had decreasing investments in science in the recent decades (Arora et al 2017, Fleming et al. 2019). This, along with the NIH being a quantitatively important funder of research due to its size and scope, suggests hesitancy by the NIH to fund risky early-stage basic science research may not easily be compensated for by other funding sources and may lead to a slower overall pace of technical advancement (Foster et al. 2015). Similarly, the NIH may play an important role in sponsoring clinical studies for which there is limited ability for industry to appropriate value, resulting in a potential under-provision of risk taking in this type of research (Greenblatt et al. 2023). While private philanthropies provide an alternative source of research funding, their priorities and areas of support may diverge from those with the most compelling public benefit (Murray 2012).[23]

The challenge of supporting high-risk research has received increasing attention in recent years, including from the NIH. The NIH's efforts have often centered on creating distinct funding mechanisms with explicit mandates around high-risk, high-reward research. This includes the Pioneer Awards, Director's Transformative R01 Awards, and New Innovator Awards (NIH 2019). In 2021, the NIH funded 106 such awards with an expected accompanying outlay of $329 million over five years (NIH 2021).[24] A number of private philanthropic science funders have dedicated opportunities to support high-risk research, including the Howard Hughes Medical Institute, Open Philanthropy Project, Fast Grants, Chan Zuckerberg Initiative, and Wellcome Trust.

While the focus here has been on how peer review shapes selection of projects by public funders, our work has implications for the selection of high-risk innovation projects in other settings as well. First, rather than being considered as an afterthought or as simply words of admonition to project evaluators, how to support the desired level of risk should be built into the design of supporting processes and institutions. While dedicated funding is an important step forward, it still must contend with bias against

[23] Researchers who do not receive NIH funding may be able to substitute their research with other fund sources. For instance, Chuban and Hackett (1990) find that many investigators are able to pursue their proposed research even if they do not receive the focal grant. Similarly, Jacob and Lefgren (2011) find the marginal impact of receiving an NIH grant to be only about one additional publication over five years, in part because researchers can substitute to other funding sources. Yet NIH punishing risk taking may have an indirect effect on what proposals investigators take to these sources. It remains an open question the extent to which these other sources themselves punish scientific risk taking.

[24] This represents 1.4% by number of awards and 1.8% by funding amount of all R01-equivalent awards that year at the NIH (Lauer 2022).

novelty or an over-reliance on criteria that are easy to measure but may be poorly correlated with promising risk taking. For instance, as high-risk projects may have greater variation in evaluator assessments of quality, moving from a consensus-based to a champion-based approach or emphasizing the degree of enthusiastic support rather than mean score may facilitate risk taking (Azoulay and Li 2022, Franzoni et al. 2022). Processes which encourage independent assessment of projects by the entire population of evaluators may help better identify promising risky proposals than those in which only a subset of the panel evaluates each application in detail. Similarly, who comprises the evaluation panel may affect the evaluation. For instance, expert evaluators may have more pronounced bias against novelty in their own field (Wang et al. 2017). Having greater diversity of expertise and traditions, as well as explicitly considering an evaluator's own prior risk taking behavior, when consider whom to appoint to select innovation proposals may also reduce risk penalization (Nicholson and Ioannidis 2012). Finally, a greater emphasis on selection and outcomes evaluation at the level of the portfolio of projects rather than the individual project level may facilitate greater acceptance of strategic risk taking.

Second, a fundamental challenge of risky innovation is ex post, a failed but highly promising risky research project and a low-quality idea can look much the same, with no publication, patent, or product to show for the effort. While innovators can be incentivized by status and financial awards for successful risky innovation, they may be reluctant to do so when failed risky projects might lead to job loss or lab closure. While it is important to maintain high standards and ultimate accountability (Manso 2011), this may mean adjusting the criteria upon which innovators are evaluated to focus more on innovation inputs and the level of risk rather than simply the outputs of innovative efforts.

Finally, we encourage greater efforts to build evaluation into the design of programs to support risk taking. Such efforts to turn the scientific method upon the design of institutions and processes to support risk taking is likely to allow funders to more effectively develop the portfolio of projects most likely to meet their goals. We believe such efforts are a risk worth taking.

References

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. "Academic freedom, private-sector focus, and the process of innovation." *RAND Journal of Economics*, 39(3):617-635.

Alberts, Bruce, Marc W. Kirschner, Shirley Tilghman, and Harold Varmus. 2014. "Rescuing US biomedical research from its systemic flaws." *Proceedings of the National Academy of Sciences*, 111(16):5773-5777.

Althaus, Catherine E. 2005. "A disciplinary perspective on the epistemological status of risk." *Risk Analysis*, 25(3):567-588

Angrist, Joshua D., and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Arora, Ashish, Sharon Belenzon, and Andrea Patacconi. 2017. "The decline of science in corporate R&D." *Strategic Management Journal*, 39(1):3-32.

Arts, Sam, and Lee Fleming. 2018. "Paradise of Novelty—Or Loss of Human Capital? Exploring New Fields and Inventive Output" *Organization Science*, 29(6):1074-1092.

Ayoubi, Charles, Michele Pezzoni, and Fabiana Visentin. 2021. "Does it Pay to do Novel Science? The Selectivity Patterns in Science Funding." *Science and Public Policy,* 48(5):635-648.

Ayoubi, Charles, Michele Pezzoni, and Fabiana Visentin. 2019. "The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions?" *Research Policy*, 48(1):84-97.

Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso. 2011. "Incentives and creativity: evidence from the academic life sciences." *RAND Journal of Economics*, 42(3):527-554.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven Sampat. 2019. "Public R&D and Private-Sector Patenting: Evidence from NIH Funding Rules." *The Reviewed of Economic Studies,* 86(1):117-152.

Azoulay, Pierre, Wesley H. Greenblatt, and Misty L. Heggeness. 2021. "Long-term effects from early exposure to research: Evidence from the NIH 'Yellow Berets.'" *Research Policy*, 50(9):104332.

Azoulay, Pierre and Danielle Li. 2022. "Scientific Grant Funding." In Goolsbee, Austan and Benjamin Jones (Eds). *Innovation and Public Policy*. Chicago, IL:University of Chicago Press.

Bachrach, C.A., and Thelma Charen. 1978. "Selection of Medline contents, the development of its thesaurus, and the indexing process." *Medical Informatics,* 3(3):237-254.

Banal-Estanol, Albert, Ines Macho-Stadler, and David Perez-Castrillo. 2019. "Evaluation in research funding agencies: Are structurally diverse teams biased against?" *Research Policy*, 48(7):1823-1840.

Bornmann, Lutz, Sitaram Devarakonda, Alexander Tekles, and George Chacko. 2020. "Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers." *Quantitative Science Studies,* 1(3):1242-1259.

Boudraeu, Kevin J., Eva C. Guinan, Karim R. Lakhani, and Christoph Riedl. 2016. "Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science." *Management Science*, 62(10):2765-2783.

Bromham, Lindell, Russell Dinnage, and Xia Hua. 2016. "Interdisciplinary research has consistently lower funding success." *Nature*, 534(7609):684-687.

Carson, Richard T., Joshua Graff Zivin, Jordan J. Louviere, Sally Sadoff, and Jeffrey G. Shrader. 2022. "The Risk of Caution: Evidence from an Experiment." *Management Science*, 68(12):9042-9060.

Chubin, Daryl E., and Edward J. Hackett. 1990. *Peerless Science: Peer Review and U.S. Science Policy.* Albany, NY: SUNY Press.

Clancy, Matt. 2023. "Biases Against Risky Research." New Things Under the Sun.

Cole, Stephen, Jonathan R. Cole, and Gary A. Simon. 1981. "Chance and Consensus in Peer Review." *Science*, 214(4523):881-886.

Coletti, Margaret H. and Howard L. Bleich. 2001. "Medical Subject Headings Used to Search the Biomedical Literature." *Journal of the American Medical Informatics Association*, 8(4):317-323.

Collison, Patrick, Tyler Cowen, and Patrick Hsu. 2021. "What We Learned Doing Fast Grants." *Future*.

Danthi, Narasimhan, Colin O. Wu, Peibei Shi, and Michael Lauer. 2014. "Percentile ranking and citation impact of a large cohort of National Heart, Lung, and Blood Institute-funded cardiovascular R01 grants." *Circulation Research*, 114(4):600-606.

Dasgupta, Partha, and Paul A. David. 1994. "Toward a new economics of science." *Research Policy*, 23:487-521.

Fang, Ferric C., and Arturo Casadevall. 2015. "Competitive science: Is competition ruining science?" *Infection and Immunity*, 83(4):1229-1233.

Fields, Stanley. 2014. "Would Fred Sanger get funded today?" *Genetics*, 197(2):435-439.

Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science*, 47(1):117-132.

Fleming, Lee, Hillary Greene, Guan-Cheng Li, Matt Marx, and Dennis Yao. 2019. "Government-funded research increasingly fuels innovation." *Science*, 364(6446):1139-1141.

Foster, Jacob G., Andrey Rzhetsky, and James A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review,* 80(5):875-908.

Fox, Craig R., and Amos Tversky. 1995. "Ambiguity Aversion and Comparative Ignorance." *Quarterly Journal of Economics,* 110(3):585-603.

Franzoni, Chiara, and Paula Stephan. 2023. "Uncertainty and risk-taking in science: Meaning, measurement and management in peer review of research proposals." *Research Policy*, 52(3):104706.

Franzoni, Chiara, Paula Stephan, and Reinhilde Veugelers. 2022. "Funding Risky Research." *Entrepreneurship and Innovation Policy, 1:103-133.*

Funk, Mark E., and Carolyn Anne Reid. "Indexing consistency in MEDLINE." *Bulletin of the Medical Library Association*, 71(2):176-183.

Funk, Russell J., and Jason Owen-Smith. 2017. "A Dynamic Network Measure of Technological Change." *Management Science*, 63(3):791-817.

Gerin, William, Christine K. Kinkade, and Niki L. Page. 2018. *Writing the NIH Grant Proposal: A Step-by-Step Guide.* Thousand Oaks, CA: Sage Publications.

Graff Zivin, Joshua, and Elizabeth Lyons. 2021. "The Effects of Prize Structures on Innovative Performance." *AEA Papers and Proceedings*, 111:577-581.

Greenblatt, Wesley, Charu Gupta, and Jennifer Kao. 2023. "Drug Repurposing During the COVID-19 Pandemic: Lessons for Expediting Drug Development and Access." Health Affairs, 42(3):424-432.

Higgins, Monica C. 2005. *Career Imprints: Creating Leaders Across an Industry.* San Francisco: Jossey-Bass/Wiley.

Hill, Ryan, Yian Yin, Carolyn Stein, Dashun Wang, and Benjamin F. Jones. 2021. "Adaptability and the Pivot Penalty in Science." Working paper.

Jacob, Brian A. and Lars Lefgren. 2011. "The Impact of research grant funding on scientific productivity." Journal of Public Economics, 95(9-10):1168-1177.

Kent, David. 2018. "Scientists can't take risks until they are in their 40s." *University Affairs.*

Kerr, William R., Ramana Nanda, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as Experimentation." *Journal of Economic Perspectives,* 28(3):25-48.

Krieger, Joshua, Danielle Li, and Dimitris Papanikolaou. 2022. "Missing Novelty in Drug Development." *The Review of Financial Studies*, 35(2):636-679.

Kolata, Gina. 2009. "Grant System Leads Cancer Researchers to Play it Safe." *New York Times.*

Kornberg, Roger D. 2007. "2006 Novel Laureates." U.S. Senate Committee on Commerce, Science, and Transpiration.

Krieger, Joshua, and Ramana Nanda. 2022. "Are Transformational Ideas Harder to Fund? Resource Allocation to R&D Projects at a Global Pharmaceutical Firm." Harvard Business School Working Paper No. 23-014.

Lane, Jacqueline N., Zoe Szajnfarber, Jason Crusan, Michael Menietti, and Karim R. Lakhani. 2022a. "Are Experts Blinded by Feasibility? Experimental Evidence from a NASA Robotics Challenge." Harvard Business School Working Paper No. 22-071.

Lane, Jacqueline N., Misha Teplitskiy, Gary Gray, Hardeep Ranu, Michael Menietti, Eva C. Guinan, and Karim R. Lakhani. 2022b. "Conservatism Gets Funded? A Field Experiment on the Role of Negative Information in Novel Project Evaluation." *Management Science*, 68(6):4478-4495.

Langer, James S. 2012. "Enabling Scientific Innovation." *Science*, 338(6104):171.

Lauer, Mike. 2022. "FY 2021 by the Numbers: Extramural Grant Investments in Research." National Institutes of Health Office of Extramural Research.

Lauer, Mike. 2016. "Are Attempts at Renewal Successful?." National Institutes of Health Office of Extramural Research.

Lee, You-Na, John P. Walsh, and Jian Wang. 2015. "Creativity in scientific teams: Unpacking novelty and impact." *Research Policy*, 44(3):684-697.

Lercehmueller, Marc J. and Olav Sorenson. 2016. "Author Disambiguation in PubMed: Evidence on the Precision and Recall of Author-ity among NIH-Funded Scientists." *PLoS One,* 11(7):e0158731.

Lerner, Josh and Julie Wulf. 2007. "Innovation and Incentives: Evidence from Corporate R&D." *Review of Economics and Statistics*, 89(4):634-644.

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

Lorsch, Jon R. 2015. "Maximizing the return on taxpayers' investments in fundamental biomedical research." *Molecular Biology of the Cell*, 26(9):1578-1582.

Manso, Gustavo. 2011. "Motivating Innovation." *Journal of Finance*, 66(5):1823-1860.

March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science*, 2(1):71-87.

Mishra, Shubhanshu and Vetle I. Torvik. 2016. "Quantifying Conceptual Novelty and the Biomedical Literature." *Dlib Mag.*, 22(9-10).

Moses 3rd, Hamilton, David H. M. Matheson, Sarah Cairns-Smith, Benjamin P. George, Chase Palisch, and E. Ray Dorsey. 2015. "The anatomy of medical research: US and international comparisons." *Journal of the American Medical Association*, 313(2):174-89.

Mueller, Jennifer S., Shimul Melwani, and Jack A. Goncalo. 2012. "The Bias Against Creativity: Why People Desire but Reject Creative Ideas." *Psychological Science*, 23(1):13-17.

Murray, Fiona. 2012. "Evaluating the Role of Science Philanthropy in American Universities." *Innovation Policy and the Economy*, 13:23-59.

