Information, Bias, and Efficiency in Expert Evaluation: Evidence from the NIH *

Danielle Li
Northwestern University†

First draft: Aug 1, 2011
This draft: Sept 25, 2012

Abstract

Experts may have more information about the potential of projects in their area, but may also be biased. This paper develops a framework for separately identifying the effects of bias and information on expert evaluation and applies it in the context of peer review at the National Institutes of Health (NIH). I find that while reviewers are biased in favor of applications from their own subfield, they are also more informed about their quality. On net, the benefits of information tend to dominate, indicating that policies designed to reduce conflicts of interest may also reduce the quality of funding decisions.

*I am very grateful to Pierre Azoulay, Michael Greenstone, and especially David Autor, for detailed feedback on this project. I also thank Jason Abaluck, Daron Acemoglu, Leila Agha, Josh Angrist, David Berger, David Chan, Esther Duflo, Brigham Frandsen, Richard Freeman, Bob Gibbons, Nathan Hendren, Xiao Yu May Wang, Ziad Obermeyer, Amanda Pallais, Chris Palmer, Michael Powell, Pian Shu, Scott Stern, Heidi Williams, and numerous participants at Berkeley Haas, Boston University, Columbia GSB, Dartmouth, Harvard Business School, Harvard Kennedy School, the London School of Economics, MIT, the NBER Productivity Lunch, and Northwestern Kellogg, for helpful comments and suggestions. I am grateful to George Chacko, Raviv Murciano-Goroff, Joshua Reyes, and James Vines for assistance with data. All errors are my own.

†danielle-li@kellogg.northwestern.edu

1 Introduction

When decisions are complex and technical, it is natural to turn to experts for advice. This is true in a wide variety of settings: lawmakers, corporate boards, venture capital groups, and regulatory bodies, for instance, all seek input from industry insiders. But how much should we trust their advice? While experts may have valuable insights about a project's potential, they may also have preferences that compromise their objectivity. As a result, attempts to limit bias by reducing conflicts of interest can come at the direct cost of reducing information.

Understanding how experts shape investment decisions is particularly crucial in the innovative sector, where the payoffs to specific investments are notoriously uncertain (Arrow, 1962). Because ideas are so difficult to assess and because their value may take years or even decades to be realized, there is both greater value placed on expertise and greater scope for obfuscation.[1]

I develop a framework for separately identifying the effect of bias on decision-making from that of information, and provide the first empirical estimate of the efficiency tradeoff between bias and information in expert evaluation. I do so in a context that is extremely important for medical innovation: grant funding at the National Institutes of Health (NIH). With an annual budget of 30 billion dollars, the NIH is the world's largest funder of biomedical research, spending nearly half as much on basic and applied science as the entire US pharmaceutical industry combined.[2] NIH-sponsored research plays a role in the development of over half of all FDA approved drugs, including path-breaking treatments such as Gleevec, the first drug therapy to selectively target cancerous cells, and Lipitor, one of the most prescribed drugs in America.[3]

The majority of NIH funds are allocated via a non-blind review process in which individual scientists propose research projects that are then evaluated by committees of their peers. Peer review is the key institution responsible for consolidating thousands of investigator-initiated submissions into a concrete, publicly funded research agenda. The success of this system, then, depends in large part on the ability of reviewers to identify and fund the most promising ideas in their areas of specialty.

This paper examines the role that potentially biased reviewers play in NIH peer review. Reviewers may be more qualified to evaluate the merit of proposals in their own area of expertise, but they may also have personal preferences that distort their assessments. I formalize this intuition with a model of strategic communication in review meetings. In this model, reviewers are biased, meaning that they receive an additional payoff from funding related applicants, independent of that applicant's quality. Reviewers, however, may also improve the quality of funding decisions by introducing better information about the quality of proposals from these related applicants. In equilibrium, a grant proposal's likelihood of being funded can be expressed as a function of its

[1] See, for example, Aghion and Tirole (1994) and David, Mowery, and Steinmueller (1992).

[2] In 2006, pharmaceutical companies spent close to 50 billion dollars on R&D. CBO "Research and Development in the Pharmaceuticals Industry" (2006).

[3] Over two-thirds of FDA priority review drugs cite NIH-funded research. See Sampat and Lichtenberg (2011).

quality, the relatedness of the applicant to the committee, and their interaction. The effect of reviewer bias on funding decisions comes through the level effect of relatedness, while the effect of better information comes through the interaction effect.

The intuition behind this result is simple and underlies my empirical work: if committees use reviewer-applicant relationships to make inferences about quality, then the effect of being related to a reviewer should differ for high- and low-quality applicants. In particular, high-quality applicants should benefit from being related to reviewers who can more accurately observe their quality while low-quality applicants should be hurt. Reviewers are biased, on the other hand, if they are systematically more (or less) likely to fund related applicants regardless of quality.

Peer review at the NIH presents a rare opportunity to get empirical traction on these issues. To do so, I have assembled a new, comprehensive dataset linking almost 100,000 NIH grant applications to the committees in which they were evaluated. I observe many characteristics of the application, including the application's final score, the name of the applicant, demographic information, grant history, and publication history. For each review meeting, I observe the names of all reviewers who attend and the capacity in which they serve. Using names of applicants and reviewers, I create measures of a reviewer's familiarity with an applicant, as measured by whether the reviewer has cited him in the past.

In order to separately identify bias and information, I need two empirical ingredients: 1) a measure of the quality of all grant applications; and 2) exogenous variation in relatedness. But how does one measure the quality of an application that is not funded? The key in this setting is that large NIH grants require applicants to provide very substantial preliminary results. As a result, researchers often publish the research outlined in a grant proposal even if the application goes unfunded.

I construct my measure of application quality using a novel text-matching approach that links grant application titles with the titles and abstracts of semantically related publications by the same applicant (see Section 5 for details). In addition to restricting to publications that are on the same topic as a grant, I also restrict to articles published so soon after grant review that they are unlikely to be directly affected by any grant funds (See Section 5.1 and Appendix C for discussion and robustness tests.) This approach allows me to apply the same algorithm to identify publications related to both funded unfunded applications in a way that standard publication-grant acknowledgement data cannot.

Finally, remaining measurement error in application quality can still affect my estimates of bias. This could happen if, for example, prominent scientists are more likely to be related to reviewers and their applications are more likely to be funded because of bias. This bias will be difficult to identify because the same bias that leads reviewers to favor prominent scientists in terms of funding may also lead journal editors and other scientists to favor them in terms of publications and citations.[4] In this example, using citations as a measure of quality, even accounting for any direct effect of funding, would still lead me to me to underestimate bias in grant allocation.

[4]This phenomenon is known as the "Matthew Effect." See Merton, 1986 and Azoulay, Stuart, and Wang, 2011.

I address these and other related concerns by exploiting the institutional structure of review committees to generate exogenous variation in relatedness. In particular, the NIH review committees that I study consist of two types of members, "permanent" and "temporary," who have similar qualifications as scientists but substantially different levels of influence in the committee.[5] Instead of comparing applicants related to reviewers with those who are not—these scientists may differ in their unobserved quality or prominence—I compare scientists who are related to the same *total* number of reviewers, but who differ in their number of related *permanent* reviewers. I thus identify the causal effect of relationships on committee decisions under the assumption that, conditional on a large set of observables, the quality of applications from applicants related to the same total number of reviewers is on average similar. Section 5.2 provides evidence for this assumption.

Together, my measures of quality and my use of exogenous variation in relatedness allow me to 1) estimate the effect of being related to a reviewer on an applicant's scores or likelihood of being funded; 2) assess the role of related reviewers both in terms of how they may bias or inform NIH funding decisions; and finally 3) quantify the efficiency consequences of relationships in terms of the quality of research that the NIH supports.

My paper has three primary findings. First, I show that, holding quality constant, every additional permanent member an applicant is related to increases her chances of being funded by 2.9 percent, the equivalent of a one-fifth standard deviation increase in application quality. Second, I show that reviewers shape committee decisions by both increasing bias and improving information. In particular, while bias increases the average likelihood that related applicants are funded, the expertise that reviewers have about those applicants improves the ability to committees to identify high-quality research. I find that the correlation between scores and funding outcomes is 30 percent higher for applicants related to permanent members than it is for those who are related to no permanent reviewers but to the same number of total reviewers. Finally, on net, I show that the gains associated with review by potentially biased experts dominate the losses. Treating related applicants as if they were unrelated—thereby eliminating both bias and information—would reduce the quality of the NIH-supported research portfolio by two to three percent, as measured by future citations and publications. In addition to quantifying the role that bias and information play on average, I also document substantial and persistent variation in how well grant review committees perform. In particular, I show that some of this variation is attributable to differences in how well committees make use of biased experts.

My empirical work is particularly relevant for innovation policy. While large literatures investigate the inputs and outputs of innovative investments more generally, little is known about how—and how successfully—organizations make research investments.[6] Different organizations, moreover, allocate funds in different ways. For example, NIH's reliance on peer review of individual grants stands in contrast to major European funding agencies, which often support large

[5] "Permanent" members are not actually permanent; they serve four-year terms. See Section 5.2 for a discussion of permanent versus temporary reviewers.

[6] See Acemoglu, 2008; Kremer and Williams, 2010; Griliches, 1992; and Cockburn and Henderson, 2000 for surveys. One recent exception is Hegde (2009), which considers the political economy of NIH congressional appropriations.

groups of scientists and guarantee their salary. Understanding the strengths and weaknesses of these models is of particular importance because, by making investments in specific people, labs, and ideas, funding not only affects near-term scientific output but may also shape the allocation of future research attention and resources.

The tradeoff between information and bias is also important in many empirical settings outside of innovation: social ties in the workplace may discipline employees or help them find better jobs but may also lead to nepotism; academics are more informed about the quality of their colleagues but may show bias when making promotion and editorial decisions. A large literature including Bewley (1999), Bayer, Ross, and Topa (2008), and Bandiera, Barankay, and Rasul (2009) examines the role that networks play in the labor market and a smaller literature studies the role of connections in academic promotion and publishing (Bagues and Zinovyeva, 2012 and Brogaard, Engleberg, and Parsons, 2012, respectively). A first-order challenge in these literatures is that it is difficult to attribute differences in treatment to bias or better information if one does not observe the quality of workers who are not hired or academics who are not published or promoted. This paper contributes by studying these issues in a new empirical context where these challenges can be better overcome.

The remainder of this paper proceeds as follows. In the next section, I discuss the details of NIH grant review. I discuss my conceptual and statistical frameworks in Sections 3 and 4, respectively. Section 5 explains how I construct my dataset and variables in order to identify the role of bias and information. Main results are presented in Section 6. Section 7 discusses implications for efficiency, and the final section concludes.

2 Institutional Context

Each year, thousands of scientists travel to Bethesda, Maryland where they read approximately 20,000 grant applications and allocate over 20 billion dollars in federal grant funding. During this process, more than 80 percent of applicants are rejected even though, for the vast majority of biomedical researchers, winning and renewing NIH grants is crucial for being an independent investigator, maintaining a lab, earning tenure, and paying salaries (Stephan, 2012).

The largest and most established of these grant mechanisms is the R01, a project-based renewable research grant that constitutes half of all NIH grant spending and is the primary funding source for most academic biomedical labs in the United States. There are currently 27,000 outstanding awards, with 4,000 new projects approved each year. The average size of each award is 1.7 million dollars spread over 3 to five years.

At the NIH, applications are assigned to a review committee, called a "study section," for scoring and to an Institute or Center (IC) for funding. Study sections assess the scientific merit of applications by assigning them a "priority score," which, during the period my data come from, ranged from 1.0 for the best application to 5.0 for the worst, in increments of 0.1. Up to three reviewers read the application and present their initial scores. All study section members then discuss and anonymously vote on the application using the scores of initial reviewers as a guide.

The final score is the average of all member scores. This priority score is then converted into a percentile from 1 to 99, where a percentile reflects the percentage of applications from the same study section and reviewed in the same year that received a better priority score. However, for ease of exposition and intuition, I report percentiles to mean the percentage of applications that are worse, so that higher percentiles are better. For more details, see Gerin (2006).

Once an application has been scored, it is funded in order of score by the IC to which it was assigned, until that IC's budget is exhausted. The lowest percentile score that is funded is known as the payline. A grant's score affects its chances of being funded, but not its actual funding amount; NIH will choose to fund one large grant instead of two or three smaller grants as long as the larger grant has a better score, even if it is only marginally better. Scores are never made public.

The bulk of R01 applications are assigned to one of about 180 "chartered" study sections.[7] These are standing review committees organized around a particular theme, for instance "Cellular Signaling and Regulatory Systems" or "Clinical Neuroplasticity and Neurotransmitters." My analysis focuses on these committees. Chartered study sections meet three times a year in accordance with NIH's three annual funding cycles. During each meeting, they review, on average, 40 to 80 grant applications. Chartered study sections are typically composed of 15 to 30 "permanent" members who are elected to serve four-year terms and 10 to 20 "temporary" reviewers who are called in as needed. The division of committees into permanent and temporary members plays an important role in my identification strategy and I discuss this in greater detail in Section 5.2.

3 How do Relationships Impact Funding Decisions? Conceptual Framework

The following model of decision-making defines what I mean by reviewer bias and by reviewer expertise. It then illustrates how the biases and expertise of an individual reviewer can affect grant allocation through strategic communication. The equilibrium of this model will then be used to motivate my empirical strategy, discussed in Section 4. In this model, committees want to fund the best grant applications, but must rely on the recommendation of a reviewer who is potentially biased.

Grant applications have some true quality Q^* that is unobserved by the committee, but which can be observed with varying noise by the reviewer. A reviewer is either related or unrelated. A related reviewer sees a signal $Q_R = Q^* + \varepsilon_R$ about the quality of the grant and an unrelated reviewer sees the signal $Q_{UR} = Q^* + \varepsilon_{UR}$. I assume that $\text{Var}(\varepsilon_{UR}) > \text{Var}(\varepsilon_R)$, meaning that a related reviewer is more informed about the true quality of the grant. A related reviewer, however, may be biased: if the grant is funded, he receives a payoff $P^R = Q^* + B$, where B is known. Without loss of generality, I assume that $B > 0$. Neither the committee nor the unrelated reviewer are biased; they receive payoffs of $P^C = Q^*$ and $P^{UR} = Q^*$, respectively. If the grant goes unfunded,

[7] The NIH restructured chartered study sections during my sample period and my data include observations from 250 distinct chartered study sections. These changes do not affect my estimation because I use within meeting variation only.

all parties receive a common outside option U. The committee can observe whether a reviewer is related or unrelated. I assume that the committee acts as a single unit.

The timing works as follows:

1. Nature draws true quality Q^* and the signals Q_R and Q_{UR}.

2. The reviewer, knowing her posterior, makes a costless and unverifiable recommendation $M \in \mathbf{M} = \{M_1, \ldots, M_K\}$ to the committee.

3. The committee observes M and takes a decision $D \in \{0, 1\}$ of whether or not to fund the grant.

4. True quality is revealed and the reviewer and committee both receive their payoffs.

Proposition 3.1 *The Perfect Bayesian equilibria of this game are given by:*
Case 1: If $R = 0$, then all informative equilibria are payoff-equivalent to a full-revelation equilibrium in which:

1. *The reviewer truthfully reports her posterior $E(Q^*|Q^* + \varepsilon_{UR})$.*

2. *The committee funds the grant if $E(Q^*|Q^* + \varepsilon_{UR}) > U$.*

Case 2: If $R = 1$ then:
For $E(Q^*|E(Q^*|Q^* + \varepsilon_R) > U - B) > U$, the unique informative equilibrium is partially-revealing:

1. *With probability one, the reviewer sends a signal Y if $E(Q^*|Q^* + \varepsilon_R) > U - B$ and N otherwise.*[8]

2. *The committee funds the grant if and only if it receives the signal Y.*

In all cases where an informative equilibrium exists, there also exist uninformative equilibria where the grant is never funded.
For $E(Q^*|E(Q^*|Q^* + \varepsilon_R) > U - B) < U$, only uninformative equilibria exist and the grant is never funded.

Proof: See Appendix 3.

Reviewers in this equilibrium signal according to their preferences but, as in Crawford and Sobel (1982), information is distorted because the committee is unable to distinguish, in some cases, a situation when an application reviewed by a related reviewer should be funded (e.g. when $E(Q^*|Q^* + \varepsilon) > U$) from one in which it should not (e.g. when $U > E(Q^*|Q^* + \varepsilon) > U - B$). In order for an informative equilibrium to exist, committees must believe that enough information about the true quality of the grant is communicated in spite of the reviewer's bias.

[8] Although the type space of the messages is not restricted to be binary, all informative equilibria will be payoff-equivalent to one in which they are.

I will focus on the informative equilibrium both in cases when $R = 0$ and in cases when $R = 1$.[9] The equilibrium message strategy is given by:

$$M(Q) = \begin{cases} Y & \text{if } E(Q^*|Q^* + \varepsilon_{UR}) > U \text{ and } R = 0 \\ Y & \text{if } E(Q^*|Q^* + \varepsilon_R) > U - B \text{ and } R = 1 \\ N & \text{otherwise} \end{cases}$$

and the equilibrium decision strategy is given by:

$$D(M) = \begin{cases} Y & \text{if } M = Y \\ N & \text{otherwise} \end{cases}$$

The equilibrium decision rule can be more succinctly expressed as:

$$D = \underbrace{\mathbb{I}(E(Q^*|Q^* + \varepsilon_{UR}) > U)}_{\text{baseline for unrelated}} + \underbrace{[\mathbb{I}(U > E(Q^*|Q^* + \varepsilon_R) > U - B)]}_{\text{bias for related (+)}} R \quad (1)$$

$$+ \underbrace{[\mathbb{I}(E(Q^*|Q^* + \varepsilon_R) > U) - \mathbb{I}(E(Q^*|Q^* + \varepsilon_{UR}) > U)]}_{\text{additional information for related (+/-)}} R \quad (2)$$

In this model, committees listen to the advice of related reviewers even if they are biased because committees value expertise. Equation (2) shows that committees have some baseline performance that is captured by how well unrelated reviewers assess the quality of a grant. Advice from related reviewers can improve committee decisions because it increases the chances that a qualified related applicant is funded ahead of an unqualified related applicant. Related reviewers, however, can worsen committee performance by increasing the probability that an unqualified related applicant is funded ahead of a qualified unrelated applicant. The net effect of related reviewers on the quality of decisions is thus ambiguous.

Many popular critiques of NIH peer review assume that differences in funding likelihood among applicants with the same quality must be due to bias (see Ginther et. al., 2011). Equation (5) shows, however, that this need not be the case. In particular, the difference in expected funding likelihood between applicants with the same quality Q^* but different relatedness R is given by:

$$\begin{aligned} E[D|Q^*, R = 1] - E[D|Q^*, R = 0] &= \Pr(U > E(Q^*|Q^* + \varepsilon_R) > U - B) \\ &\quad + \Pr(E(Q^*|Q^* + \varepsilon_R) > U) - \Pr(E(Q^*|Q^* + \varepsilon_{UR}) > U) \end{aligned}$$

This expression will be non-zero if reviewers are biased ($B \neq 0$). Yet, even without bias, this term need not be zero. The intuition is simple: a committee may fund a related applicant ahead of an equally qualified unrelated applicant simply because they know the related applicant and

[9] If the equilibrium were not informative, then advice from related reviewers would not be taken; I would find no effect of bias and perhaps a lower correlation between scores and quality for applications reviewed by related reviewers. My results are not consistent with a non-informative equilibrium.

know that she is high-quality; they may not know that the unrelated applicant is just as qualified. Here, the difference in funding likelihood between similar applicants is attributable to differences in information, not to bias. As such, distinguishing between these explanations is important because they have different implications for whether relatedness enhances the quality of peer review.

4 How do Relationships Impact Funding Decisions? Statistical Framework

Next, I assume that the committee decisions I observe are generated by the equilibrium decision rule described by Equation (2) in Section 3. Under the assumption that ε is uniform ($\varepsilon_{UR} \sim U[-a_{UR}, a_{UR}], \varepsilon_R \sim U[-a_R, a_R]$) and that $E(Q^*|Q^*+\varepsilon)$ is approximated by $\lambda(Q^*+\varepsilon)$ for any constant λ, the conditional mean of D is given by:[10]

$$\begin{aligned}
E[D|Q^*,R] &= \Pr(\lambda_{UR}(Q^*+\varepsilon_{UR}) > U) + \Pr(U > \lambda_R(Q^*+\varepsilon_R) > U-B)R && (3)\\
&\quad + \left[\Pr(\lambda_R(Q^*+\varepsilon_R) > U) - \Pr(\lambda_{UR}(Q^*+\varepsilon_{UR}) > U)\right] R && (4)\\
&= \frac{1}{2a_{UR}}\left[a_{UR} - U/\lambda_{UR} + Q^*\right] + \frac{B}{2a_R\lambda_R}R \\
&\quad + \left(\frac{1}{2a_R}\left[a_R - U/\lambda_R + Q^*\right] - \frac{1}{2a_{UR}}\left[a_{UR} - U/\lambda_{UR} + Q^*\right]\right) R \\
&= \frac{1}{2} + \underbrace{\frac{1}{2a_{UR}}}_{\text{Quality corr.}} Q^* + \underbrace{\frac{B}{2a_R\lambda_R}}_{\text{Bias term}} R + \underbrace{\left[\frac{1}{2a_R} - \frac{1}{2a_{UR}}\right]}_{\text{Add. corr. for related}} RQ^* \\
&\quad - \frac{U}{2a_{UR}\lambda_{UR}} + \left[\frac{1}{2a_{UR}\lambda_{UR}} - \frac{1}{2a_R\lambda_R}\right] RU && (5)
\end{aligned}$$

Equation (5) says that the effect of relatedness coming from bias and from information can be separately identified. First, the coefficient on R tests for reviewer bias: it is non-zero if and only if $B \neq 0$. Second, the coefficient on RQ tests for information. To see this, notice that the coefficient on Q^* captures how well the committee identifies high-quality research among unrelated applicants. Specifically, a high coefficient on Q^* means committees are more likely to fund high-quality unrelated applicants over low-quality unrelated applicants. The coefficient on RQ^*, meanwhile, captures the additional correlation between quality and likelihood of funding for related applicants. A high coefficient on RQ means that a committee is more sensitive to increases in the quality of related applicants than it is to increases in the quality of unrelated applicants. This effect of information is larger when the difference between the precisions of related and unrelated beliefs, $\frac{1}{2a_R} - \frac{1}{2a_{UR}}$ is greater.

The intuition for separately identifying information from bias is simple: if reviewers make different funding decisions for related applicants because they have more information about their quality, then the effect of relatedness on funding likelihood should differ for high- and low-quality

[10] See the Appendix C for discussion and empirical tests related to these assumptions.

applicants. If committees were influenced by the bias of related reviewers, then related applicants should be more likely to be funded regardless of quality.

Finally, the terms U and RU control for the degree of selectivity; when the cutoff U is high, there is little correlation between funding and quality—even in the absence of bias or differential information—because it is difficult to distinguish quality when all funded applicants are very high-quality. In the model, there is no limit to the number of grants that are funded so that relationships can also affect the generosity of committees. The RU term ensures that relationships are not credited for changing the correlation between funding and quality simply by lowering the threshold at which grants are funded.

The exact form of Equation (5) depends on linearity assumptions but, in practice, my results are robust to allowing for non-linear effects of relatedness and quality measures. These results are discussed in Appendix C and Appendix Table E.

Equation (5) has a somewhat surprising feature: it says that, as long as Q^* is perfectly observed, I do not need exogenous variation in relatedness to identify the presence of bias. This is because exogenous variation in relationships matters only if application quality is an omitted variable. If, however, quality is observed, then exogenous variation in relatedness would not be necessary because I would be able to directly control for quality.

In practice, though, I do not observe a grant's true quality Q^*. Instead, I observe a signal of quality $Q = Q^* + v$. Thus, while the model suggests the following equation:

$$S = \alpha_0 + \alpha_1 Q^* + \alpha_2 R + \alpha_3 RQ^* + \alpha_4 U + \alpha_5 RU + X\beta + \varepsilon \tag{6}$$

I can only estimate:

$$S = a_0 + a_1 Q + a_2 R + a_3 RQ + a_4 U + a_5 RU + Xb + e. \tag{7}$$

where, in both equations, X includes other relevant variables that I can condition on.

Proposition 4.1 *Given observed quality $Q = Q^* + v$, the bias parameter α_2 in Equation (6) is consistently estimated by a_2 in Equation (7) as long as the following conditions are met:*

1. $Cov(R, Q^*|U, RU, X) = 0$ *and* $Cov(R^2, Q^*|U, RU, X) = 0$
2. $E(v|U, RU, X) = 0$
3. $Cov(v, R|U, RU, X) = 0$

Proof: See Appendix B.

These are my identifying conditions. Condition 1 requires that my measure of relatedness not be correlated with true application quality, conditional on some set of observables. If this were not the case, any mismeasurement in true quality Q^* would bias estimates of α_2 through the correlation between Q^* and my relatedness measure R. Thus, in my study, exogenous variation in relatedness is required only to deal with measurement error.

Condition 2 requires that measurement error be mean zero conditional on observables. This means that, after conditioning on observable traits of the grant application, my measure of quality cannot be systematically different from what committees themselves are trying to maximize. Otherwise, I may conclude that committees are biased when they are actually prioritizing something I do not observe, but which is not mean zero different from my quality measure. In order to address this concern, I include very detailed controls for almost all of the salient characteristics that a committee can observe: demographics, institutional affiliation, and past publication characteristics. I also use different measures of application quality, not just number of citations. This allows my framework to identify bias even if, for instance, committees are attempting to maximize the number of big hit publications instead of total citations, or if they are also taking applicant demographics into account when assessing quality. Finally, even if committees are maximizing something unobservably and systematically different, my estimate of the efficiency tradeoff of related reviewers with respect to the number of citations and top hit publications produced by the NIH will still be unbiased (see Section 7). This in itself is a metric of decision-making quality that is relevant for policy.

Finally, Condition 3 says that the extent of measurement error should not depend, conditional on observables, on whether an applicant is related to a reviewer. This may not be satisfied if related applicants are more likely to be funded and funding itself affects my measure of quality. Suppose, for instance, that two scientists apply for a grant using proposals that are of the same quality. One scientist is related to a reviewer and is funded because of bias. The funding, however, allows her to publish more articles, meaning that my measure of quality—future citations—may mistakenly conclude that her proposal was better than the other scientist's to begin with. Mismeasurement of ex ante grant quality makes it *less* likely that I would find an effect of bias.

Another important reason why Condition 3 may not be satisfied is given by the Matthew Effect, a sociological phenomenon wherein credit and citations accrue to established investigators simply because they are established (see Merton, 1986; Azoulay, Stuart, and Wang, 2011). Were this the case, more related applicants would receive more citations regardless of the true quality of their work, meaning that measurement error v would be correlated with relatedness. The Matthew Effect would also make it less likely that I would find an effect of bias; related applicants may get higher scores simply for being established, but this bias would look justified by my measure of quality (which reflects bias in the scientific community at large).

Together, Conditions 1-3 are weaker than assuming classical measurement error. In the next section, I discuss how my sample and variables are constructed in order to separate the effect of bias from the effect of information. I pay particular attention to describing how I define and measure relatedness and quality in order to meet my identifying conditions, described above.

5 Data and Empirical Strategy

In order to understand how relatedness affects committee decisions, I have constructed a new dataset describing grant applications, review committee members, and their relationships for almost 100,000 applications evaluated in more than 2,000 meetings of 250 chartered study sections. My analytic file combines data from three sources: NIH administrative data for the universe of R01 grant applications, attendance rosters for NIH peer review meetings, and publication databases for life sciences research. Figure 1 summarizes how these data sources fit together and how my variables are constructed from them.

I begin with two primary sources: the NIH IMPAC II database, which contains administrative data on grant applications and a series of study section attendance rosters obtained from NIH's main peer review body, the Center for Scientific Review. The application file contains information on an applicant's full name and degrees, the title of the grant project, the study section meeting to which it was assigned for evaluation, the score given by the study section, and the funding status of the application. The attendance roster lists the full names of all reviewers who were present at a study section meeting and whether a reviewer served as a temporary member or as a permanent member. These two files can be linked using meeting-level identifiers available for each grant application. Thus, for my sample grant applicants, I observe the identity of the grant applicant, the identity of all committee members, and the action undertaken by the committee.

Next, I construct detailed measures of applicant demographics, grant history, and prior publications. Using an applicant's first and last name, I construct probabilistic measures of gender and ethnicity (Hispanic, East Asian, or South Asian).[11] I also search my database of grant applications to build a record of an applicant's grant history as measured by how many new and renewal grants an applicant has applied for in the past, and how many he has received. This includes data on non-R01 NIH grants such as post-doctoral fellowships and career training grants. To get measures of an applicant's publication history, I use data from Thomson-Reuters Web of Science (WoS) and the National Library of Medicine's PubMed database. From these, I construct information on the number of research articles that an applicant has published in the five years prior to submitting her application, her role in those publications (in the life sciences, this is discernable from author position), and the impact of those publications as measured by citations. In addition to observing total citations, I can also identify a publication as "high-impact" by comparing the number of citations it receives with the number of citations received by other life science articles that were published in the same year.

My final sample consists of 93,558 R01 applications from 36,785 distinct investigators over the period 1992-2005. Of these applications, approximately 25 percent are funded and 20 percent are from new investigators, those who have not received an R01 in the past. This sample is derived from the set of grant applications that I can successfully match to meetings of study sections for which I have attendance records, which is about half of all R01 grants reviewed in chartered study sections.