National Institutes of Health. 2021. "NIH supports 106 grants featuring high-risk, high-reward research." https://www.nih.gov/news-events/news-releases/nih-supports-106-grants-featuring-high-risk-high-reward-research. Press release.

National Institutes of Health. 2019. "Report of the ACD Working Group on High-Risk, High-Reward Research." Advisory Committee to the Director Working Group on High-Risk High-Reward Programs, National Institutes of Health.

Nielson, Michael and Kanjun Qiu. 2022. "The trouble in comparing different approaches to science funding." The Science++ Project.

Nielsen, Michel (michael_nielsen). 2022. "(A brief afternoon rant.)" Tweet.

Neyman, Jerzy, and Elizabeth L. Scott. 1948. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica* 16(1):1-32.

Nicholson, Joshua M. and John P. A. Ioannidis. 2012. "Conform and be funded." *Nature*, 492(7427):34-36.

Packalen, Mikko, and Jay Bhattacharya. 2020. "NIH funding and the pursuit of edge science." *Proceedings of the National Academy of Sciences*, 117(22):12011-12016.

Petsko, Gregory A. 2012. "Goodbye, Columbus." *Genome Biology*, 13(5):155.

Pier, Elizabeth L. Markus Brauer, Amarette Filut, Anna Kaatz, Joshua Raclaw, Mitchell J. Nathan, Cecilia E. Ford, and Molly Carnes. 2018. "Low agreement among reviewers evaluating the same NIH grant applications." *Proceedings of the National Academy of Sciences,* 115(2):2952-2957.

Rosenberg, Nathan. 1995. "Uncertainty and Technological Change." *Federal Reserve Bank of Boston Conference Proceedings*, 40(Jun):91-125.

Rothwell, Peter M., and Christopher N. Martyn. 2000. "Reproducibility of peer review in clinical neuroscience." *Brain*, 123:1964-1969.

Ruben, Adam. 2017. "Another tenure-track scientist bites the dust." Experimental Error, *Science*, https://www.science.org/content/article/another-tenure-track-scientist-bites-dust.

Rzhetsky, Andrey, Jacob G. Foster, Ian T. Foster, and James A. Evans. 2015. "Choosing experiments to accelerate collective discovery." *Proceedings of the National Academy of Sciences*, 112(47):14569-74.

Schilling, Melissa A. and Elad Green. 2011. "Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences." *Research Policy*, 40(10):1321-1331.

Shibayama, Sotaro, Deyun Yin, and Kuniko Matsumoto. 2021. "Measuring novelty in science with word embedding. *PLoS ONE*, 16(7):e0254034.

Shibayama, Sotaro and Jian Wang. 2020. "Measuring originality in science." *Scientometrics*, 122:409-427.

Sutter, Paul M. 2022. "Risk Aversion Is Ruining Science." *Undark*.

Stephan, Paula E. 2012. *How Economics Shapes Sciences.* Cambridge, MA: Harvard University Press.

Teplitskiy, Misha, Hao Peng, Andrea Blasco, and Karim R. Lakhani. 2022. "Is novel research worth doing? Evidence from peer review at 49 journals." *Proceedings of the National Academy of Sciences*, 119(47):e2118046119.

Tian, Xuan and Tracy Yue Wang. 2014. "Tolerance for Failure and Corporate Innovation." *Review of Financial Studies*, 27(1):211-255.

Torvik, Vetle I. and Neil R. Smalheiser. 2009. "Author name disambiguation in MEDLINE." *ACM Transactions on Knowledge Discovery from Data*, 3(3):11.

Torvik, Vetle I. and Neil R. Smalheiser. 2021. Author-ity 2018 - PubMed author name disambiguated dataset. University of Illinois at Urbana-Champaign.

Torvik, Vetle I, Marc Weeber, Don R. Swanson, and Neil R. Smalheiser. 2005. "A Probabilistic Similarity Metric for Medline Records: A Model for Author Name Disambiguation." *Journal of the American Society for Information Science and Technology*, 56(2):140-158.

Uzzi, Brian, Satyam Mukherjee Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science*, 342(6157):468-472.

Verhoeven, Dennis, Jurrien Bakker, and Reinhilde Veugelers. 2016. "Measuring technological novelty with patent-based indicators." *Research Policy*, 45(3):707-723.

Veugelers, Reinhilde, Jian Wang, and Paula Stephan. 2022. "Do Funding Agencies Select and Enable Risky Research: Evidence from ERC Using Novelty as a Proxy of Risk Taking." National Bureau of Economic Research Working Paper #30320.

Veugelers, Reinhilde and Jian Wang. 2019. "Scientific novelty and technological impact." *Research Policy*, 48:1362-1372.

Wang, Jian, Reinhilde Veugelers, and Paula Stephan. 2017. "Bias against novelty in science: A cautionary tale for users of bibliometric indicators." *Research Policy,* 46:1416-1436.

Woolston, Chris. 2014. "Sanger's legacy stirs up digital chatter." *Nature*, 511(9).

Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large teams develop and small teams disrupt science and technology." *Nature*, 566(7744):378-382.

Table 1a. Descriptive statistics: Investigator characteristics

	Mean	Median	Std. Dev.	Min.	Max.
Female	0.25	0	0.44	0	1
Degree year	1982.15	1982	12.26	1932	2012
Degree: PhD	0.68	1	0.47	0	1
Degree: MD	0.20	0	0.40	0	1
Degree: MD/PhD	0.10	0	0.31	0	1
Degree: Other	0.02	0	0.13	0	1
Nb of R01-equivalent grants	1.72	1	1.16	1	14
Nb of R01-equivalent grant cycles	3.05	2	2.78	1	37
Career NIH R01 funding ($2020)	10,268,954	6,953,402	10,515,349	133,632	134,399,904
Career NIH funding ($2020)	16,717,503	8,803,164	27,005,356	133,632	1,743,186,688
Career elite status	0.05	0	0.22	0	1

Note: Number of grants and cycles refers to only those grants that started 1980-2015. Career elite status identifies investigators who were ever won a Nobel Prize, Lasker Award, Howard Hughes Medical Institute Investigatorship, or membership in the Institute of Medicine or National Academy of Sciences. N=37,222 investigators.

Table 1b. Descriptive statistics: Grant cycle

	Renewed grant cycles					Not renewed grant cycles				
	Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
Grant cycle number	1.94	1	1.28	1	10	1.69	1	1.15	1	11
First cycle of grant	0.51	1	0.50	0	1	0.62	1	0.49	0	1
Year started	1997.85	1999	9.16	1980	2015	2000.38	2002	9.82	1980	2015
Cycle length (years)	4.37	4	1.59	1	19	3.16	3	1.11	1	16
Time, degree to cycle start (years)	17.62	16	8.48	-22	61	19.57	18	9.42	-18	66
Investigator pubs. prior to cycle	53.35	36	57.52	0	1148	58.66	39	63.74	0	1196
Investigator citations prior to cycle	2,118.12	887	3,952.72	0	89,857	2,329.03	967	4,443.51	0	184,460
Publications acknowledging grant										
Nb of publications	8.06	6	7.87	1	184	7.01	5	7.34	1	164
Journal impact factor	6.28	5	4.00	0	72	5.36	5	3.51	0	43
Citations	662.91	355	1,146.86	1	6,4421	421.90	201	797.27	1	67,451
Highly selective journal	0.12	0	0.33	0	1	0.07	0	0.25	0	1
Publications not acknowledging grant										
Nb of publications	9.38	5	13.25	0	328	11.33	6	15.26	0	258
Journal impact factor	6.08	5	4.25	0	43	5.19	4	3.42	0	43
Citations	824.09	291	1,764.14	0	50,927	780.04	287	1,669.59	0	65,447
Highly selective journal	0.17	0	0.37	0	1	0.12	0	0.33	0	1
Fraction with any NIH funding	0.58	1	0.33	0	1	0.60	1	0.32	0	1

Note: Unit of analysis is at the investigator-grant-cycle level. N=103,164 (46,619 renewed and 56,545 not renewed investigator-grant-cycles). Grant cycle number minus one is the number of time the grant has previously been competitively renewed. Time from degree to cycle is measured from the last terminal degree earned by the investigator; 55 individuals (92 investigator-grant-cycles) earned an R01-equivalent grant prior to their last degree. This largely reflects investigators with multiple doctorates, frequently with large time gaps between them. Publications acknowledging a grant reflect those publications citing funding from the focal grant which were published up to one year after the conclusion of the grant cycle; publications not acknowledging the focal grant represents the investigator's other publications during the same time period. Citations are measured through 2020.

Table 1c. Descriptive statistics: Measures of risk taking and novelty

	Renewed grant cycles					Not renewed grant cycles				
	Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
Extreme tail outcomes	46.69	47	23.80	0	100	47.40	48	23.79	0	100
Disruption	62.43	63	21.35	0	100	64.59	66	20.77	0	100
Pivoting	0.73	1	0.14	0	1	0.74	1	0.14	0	1
Standing out	0.87	1	0.11	0	1	0.85	1	0.11	0	1
Novelty	0.08	0	0.06	0	1	0.07	0	0.06	0	1

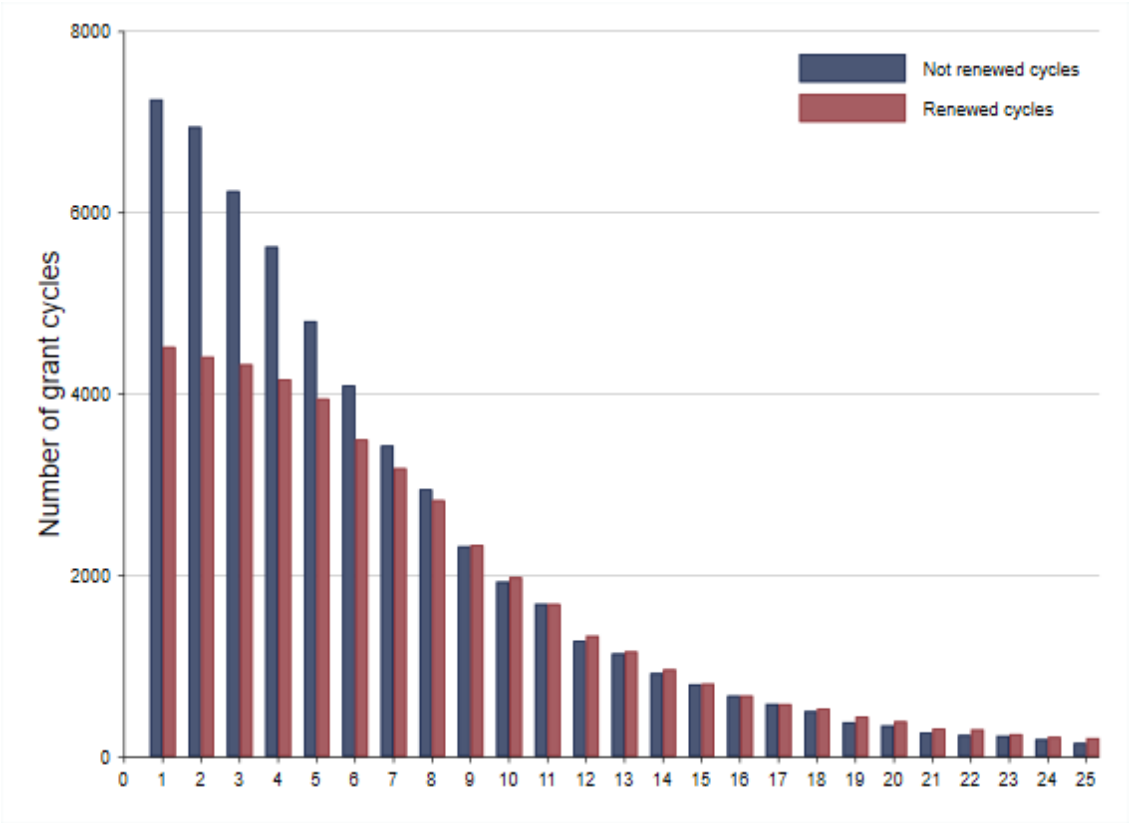
Note: Outcomes are measured at the investigator-grant-cycle level (N=103,164 investigator-grant-cycles). Measures are only defined for those cycles with qualifying associated publications (see footnote 7 for details).

Table 1d. Descriptive statistics: Correlations across measures of risk taking and novelty

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Extreme tail outcomes	1.000						
(2) Disruption	0.399	1.000					
(3) Pivoting	-0.062	0.029	1.000				
(4) Standing out	0.099	0.036	0.118	1.000			
(5) Novelty	-0.055	-0.030	0.317	0.053	1.000		
(6) Nb. of publications	0.508	0.351	-0.125	0.119	-0.021	1.000	
(7) Log citations	0.353	0.192	-0.026	0.131	0.103	0.613	1.000

Note: N=103,164 investigator-grant-cycles. Measures are only defined for those cycles with qualifying associated publications (see footnote 7 for details). Number of publications and citations are based solely on publications acknowledging funding from the focal grant during the grant cycle. Citations are measured through 2020.

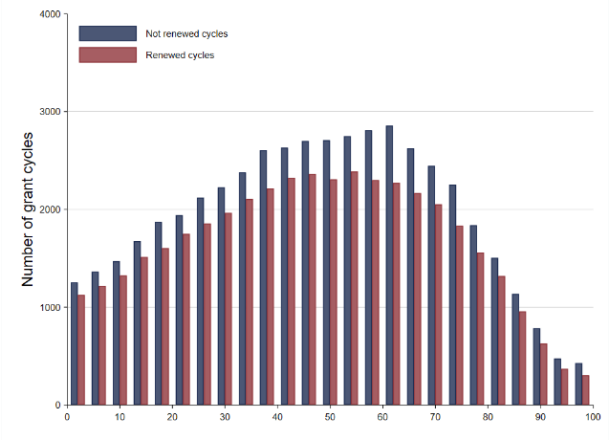
Figure 1. Distribution of grant cycle publications by renewal status



Note: Histogram of the number of publications acknowledging the grant within the grant cycle. Unit of analysis is at the investigator-grant-cycle level. Two thousand nine hundred fifty-two (2.9% of the sample) outliers with over 25 grant cycle publications are not shown.