[11] For more details, see Kerr (2008).

Table 1 shows that my sample appears to be comparable to the universe of R01 applications that are evaluated in chartered study sections.

So far, I have discussed how I measure the prior qualifications of an applicant. As Conditions 1-3 of Section 4 indicate, however, I also need a direct measure of grant quality and a measure of relatedness that is conditionally independent of quality. I discuss each of these requirements in turn.

5.1 Measuring Quality

A major strength of this project lies in my ability to go beyond using past applicant characteristics to assess application quality. Instead, I observe a direct measure of application quality by examining the publications and citations it produces in the future. Due to the nature of the R01 grant application process, grant applications are likely to produce publications even when the application is not funded. This is because R01s are intended for projects that have demonstrated a substantial likelihood of success, meaning that R01 applicants are required to produce substantial "preliminary results" as a part of their grant application. In practice these stringent requirements mean that preliminary results are often developed fully enough to be published as standalone articles even if the grant application itself goes unfunded. In fact, the bar for preliminary results is so high that the NIH provides a separate grant mechanism, the R21, for pursuing exploratory research leading to an R01 application.

For every grant application I observe, I find articles published by that grant's primary investigator around the time when the grant was reviewed. These publications, and the citations that they generate, form the basis of my measure of grant quality. As discussed in Section 4, however, measurement error in the quality of applications poses several challenges. In particular, I need to find a quality measure that is consistent for funded and unfunded grants and not directly affected by funding.

I tackle the first concern by devising a way to link grant applications to their related publications using only information that would exist for both funded and unfunded grants. This means that I cannot make use of explicit grant acknowledgements because they are available only for funded grants. Instead, I compare the titles and abstracts of an applicant's publications with the title of her grant proposal to determine which publications are related. For instance, if I see a grant application titled "Traumatic brain injury and marrow stromal cells" reviewed in 2001 and an article by the same investigator titled "Treatment of traumatic brain injury in female rats with intravenous administration of bone marrow stromal cells," published around this time, I conclude that this publication and its future citations can be used as a measure of the quality of the grant application. Text-matching ensures that I can measure quality using the same procedure for all grant applications.

The second challenge in assessing quality is to make sure that my measure of quality is not directly affected by funding. Grant funding, for instance, can be used to start new experiments related to the proposed project or to subsidize research on unrelated projects. Existing evidence on the effect of grant funding on research outcomes suggests that this effect is likely to be small; using a regression-discontinuity approach, Jacob and Lefgren (2011) find that receiving an R01 increases the number of articles a PI publishes in the next five years by 0.85, from a mean of 14.5. This figure includes all publications by a PI, including ones that may be on a different topic from the original application. Jacob and Lefgren's analysis, however, only documents the effect of grant receipt for marginal applicants. The effect of funding on future publications and citations could be larger elsewhere in the distribution and I take additional precautions to create a measure of quality not affected by funding.

Text-matching limits the set of publications I use to infer application quality to those that are on the same topic as the grant. This reduces the possibility that my measure of application quality is contaminated by unrelated research that the grant is used to subsidize. Funding, however, may also increase the number of publications on the same topic as the grant. To address this concern, I also restrict my quality calculations to articles published in a short time window surrounding grant review. These articles are likely to be based on research that was already completed or underway at the time the grant application was written. To compute the appropriate window, I consider funding, publication, and research lags. A grant application is typically reviewed four months after it is formally submitted and, on average, another six months elapse before it is officially funded.[12] In addition to this ten-month funding lag, publication lags in the life sciences (the time between first submission and publication) typically range from three months to well over a year. Because running experiments, analyzing data, and writing drafts also takes time, it is unlikely that articles published up to two years after a grant's review would have been directly supported by that grant. I also include related publications published one year before a grant is reviewed because these publications likely contribute to the research that is proposed in the application.

Figure 2 confirms that unfunded grant applications produce related publications. In fact, using my measure of quality described above, I find that funded and unfunded grants are almost equally represented among the subset of grant applications that generate many citations. Figure 2 also shows, however, that unfunded grants are more likely to produce few citations. There are two possible explanations for this finding: 1) unfunded applications are of lower quality, and should thus be expected to produce fewer citations; or 2) funding directly improves research output, meaning that my measure of grant quality is not consistent for funded and unfunded grants.

I distinguish between these explanations by using year-to-year and subject-to-subject variation in whether grant applications with the same score are funded. If funding has a direct impact on my measure of quality, then I should mistakenly attribute higher quality to funded grants than to unfunded grants with the same score. Figure 3 shows that this is not the case. Each dot represents the mean number of citations associated with grant applications that received a particular percentile score, regression adjusted to account for differences across fields and years. The dots represent outcomes and scores for funded grants, the crosses for unfunded grants. The dots and crosses overlap because budgets vary across time and across fields, meaning that similarly ranked grants

[12]See http://grants.nih.gov/grants/grants_process.htm.

are sometimes funded and sometimes not. In these areas, outcomes for funded and unfunded grants with the same score are similar. There is no evidence that my measure of quality is directly affected by funding.

The accompanying statistical test is reported in Table 2. I compare measured quality for funded and unfunded grant applications with similar scores from applicants that have similar characteristics. Funding status can vary if some grants are funded out of scoring order, or it can vary because budgets fluctuate across fields and years. Columns 1 and 2 show that awarded grants tend to be higher quality, but this effect goes away once I control for a smooth function of scores. Together with Figure 3, this finding mitigates concerns that my measure of quality is directly affected by funding.

I discuss several more robustness tests in Appendix C. First, I show that my results hold if I restrict publications associated with grants to those published one year before and one year after grant review. This short time window means that it would be highly unlikely that an article could be directly supported by grant funding because funding and publication lags together may run more than a year. Appendix C also reports another test of the validity of my quality measure. If my results are driven by changes in measured grant quality near the payline, then I should find no effect of relatedness on scores for in the subset of applications that are either well above or well below the payline. However, in both of these subsamples, I find evidence that being related to a permanent member increases scores and increases the correlation between scores and quality. Because relatedness cannot affect actual funding status in these subsamples, the effect I find cannot be driven by differences in how well quality is measured.

It is also worth emphasizing that, as discussed in Section 4, overcrediting funded applications relative to unfunded applications would lead me to *underestimate* the extent of bias.

5.2 Identifying Relationships

Next, I determine whether an applicant and a reviewer are related using their citation history. Specifically, using data from Web of Science, I define an applicant to be related to a reviewer if the reviewer has cited the applicant in the five years prior to the review meeting. Citation relationships capture the extent to which reviewers are aware of an applicant's prior work and whether they find that work useful for their own research. In particular, I assume that reviewers are more likely to be familiar with the work or subfield of authors they cite than authors they do not cite.

Table 3 describes applicant-reviewer relationships in my sample study sections. In total, I observe 18,916 unique reviewers. On average, 30 reviewers attend each meeting, 17 of whom are permanent and 13 of whom are temporary. The average applicant has been cited by two reviewers, one temporary and one permanent. The average permanent and average temporary reviewer both cite four applicants. This relatively low amount of relatedness indicates that citations are capturing a finer measure of expertise than simple field overlap. Because the review committees I study are highly focused, most reviewers in a committee are drawn from the same academic departments (molecular biology, surgery, etc.). Using citations allows me to more finely measure the type of work that reviewers are familiar with and thus generate more variation in relatedness. Appendix C discusses the robustness of my results to alternative measures of relatedness.[13]

Whether an applicant has been cited by a reviewer is likely to be correlated with the applicant's quality. Applicants who are prominent scientists may be more likely to be cited by reviewers and they may also be more likely to receive higher scores. This correlation would violate Condition 1 of Section 4. I exploit the structure of chartered NIH study sections in order to find exogenous variation in reviewer-applicant relatedness. As discussed in Section 2, the review committees I study consist of "permanent" and "temporary" members. Permanent members and temporary members are comparable as scientists. Figure 4 and Table 4 show that they have similar publication histories and demographics. In fact, Table 4 indicates that they are often the same people; 35 percent of current permanent members will work as temporary members in the future and 40 percent of current temporary members will work as permanent members in the future.

Permanent members are also likely to have more influence in general. Because they serve as reviewers for more grants, permanent members exert greater influence over committee decisions by providing more initial scores. Temporary members, moreover, vote on fewer proposals because they are often not expected to be present except on those meeting days when their assigned grants are to be discussed. Finally, permanent members are more likely to have relationships with each other because they work together over the course of four years or twelve committee meetings. A test of the assumption that permanent members have more influence is reported in Appendix C.

Given this, I identify the effect of relationships by examining how the number of permanent members an applicant is related to, call this R^P, affects the committee decision, conditional on the *total* number of a related reviewers, R. My identification compares the outcomes of scientists whose applications are reviewed in the same meeting, who have similar past performance, and who are related to the same total number of reviewers, but who are related to different numbers of permanent reviewers.

Using relatedness to permanent members also addresses concerns about the Matthew Effect. Because my identification holds scientific esteem as measured by total relationships constant, there is no reason to believe that applicants related to permanent members would be more or less likely to be cited than applicants related to temporary members.

Figure 5 provides general evidence that the number of permanent members an applicant is related to is not correlated with her quality, conditional on total relatedness. The first panel shows the distribution of application quality as measured by future citations for applicants related to exactly one reviewer. The solid line shows the distribution for applicants related to one permanent member; the dotted line shows the distribution for those related to one temporary member. These distributions are essentially identical. Similarly, Panel 2 shows that the distribution of application quality is the same whether an applicant is related to two temporary, two permanent, or one temporary and one permanent member.

[13]Specifically, I measure relatedness using mutual citations and by restricting citation linkages to publications in which both the reviewer and applicant were primary (first, second, or last) authors.

5.3 Estimating Equations

Taking these specific measures of quality and relatedness, my schematic regression from Section 4 translates into the following set of estimating equations:

First, using variation in relatedness to permanent members, the effect of relatedness on an applicant's likelihood of funding can be estimated from the following regression:

$$D_{icmt} = a_0 + a_1 R^P_{icmt} + a_2 R_{icmt} + \mu X_{icmt} + \delta_{cmt} + e_{icmt}. \tag{8}$$

D_{icmt} is a variable describing the decision (either the score or the funding status) given to applicant i whose proposal is evaluated by committee c in meeting m of year t. R^P_{icmt} is the number of permanent reviewers an applicant is related to and R_{icmt} is the total number. The covariates X_{icmt} include indicators for sex; whether an applicant's name is Hispanic, East Asian, or South Asian; quartics in an applicant's total number of citations and publications over the past five years; indicators for whether an applicant has an M.D. and/or a Ph.D.; and indicators for the number of past R01 and other NIH grants an applicant has won, and indicators for how many she has applied to. The δ_{cmt} are fixed effects for each committee-meeting so that my analysis compares outcomes for grants that are reviewed by the same reviewers in the same meeting. Standard errors are clustered at the committee-fiscal year level. Given these controls, a_1 captures the effect of being related to an additional permanent reviewer on the likelihood that an applicant is funded.

The full effect of relationships on funding decisions, however, is more nuanced. The model in Section 3 predicts that both the level likelihood of funding and its slope with respect to quality will be higher for related applicants. To test these predictions, I estimate:

$$\begin{aligned} D_{icmt} &= a_0 + a_1 R^P_{icmt} + a_2 Q_{icmt} \times R^P_{icmt} + a_3 Q_{icmt} \\ &+ a_4 R_{icmt} + a_5 R_{icmt} \times Q_{icmt} + \mu X_{icmt} + \delta_{cmt} + \varepsilon_{icmt} \end{aligned} \tag{9}$$

Equation (9) uses the same controls as in Equation (8) and adds several variables describing the quality of the grant application. $Q_{icmt} \times R^P_{icmt}$ is the interaction between number of permanent reviewers and quality, and Q_{icmt} is the level effect of quality on the committee decision D_{icmt}. Equation (9) includes a control for the total number of related reviewers interacted with quality, $R_{icmt} \times Q_{icmt}$. This is necessary because the total number of reviewers who cite an applicant may be correlated with an applicant's quality; without this control, the variable of interest $R^P_{icmt} \times Q_{icmt}$ may simply be capturing the difference in correlation between quality Q_{icmt} and committee decisions D_{icmt} for high-quality applicants (those cited by more reviewers). For instance, the correlation between scores and quality for well-cited candidates may be mechanically lower than for poorly-cited candidates because it may simply be harder to distinguish among high-quality applications. Controlling for $R_{icmt} \times Q_{icmt}$ accounts for this possibility.

In Equation (9), the coefficient a_1 is the effect of being related to an additional permanent member on funding that is attributable to bias. The coefficient a_2 measures the information effect of being related to a permanent member. That is, comparing two scientists related to the same total number of reviewers, a_2 captures the additional change in the likelihood of funding for the applicant related to a permanent member, for the same one unit increase in quality. Equation (9) says that if committees are using relationships to make better inferences about the quality of an application, then the effect of relationships should be captured by the interaction of quality and relatedness, $Q_{icmt} \times R^P_{icmt}$. Any remaining level effect of relationships is then attributable to bias.

6 Main Results

Table 5 considers the effect of being related to a committee member on scores and funding. The first column reports the raw within-meeting association between the number of permanent related reviewers and an applicant's likelihood of being funded. Without controls, each additional related permanent member is associated with a 3.3 percentage point increase in the probability of funding, off an average of 21.4 percent. This translates into a 15.3 percent increase. Most of this correlation, however, reflects differences in the quality of applications; applicants may be more highly cited by reviewers simply because they are better scientists. Column 2 adds controls for applicant characteristics such as past publication and grant history. This reduces the effect of an additional permanent related reviewer on funding probability to 1.5 percentage points, or 7.1 percent. Even with these controls, relatedness may still be proxying for some unobserved aspect of application quality. Finally, I control for the total number of reviewers each applicant has been cited by. Given this, my identification comes from variation in the composition of an applicant's related reviewers; I am comparing outcomes for two scientists with similar observables, who are cited by the same total number of reviewers, but by different numbers of influential reviewers. In Column 3, I find that an additional permanent related reviewer increases an applicant's chances of being funded by 0.6 percentage points, or 2.9 percent. This is my preferred specification because it isolates variation in relatedness that is plausibly independent of an application's quality. I find similar effects when an applicant's score is the dependent variable.

The estimates in Table 5 do not distinguish between the impact of bias and the impact of information. Table 6 reports my main regressions, decomposing these effects. Columns 1 and 3 reproduce the estimates of the level effect of relatedness on funding and scores from Table 5. Column 2 reports estimates of the coefficients from Equation (9). I show that each additional applicant still increases the likelihood that a grant is funded by 0.6 percentage points, or 2.9 percent. Since I also include controls for an application's quality and its quality interacted with relatedness, this figures means that the entire level effect of relationships on funding is likely due to bias.

Column 2 also shows that the review committee does a better job of discerning quality when an applicant is related to a permanent member, conditional on the total number of related reviewers. To see this, consider an applicant who is related to one permanent member versus an applicant who is related to one temporary member. A one standard deviation increase in quality for the former applicant increases her likelihood of funding by 1.06+3.15-0.16 = 4.05 percentage points or 4.05/21.4 = 18.9 percent compared with 3.14-0.16 = 2.99 percentage points or 2.99/21.2=14.0 percent for the latter applicant. Being related to a permanent member, then, increases the ability of the committee to predict application quality by more than 30 percent. Thus, despite overall positive bias in favor of related applicants, being related to a permanent member may not be beneficial for all applicants. Because reviewers have more information about the quality of related applicants, related applicants with lower quality proposals end up receiving lower scores. These results are consistent with the predictions of my model: relationships decrease the variance of the committee's signal of quality but also increase the distortion arising from bias.

Column 2 also reports the increase in funding likelihood associated with an increase in application quality. The figure of 0.0315 means that a one standard deviation increase in application quality is associated with a 3.2 percentage point, or 3.2/21.4=14.9 percent, increase funding probability for applicants who are not related to any reviewers at all. The sensitivity of committees to changes in application quality highlights the magnitude of the bias effects that I find: being related to an additional permanent reviewer increases an applicant's chances of being funded by as much as a one-fifth standard deviation increase in quality.

The coefficient on total related reviewers interacted with quality is estimated to be negative. This means that the correlation between quality and funding is lower for applicants related to more reviewers. If total related reviewers were proxying for quality, this result would not be unexpected; it may be harder to distinguish quality among grant proposals from high-quality scientists than from low-quality scientists, where the variance in quality may be higher overall.

Finally, looking at Column 4, a similar though noisier pattern can be seen for scores. While being related to a reviewer increases the level score that one receives for reasons due to bias, it also improves the correlation between an application's quality and its chances of being funded.

In Table 7, I consider how the role of relationships may differ for new and experienced investigators and for new and competing renewal applications. Approximately 20 percent of grant applications are submitted by scientists who have no prior R01 grants. Understanding how applications from new investigators are treated is of particular importance for identifying and supporting promising young scientists.

Even though they are applying for their first R01 grant, new investigators are not entirely unknown to study sections. Forty percent of would-be new investigators have been cited by a reviewer in the past, indicating that the reviewer may be familiar with their work, or at least the work coming out of their lab. Columns 1 and 2 show that there appears to be little bias in the evaluation of new investigators. Although related reviewers appear to be more informed about the quality of experienced investigators, they do not have better information about the quality of new investigators. In fact, the entire effect of bias and information estimated in Table 6 appears to be driven by the evaluation of experienced investigators.

Columns 5-8 of Table 7 consider the effect of relatedness for new versus competing renewal applications. I find that related reviewers have fewer insights or biases about the quality of new grants. In both the case of new investigators and new proposals, the bias and information effects of relationships are driven by the subset of grants for which there already may be more information. Because there are substantially more experienced investigators but substantially fewer renewal grants in my sample, this effect is not driven by larger sample sizes or more precise estimates.

7 How Do Relationships Affect the Efficiency of Grant Provision?

My main results show that relationships affect committee decisions by increasing bias and increasing information. In this section, I embed my analysis of the effect of relationships on *decisions* into a broader analysis of their effect on *efficiency*. In particular, I estimate the net effect of relationships on the quality of decision-making, assuming that policy makers care about maximizing the number of publications and citations associated with NIH-funded research.

I begin by comparing the actual funding decision for an application to the counterfactual funding decision that would have obtained in the absence of relationships. Specifically, I define:

$$\begin{aligned} D_{icmt}^{\text{Benchmark}} &= D_{icmt} \text{ (actual funding)} \\ D_{icmt}^{\text{No Relationship}} &= D_{icmt} - \widehat{a_1} R_{icmt}^{P} + \widehat{a_2} Q_{icmt} \times R_{icmt}^{P} \end{aligned}$$

where $\widehat{a_1}$ and $\widehat{a_2}$ are estimated from Equation (9) of Section 5.3.[14] The counterfactual funding decision represents what the committee would have chosen had applicants related to permanent members been treated as if they were unrelated.

I summarize the effect of relationships by comparing the quality of the proposals that would have been funded had relationships not been taken into account with the quality of those that actually are funded. Specifically, I consider all applications that are funded and sum up the number of publications and citations that accrue to this portfolio. This is my benchmark measure of the quality of NIH peer review. I then simulate what applications would have been funded were relationships not taken into account. To do this, I fix the total number of proposals that are funded in each committee meeting but reorder applications by their counterfactual funding probabilities. I sum up the number of publications and citations that accrue to this new portfolio of funded grants. The difference in the quality of the benchmark and counterfactual portfolio provides a concrete, summary measure of the effect of relationships on the quality of research that the NIH supports.

To get a fuller sense of how committees affect decision-making, I create a measure of committee-specific performance and examine how relationships affect the distribution of performance among NIH peer review committees. First, I define a committee's *value-added*. Suppose two scientists submit applications to the same committee meeting. A good committee is one that systematically funds the application that is higher quality. Good committees, moreover, should bring insights beyond what can simply be predicted by objective measures of an applicant's past performance. In particular, suppose now that two scientists with identical objective qualifications submit applications to

[14] Even though $D_{icmt}^{\text{No Relationship}}$ is constructed using estimates from Equation (9), it does not rely on the model to interpret those coefficients.

the same committee meeting. A committee with high value-added is one that systematically funds the application that subsequently generates more citations, even though the applications initially look similar. My measure of committee value-added formalizes this intuition:

$$D_{icmt} = a + b_{cmt} Q_{icmt} + \mu X_{icmt} + \delta_{cmt} + e_{icmt}. \tag{10}$$

Here, the dependent variable is either an application's actual funding status $D_{icmt} = D_{icmt}^{\text{Benchmark}}$ or its counterfactual funding status $D_{icmt} = D_{icmt}^{\text{No Relationship}}$. The committee fixed effects δ_{cmt} restrict comparisons of applications to those evaluated in a single meeting and the X_{icmt} control for past applicant qualifications. The coefficients of interest are the b_{cmt}. These are meeting-specific slopes that capture the relationship between an application's quality Q_{icmt} and its likelihood of being funded D_{icmt}. Each b_{cmt} is interpreted as the percentage point change in the likelihood that an application is funded for a one unit increase in quality. This forms the basis of my committee value-added measure.

This concept of committee value-added differs from the classical notion of value-added commonly used in the teacher or manager performance literature (see Kane, Rockoff, and Staiger 2007, and Bertrand and Schoar 2003). Teacher value-added, for instance, is typically estimated by regressing student test scores on lags of test scores, school fixed effects, and teacher fixed effects. A teacher's fixed effect, the average performance of her students purged of individual, parental, and school-wide inputs, is taken to be the basic measure of quality.

This traditional measure, however, does not capture value-added in my setting. Good committees are not ones in which all applications are high-performing; after all, committees have no control over what applications get submitted. Rather, good committees are ones in which funded grants perform better than unfunded grants. I measure a committee's performance by the relationship between an applicant's quality and its likelihood of getting funded because, unlike a teacher, a committee's job is not to *improve* the quality of grant applications but to *distinguish* between them.

One concern with the estimated $\hat{b}_{cmt}$ is that idiosyncratic variation in grant performance may lead me to conclude that some committee meetings do an excellent job of identifying high-quality applications when in fact they are simply lucky. I correct for this by modeling $\hat{b}_{cmt}$ as a combination of the committee's true value-added plus a noise term, which I assume to be independent and normal:

$$\hat{b}_{cmt} = b^*_{cmt} + \nu_{cmt} \tag{11}$$

Using an empirical Bayes estimator, I adjust $\hat{b}_{cmt}$ for sampling variation so that I define committee quality based only on the portion of $\hat{b}_{cmt}$ that is correlated across multiple meetings; an estimate $\hat{b}_{cmt}$ is taken seriously only if it is consistent across multiple meetings of that committee within the same fiscal year. Otherwise, the Bayesian shrinkage estimator reweights that observation toward the mean. Appendix D describes this procedure in more detail.

7.1 Results

Table 8 estimates the effect of relationships on the quality of research that the NIH supports. In effect, I ask what the NIH portfolio of funded grants would have been had committees treated applicants who are related to permanent members as if they were not, holding all else fixed. In my sample, I observe 93,558 applications, 24,404 of which are funded. Using this strategy, I find that 2,166, or 2.3 percent, of these applications change funding status under the counterfactual.

On average, relationships help applicants get funded; ignoring them would decrease the number of related applicants who are funded by 3.5 percent. These applications from related reviewers, however, are on average better than the applications that would have been funded had relationships not mattered. The overall portfolio of funded grants under the counterfactual produces two to three percent fewer citations, publications, and high-impact publications.

This pattern is underscored by Figure 6, which graphs the distribution of value-added under the benchmark and counterfactual cases. Under the benchmark, a one standard deviation increase in the quality of an application evaluated by the median committee would increase its likelihood of funding by approximately 14.5 percent. When relationships are ignored, this figure falls to 11.1 percent.

Figure 6 also shows that there is significant variation in the ability of committees to identify grant applications that subsequently produce high-impact research. Regardless of whether relationships inform committee decisions, the bottom one-quarter to one-third of committees actively subtract value, meaning that increases in quality are correlated with *decreases* in the likelihood that an application is funded. As explained in Section 7, these figures account for sampling variation so that a committee is deemed to have negative value-added only if it systematically does so from meeting to meeting.

Table 9 presents preliminary evidence that good committees are able to make better use of expert information while limiting the extent of bias. In this table, I run my main regressions on separate samples of high- and low-performing committees according to the value-added measure discussed in Section 7. All results are weighted by the precision of the value-added estimate. Columns 1 and 2 present the main results on bias and information estimated separately for above and below median committee meetings. Although the standard errors are large, relationships appear to affect the decisions of below median committees by increasing bias but not increasing information. This pattern is seen more clearly in Columns 3 and 4, which consider bottom and top tercile committees separately. In Column 3, the correlation between quality and funding is zero or possibly even negative in committees with low value-added. In contrast, in committees that rank in the top tercile of value-added, the effect of relationships on decision-making that comes through information is positive and significant.

It is important to note that this effect is not a mechanical artifact of the way committee value-added is defined; committees are deemed to perform well if increases in applicant quality translate into increases in funding (see Equation (10)). This effect is captured by the coefficient on application quality alone, which indeed is higher for high value-added committees than for low value-added committees. My results in Table 9 say that, in addition, a high-performing committee has more information about a scientist related to a permanent member than one who is not, holding constant their total relatedness to committee members. This is captured by the interaction between application quality and whether an applicant is related to a permanent reviewer. Better performing committees not only have higher correlation between quality and funding overall, but also appear to make more use of the information that permanent members have.

Looking again at Figure 6, ignoring relationships appears to be least harmful in the most poorly performing committees. This is consistent with the finding in Table 9 that bias tends to be higher in poorly performing committees and information tends to be lower. The magnitudes of these effects, however, are not large; regardless of whether relationships are taken into account, the distribution of committee performance is substantial. Understanding other reasons for this dispersion is an important area for future research.

8 Conclusion

This paper develops a conceptual and statistical framework for understanding the tradeoff between bias and information in expert evaluation. In particular, I make use of exogenous variation in reviewer assignments and detailed data on grant application quality to separately identify the effects of bias and information. My results show that, as a result of bias, each additional related permanent reviewer increases an application's chances of being funded by 2.9 percent. Viewed in terms of how committees respond to increases in application quality, being related to a reviewer increases the chances that an application is funded by the same amount as would be predicted by a one-fifth standard deviation increase in its quality. Related reviewers, however, also bring expertise to the committee. I show that their information increases the correlation between quality and funding decisions by more than 30 percent. On net, ignoring relationships reduces the quality of the NIH-funded portfolio as measured by numbers of citations and publications by two to three percent.

My results suggest that there may be scope for improving the quality of peer review. I document significant and persistent dispersion in the ability of committees to fund high-quality research. Finding ways to eliminate the lower tail of committees, for which increases in quality are actually associated with *decreases* in funding likelihood, could lead to large improvements in the quality of NIH-funded research as measured by citations. The magnitude of these potential benefits are not small when viewed in dollar terms. NIH spending for my sample of approximately 25,000 funded grants totaled more than 34 billion dollars (2010 dollars). These grants generated approximately 170,000 publications and 6.8 million citations.[15] This means that, in my sample, the NIH spent about 250,000 dollars per publication, or about 5,000 dollars per single citation. Even

[15] I have 170,000 publications linked to grants via formal grant acknowledgments computed from the PubMed database. PubMed, however, undercounts citations because it only counts citations from a subset of articles archived in PubMed Central. To arrive at the 6.8 million citations figure, I use total publications calculated via text-matching (about 100,000 publications) and the total citations accruing to those publications (4.3 million) to compute the average number of citations per publication. I then scale this by the 170,000 publications found in PubMed.

if these numbers do not represent the social value of NIH-funded research, they suggest that the value generated by high-quality peer review can be substantial.

A small part of this overall dispersion can be explained by my finding that high value-added committees extract more information from related reviewers but are less susceptible to bias. Understanding and quantifying other factors affecting committee performance is an important area for future work. Here, the uniformity of NIH's many chartered study sections is helpful because it allows for the possibility of targeted randomized experiments, holding other institutional features constant. For instance, to understand the impact of committee composition on peer-review quality, applicants could be assigned to intellectually broad or narrow committees. Answers to these questions can provide insights on how to improve project evaluation at the NIH and elsewhere.

References

[1] Acemoglu, Daron. (2008) *Introduction to Modern Economic Growth*, Princeton University Press.

[2] Aghion, Philippe and Jean Tirole. (1994) "The Management of Innovation." *Quarterly Journal of Economics*, 109(4), 1185-1209.

[3] Arrow, Kenneth. (1962). "Economic welfare and the allocation of resources for invention." in Richard Nelson, ed., *The Rate and Direction of Inventive Activity*, Princeton University Press.

[4] Autor, David and David Scarborough. (2008) "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments." *Quarterly Journal of Economics*, 123(1): 219-277.

[5] Azoulay, Pierre, Toby Stuart, and Yanbo Wang. (2011) "Matthew: Fact or Fable?" Working paper. Available online: http://pazoulay.scripts.mit.edu/docs/shmatus.pdf

[6] Bagues, Manuel and Natalia Zinovyeva. (2012) "The Role of Connections in Academic Promotions." IZA Discussion Paper #6821.

[7] Bayer, Patrick, Stephen Ross and Giorgio Topa. (2008) "Place of work and place of residence: informal hiring network and labor market outcomes." *Journal of Political Economy*, 116(6), 1150-1195.

[8] Bandiera, Oriana, Iwan Barankay, and Imran Rasul. (2009) "Social Connections and Incentives in the Workplace: Evidence from Personnel Data." *Econometrica*, 77(4): 1047-1094.

[9] Bewley, Truman. (1999) *Why Wages Dont Fall During a Recession*. Cambridge, MA: Harvard University Press.