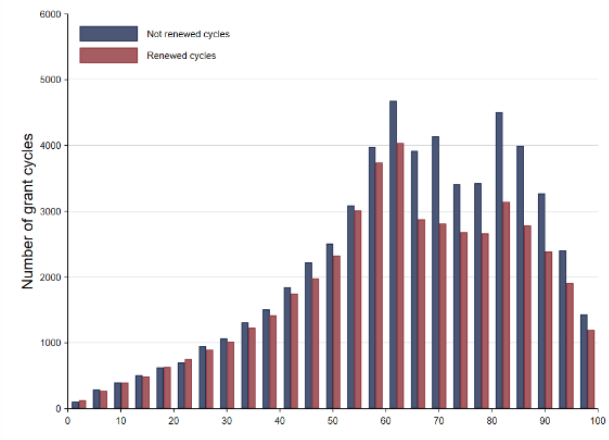
Figure 2. Distribution of risk-taking measures by renewal status

A. Extreme tail outcomes



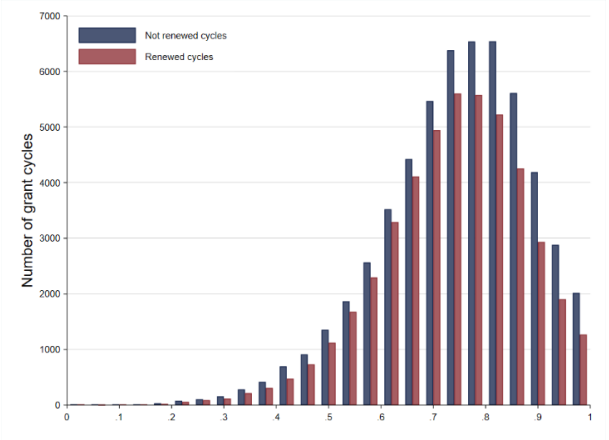
Note: Histogram of range between the maximum and minimum vintage-adjusted citation percentile for grant cycle publications. Unit of analysis is at the investigator-grant-cycle level.

B. Disruption



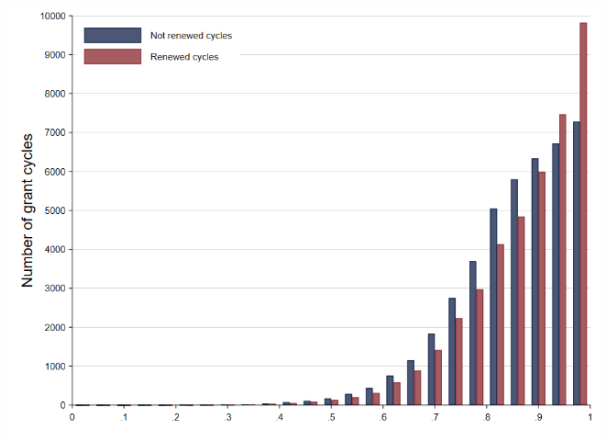
Note: Histogram of the maximum disruption index percentile among grant cycle publications. Unit of analysis is at the investigator-grant-cycle level.

C. Pivoting



Note: Histogram of the fraction of MeSH term pairs for grant cycle publications that were not used in the investigator's publications in the 5 years preceding the grant. Unit of analysis is at the investigator-grant-cycle level.

D. Standing out



Note: Histogram of the fraction of MeSH term pairs for grant cycle publications that were not used by funded grant proposal abstracts in the same NIH IC-study section-year. Unit of analysis is at the investigator-grant-cycle level.

Table 2. Effect of risk taking on grant renewal

A. Extreme tail outcomes

	All renewals			First renewal			Second and future renewals		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<10%ile, max-min citation percentile	0.054**	0.051**	0.022**	0.061**	0.055**	0.028**	0.064**	0.057**	0.018*
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.008)	(0.008)	(0.008)
>90%ile, max-min citation percentile	-0.039**	-0.044**	-0.021**	-0.039**	-0.040**	-0.018**	-0.047**	-0.048**	-0.020**
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Investigator demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year and length	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IC x study section	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institutional quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Investigator quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Pub. quantity, quality during cycle	No	No	Yes	No	No	Yes	No	No	Yes
Mean of dependent variable	0.463	0.463	0.463	0.413	0.413	0.413	0.523	0.523	0.523
Std. Dev. of dependent variable	0.499	0.499	0.499	0.492	0.492	0.492	0.499	0.499	0.499
Effect bottom %ile group, in s.d. units	0.108	0.103	0.044	0.123	0.111	0.057	0.129	0.115	0.037
Effect top %ile group, in s.d. units	-0.078	-0.088	-0.041	-0.080	-0.082	-0.037	-0.093	-0.095	-0.040
Adjusted R^2	0.4685	0.5026	0.5175	0.5055	0.5267	0.5370	0.4572	0.4983	0.5166
Nb. of investigators	33,409	33,403	33,403	30,446	30,445	30,445	17,274	17,260	17,260
Nb. of investigator-grant-cycles	90,123	90,075	90,075	48,530	48,529	48,529	40,859	40,810	40,810

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns 1-3 includes the full sample, columns 4-6 limit the sample to the first cycle for each investigator-grant, and columns 7-9 limit it to cycles 2 and above for each investigator-grant. The independent variables are indicator variables for <10%ile and >90%ile for the difference between the maximum and minimum vintage-adjusted citation percentile of grant cycle publications, with the 10-90%iles as the omitted category. Columns 1, 4 and 7 include indicator variables for gender, investigator degree, investigator career age, year of grant evaluation, length of the grant cycle, and a set of 2,911 indicator variables corresponding to each NIH institute or center and study section. Columns 2, 5, and 8 include all of the previous controls, but also include log NIH funding to the investigator's institution, log investigator prior career citations, log investigator prior NIH grant funding, and an indicator for no prior NIH grant funding. Finally, columns 3, 6, and 9 include all previous controls, but also add log mean journal impact factor, indictor variables for grant and non-grant publications in one of four highly selective journals, and a set of 21 indicator variables separately for the number of publications during the grant cycle acknowledging and not acknowledging the grant. Robust standard errors were used, clustering at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

B. Disruption

	All renewals			First renewal			Second and future renewals		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<10%ile, maximum disruption index	0.091**	0.089**	0.054**	0.094**	0.089**	0.057**	0.080**	0.078**	0.037
	(0.012)	(0.012)	(0.012)	(0.014)	(0.013)	(0.013)	(0.024)	(0.023)	(0.023)
>90%ile, maximum disruption index	-0.019**	-0.022**	0.008†	-0.020**	-0.020**	0.008	-0.023**	-0.024**	0.009
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Investigator demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year and length	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IC x study section	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institutional quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Investigator quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Pub. quantity, quality during cycle	No	No	Yes	No	No	Yes	No	No	Yes
Mean of dependent variable	0.454	0.454	0.454	0.405	0.405	0.405	0.519	0.519	0.519
Std. Dev. of dependent variable	0.498	0.498	0.498	0.491	0.491	0.491	0.500	0.500	0.500
Effect bottom %ile group, in s.d. units	0.183	0.179	0.108	0.192	0.180	0.116	0.159	0.156	0.075
Effect top %ile group, in s.d. units	-0.039	-0.044	0.016	-0.040	-0.041	0.016	-0.045	-0.048	0.017
Adjusted R^2	0.4481	0.4833	0.5017	0.4798	0.5028	0.5169	0.4427	0.4861	0.5074
Nb. of investigators	36,876	36,876	36,876	34,889	34,889	34,889	18,196	18,196	18,196
Nb. of investigator-grant-cycles	102,250	102,250	102,250	57,667	57,667	57,667	43,832	43,832	43,832

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns 1-3 includes the full sample, columns 4-6 limit the sample to the first cycle for each investigator-grant, and columns 7-9 limit it to cycles 2 and above for each investigator-grant. The independent variables are indicator variables for <10%ile and >90%ile for the maximum of the disruption index for grant cycle publications, with the 10-90%iles as the omitted category. Columns 1, 4 and 7 include indicator variables for gender, investigator degree, investigator career age, year of grant evaluation, length of the grant cycle, and a set of 2,911 indicator variables corresponding to each NIH institute or center and study section. Columns 2, 5, and 8 include all of the previous controls, but also include log NIH funding to the investigator's institution, log investigator prior career citations, log investigator prior NIH grant funding, and an indicator for no prior NIH grant funding. Finally, columns 3, 6, and 9 include all previous controls, but also add log mean journal impact factor, indictor variables for grant and non-grant publications in one of four highly selective journals, and a set of 21 indicator variables separately for the number of publications during the grant cycle acknowledging and not acknowledging the grant. Robust standard errors were used, clustering at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

C. Pivoting

	All renewals			First renewal			Second and future renewals		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<10%ile, fraction new MeSH pairs	0.068**	0.053**	0.059**	0.052**	0.042**	0.044**	0.071**	0.057**	0.063**
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
>90%ile, fraction new MeSH pairs	-0.039**	-0.024**	-0.043**	-0.019**	-0.012**	-0.030**	-0.052**	-0.039**	-0.062**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.009)	(0.008)	(0.008)
Investigator demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year and length	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IC x study section	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institutional quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Investigator quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Pub. quantity, quality during cycle	No	No	Yes	No	No	Yes	No	No	Yes
Mean of dependent variable	0.455	0.455	0.455	0.406	0.406	0.406	0.519	0.519	0.519
Std. Dev. of dependent variable	0.498	0.498	0.498	0.491	0.491	0.491	0.500	0.500	0.500
Effect bottom %ile group, in s.d. units	0.137	0.107	0.119	0.106	0.085	0.090	0.141	0.115	0.127
Effect top %ile group, in s.d. units	-0.079	-0.048	-0.086	-0.040	-0.025	-0.060	-0.105	-0.077	-0.124
Adjusted R^2	0.4499	0.4847	0.5039	0.4802	0.5032	0.5181	0.4453	0.4887	0.5103
Nb. of investigators	36,544	36,539	36,539	34,497	34,497	34,497	18,127	18,114	18,114
Nb. of investigator-grant-cycles	101,389	101,338	101,338	57,008	57,007	57,007	43,629	43,577	43,577

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns 1-3 includes the full sample, columns 4-6 limit the sample to the first cycle for each investigator-grant, and columns 7-9 limit it to cycles 2 and above for each investigator-grant. The independent variables are indicator variables for <10%ile and >90%ile for the fraction of MeSH term pairs from grant cycle publications that were not used by the investigator in the 5 years preceding the grant cycle, with the 10-90%iles as the omitted category. Columns 1, 4 and 7 include indicator variables for gender, investigator degree, investigator career age, year of grant evaluation, length of the grant cycle, and a set of 2,911 indicator variables corresponding to each NIH institute or center and study section. Columns 2, 5, and 8 include all of the previous controls, but also include log NIH funding to the investigator's institution, log investigator prior career citations, log investigator prior NIH grant funding, and an indicator for no prior NIH grant funding. Finally, columns 3, 6, and 9 include all previous controls, but also add log mean journal impact factor, indictor variables for grant and non-grant publications in one of four highly selective journals, and a set of 21 indicator variables separately for the number of publications during the grant cycle acknowledging and not acknowledging the grant. Robust standard errors were used, clustering at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

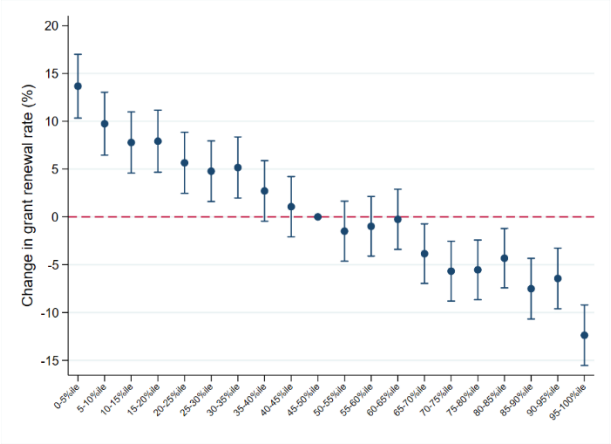
D. Standing out

	All renewals			First renewal			Second and future renewals		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<10%ile, fraction unique MeSH pairs	0.040**	0.038**	0.025**	0.037**	0.035**	0.022**	0.048**	0.044**	0.031**
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.009)	(0.008)	(0.008)
>90%ile, fraction unique MeSH pairs	-0.022**	-0.023**	-0.027**	-0.003	-0.003	-0.010	-0.035**	-0.041**	-0.041**
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Investigator demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grant year and length	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IC x study section	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institutional quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Investigator quality	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Pub. quantity, quality during cycle	No	No	Yes	No	No	Yes	No	No	Yes
Mean of dependent variable	0.491	0.491	0.491	0.443	0.443	0.443	0.551	0.551	0.551
Std. Dev. of dependent variable	0.500	0.500	0.500	0.497	0.497	0.497	0.497	0.497	0.497
Effect bottom %ile group, in s.d. units	0.081	0.077	0.050	0.075	0.070	0.044	0.096	0.088	0.061
Effect top %ile group, in s.d. units	-0.043	-0.046	-0.053	-0.006	-0.007	-0.020	-0.071	-0.083	-0.083
Adjusted R^2	0.4294	0.4688	0.4905	0.4690	0.4941	0.5106	0.4200	0.4694	0.4946
Nb. of investigators	32,370	32,364	32,364	30,209	30,209	30,209	16,041	16,031	16,031
Nb. of investigator-grant-cycles	83,005	82,962	82,962	46,902	46,901	46,901	35,650	35,609	35,609

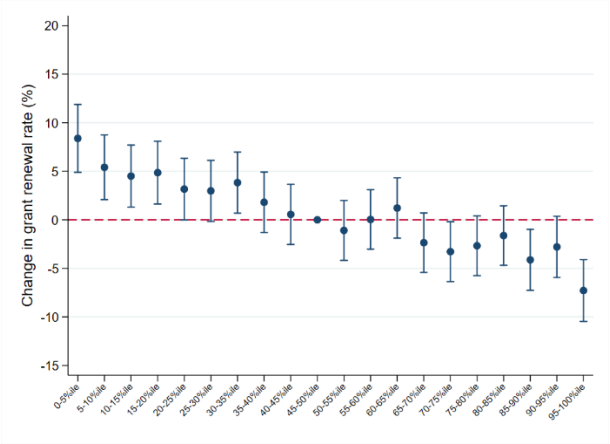
Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns 1-3 includes the full sample, columns 4-6 limit the sample to the first cycle for each investigator-grant, and columns 7-9 limit it to cycles 2 and above for each investigator-grant. The independent variables are indicator variables for <10%ile and >90%ile for the fraction of MeSH term pairs from grant cycle publications that were not used by funded grant proposal abstracts in the same NIH IC-study section-year, with the 10-90%iles as the omitted category. Columns 1, 4 and 7 include indicator variables for gender, investigator degree, investigator career age, year of grant evaluation, length of the grant cycle, and a set of 2,911 indicator variables corresponding to each NIH institute or center and study section. Columns 2, 5, and 8 include all of the previous controls, but also include log NIH funding to the investigator's institution, log investigator prior career citations, log investigator prior NIH grant funding, and an indicator for no prior NIH grant funding. Finally, columns 3, 6, and 9 include all previous controls, but also add log mean journal impact factor, indictor variables for grant and non-grant publications in one of four highly selective journals, and a set of 21 indicator variables separately for the number of publications during the grant cycle acknowledging and not acknowledging the grant. Robust standard errors were used, clustering at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure 3. Effect of risk taking on grant renewal: Extreme tail outcomes