[10] Brogaard, Jonathan, Joseph Engelberg and Christopher Parsons. (2012) "Network Position and Productivity: Evidence from Journal Editor Rotations." mimeo.

[11] Cockburn, Iain, and Rebecca Henderson. (2000) "Publicly Funded Science and the Productivity of the Pharmaceutical Industry." *Innovation Policy and the Economy*, 1: 1-34.

[12] Congressional Budget Office. (2006) "Research and Development in the Pharmaceuticals Industry." Available online at: http://www.cbo.gov/ftpdocs/76xx/doc7615/10-02-DrugR-D.pdf

[13] Crawford, Vincent and Joel Sobel. (1982) "Strategic Information Transmission." *Econometrica*, 50(6):1431-1451.

[14] David, Paul, Mowery David, and W.E. Steinmueller. (1992) "Assessing the Economic Payoff from Basic Research." *Economics of Innovation and New Technology* 73-90.

[15] Ellison, Glenn. (2011) "Is Peer Review in Decline?" *Economic Inquiry*, 49(3): 635-657.

[16] Gerin, William. (2006) *Writing the NIH grant proposal: a step-by-step guide.* Thousand Oaks, CA: Sage Publications.

[17] Ginther, Donna, Walter Schaffer, Joshua Schnell, Beth Masimore, Faye Liu, Laurel Haak, Raynard Kington. (2011) "Race, Ethnicity, and NIH Research Awards." *Science*, 333(6045): 1015-1019.

[18] Griliches, Zvi. (1992) "The search for R&D spillovers." *Scandinavian Journal of Economics*, 94 (supplement), S29-S47.

[19] Hall, Bronwyn. (1994) "R&D Tax Policy During the Eighties: Success or Failure?" NBER Working Paper #4240

[20] Hegde, Deepak. (2009) "Political Influence behind the Veil of Peer Review: An Analysis of Public Biomedical Research Funding in the United States" *Journal of Law and Economics*, 52(4): 665-690.

[21] Jacob, Brian and Lars Lefgren. (2011) "The Impact of Research Grant Funding on Scientific Productivity." *Journal of Public Economics* 95(9-10): 1168-1177.

[22] Kane, Thomas, Jonah Rockoff, and Doug Staiger. (2007) "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27(6), 615-631.

[23] Kerr, William. (2008) "Ethnic Scientific Communities and International Technology Diffusion." *The Review of Economics and Statistics*, 90(3): 518-537.

[24] Kremer, Michael and Heidi Williams. (2010) "Incentivizing innovation: Adding to the toolkit," in Josh Lerner and Scott Stern, eds., *Innovation Policy and the Economy*, Vol. 10. Chicago, IL: University of Chicago Press.

[25] Lamont, Michele. (2010) *How Professors Think: Inside the Curious World of Academic Judgment.* Cambridge, MA: Harvard University Press.

[26] Lerner, Josh. (1999) "The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program." *The Journal of Business*, 72(3): 285-318.

[27] Merton, Robert. (1968) "The Matthew Effect in Science" *Science* 159(3810): 5663.

[28] National Institutes of Health. (2008) Office of Extramural Research. Peer Review Process. http://grants.nih.gov.libproxy.mit.edu/grants/peer_review_process.htm.

[29] Nelson, Richard. (1959) "The simple economics of basic scientific research." *Journal of Political Economy*, 67 (3), 297-306.

[30] Romer, Paul. (1990) "Endogenous Technological Change." *Journal of Political Economy*, 98, S71-S102.

[31] Sampat, Bhaven and Frank Lichtenberg. (2011) "What are the Respective Roles of the Public and Private Sectors in Pharmaceutical Innovation?" *Health Affairs*, 30(2): 332-339.

[32] Stephan, Paula. (2012) *How Economics Shapes Science.* Cambridge, MA: Harvard University Press.