A. Base controls



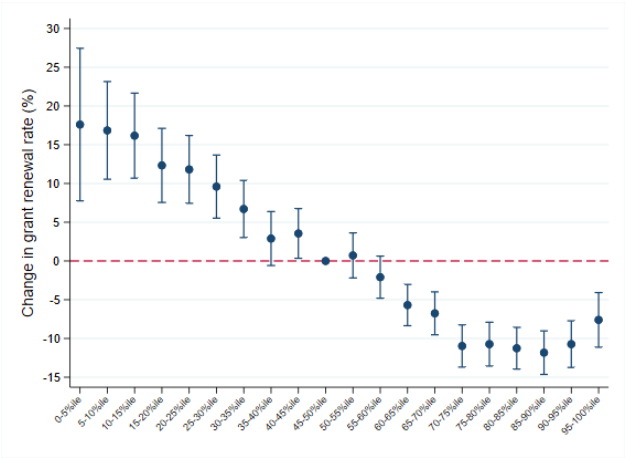
B. Controlling for grant publication quantity and quality



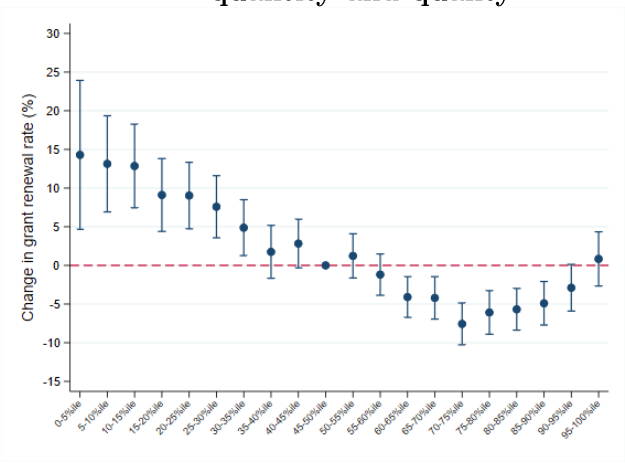
Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Panel A was modeled on Table 2, column 2, while panel B was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, the difference between the maximum and minimum vintage-adjusted citation percentile among grant cycle publications. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

Figure 4. Effect of risk taking on grant renewal: Disruption

A. Base controls

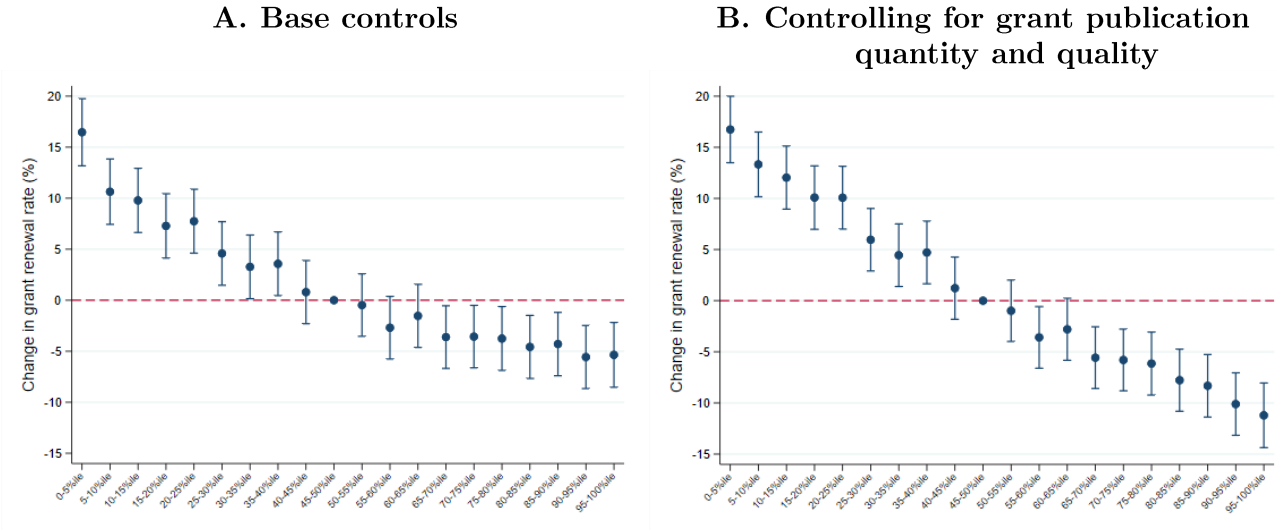


B. Controlling for grant publication quantity and quality



Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Panel A was modeled on Table 2, column 2, while panel B was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, the maximum of the disruption index percentile for grant cycle publications. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

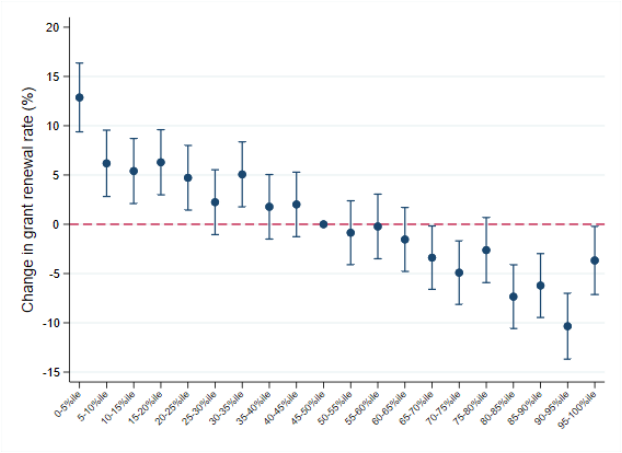
Figure 5. Effect of risk taking on grant renewal: Pivoting



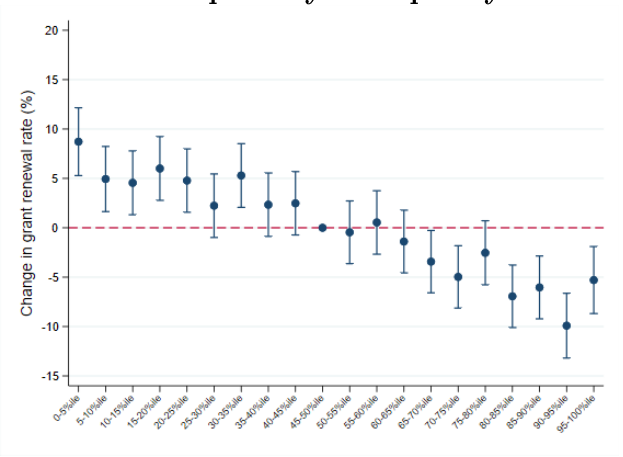
Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Panel A was modeled on Table 2, column 2, while panel B was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, the fraction of MeSH term pairs from grant cycle publications that were not used by the investigator in the 5 years preceding the grant cycle. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

Figure 6. Effect of risk taking on grant renewal: Standing out

A. Base controls

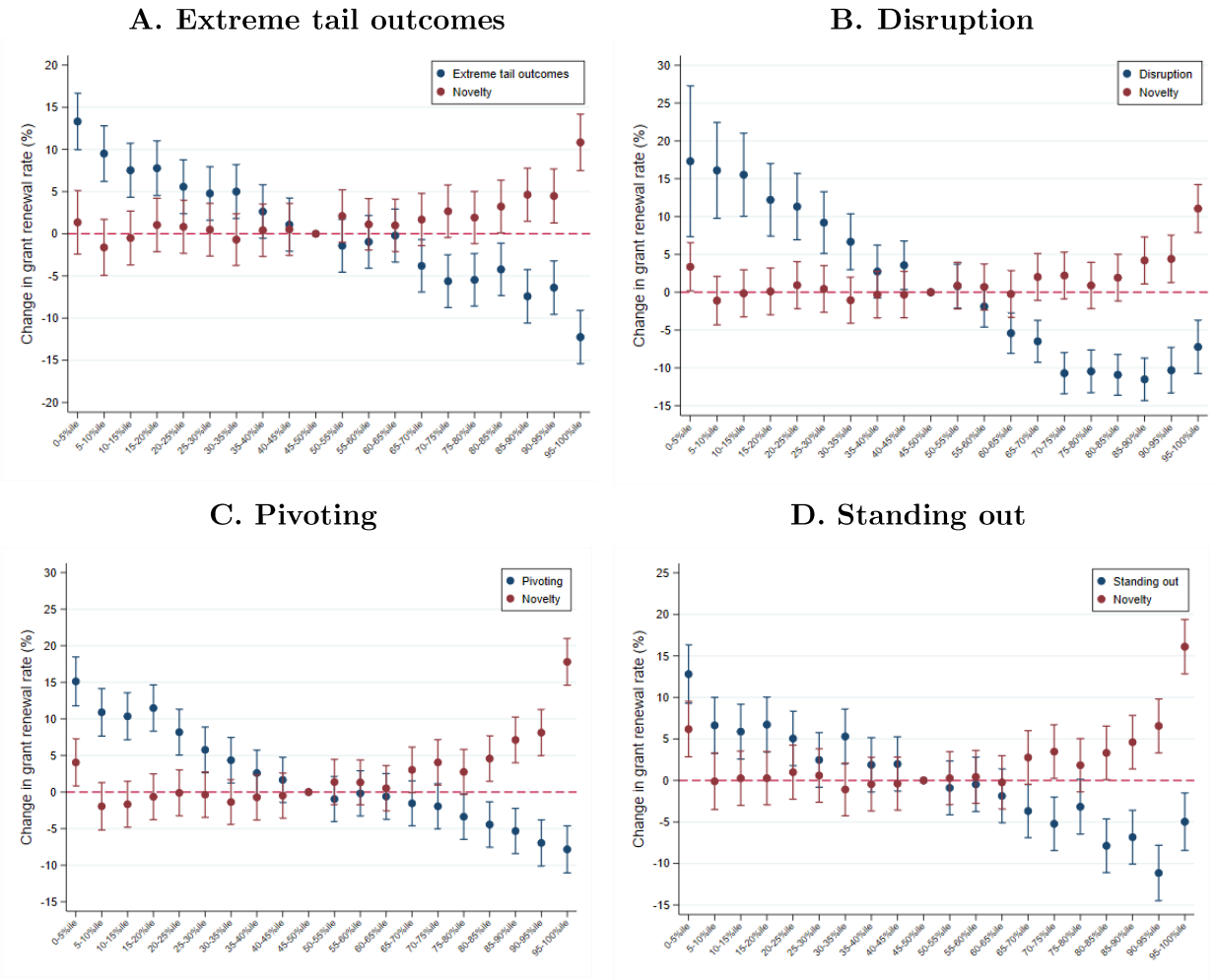


B. Controlling for grant publication quantity and quality



Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Panel A was modeled on Table 2, column 2, while panel B was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, the fraction of MeSH term pairs from grant cycle publications that were not used by funded grant proposal abstracts in the same NIH IC-study section-year. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

Figure 7. Effect of risk taking and novelty on grant renewal



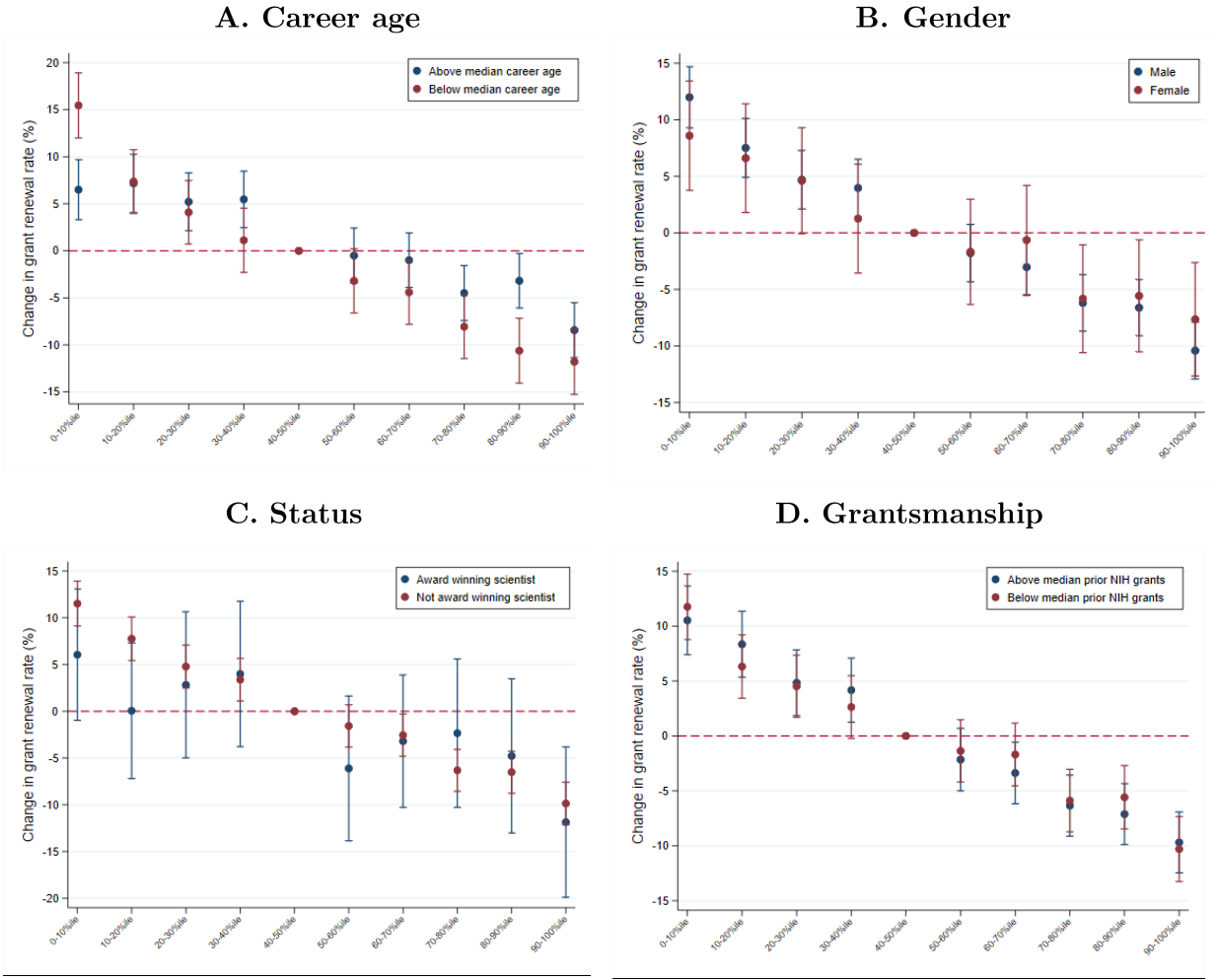
Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed modeled after Table 2, column 2. The unit of analysis is the investigator-grant-cycle. Each regression includes a set of 19 indicator variables for each ventile of the risk-taking measure, and 19 indicator variables for each ventile of novelty. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

Table 3. Effect of novelty on the risk taking grant renewal penalty

	Base controls				Grant output controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Novelty >50%ile	0.0254**	0.0211**	0.0922**	0.1137**	0.0284**	0.0191*	0.1200**	0.1252**
	(0.0057)	(0.0077)	(0.0142)	(0.0230)	(0.0056)	(0.0076)	(0.0139)	(0.0225)
Extreme tail outcomes	-0.0117**				-0.0058**			
	(0.0008)				(0.0009)			
Extreme tail outcomes × Novelty >50%ile	-0.0025*				-0.0025*			
	(0.0010)				(0.0010)			
Disruption		-0.0145**				-0.0078**		
		(0.0008)				(0.0008)		
Disruption × Novelty >50%ile		-0.0015				-0.0007		
		(0.0011)				(0.0011)		
Pivoting			-0.1601**				-0.2124**	
			(0.0121)				(0.0119)	
Pivoting × Novelty >50%ile			-0.0922**				-0.1178**	
			(0.0184)				(0.0181)	
Standing out				-0.1897**				-0.1513**
				(0.0201)				(0.0198)
Standing out × Novelty >50%ile				-0.1135**				-0.1215**
				(0.0261)				(0.0255)

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Each estimate includes a risk-taking measure as a continuous variable: extreme tail outcomes scaled so a change of 1 in the regressor reflects a change of 10 in the long-run vintage-adjusted citation percentile spread (columns 1 and 5), disruption scaled so a change of 1 in the regressor reflects a change of 10 in the maximum disruption index percentile (columns 2 and 6), pivoting (columns 3 and 7), and standing out (columns 4 and 8). All estimates include an indicator variable for above median novelty as well as an interaction term between this indicator variable and the risk-taking measure. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure 8. Heterogeneity in penalty for risk taking

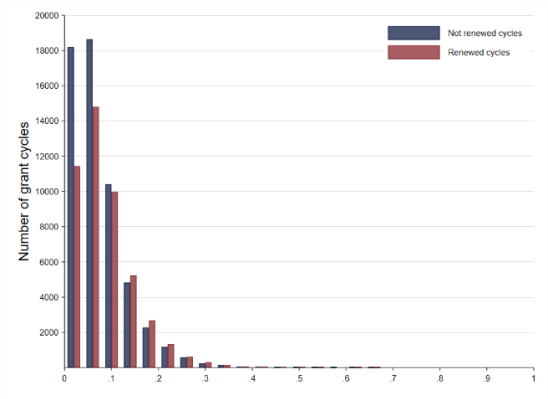


Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed modeled after Table 2, column 2. The unit of analysis is the investigator-grant-cycle. In all cases the risk-taking measure is extreme tail outcomes. Each regression includes a set of 9 indicator variables for each decile of risk taking interacted with an indicator for above median career age and another 9 indicator variables for each decile of risk taking interacted with an indicator for below median career age (Panel A). Panel B similarly interacts on gender, panel C on elite status as measured by prominent scientific awards, and panel D on NIH funding prior to the grant cycle. The omitted category is the 40-50%ile. Robust standard errors were used, clustered at the investigator and grant.

Appendix A
Ancillary Results and Robustness Checks

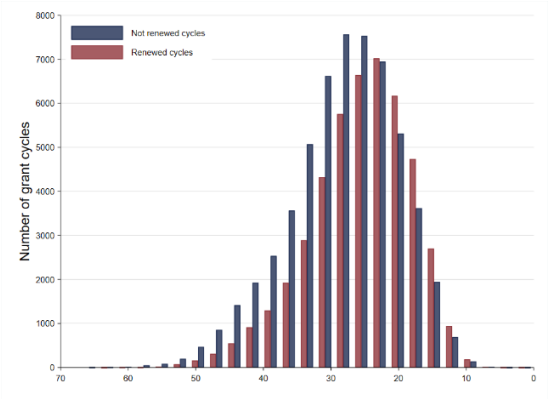
Figure A1. Distribution of measures of novelty

A. Fraction novel MeSH pairs



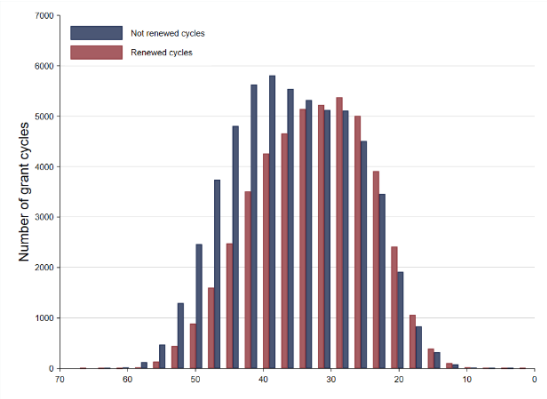
Note: Histogram of the fraction of MeSH pairs among grant cycle publications that were first used by any PubMed publication within the prior 3 years. Unit of analysis is at the investigator-grant-cycle level.

B. MeSH pair age, mean



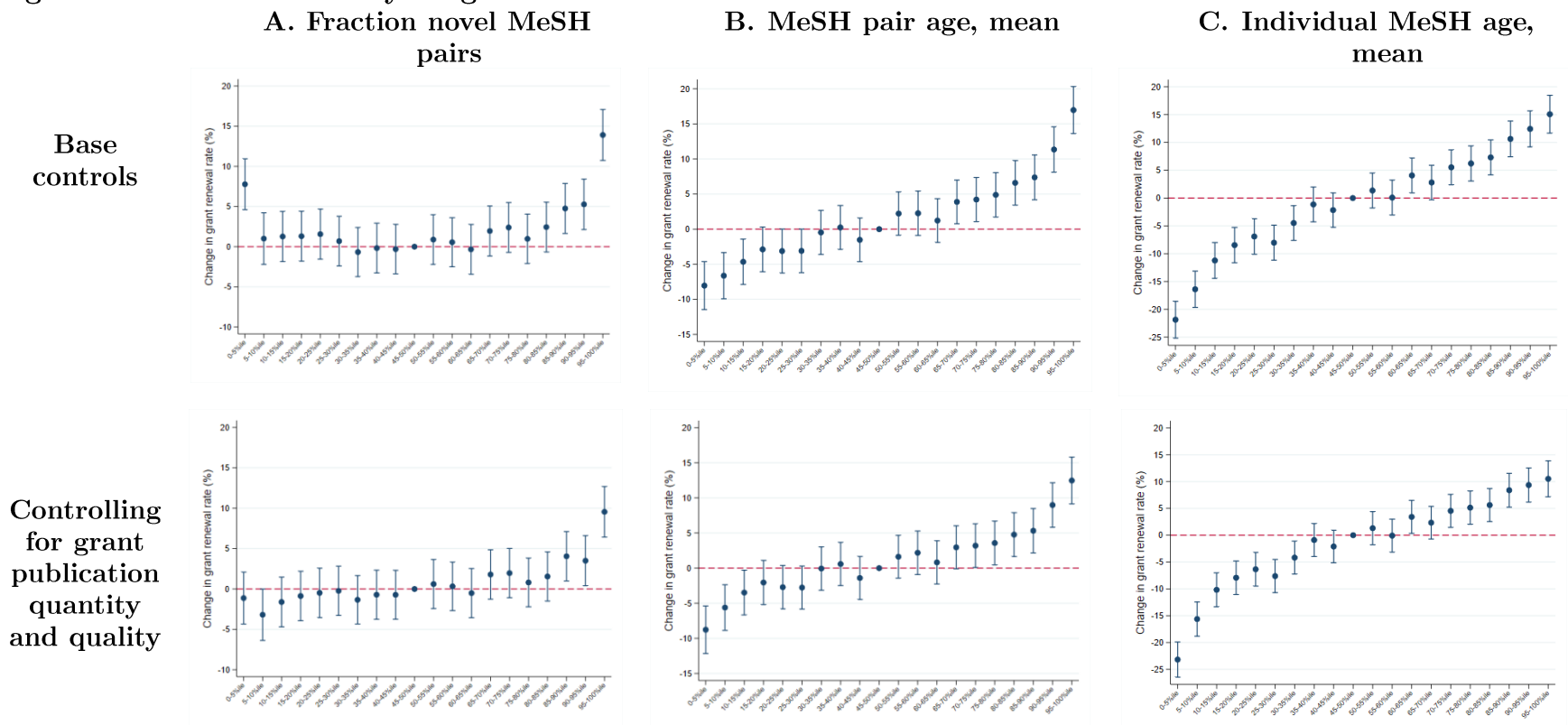
Note: Histogram of the mean MeSH pair age among all grant cycle publication MeSH pairs. Age is defined as the difference between the year of the focal publication and the year the MeSH pair was first used by any publication in PubMed. Unit of analysis is at the investigator-grant-cycle level.

C. MeSH age, mean



Note: Histogram of the mean MeSH age among all grant cycle publication MeSH. Age is defined as the difference between the year of the focal publication and the year the MeSH was first used by any publication in PubMed. Unit of analysis is at the investigator-grant-cycle level.

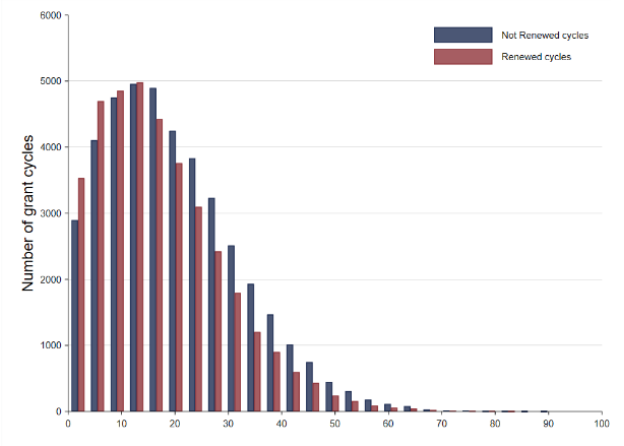
Figure A2. Effect of novelty on grant renewal



Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Row 1 was modeled on Table 2, column 2, while row 2 was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, which is the fraction of MeSH pairs first used in PubMed within the prior 3 years (Panel A), the mean MeSH pair age (Panel B) and mean MeSH age (Panel C) among all MeSH and MeSH pairs for grant cycle publications. For ease of comparison to other figures, age is inverted so that higher percentiles represent grant cycles with younger mean ages. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

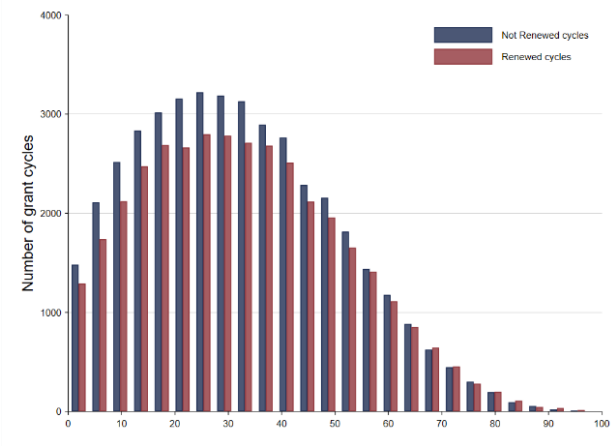
Figure A3. Distribution of alternative measures of extreme tail outcomes

A. Range, max-median citation percentile



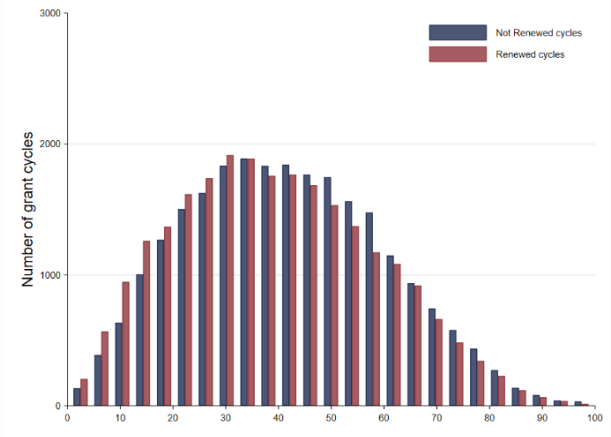
Note: Histogram of the range between the maximum and median vintage-adjusted citation percentile for grant cycle publications. Unit of analysis is at the investigator-grant-cycle level.

B. Range, median-min citation percentile



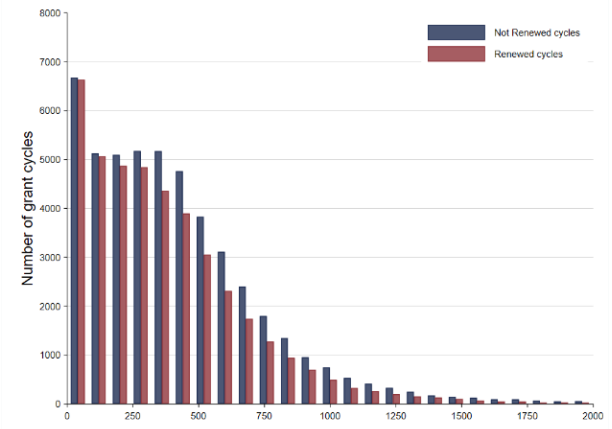
Note: Histogram of the range between the median and minimum vintage-adjusted citation percentile for grant cycle publications. Unit of analysis is at the investigator-grant-cycle level.