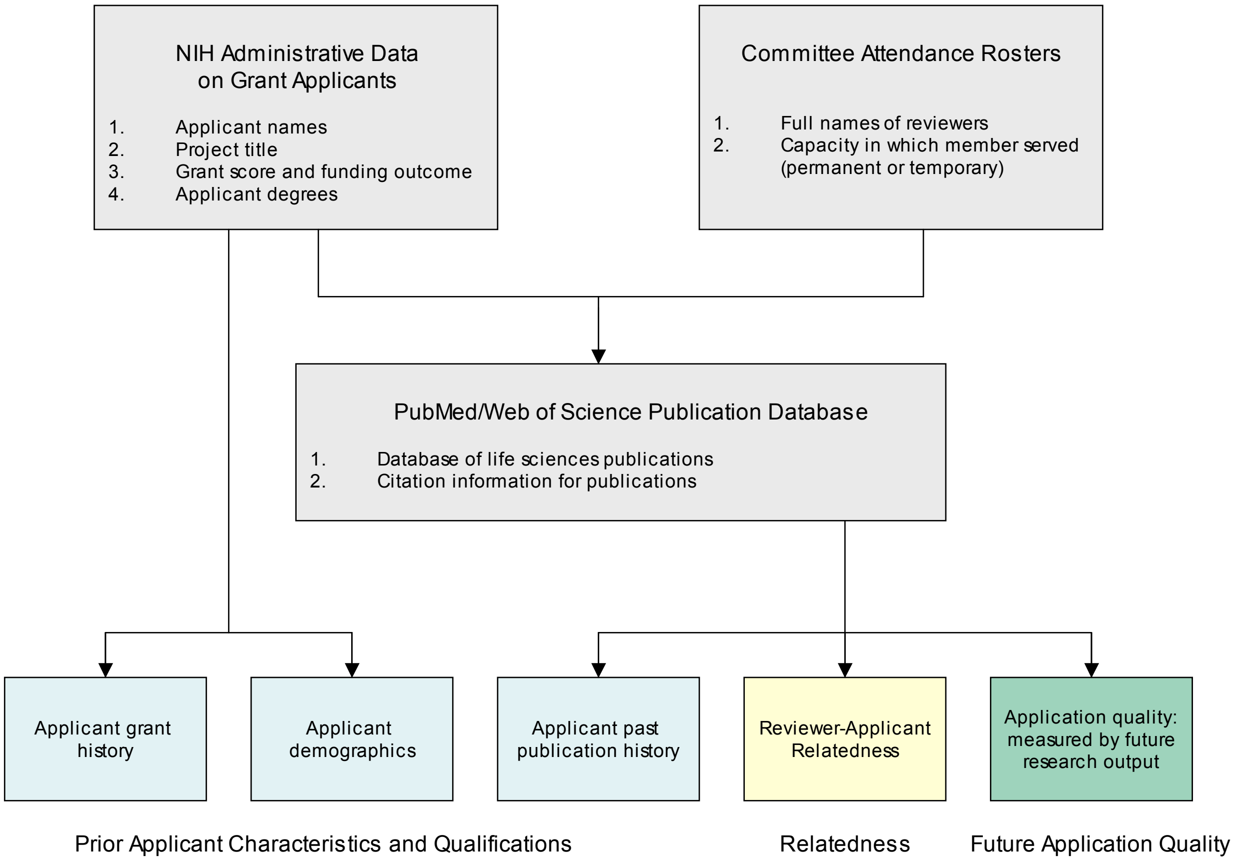

Figure 1: Data sources and variable construction

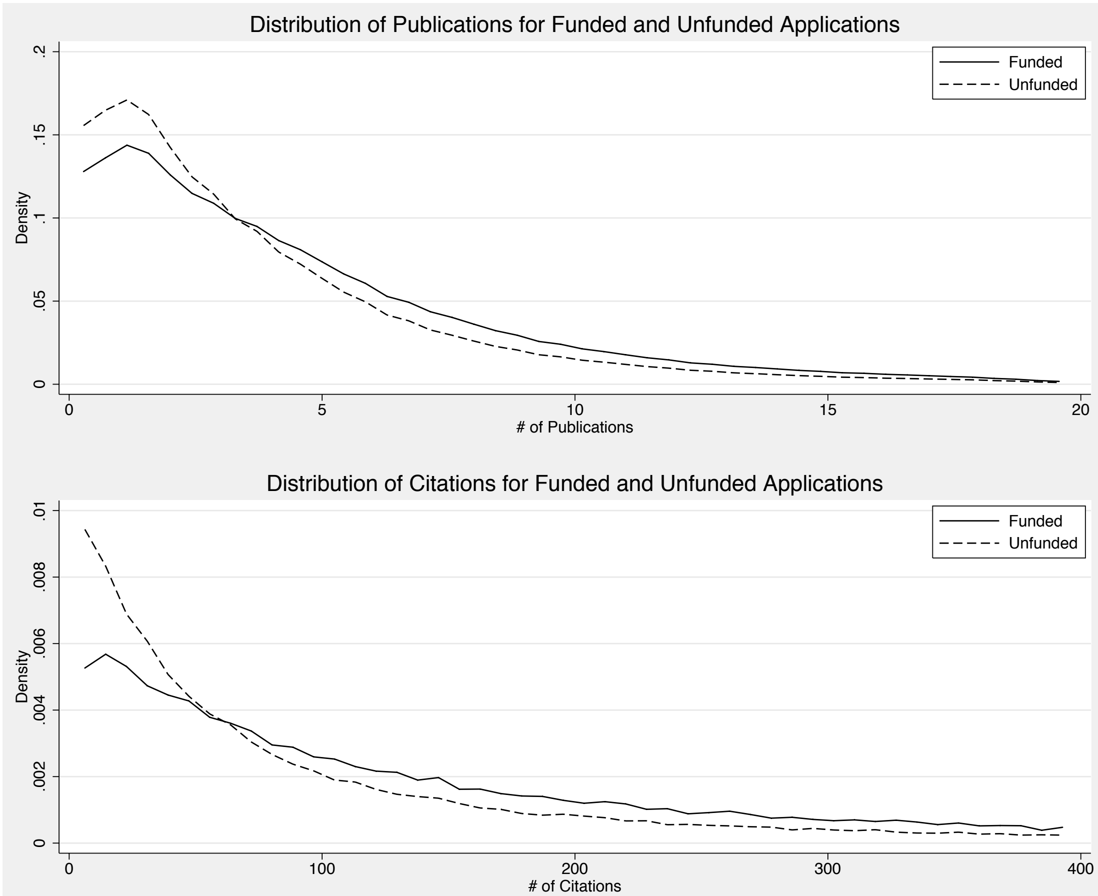

Figure 2: Distribution of application quality: funded and unfunded grants

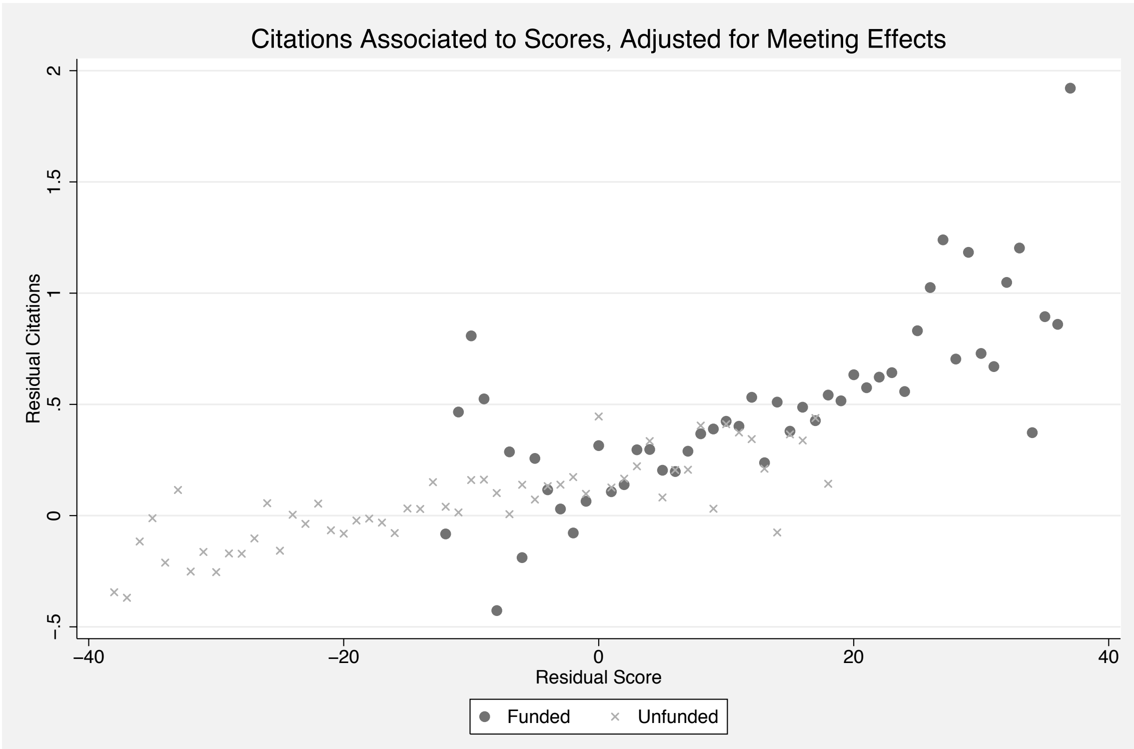

Figure 3: Mean application quality by score: funded and unfunded grants

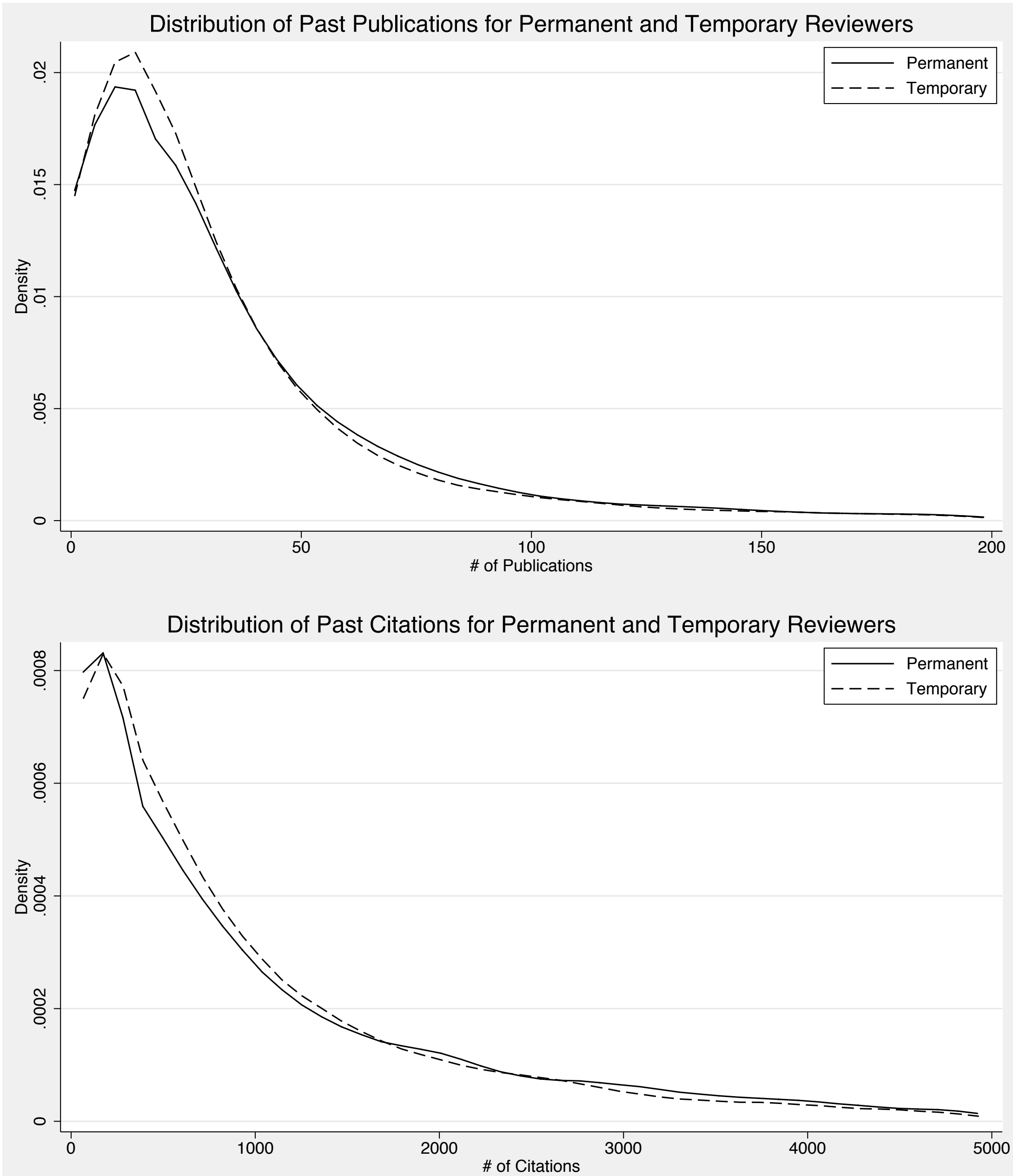

Figure 4: Distribution of past citations: permanent and temporary reviewers

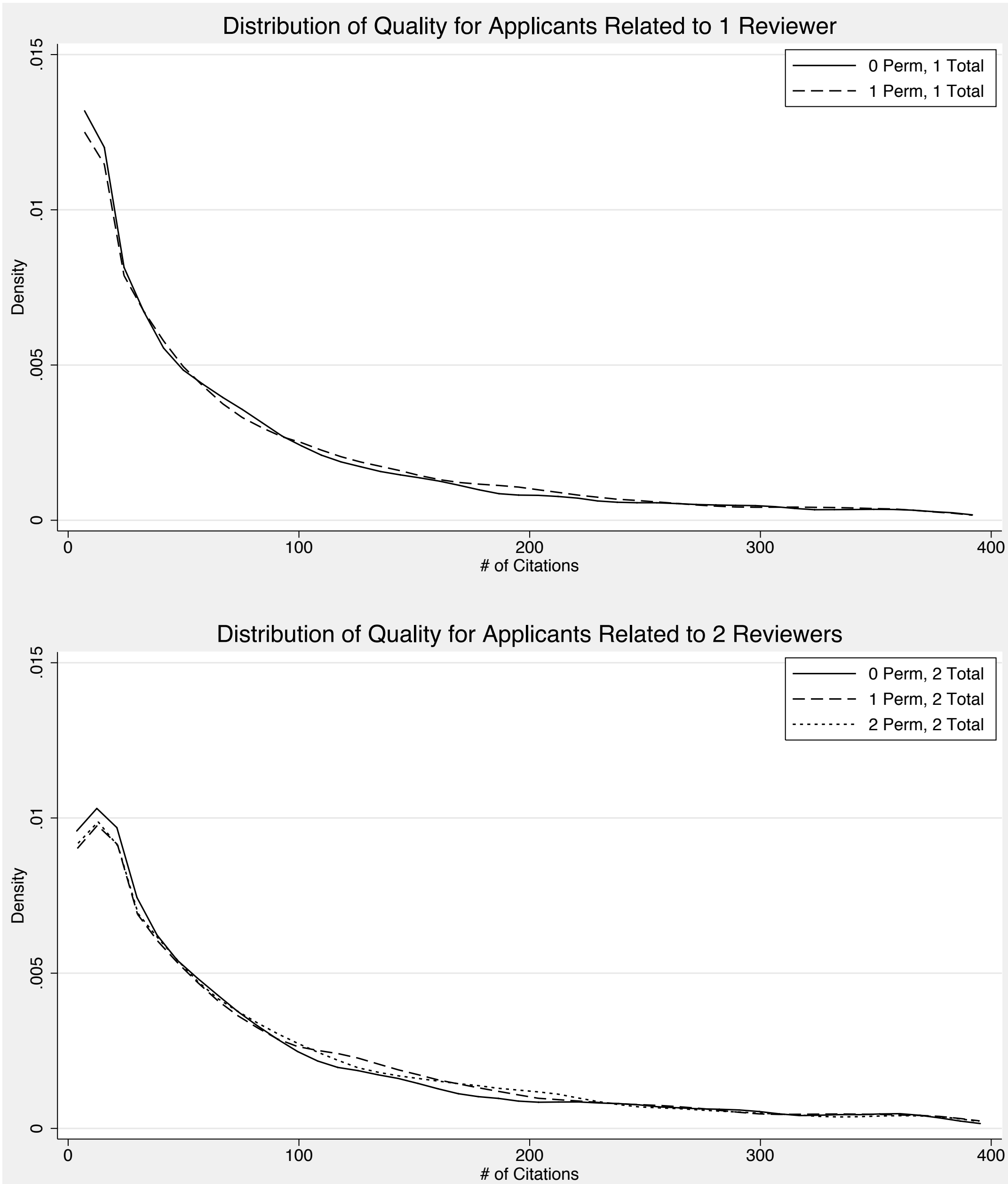

FIGURE 5: APPLICATION QUALITY CONDITIONAL ON TOTAL RELATED REVIEWERS

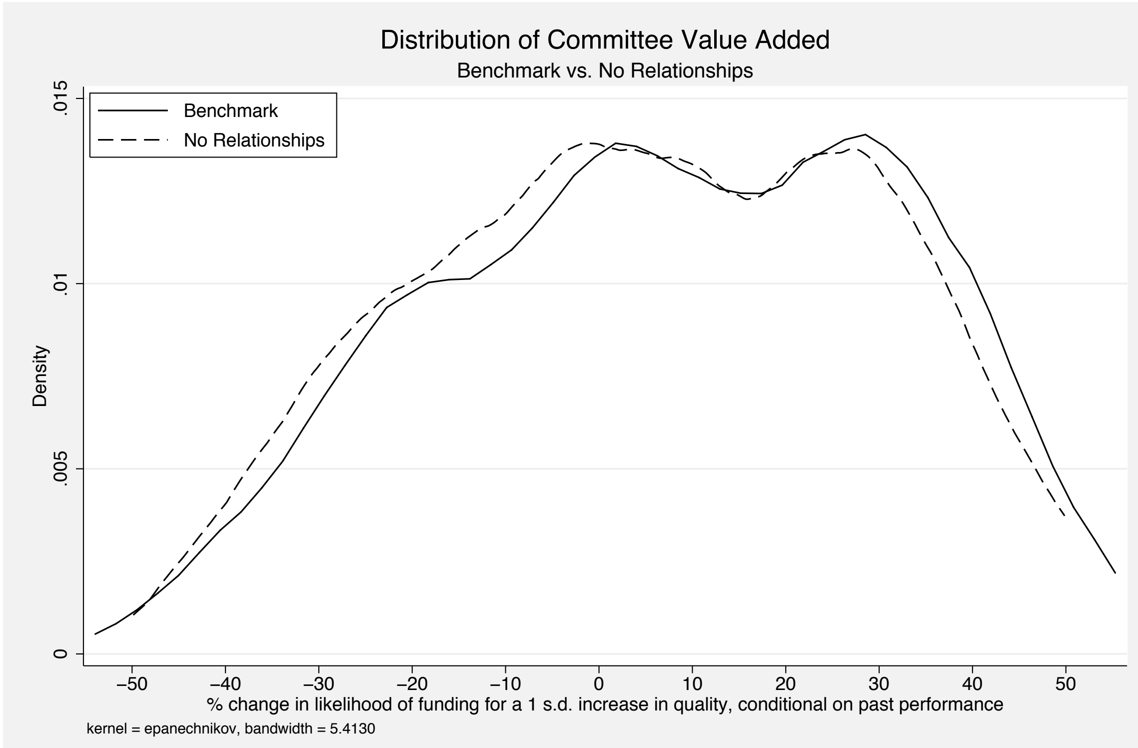

Figure 6: Distribution of meeting-level value-added

Table 1: Applicant Characteristics

	Roster-Matched Sample		Full Sample	
Sample Coverage		Std. Dev.		Std. Dev.
# Grants	93,558		156,686	
# Applicants	36,785		46,546	
Years	1992-2005		1992-2005	
# Study Sections	250		380	
# Study Section Meetings	2,083		4,722	
Grant Characteristics				
% Awarded	26.08		30.48	
% Scored	61.58		64.04	
% New	70.31		71.21	
Percentile Score	70.05	18.42	71.18	18.75
# Publications, grant-publication matched (median)	2	5	2	5
# Citations, grant-publication matched (median)	36	265	38	302
PI Characteristics				
% Female	23.21		22.58	
% Asian	13.96		13.27	
% Hispanic	5.94		5.79	
% M.D.	28.72		29.26	
% Ph.D.	80.46		79.69	
% New investigators	19.70		20.02	
# Publications, past 5 years	15	60	15	55
# Citations, past 5 years	416	1431	423	1474

Notes: The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. Future publications refers to the number of research articles that the grant winner publishes in the 2 years following the grant which share at least one salient word overlap between the grant project title and the publication title. Past publications include any first, second, and last authored articles published in the five years prior to applying for the grant. The full sample includes data from any new or competing R01 grant evaluated in chartered study sections from 1992 to 2005. Investigators with common names are dropped as are any for which the covariates are missing. Social science study sections are dropped.

Table 2: Does being funded directly affect my measure of quality?

Dep var: Grant Quality	(1) No score controls	(2) Controls for smooth function of score
1(Grant is funded)	0.0486***	0.0054
	(0.0053)	*(0.0104)*
Observations	100276	100276
R-squared	0.3329	0.3335
Past Performance, Past Grants, and Demographics	X	X

Notes: Coefficients are reported from a regression of grant quality on an indicator for whether the grant was funded and controls for applicant characteristics. Column (2) includes controls for quartics in the applicant score. Column (2) compares grant applications with the same score and the same characteristics but which differ in funding status. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to.

Table 3: Committee Descriptives

	Roster Matched Sample	
Reviewer Characteristics		Std. Dev.
# Reviewers	18,916	
# Permanent reviewers per meeting	17.23	4.52
# Temporary reviewers per meeting	12.35	7.44
# Meetings per permanent reviewer	3.69	3.03
# Meetings per temporary reviewer	1.78	1.30
# Applications	53.73	17.31
Relationship Characteristics		
# Reviewers who cite applicant	1.94	2.81
# Permanent reviewers who cite applicant	1.11	1.73
# Applicants cited by permanent reviewers	4.12	5.32
# Applicants cited by temporary reviewers	4.12	5.09

Notes: The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. Future publications refers to the number of research articles that the grant winner publishes in the 2 years following the grant which share at least one salient word overlap between the grant project title and the publication title. Past publications include any first, second, and last authored articles published in the five years prior to applying for the grant. Investigators with common names are dropped as are any for which the covariates are missing. Social science study sections are dropped.

Table 4: Characteristics permanent and temporary members

	Permanent	Temporary
Number of reviewers	9371	14067
Reviewer Characteristics		
% Female	31.68	24.28
% Asian	14.99	13.08
% Hispanic	6.40	5.05
% M.D.	27.42	25.85
% Ph.D.	79.45	80.99
# Publications, past 5 years (median)	22	21
# Citations, past 5 years (median)	606	590

Reviewer Transitions

	% Permanent in the Past	% Permanent in the Future	% Temporary in the Past	% Temporary in the Future
Current Permanent Members	61.87	63.71	38.11	35.45
Current Temporary Members	16.25	41.30	32.73	50.13

Notes: The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. Future publications refers to the number of research articles that the grant winner publishes in the 2 years following the grant which share at least one salient word overlap between the grant project title and the publication title. Past publications include any first, second, and last authored articles published in the five years prior to applying for the grant. Investigators with common names are dropped as are any for which the covariates are missing. Social science study sections are dropped. Transitions are calculated based on whether a reviewer is present in the roster database during the full sample years from 1992 to 2005. Means are taken for the years 1997 to 2002 in order to allow time to observe members in the past and future within the sample.

Table 5: What is the effect of being related to a reviewer on an applicant's likelihood of funding?

	(1)	(2)	(3)	(4)	(5)	(6)
	1(Score is above the payline)			**Score**		
	Mean = 0.214, SD = 0.410			Mean = 71.18, SD = 18.75		
Related Permanent Reviewers	0.0328***	0.0153***	0.0063***	1.1083***	0.5184***	0.2285**
	(0.0013)	*(0.0012)*	*(0.0020)*	*(0.0542)*	*(0.0517)*	*(0.0926)*
Total Related Reviewers			0.0067***			0.2163***
			(0.0014)			*(0.0601)*
Observations	93558	93558	93558	57613	57613	57613
R-squared	0.0630	0.0947	0.0950	0.1186	0.1433	0.1436
Committee $\times$ Year $\times$ Cycle FE	X	X	X	X	X	X
Past Performance, Past Grants, and		X	X		X	X

Notes: Coefficients are reported from a regression of committee decisions (score or funding status) on the number of permanent members related to an applicant, controlling for meeting level fixed effects. Column 2 includes indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to. Column 3 includes an additional control for the total number of related reviewers. The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. A reviewer is related to an applicant if the reviewer has cited any of the applicant's previous research in the 5 years prior to grant review.

Table 6: What is the contribution of bias and information?