C. Range, winsorized max-min citation percentile



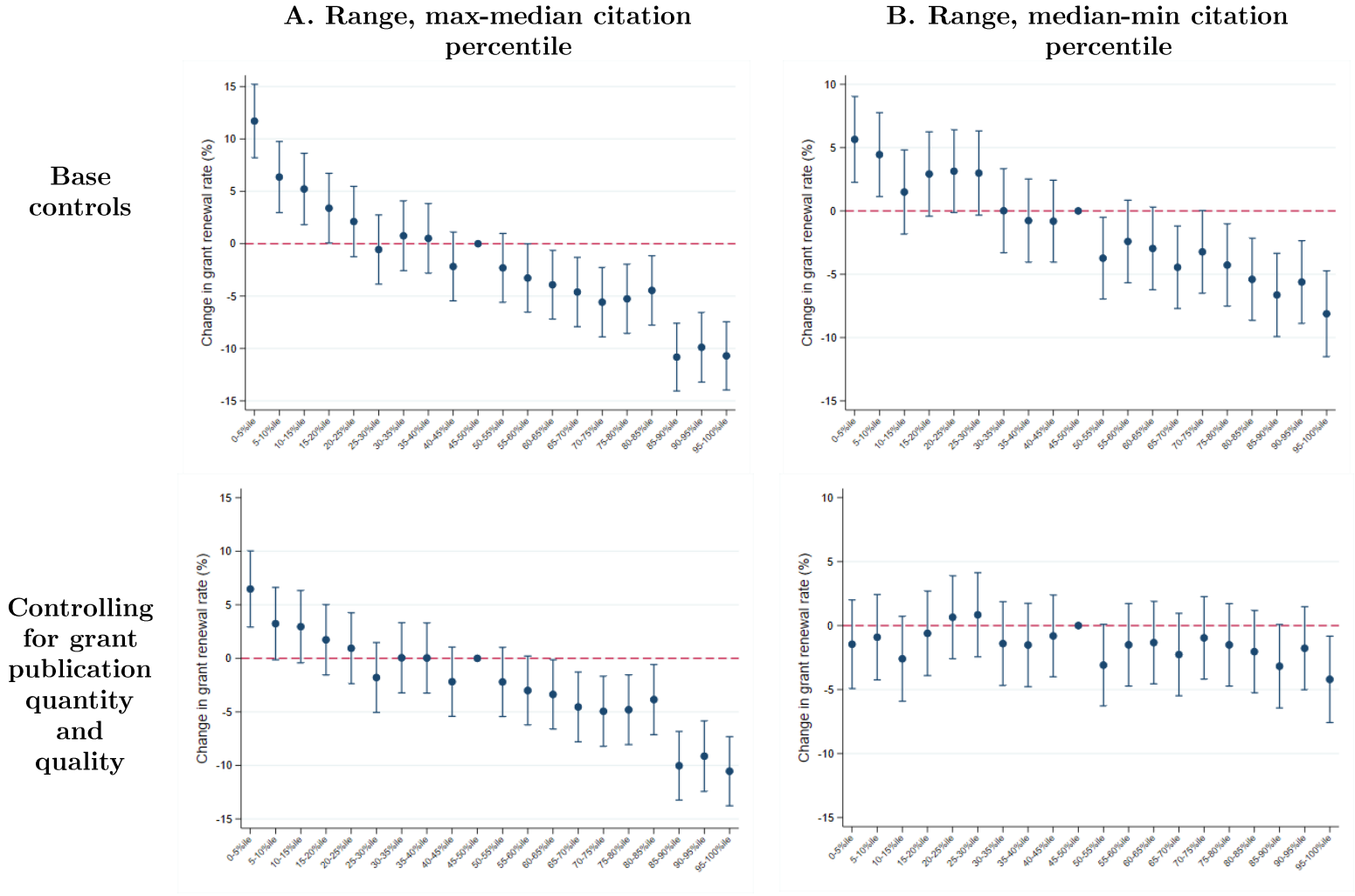
Note: Histogram of the range between the winsorized maximum and minimum vintage-adjusted citation percentile for grant cycle publications. Unit of analysis is at the investigator-grant-cycle level.

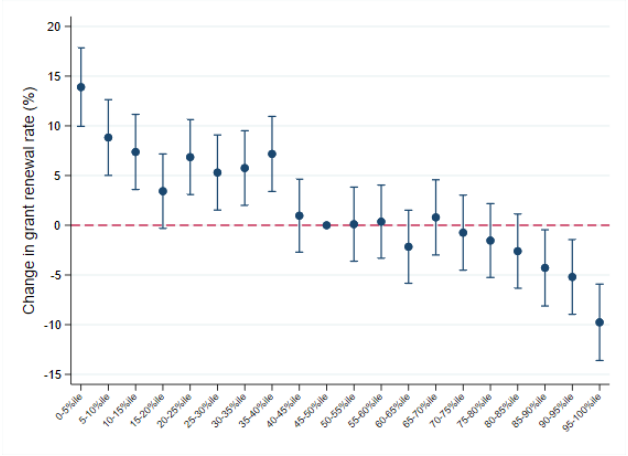
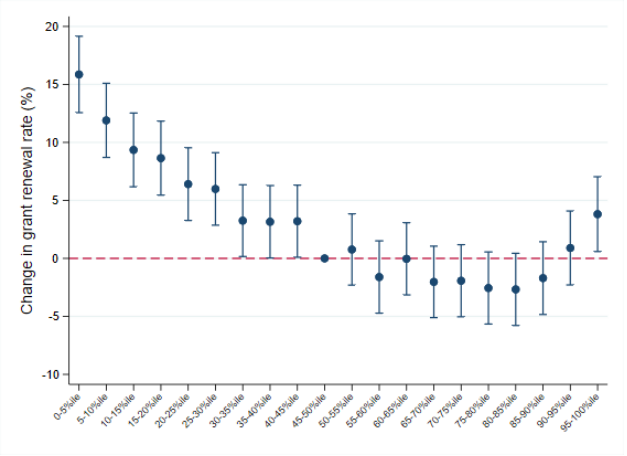
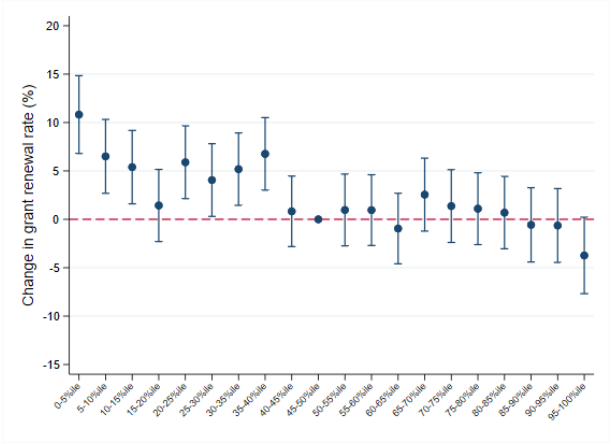
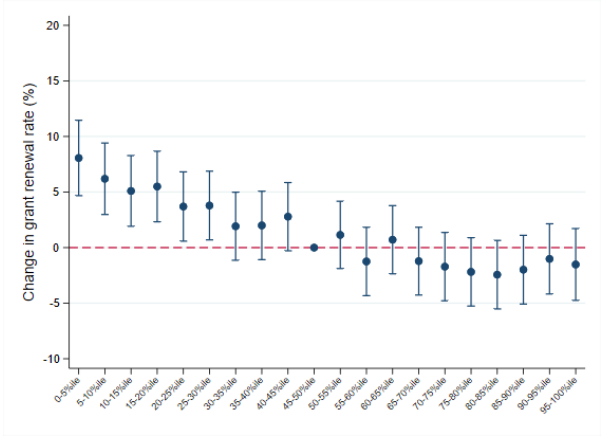
D. Variance, citation percentile



Note: Histogram of the variance in vintage-adjusted citation percentile for grant cycle publications. Unit of analysis is at the investigator-grant-cycle level. Four hundred seventy four outliers with a variance over 2,000 are not shown.

Figure A4. Effect of risk taking on grant renewal: Alternative measures of extreme tail outcomes

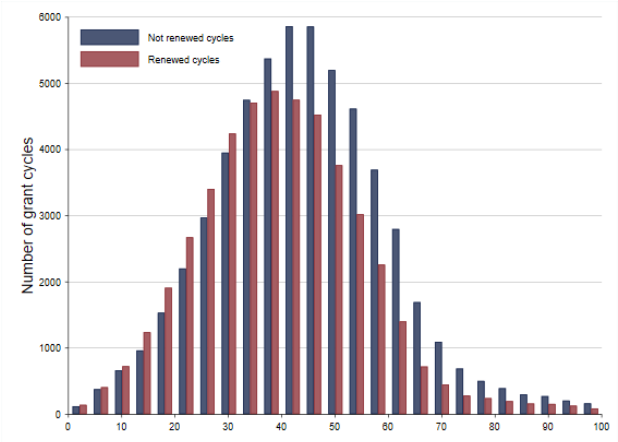


	C. Range, winsorized max-min citation percentile	D. Variance, citation percentile
Base controls		
Controlling for grant publication quantity and quality		

Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Row 1 was modeled on Table 2, column 2, while row 2 was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest: the difference between the maximum and median vintage-adjusted citation percentile (Panel A), difference between the median and minimum vintage-adjusted citation percentile (Panel B), difference between the winsorized maximum and minimum vintage-adjusted citation percentile (Panel C), and variance in vintage-adjusted citation percentile (Panel D) among grant cycle publications. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

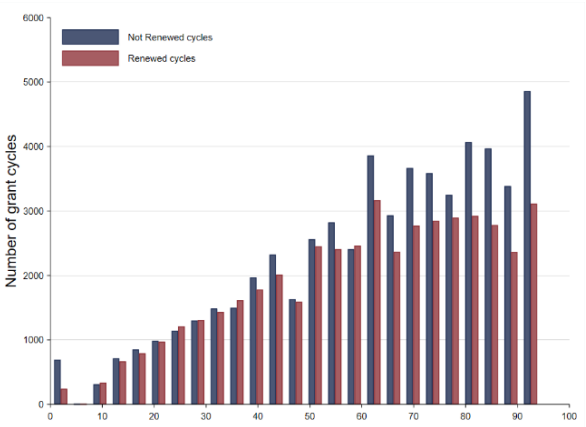
Figure A5. Distribution of alternative measures of disruption

A. Disruption index percentile, mean



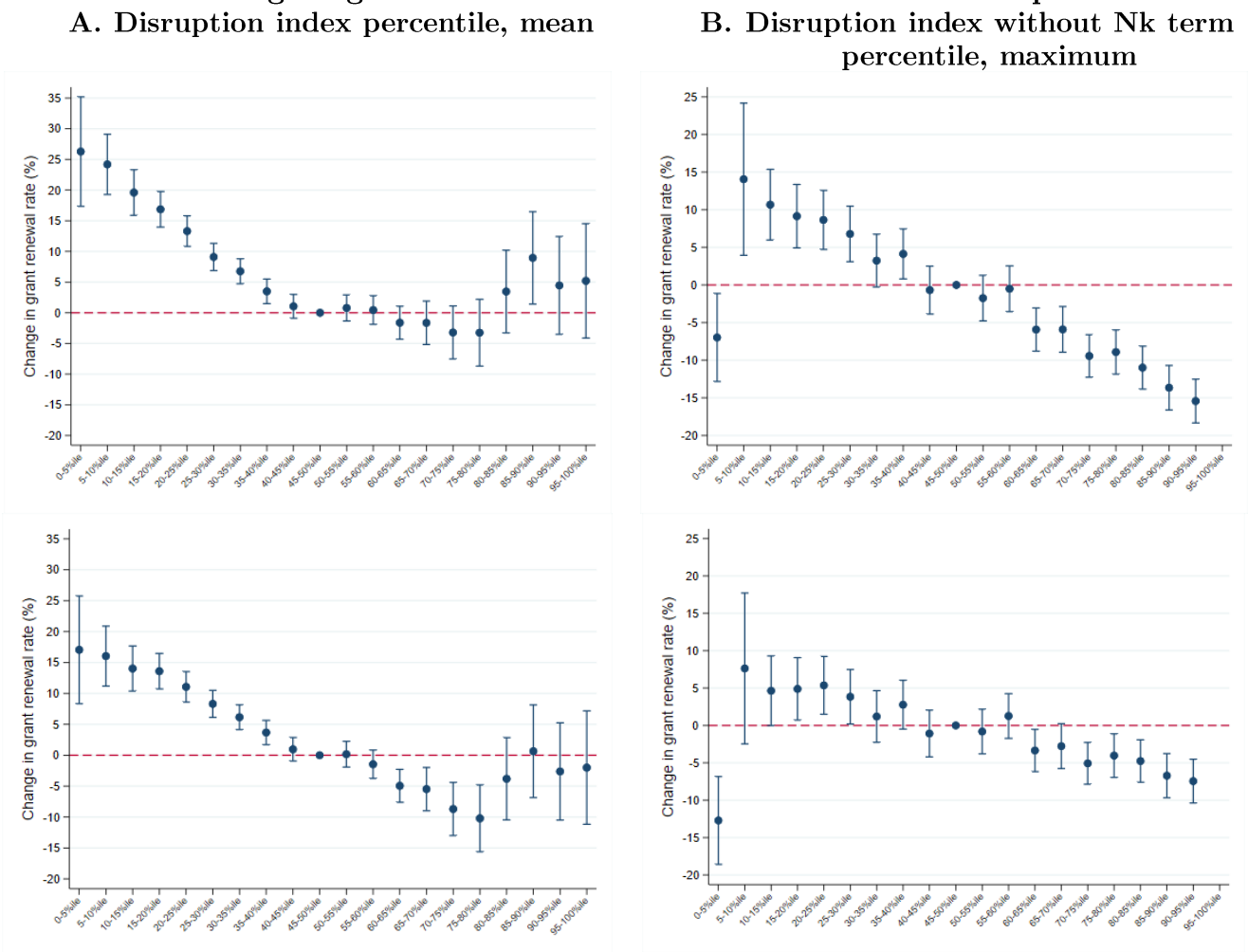
Note: Histogram of the mean disruption index percentile among grant cycle publications. Unit of analysis is at the investigator-grant-cycle level.