	(1)	(2)	(3)	(4)
	1(Score is above the payline)		**Score**	
	Mean = 0.214, SD = 0.410		Mean = 71.18, SD = 18.75	
Related Permanent Reviewers	0.0063***	0.0061***	0.2285**	0.2102**
	(0.0020)	*(0.0020)*	*(0.0926)*	*(0.0926)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations		0.0106**		0.2202
		(0.0049)		*(0.2230)*
Standardized Future Citations		0.0315***		1.1674***
		(0.0039)		*(0.1812)*
Total Related Reviewers × Standardized Future Citations		-0.0016**		-0.0524**
		(0.0006)		*(0.0236)*
Total Related Reviewers	0.0067***	0.0072***	0.2163***	0.2403***
	(0.0014)	*(0.0014)*	*(0.0601)*	*(0.0608)*
Observations	93558	93558	57613	57613
R-squared	0.0950	0.0980	0.1436	0.1453
Committee × Year × Cycle FE	X	X	X	X
Past Performance, Past Grants, and Demographics	X	X	X	X

Notes: Coefficients are reported from a regression of committee decisions (score or funding status) on the variables reported, controlling for meeting level fixed effects and detailed applicant characteristics. Columns 1 and 3 reproduce Columns 3 and 6 from Table 5. Columns 2 and 4 add controls for application quality and application quality interacted with relatedness to permanent and all reviewers. The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. A reviewer is related to an applicant if the reviewer has cited any of the applicant's previous research in the 5 years prior to grant review. Future citations are standardized to be mean zero, standard deviation 1 within each committee-year. Future citations are calculated using all publications by an applicant in the -1 to 2 years after grant review, with text matching. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to.

Table 7: What is the contribution of bias and information? Heterogeneity in applicant and grant type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1(Score is above the payline)				**Score**			
	Mean = 0.214, SD = 0.410				Mean = 71.18, SD = 18.75			
	New Investigators		Experienced Investigators		New Grants		Renewal Grants	
Related Permanent Reviewers	-0.0023	-0.0023	0.0071***	0.0068***	0.0040*	0.0038	0.0079**	0.0074**
	(0.0056)	*(0.0057)*	*(0.0022)*	*(0.0022)*	*(0.0024)*	*(0.0025)*	*(0.0036)*	*(0.0035)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations		-0.0072		0.0120**		0.0029		0.0189*
		(0.0139)		*(0.0053)*		*(0.0059)*		*(0.0098)*
Standardized Future Citations		0.0361***		0.0305***		0.0311***		0.0234***
		(0.0090)		*(0.0043)*		*(0.0045)*		*(0.0084)*
Total Related Reviewers × Standardized Future Citations		0.0011		-0.0017***		-0.0002		-0.0023**
		(0.0030)		*(0.0007)*		*(0.0008)*		*(0.0009)*
Total Related Reviewers	0.0101***	0.0100***	0.0063***	0.0069***	0.0071***	0.0071***	0.0036	0.0047*
	(0.0036)	*(0.0036)*	*(0.0015)*	*(0.0015)*	*(0.0015)*	*(0.0015)*	*(0.0024)*	*(0.0024)*
Observations	18428	18428	75130	75130	65776	65776	27782	27782
R-squared	0.1768	0.1797	0.0964	0.0992	0.0807	0.0836	0.1622	0.1643
Committee × Year × Cycle FE	X	X	X	X	X	X	X	X
Past Performance, Past Grants, and Demographics	X	X	X	X	X	X	X	X

Notes: See notes to Table 6. Coefficients are reported from a regression of committee decisions (score or funding status) on the variables reported, controlling for meeting level fixed effects and detailed applicant characteristics. Columns 1 and 3 reproduce Columns 3 and 6 from Table 5. Columns 2 and 4 add controls for application quality and application quality interacted with relatedness to permanent and all reviewers. New Investigators are those who have not received an R01 in the past. New grants are those that are about a new subject, not a renewal of an existing grant.

Table 8: What is the effect of relationships on the quality of research that the NIH supports?

	Benchmark	**No Relationships**
Number of Funded Grants	24,404	24,404
Number of Grants that Change Funding Status	2,166	2,166
Total # Citations	6,680,590	6,547,750
(% change relative to benchmark)		*-1.99*
Total # Publications	149,600	145,331
(% change relative to benchmark)		*-2.85*
Total # in Top 99% of Citations	10,035	9,815
(% change relative to benchmark)		*-2.19*
Total # in Top 90% of Citations	58,149	56,724
(% change relative to benchmark)		*-2.45*
Total # in Top 50% of Citations	132,490	128,980
(% change relative to benchmark)		*-2.65*
Total # Related Applicants Funded	18,059	17,431
(% change relative to benchmark)		*-3.48*

Notes: Benchmark refers to characteristics of grants ordered according to their predicted probability of funding, using the main regression in Table 6 of funding status on relationships and other characteristics. No relationships refers to ordering of grants under the assumption that relatedness to permanent members and relatedness to permanent members interacted with quality do not matter (their coefficients are set to zero). Expected citations are calculated as fitted values from a regression of citations on relationships, past performance, demographics, and meeting fixed effects. The number of projects that are funded is kept constant within meeting. See text for details.

Table 9: Do highly performing committees make better use of related reviewers?

	(1)	(2)	(3)	(4)
Dep var: 1(Score > payline) Mean = 0.214, SD = 0.410	**Value-added < Median**	**Value-added > Median**	**Value-added bottom tercile**	**Value-added top tercile**
Related Permanent Reviewers	0.0066	0.0017	0.0044	0.0033
	(0.0043)	*(0.0053)*	*(0.0067)*	*(0.0069)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations	0.0034	0.0123	-0.0081	0.0307**
	(0.0106)	*(0.0101)*	*(0.0114)*	*(0.0142)*
Standardized Future Citations	-0.0073	0.0635***	-0.0126	0.0772***
	(0.0064)	*(0.0087)*	*(0.0076)*	*(0.0143)*
Total Related Reviewers × Standardized Future Citations	0.0007	-0.0039***	0.0009	-0.0064***
	(0.0012)	*(0.0014)*	*(0.0015)*	*(0.0013)*
Total Related Reviewers	0.0051	0.0121***	0.0055	0.0091**
	(0.0032)	*(0.0031)*	*(0.0055)*	*(0.0038)*
Observations	34494	34385	22962	23129
R-squared	0.0842	0.1101	0.0845	0.1173
Committee × Year × Cycle FE	X	X	X	X
Past Performance, Past Grants, and Demographics	X	X	X	X

Notes: Coefficients are reported from a regression of committee decisions (score or funding status) on the variables reported, controlling for meeting level fixed effects and detailed applicant characteristics. The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. I make the additional restriction that the sample be limited to those committees for which I have value-added data. These are typically committees that I observe meeting at least three times. A reviewer is related to an applicant if the reviewer has cited any of the applicant's previous research in the 5 years prior to grant review. Future citations are standardized to be mean zero, standard deviation 1 within each committee-year. Future citations are calculated using all publications by an applicant in the -1 to 2 years after grant review, with text matching. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to.

APPENDIX MATERIALS: FOR ONLINE PUBLICATION

A Proof of Proposition 4.1

Nature has drawn true quality Q^*, and types $Q = \begin{cases} Q_R = Q^* + \varepsilon_R & \text{if } R = 1 \\ Q_{UR} = Q^* + \varepsilon_{UR} & \text{if } R = 0 \end{cases}$

Given this, the Perfect Bayesian equilibrium for this game is characterized by:

1. A set of beliefs that the committee has about true quality Q^* given the message M: $\mu(Q^*|M)$.
2. A message strategy $M(Q)$ for a reviewer, given his or her posterior Q.
3. A decision strategy $D(M)$ for the committee, given the reviewer's message.

These strategies and beliefs must be optimal in the following sense:

1. For each Q^*, $\int_{M\in\mathbf{M}} \mu(Q^*|M)dM = 1$.
2. For each message M, the committee's decision $D(M)$ must maximize its expected payoffs given their beliefs $\mu(Q^*|M)$:

$$D \in \text{argmax} \int_{Q^*\in\mathbf{Q}^*} P^C(D, Q^*)\mu(Q^*|M)dQ^*$$

3. For each posterior Q, the reviewer's message $M(Q)$ must maximize his/her payoffs given the committee's strategy:

$$M \in \text{argmax} \int_{Q^*\in\mathbf{Q}^*} P(D(M), Q^*)f(Q^*|Q)dQ^*, \qquad \text{for } P = \{P^{UR}, P^R\}$$

   where $f(\cdot|Q)$ is the density of Q^* given Q.

4. For all reviewer posteriors $Q \in \mathbf{Q}^M$ that induce message M to be sent with positive probability, committee beliefs $\mu(Q^*|M)$ must follow from Bayes' Rule:

$$\mu(Q^*|M) = \frac{\int_{Q^*\in\mathbf{Q}^*} M(Q)f(Q^*|Q)dQ^*}{\int_{Q\in\mathbf{Q}^M}\int_{Q^*\in\mathbf{Q}^*} M(Q)f(Q^*|Q)dQ^*dQ}$$

Having defined the equilibrium concept, I proceed with the proof.

Case 1. Suppose that the reviewer reports her exact posterior and the committee to believes it. In this case, the committee maximizes its utility by funding the proposal if and only if $Q^* + \varepsilon_{UR} > U$. The reviewer has no incentive to deviate from this strategy because she is receiving her highest payoff as well.

Suppose, now, that there were another informative equilibrium. Each message $M \in \mathbf{M}$ induces a probability of funding $D(M)$. Let the messages be ordered such that $D(\mathbf{M}_1) \leq \cdots \leq D(\mathbf{M}_K)$

where $\mathbf{M}_i$ are the set of messages M_i that induce the same probability of funding $D(M_i)$. For reviewers of type $Q^* + \varepsilon_{UR} > U$, the reviewer strictly prefers that the grant be funded. She thus finds it optimal to send the message $\mathbf{M}_K$ that maximizes the probability that the grant is funded. Call this set Y. For $Q^* + \varepsilon_{UR} < U$ the reviewer strictly prefers that the grant be unfunded and sends messages in $\mathbf{M}_1$. Call this set N. The only reviewer who sends any other message is one for which $Q^* + \varepsilon_{UR} = U$. This occurs with probability zero. Thus, with probability one, the space of possible messages is equivalent to $\mathbf{M} = \{Y, N\}$. For this equilibrium to be informative, it must be that $D(N) < D(Y)$.

Given this, the committee's optimal reaction is to fund when $M = Y$ and to reject otherwise. Thus, this equilibrium is payoff equivalent to the first equilibrium. If the we allow uninformative equilibria, $D(\mathbf{M}_1) = \cdots = D(\mathbf{M}_K)$ and any reviewer message is permissible. It must be that $D(M_i) = 0$ for all M_i because the outside option U is assumed to be greater than the committee's prior on quality.

Case 2.

Now consider the case when the reviewer is related and biased. As in Case 1, the set of messages is equivalent, with probability one, to $\mathbf{M} = \{Y, N\}$. In this case, however, reviewers of type $E(Q^*|Q^* + \varepsilon_R) > U - B$ send $M = Y$ and reviewers of type $E(Q^*|Q^* + \varepsilon_R) < U - B$ send $M = N$. The only reviewer who sends any other message is one for which $E(Q^*|Q^* + \varepsilon_R) = U - B$.

$E(Q^*|E(Q^*|Q^* + \varepsilon_R) > U - B) > U$

Under this strategy, the committee's expectation of Q^* given $M = N$ is $E(Q^*|E(Q^*|Q^* + \varepsilon_R) < U - B)$. Since this is less than U, the grant goes unfunded. The committee's expectation of Q^* given $M = Y$ is $E(Q^*|E(Q^*|Q^* + \varepsilon_R) < U - B)$. When this is larger than U, the committee listens to the reviewer's recommendation and we can verify that $D(Y) > D(N)$. There also exists an uninformative equilibria where all grants are rejected.

When $E(Q^*|E(Q^*|Q^* + \varepsilon_R) < U - B) < U$, the grant is never funded: $D(Y) = D(N) = 0$. In this case, only babbling equilibria exist.

B Proof of Proposition 5.1

Measurement error in Q^* can potentially affect the estimation of α_2 in Equation (6). The presence of U, RU, and X, however, will not affect consistency; for simplicity, I rewrite both the regression suggested by the model and the actual estimating equation with these variables partialed out. The remaining variables should then be thought of as conditional on U, RU, and X

$$D = \alpha_0 + \alpha_1 Q^* + \alpha_2 R + \alpha_3 RQ^* + \varepsilon \tag{12}$$

$$\begin{aligned} D &= a_0 + a_1 Q + a_2 R + a_3 RQ + e \\ &= a_0 + W + a_2 R + e, W = a_1 Q + a_3 RQ \end{aligned}$$

The coefficient a_2 is given by:

$$a_2 = \frac{\text{Var}(W)\text{Cov}(D,R) - \text{Cov}(W,R)\text{Cov}(D,W)}{\text{Var}(W)\text{Var}(R) - \text{Cov}(W,R)^2} \tag{13}$$

Consider $\text{Cov}(W,R)$:

$$\begin{aligned} \text{Cov}(W,R) &= \text{Cov}(a_1(Q^*+v) + a_3 R(Q^*+v), R) \\ &= a_1\text{Cov}(Q^*,R) + a_1\text{Cov}(v,R) + a_3\text{Cov}(RQ^*,R) + a_3\text{Cov}(Rv,R) \end{aligned}$$

Under the assumption that R and Q^* are conditionally independent, this yields:

$$\begin{aligned} \text{Cov}(W,R) &= a_3\text{Cov}(RQ^*,R) + a_3\text{Cov}(Rv,R) \\ &= a_3\left[E(R^2Q^*) - E(RQ^*)E(R)\right] + a_3\left[E(R^2v) - E(Rv)E(R)\right] \\ &= a_3\left[E(R^2)E(Q^*) - E(R)^2E(Q^*)\right] + a_3\left[E(R^2)E(v) - E(R)^2E(v)\right] \\ &= a_3\left[E(R^2)0 - E(R)^2 0\right] + a_3\left[E(R^2)0 - E(R)^2 0\right] && (14) \\ &= 0 && (15) \end{aligned}$$

With this simplification, the expression for the estimated coefficient on a_2 becomes:

$$\begin{aligned} a_2 &= \frac{\text{Var}(W)\text{Cov}(D,R) - \text{Cov}(W,R)\text{Cov}(D,W)}{\text{Var}(W)\text{Var}(R) - \text{Cov}(W,R)^2} \\ &= \frac{\text{Var}(W)\text{Cov}(D,R)}{\text{Var}(W)\text{Var}(R)} \\ &= \frac{\text{Cov}(D,R)}{\text{Var}(R)} \\ &= \frac{\text{Cov}(\alpha_0 + \alpha_1 Q^* + \alpha_2 R + \alpha_3 RQ^* + \varepsilon, R)}{\text{Var}(R)} \\ &= \frac{\alpha_2\text{Var}(R) + \alpha_3\text{Cov}(RQ^*,R)}{\text{Var}(R)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\alpha_2 \mathrm{Var}(R) + \alpha_3 \left[E(R^2)E(Q^*) - E(R)^2 E(Q^*)\right]}{\mathrm{Var}(R)} \\
&= \alpha_2
\end{aligned}
$$

C Robustness Checks

Appendix Table A addresses concerns that funding may directly influence the number of citations produced by a grant. Instead of including articles published up to two years after a grant is reviewed, Appendix Table A restricts my analysis to articles published one year before a grant is reviewed up to one year afterward. These publications are highly likely to be based off research that existed before the grant was reviewed. Using this metric, I find nearly identical measures of bias and information.