B. Disruption index without Nk term percentile, maximum



Note: Histogram of the maximum disruption index percentile among grant cycle publications. This alternative specification of the disruption index follows Bornmann et al. (2020) in excluding the Nk term, which adjusts the disruption index for citations to a focal paper's references that do not cite the focal paper itself. Unit of analysis is at the investigator-grant-cycle level.

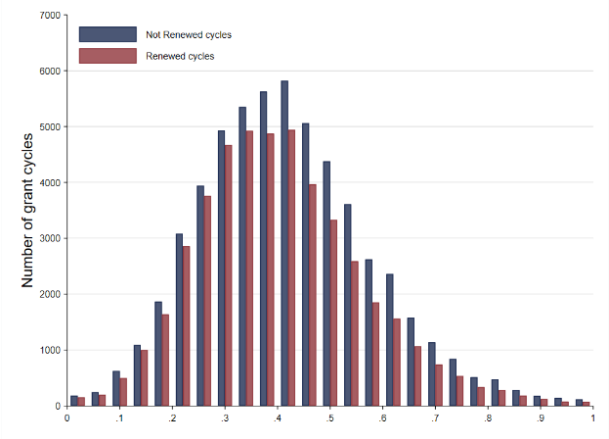
Figure A6. Effect of risk taking on grant renewal: Alternative measures of disruption



Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Row 1 was modeled on Table 2, column 2, while row 2 was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, which is the mean disruption index percentile (Panel A) and maximum disruption index percentile following Bormann et al. (2020) in excluding the Nk term (Panel B) among grant cycle publications. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

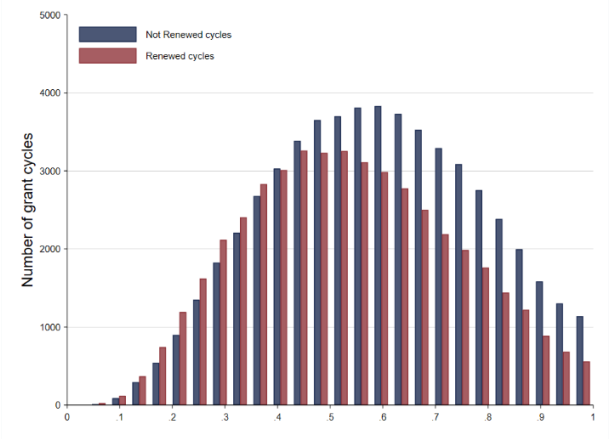
Figure A7. Distribution of alternative measures of pivoting

A. Fraction new MeSH vs. investigator's prior work



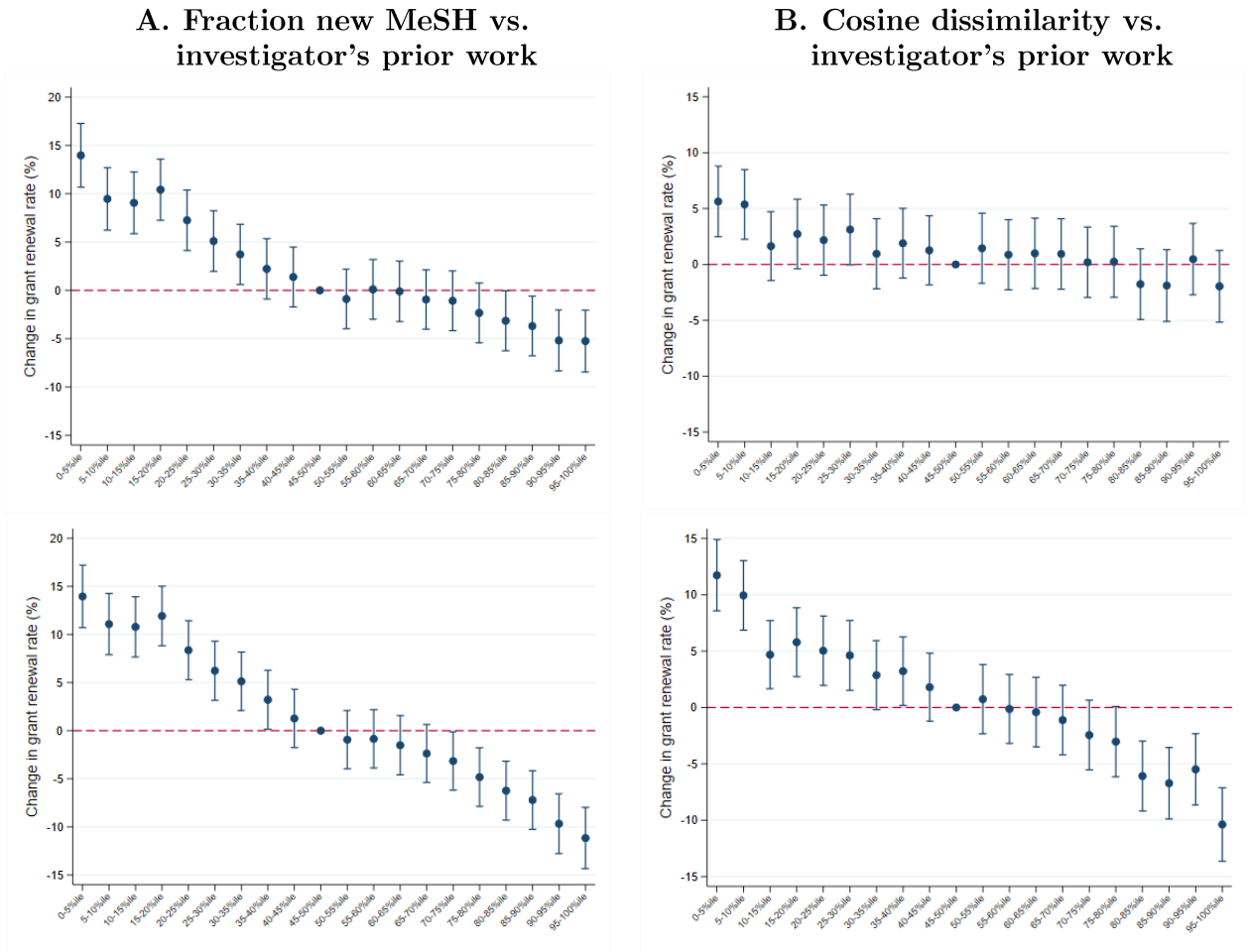
Note: Histogram of the fraction of MeSH terms from grant cycle publications that were not used in the MeSH for the investigator's publications in the 5 years preceding the grant cycle. Unit of analysis is at the investigator-grant-cycle level.

B. Cosine dissimilarity vs. investigator's prior work



Note: Histogram of the cosine dissimilarity between the MeSH of grant cycle publications and MeSH for the investigator's publications in the 5 years preceding the grant. MeSH were weighted using term-frequency-inverse document frequency. Unit of analysis is at the investigator-grant-cycle level.

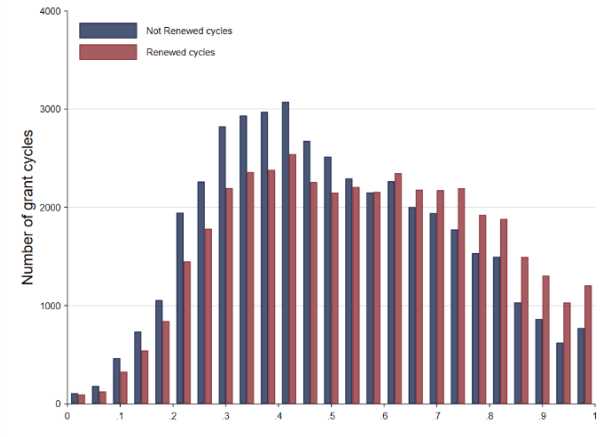
Figure A8. Effect of risk taking on grant renewal: Alternative measures of pivoting



Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Row 1 was modeled on Table 2, column 2, while row 2 was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, which is the fraction of new MeSH (Panel A) and cosine dissimilarity with term frequency-inverse document frequency weighting (Panel B) comparing MeSH from grant cycle publications to those for the investigator in the 5 years preceding the grant cycle. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

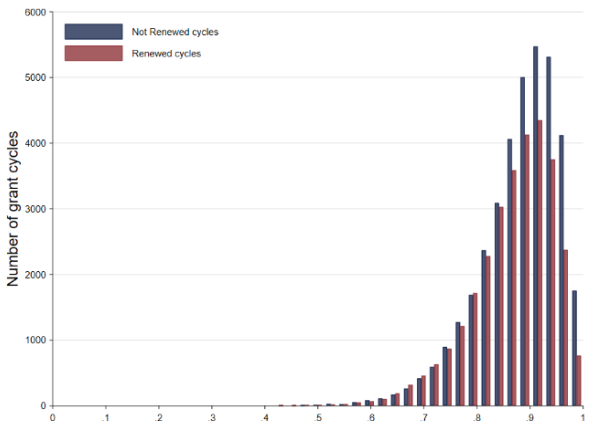
Figure A9. Distribution of alternative measures of standing out

A. Fraction unique MeSH vs. other investigators



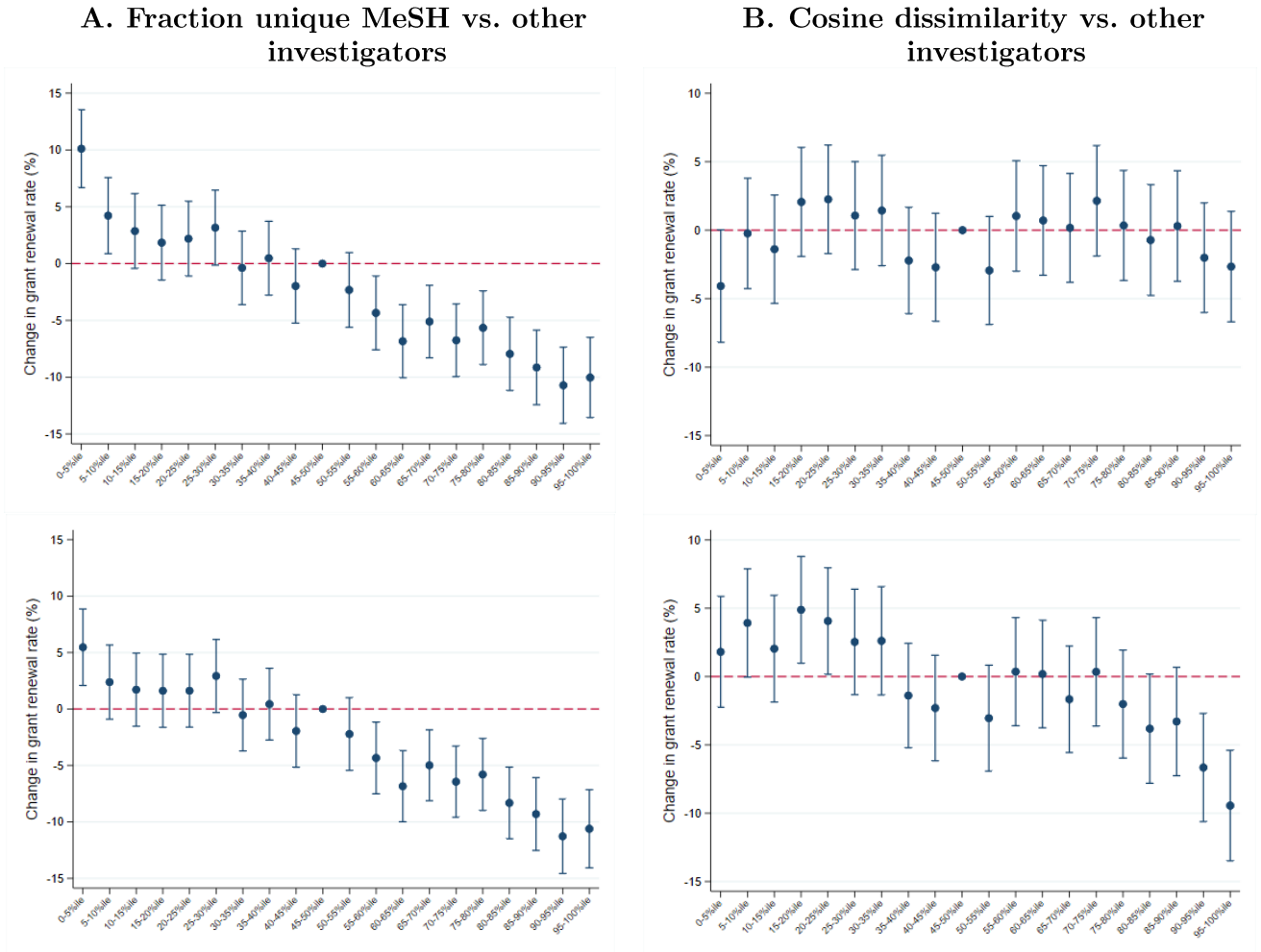
Note: Histogram of the fraction of MeSH terms for grant cycle publications that were not used by any funded grant proposal abstracts in the same NIH IC-study section-year. Unit of analysis is at the investigator-grant-cycle level.

B. Cosine dissimilarity vs. other investigators



Note: Histogram of the cosine dissimilarity between the MeSH for grant cycle publications and all funded grant proposal abstracts in the same NIH IC-study section-year. MeSH were weighted using term-frequency-inverse document frequency. Unit of analysis is at the investigator-grant-cycle level.

Figure A10. Effect of risk taking on grant renewal: Alternative measures of standing out



Note: The plots above present the coefficient estimates (and associated 95% confidence intervals) from a linear probability model where the dependent variable is an indicator variable for if the grant cycle was renewed. The unit of analysis is the investigator-grant-cycle. Row 1 was modeled on Table 2, column 2, while row 2 was modeled on Table 2, column 3. Each regression includes a set of 19 indicator variables for each ventile of the independent variable of interest, which is the fraction of unique MeSH (Panel A) and cosine dissimilarity with term frequency-inverse document frequency weighting (Panel B) comparing MeSH from grant cycle publications to those for funded grant proposal abstracts in the same NIH IC-study section-year. The omitted category is the 45-50%ile. Robust standard errors were used, clustered at the investigator and grant.

Table A1. Effect of novelty on the risk taking grant renewal penalty

A. Novelty above the 75%ile

	Base controls				Grant output controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Novelty >75%ile	0.0204**	0.0117	0.1016**	0.1666**	0.0233**	0.0095	0.1203**	0.1637**
	(0.0064)	(0.0088)	(0.0185)	(0.0293)	(0.0063)	(0.0087)	(0.0182)	(0.0286)
Extreme tail outcomes	-0.0126**				-0.0069**			
	(0.0006)				(0.0007)			
Extreme tail outcomes × Novelty >75%ile	-0.0004				-0.0007			
	(0.0012)				(0.0012)			
Disruption		-0.0153**				-0.0084**		
		(0.0007)				(0.0007)		
Disruption × Novelty >75%ile		0.0010				0.0014		
		(0.0013)				(0.0013)		
Pivoting			-0.1807**				-0.2305**	
			(0.0106)				(0.0106)	
Pivoting × Novelty >75%ile			-0.0882**				-0.1107**	
			(0.0233)				(0.0229)	
Standing out				-0.2052**				-0.1660**
				(0.0186)				(0.0183)
Standing out × Novelty >75%ile				-0.1593**				-0.1580**
				(0.0328)				(0.0321)

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Each estimate includes a risk-taking measure as a continuous variable: extreme tail outcomes scaled so a change of 1 in the regressor reflects a change of 10 in the long-run vintage-adjusted citation percentile spread (columns 1 and 5), disruption scaled so a change of 1 in the regressor reflects a change of 10 in the maximum disruption index percentile (columns 2 and 6), pivoting (columns 3 and 7), and standing out (columns 4 and 8). All estimates include an indicator variable for above the 75%ile in novelty as well as an interaction term between this indicator variable and the risk-taking measure. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

B. Novelty above the 90%ile

	Base controls				Grant output controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Novelty >90%ile	0.0206*	0.0167	0.1327**	0.2259**	0.0246**	0.0131	0.1359**	0.2085**
	(0.0090)	(0.0120)	(0.0284)	(0.0440)	(0.0089)	(0.0118)	(0.0279)	(0.0426)
Extreme tail outcomes	-0.0128**				-0.0072**			
	(0.0006)				(0.0007)			
Extreme tail outcomes × Novelty >90%ile	0.0018				0.0009			
	(0.0018)				(0.0017)			
Disruption		-0.0151**				-0.0083**		
		(0.0006)				(0.0007)		
Disruption × Novelty >90%ile		0.0024				0.0025		
		(0.0019)				(0.0018)		
Pivoting			-0.1855**				-0.2349**	
			(0.0101)				(0.0100)	
Pivoting × Novelty >90%ile			-0.1017**				-0.1136**	
			(0.0347)				(0.0342)	
Standing out				-0.2173**				-0.1777**
				(0.0181)				(0.0178)
Standing out × Novelty >90%ile				-0.2009**				-0.1911**
				(0.0488)				(0.0473)

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Each estimate includes a risk-taking measure as a continuous variable: extreme tail outcomes scaled so a change of 1 in the regressor reflects a change of 10 in the long-run vintage-adjusted citation percentile spread (columns 1 and 5), disruption scaled so a change of 1 in the regressor reflects a change of 10 in the maximum disruption index percentile (columns 2 and 6), pivoting (columns 3 and 7), and standing out (columns 4 and 8). All estimates include an indicator variable for above the 90%ile in novelty as well as an interaction term between this indicator variable and the risk-taking measure. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

C. Novelty above the 95%ile

	Base controls				Grant output controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Novelty >95%ile	0.0323**	0.0333*	0.1700**	0.2692**	0.0371**	0.0298*	0.1599**	0.2371**
	(0.0119)	(0.0152)	(0.0390)	(0.0605)	(0.0117)	(0.0148)	(0.0381)	(0.0585)
Extreme tail outcomes	-0.0127**				-0.0072**			
	(0.0006)				(0.0007)			
Extreme tail outcomes × Novelty >95%ile	0.0028				0.0014			
	(0.0025)				(0.0025)			
Disruption		-0.0149**				-0.0081**		
		(0.0006)				(0.0007)		
Disruption × Novelty >95%ile		0.0022				0.0019		
		(0.0024)				(0.0024)		
Pivoting			-0.1841**				-0.2340**	
			(0.0099)				(0.0099)	
Pivoting × Novelty >95%ile			-0.1221**				-0.1241**	
			(0.0471)				(0.0461)	
Standing out				-0.2204**				-0.1815**
				(0.0179)				(0.0177)
Standing out × Novelty >95%ile				-0.2244**				-0.2036**
				(0.0668)				(0.0647)

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Each estimate includes a risk-taking measure as a continuous variable: extreme tail outcomes scaled so a change of 1 in the regressor reflects a change of 10 in the long-run vintage-adjusted citation percentile spread (columns 1 and 5), disruption scaled so a change of 1 in the regressor reflects a change of 10 in the maximum disruption index percentile (columns 2 and 6), pivoting (columns 3 and 7), and standing out (columns 4 and 8). All estimates include an indicator variable for above the 95%ile in novelty as well as an interaction term between this indicator variable and the risk-taking measure. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table A2. Descriptive statistics: Alternative measures of risk taking and novelty

	Renewed grant cycles					Not renewed grant cycles				
	Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
Range, median-min citation percentile	32.79	31	18.58	0	97	31.98	30	18.29	0	96
Range, max-median citation percentile	17.19	15	11.68	0	84	19.76	18	12.62	0	92
Range, winsorized max-min citation percentile	39.33	38	18.92	1	99	41.65	41	18.66	1	99
Variance, citation percentile	377.65	311	334.37	0	4,579	424.31	354	369.50	0	4,950
Disruption index percentile, mean	39.71	39	15.43	0	100	43.96	44	16.10	0	100
Disruption index without Nk term percentile, max	60.28	63	22.18	0	94	62.55	66	22.57	0	94
Fraction new MeSH vs. investigator′s prior work	0.40	0	0.16	0	1	0.42	0	0.16	0	1
Cosine dissimilarity vs. investigator′s prior work	0.54	1	0.20	0	1	0.58	1	0.20	0	1
Fraction unique MeSH vs. other investigators	0.55	1	0.23	0	1	0.50	0	0.22	0	1
Cosine dissimilarity vs. other investigators	0.87	1	0.08	0	1	0.88	1	0.08	0	1
MeSH pair age, mean	25.58	25	7.48	0	63	28.30	28	8.18	1	67
MeSH age, mean	32.73	32	8.37	0	67	35.68	36	9.03	4	64

Note: Outcomes are measured at the investigator-grant-cycle level. Measures are only defined for those cycles with associated publications (see footnote 7 for details).

Table A3. Heterogeneity by degree of pivoting in proposed grant abstract

	Full sample		Greater overlap between proposal and prior work		Less overlap between proposal and prior work	
	Base controls	Grant output controls	Base controls	Grant output controls	Base controls	Grant output controls
<10%ile, fraction new MeSH pairs	0.053**	0.059**	0.054**	0.057**	0.063**	0.059**
	(0.004)	(0.004)	(0.006)	(0.005)	(0.008)	(0.008)
>90%ile, fraction new MeSH pairs	-0.024**	-0.043**	-0.032**	-0.062**	-0.017**	-0.032**
	(0.004)	(0.004)	(0.010)	(0.010)	(0.005)	(0.005)
Mean of dependent variable	0.455	0.455	0.409	0.409	0.395	0.395
Std. Dev. of dependent variable	0.498	0.498	0.492	0.492	0.489	0.489
Effect bottom %ile group, in s.d. units	0.107	0.119	0.109	0.116	0.129	0.121
Effect top %ile group, in s.d. units	-0.048	-0.086	-0.064	-0.126	-0.035	-0.065
Adjusted R^2	0.4847	0.5039	0.4720	0.4921	0.4734	0.4911
Nb. of investigators	36,539	36,539	21,618	21,618	25,374	25,374
Nb. of investigator-grant-cycles	101,338	101,338	44,229	44,229	43,004	43,004

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Each column reflects the results of a separate regression. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Columns 3 and 4 limit the sample to those investigator-grant-cycles with below median fraction new MeSH terms between the grant proposal abstract and an investigator's publications in the preceding 5 years, while Columns 5 and 6 limit the sample to above the median fraction new MeSH. The independent variables are indictor variables for <10%ile and >90%ile for pivoting, with 10-90%ile as the omitted category. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table A4. Robustness: Investigator name frequency

		Full sample		Below median name frequency		Above median name frequency	
		Base controls	Grant output controls	Base controls	Grant output controls	Base controls	Grant output controls
Extreme tail outcomes							
	<10%ile	0.051**	0.022**	0.051**	0.019**	0.055**	0.025**
		(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.007)
	>90%ile	-0.044**	-0.021**	-0.044**	-0.018**	-0.045**	-0.022**
		(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Disruption							
	<10%ile	0.089**	0.054**	0.109**	0.072**	0.072**	0.037*
		(0.012)	(0.012)	(0.016)	(0.016)	(0.017)	(0.017)
	>90%ile	-0.022**	0.008†	-0.027**	0.004	-0.019**	0.011†
		(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Pivoting							
	<10%ile	0.053**	0.059**	0.050**	0.056**	0.057**	0.062**
		(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
	>90%ile	-0.024**	-0.043**	-0.026**	-0.046**	-0.021**	-0.040**
		(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Standing out							
	<10%ile	0.038**	0.025**	0.043**	0.027**	0.035**	0.024**
		(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
	>90%ile	-0.023**	-0.027**	-0.019**	-0.023**	-0.029**	-0.032**
		(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Each column and independent variable measure reflect the results of a separate regression. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Columns 3 and 4 limit the sample to those investigator-grant-cycles with below median name frequency, calculated relative to the corpus of PubMed, while Columns 5 and 6 limit the sample to above median name frequency. The independent variables are indictor variables for <10%ile and >90%ile for each measure of risk taking, with 10-90%ile as the omitted category. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table A5. Robustness: Reliance on NIH funding

	Full sample		Below median fraction pubs. with NIH support		Above median fraction pubs. with NIH support	
	Base controls	Grant output controls	Base controls	Grant output controls	Base controls	Grant output controls
Extreme tail outcomes						
<10%ile	0.051**	0.022**	0.047**	0.021**	0.059**	0.024**
	(0.004)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
>90%ile	-0.044**	-0.021**	-0.044**	-0.021**	-0.045**	-0.021**
	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)	(0.006)
Disruption						
<10%ile	0.089**	0.054**	0.083**	0.053**	0.101**	0.050**
	(0.012)	(0.012)	(0.015)	(0.015)	(0.019)	(0.019)
>90%ile	-0.022**	0.008†	-0.019**	0.009	-0.025**	0.003
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Pivoting						
<10%ile	0.053**	0.059**	0.047**	0.053**	0.062**	0.065**
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
>90%ile	-0.024**	-0.043**	-0.022**	-0.041**	-0.026**	-0.046**
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)
Standing out						
<10%ile	0.038**	0.025**	0.042**	0.028**	0.035**	0.022**
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
>90%ile	-0.023**	-0.027**	-0.021**	-0.024**	-0.025**	-0.029**
	(0.005)	(0.005)	(0.007)	(0.007)	(0.006)	(0.006)

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Each column and independent variable measure reflect the results of a separate regression. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Columns 3 and 4 limit the sample to those investigator-grant-cycles with below median fraction of career publications in PubMed acknowledging any NIH funding, while Columns 5 and 6 limit the sample to those with above median fraction career publications with any NIH funding. The independent variables are indictor variables for <10%ile and >90%ile for each measure of risk taking, with 10-90%ile as the omitted category. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table A6. Robustness: Time period

	Full sample		Below median grant evaluation year		Above median grant evaluation year	
	Base controls	Grant output controls	Base controls	Grant output controls	Base controls	Grant output controls
Extreme tail outcomes						
<10%ile	0.051**	0.022**	0.064**	0.034**	0.031**	0.013*
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)
>90%ile	-0.044**	-0.021**	-0.050**	-0.014*	-0.033**	-0.022**
	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)	(0.005)
Disruption						
<10%ile	0.089**	0.054**	0.081**	0.054**	0.081**	0.039*
	(0.012)	(0.012)	(0.014)	(0.014)	(0.018)	(0.018)
>90%ile	-0.022**	0.008†	-0.022**	0.014*	-0.018**	0.002
	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)	(0.005)
Pivoting						
<10%ile	0.053**	0.059**	0.061**	0.062**	0.038**	0.046**
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.005)
>90%ile	-0.024**	-0.043**	-0.023**	-0.040**	-0.018**	-0.035**
	(0.004)	(0.004)	(0.006)	(0.006)	(0.005)	(0.005)
Standing out						
<10%ile	0.038**	0.025**	0.023**	0.008	0.052**	0.043**
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
>90%ile	-0.023**	-0.027**	-0.006	-0.011†	-0.045**	-0.048**
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)

Note: Estimates stem from a linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator variable for if the grant cycle was renewed. Each column and independent variable measure reflect the results of a separate regression. Columns with base controls are very similar to Table 2, column 2, while those with grant output controls are very similar to Table 2, column 3. Columns 3 and 4 limit the sample to those investigator-grant-cycles evaluated for renewal in 2003 and earlier, while Columns 5 and 6 limit the sample to 2004 and later. The independent variables are indictor variables for <10%ile and >90%ile for each measure of risk taking, with 10-90%ile as the omitted category. Robust standard errors were used, clustered at the investigator and grant. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Table A7. Robustness: Oster's Delta

	Base controls		Grant output controls	
	Coefficient estimate	Oster's Delta	Coefficient estimate	Oster's Delta
Extreme tail outcomes	-0.0128**	-4.550	-0.0072**	-16.076
	(0.0006)		(0.0007)	
Disruption	-0.0151**	-48.926	-0.0081**	2.675
	(0.0006)		(0.0007)	
Pivoting	-0.1727**	-23.281	-0.2262**	-9.879
	(0.0097)		(0.0097)	
Standing out	-0.2171**	-0.695	-0.1782**	-0.592
	(0.0178)		(0.0175)	

Note: Each cell in columns 1 and 3 stems from a separate linear probability model where the unit of analysis is the investigator-grant-cycle and the dependent variable is an indicator for if the grant cycle was renewed. This is very similar to columns 2 and 3 of Table 2A, 2B, 2C and 2D for the extreme tail outcomes (row 1), disruption (row 2), pivoting (row 3) and standing out (row 4), respectively, except a single continuous regressor is used for each measure of risk taking. Robust standard errors were used, clustered at the investigator and grant. Columns 2 and 4 report the corresponding δ parameter from Oster (2019), which can be interpreted as the degree of selection on unobservables relative to observables necessary to explain away the effect of risk taking on grant renewal. We follow Oster's recommendation of setting $R^{max} = 1.3 \times R^2$. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.