Another test of my assumption that citations are not directly affected by funding is to ask whether I find bias in the review of inframarginal grants, that is grants that are well above or well below the funding margin. All grants in either group have the same funding status so any bias I find cannot be attributed to differences in funding. Because I hold funding status constant, I can only assess the impact that related permanent members have on an applicant's score not on an applicant's funding status. Appendix Table B reports these results. In Columns 2 and 3, I report estimates of the effect of bias and information in the sample of funded and unfunded grants, respectively. In both cases, I still find evidence that bias exists. One concern is that relationships can still affect funding at the margin. In order to isolate a set of applications for which relationships could not have affected funding status, I consider grants that receive scores well above or well below the payline. Although my estimates on these subsamples are noisier, I still find evidence that bias exists. The magnitudes are somewhat smaller than in my main regression; because these are subsamples, there is no reason to expect that the magnitude of the effect of relationships should be the same for high- and low-quality grants as it is for the entire sample.

Publications associated with funded grants can also be matched using grant acknowledgments that are recorded in the National Library of Medicine's PubMed database. For the set of funded grants, Appendix Table C reruns my core regressions using citations to publications that explicitly acknowledge a grant as my measure of quality. This analysis differs slightly from my main results using citations because general citations cannot be computed for publications in PubMed. A limited set of citations can, however, be computed using publications in PubMed Central (PMC). PMC contains a subset of life sciences publications made available for free. While this is not as comprehensive a universe as that of Web of Science, it contains, for recent years, all publications supported by NIH dollars. Undercounting of publications would, further, not bias my result as long as it does not vary systematically by whether an applicant is related to a permanent or to a temporary member. I find results that are consistent with my primary findings. In fact, the magnitude of bias I find using explicit grant acknowledgements on the sample of funded grants is the same as the magnitude of bias I find using text-matching publications on this same subsample, as reported in Appendix Table B.

Appendix Table D provides evidence that permanent members do indeed have more influence. In my sample, I observe almost 5,000 reviewers serving both as permanent and as temporary members. For this subset of reviewers, I show that a larger proportion of the applicants whom they have cited are funded when the reviewer is permanent than when the reviewer is temporary, conditional on applicant qualifications. I also show that mean scores for applicants related to a reviewer are higher when that reviewer is permanent. These regressions include reviewer fixed effects.

Appendix Table E adds nonlinearity to Equation (9) in order to show that my results are robust to the assumption that error on the reviewer's posteriors in Section 3 is uniform and that $E(Q^*|Q^* + \varepsilon)$ is approximated by $\lambda(Q^* + \varepsilon)$. Without these assumptions, the association between relatedness and quality would, in general, be nonlinear. To show that this does not make a material difference for my results, I allow for relatedness to permanent reviewers R^P, relatedness to all reviewers R, and quality Q to vary flexibly by including controls for quadratics and cubics in Q, as well as quadratics and cubics of Q interacted with R^P and interacted with R. I find similar results, both qualitatively and quantitatively. In fact, my estimated bias parameter is almost exactly identical.

My results are robust to non-parametric controls for the total number of related applicants (meeting by number of related reviewers fixed effects) and using alternative definitions of relatedness, including using applicant-reviewer mutual citations and citations defined only on publications for which applicants and reviewers are primary authors (first, second, and last position). My results are also robust to alternative identification based on the attendance of reviewers at meetings as opposed to differences between permanent and temporary members. These and other detailed tables are available from the author.

D Estimating Committee Value-Added

I estimate committee value-added using the following regression:

$$D_{icmt} = a + b_{cmt}Q_{icmt} + \mu X_{icmt} + \delta_{cmt} + e_{icmt} \tag{16}$$

D_{icmt} is either the actual or counterfactual funding decision for applicant i reviewed during meeting m of committee c in year t. Q_{icmt} is a measure of application quality such as the number of citations it produces in the future and X_{icmt} are detailed controls for the past performance of the applicant, including flexible controls for number of past publications and citations, number and type of prior awarded grants and prior applications, and flexible controls for degrees, gender, and ethnicity. Finally, δ_{cmt} are committee meeting level fixed effects. The coefficients b_{cmt} capture, for each meeting, the correlation between decisions and quality, conditional on X_{icmt}.

Variation in b_{cmt} include sampling error so that $\hat{b}_{cmt}$ is a combination of true value-added plus a noise term. I assume this luck term to be independent and normal:

$$\hat{b}_{cmt} = b^*_{cmt} + \nu_{cmt} \tag{17}$$

Under this assumption, $\mathrm{Var}(\hat{b}_{cmt}) = \mathrm{Var}(b^*_{cmt}) + \mathrm{Var}(\nu_{cmt})$ so that the estimate of true variance is upwardly biased from the additional variance arising from estimation error. To correct for this, I note that the best estimate for b^*_{cmt} is given by $\mathrm{E}(b^*_{cmt}|\hat{b}_{cmt}) = \lambda_{ct}\hat{b}_{cmt} + (1 - \lambda_{ct})\bar{\hat{b}}_{ct}$ where $\bar{\hat{b}}_{ct}$ is the mean of meeting quality for that committee-year and $\lambda_{ct} = \frac{\sigma^2_{b^*_{cmt}}}{\sigma^2_{b^*_{cmt}} + \sigma^2_{\nu_{cmt}}}$ is a Bayesian shrinkage term constructed as the ratio of the estimated variance of true committee effects, $\sigma^2_{b^*_{cmt}}$, to the sum of estimated true variance $\sigma^2_{b^*_{cmt}}$ and estimated noise variance $\sigma^2_{\nu_{cmt}}$.

To derive this shrinkage term, I use the correlation in meeting quality across the three different funding cycles of a committee fiscal year. In particular, if meeting-specific errors are independent, then $\mathrm{Cov}(\hat{b}_{cmt}, \hat{b}_{cm't}) = \mathrm{Var}(b^*_{cmt}) = \hat{\sigma}^2_{b^*_{cmt}}$. This can be estimated at the committee-year level because a committee meets three times during the year. I construct

$$\hat{\lambda}_{ct} = \frac{\hat{\sigma}^2_{b^*_{cmt}}}{\hat{\sigma}^2_{b^*_{cmt}} + \hat{\sigma}^2_{\nu_{cmt}}} \tag{18}$$

so that the adjusted committee value-added is given by:

$$VA_{cmt} = \hat{\lambda}_{ct}\hat{b}_{cmt} \tag{19}$$

Because committee membership is not fixed across funding cycles within the same fiscal year (temporary members rotate, permanent members do not), variation in VA_{cmt} represents a conservative lower bound on the variance of committee quality.

Appendix Table A: What is the contribution of bias and information?
Quality measured by publications 1 year before to 1 year after grant

	(1)	(2)	(3)	(4)
	1(Score is above the payline)		**Score**	
	Mean = 0.214, SD = 0.410		Mean = 71.18, SD = 18.75	
Related Permanent Reviewers	0.0063***	0.0062***	0.2285**	0.2166**
	(0.0020)	*(0.0020)*	*(0.0926)*	*(0.0925)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations		0.0101**		0.1769
		(0.0048)		*(0.2057)*
Standardized Future Citations		0.0261***		0.9883***
		(0.0038)		*(0.1687)*
Total Related Reviewers × Standardized Future Citations		-0.0013**		-0.0359
		(0.0006)		*(0.0222)*
Total Related Reviewers	0.0067***	0.0071***	0.2163***	0.2317***
	(0.0014)	*(0.0014)*	*(0.0601)*	*(0.0609)*
Observations	93558	93553	57613	57608
R-squared	0.0950	0.0976	0.1436	0.1451
Committee × Year × Cycle FE	X	X	X	X
Past Performance, Past Grants, and Demographics	X	X	X	X

Notes: Coefficients are reported from a regression of committee decisions (score or funding status) on the variables reported, controlling for meeting level fixed effects and detailed applicant characteristics. Column 1 and 3 reproduce Columns 3 and 6 from Table 5. Column 2 and 4 add controls for application quality and application quality interacted with relatedness to permanent and all reviewers. The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. A reviewer is related to an applicant if the reviewer has cited any of the applicant's previous research in the 5 years prior to grant review. Future citations are standardized to be mean zero, standard deviation 1 within each committee-year. Future citations are calculated using all publications by an applicant in the -1 to 1 years after grant review, with text matching. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the pa and/or a Ph.D., and indicators for the number of past R01and other NIH grants an applicant has won and indicators for how many she has applied to.

Appendix Table B: What is the contribution of bias and information?
Inframarginal grant applications

Dep var: Score Mean = 71.18, SD = 18.75	(1) **All**	(2) **Funded**	(3) **Not Funded**	(4) **Well above payline**	(5) **Well below payline**
Related Permanent Reviewers	0.2102**	0.1252*	0.1492*	0.1118*	0.1132
	(0.0926)	*(0.0725)*	*(0.0889)*	*(0.0636)*	*(0.0821)*
Reviewers) × Standardized Future Citations	0.2202	0.3827**	-0.0396	0.0877	0.0642
	(0.2230)	*(0.1748)*	*(0.2410)*	*(0.1658)*	*(0.1939)*
Standardized Future Citations	1.1674***	0.0002	0.4974**	0.1960	0.0746
	(0.1812)	*(0.1382)*	*(0.2031)*	*(0.1323)*	*(0.1561)*
Total Related Reviewers × Standardized Future Citations	-0.0524**	-0.0266	0.0179	-0.0195	0.0029
	(0.0236)	*(0.0178)*	*(0.0261)*	*(0.0162)*	*(0.0216)*
Total Related Reviewers	0.2403***	0.0100	0.1343**	-0.0252	0.0366
	(0.0608)	*(0.0470)*	*(0.0578)*	*(0.0399)*	*(0.0523)*
Observations	57613	24395	33218	14800	22835
R-squared	0.1453	0.1747	0.1880	0.2491	0.7590
Committee × Year × Cycle FE	X	X	X	X	X
Past Performance, Past Grants, and Demographics	X	X	X	X	X

Notes: Coefficients are reported from a regression of committee decisions (score or funding status) on the variables reported, controlling for meeting level fixed effects and detailed applicant characteristics. The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. A reviewer is related to an applicant if the reviewer has cited any of the applicant's previous research in the 5 years prior to grant review. Future citations are standardized to be mean zero, standard deviation 1 within each committee-year. Future citations are calculated using all publications by an applicant in the -1 to 2 years after grant review, with text matching. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to.

Appendix Table C: What is the contribution of bias and information? Explicit grant acknowledgements for the sample of funded grants

	(1)	(2)
Dep var: Score Mean = 71.18, SD = 18.75	**Explict Grant Acknowledgements**	
Related Permanent Reviewers	0.1384*	0.1285*
	(0.0724)	*(0.0734)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations		0.0749
		(0.1004)
Standardized Future Citations		0.4806***
		(0.0770)
Total Related Reviewers × Standardized Future Citations		-0.0191*
		(0.0110)
Total Related Reviewers	-0.0074	0.0086
	(0.0456)	*(0.0472)*
Observations	24395	24395
R-squared	0.1743	0.1793
Committee × Year × Cycle FE	X	X
Past Performance, Past Grants, and Demographics	X	X

Notes: Coefficients are reported from a regression of committee decisions (score or funding status) on the variables reported, controlling for meeting level fixed effects and detailed applicant characteristics. The analytic sample includes all awarded R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. A reviewer is related to an applicant if the reviewer has cited any of the applicant's previous research in the 5 years prior to grant review. Future citations are standardized to be mean zero, standard deviation 1 within each committee-year. Future citations are calculated explicit grant acknowlegments. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to.

Appendix Table D: Do permanent reviewers have more influence?

	(1) Proportion of Related Applicants who are Funded	(2) Average Score of Related Applicants
Related Reviewer is Permanent	0.003***	0.336**
	(0.001)	*(0.144)*
Observations	15871	15870
R-squared	0.954	0.571
Reviewer FE	X	X
Past Performance, Past Grants, and Demographics	X	X

Notes: This examines how outcomes for related applicants vary by whether the related reviewer is serving in a permanent or temporary capacity. The sample is restricted to 4909 reviewers who are observed both in temporary and permanent positions. An applicant is said to be related by citations if a reviewer has cited that applicant in the 5 years prior to the meeting. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to.

Appendix Table E: What is the contribution of bias and information?
Nonlinear controls for quality and relatedness

	(1)	(2)	(3)	(4)
	1(Score is above the payline)		**Score**	
	Mean = 0.214, SD = 0.410		Mean = 71.18, SD = 18.75	
Related Permanent Reviewers	0.0063***	0.0062***	0.2285**	0.2240**
	(0.0020)	*(0.0021)*	*(0.0926)*	*(0.0948)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations		0.0188**		0.8273**
		(0.0073)		*(0.3900)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations^2		-0.0010		-0.2431
		(0.0047)		*(0.2347)*
1(1+ Related Permanent Reviewers) × Standardized Future Citations^3		-0.0002		0.0141
		(0.0007)		*(0.0300)*
Standardized Future Citations		0.0644***		2.2399***
		(0.0058)		*(0.2997)*
Standardized Future Citations^2		-0.0225***		-0.6377***
		(0.0044)		*(0.2038)*
Standardized Future Citations^3		0.0022***		0.0575**
		(0.0007)		*(0.0281)*
Total Related Reviewers × Standardized Future Citations		-0.0010		-0.0539
		(0.0014)		*(0.0660)*
Total Related Reviewers × Standardized Future Citations^2		0.0006		0.0299
		(0.0006)		*(0.0279)*
Total Related Reviewers × Standardized Future Citations^3		-0.0001		-0.0038
		(0.0001)		*(0.0026)*
Total Related Reviewers	0.0067***	0.0065***	0.2163***	0.2106***
	(0.0014)	*(0.0014)*	*(0.0601)*	*(0.0607)*
Observations	93558	93558	57613	57613
R-squared	0.0950	0.0994	0.1436	0.1464
Committee × Year × Cycle FE	X	X	X	X
Past Performance, Past Grants, and Demographics	X	X	X	X

Notes: The analytic sample includes new or competing R01 grants evaluated in charterd study sections from 1992 to 2005, for which I have study section attendance data. A reviewer is related to an applicant if the reviewer has cited any of the applicant's previous research in the 5 years prior to grant review. Future citations are standardized to be mean zero, standard deviation 1 within each committee-year. Future citations are calculated using all publications by an applicant in the -1 to 2 years after grant review, with text matching. Full controls include 1(1+ Related Permanent Reviewers) X Standardized Future Citations in cubics, Total Related Reviewers X Standardized Future Citations in cubics, and Standardized Future Citations in cubics. Applicant characteristics include indicators for sex and whether an applicant's name is Hispanic, East Asian, or South Asian, quartics in an applicant's total number of citations and publications over the past 5 years, indicators for whether an applicant has an M.D. and/or a Ph.D., and indicators for the number of past R01 and other NIH grants an applicant has won and indicators for how many she has applied to.