Association of percentile ranking with citation impact and productivity in a large cohort of *de novo* NIMH-funded R01 grants

JM Doyle[1], **K Quinn**[1], **YA Bodenstein**[1], **CO Wu**[2], **N Danthi**[3], and **MS Lauer**[4]

[1]Office of Science Policy, Planning, and Communications, National Institute of Mental Health (NIMH), Bethesda, MD, USA

[2]Office of Biostatistics Research, Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute (NHLBI), Bethesda, MD, USA

[3]Advanced Technologies and Surgery Branch, Division of Cardiovascular Sciences of the National Heart, Lung, and Blood Institute (NHLBI), Bethesda, MD, USA

[4]Office of the Director, Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute (NHLBI), Bethesda, MD, USA

Abstract

Previous reports from National Institutes of Health and National Science Foundation have suggested that peer review scores of funded grants bear no association with grant citation impact and productivity. This lack of association, if true, may be particularly concerning during times of increasing competition for increasingly limited funds. We analyzed the citation impact and productivity for 1755 de novo investigator-initiated R01 grants funded for at least 2 years by National Institute of Mental Health between 2000 and 2009. Consistent with previous reports, we found no association between grant percentile ranking and subsequent productivity and citation impact, even after accounting for subject categories, years of publication, duration and amounts of funding, as well as a number of investigator-specific measures. Prior investigator funding and academic productivity were moderately strong predictors of grant citation impact.

Introduction

National Institutes of Health (NIH) primarily rely on percentile rankings of peer-review priority impact scores to decide which investigator-initiated research grants they will fund. There is conflicting evidence, however, supporting funding decisions based on peer review alone.[1,2] Recent reports from the National Heart, Lung and Blood Institute (NHLBI)[3,4], the National Institute of General Medical Sciences[5], the National Science Foundation[6] and from

other sponsors have found either no or at best a modest[7] association between grant peer-review judgments and outcomes. Furthermore, one study from NHLBI suggests that, in contrast to peer-review percentile rankings, prior investigator productivity may be a moderately strong predictor of an R01 grant's eventual citation impact.[4] As percentile rankings are used by the National Institute of Mental Health (NIMH) to assist in funding decisions, we chose to examine whether we could replicate the NHLBI findings in an R01 grant cohort that would be entirely different in scientific focus and in investigator pool.

Materials and Methods

Grant sample and attributes

We identified all 1755 de novo ('Type 1') investigator-initiated R01 grants that met the following inclusion criteria: initial award between 2000 and 2009, receipt of a percentile ranking based on a peer-review priority score, project duration of at least 2 years and funding solely by NIMH.

We used internal tracking systems to secure data on a number of grant-based attributes, including the amount and duration of funding, submission status (whether the funded grant was a revision or not), use of animal and/or human subjects and focus on basic or applied research. At the time of award, trained NIMH staff prospectively coded grants as 'applied' using the following definition: 'The practical application of such knowledge for the purpose of meeting a recognized need. This includes the study of the detection, prevention, epidemiology, cause, treatment or rehabilitation of specific disorders or conditions. Applied research also encompasses studies of the structure, processes and effects of health services, the utilization of health resources and the analysis and evaluation of the delivery of health services, health-care costs and organizations.' Basic research was defined as, 'The gaining of fuller knowledge or understanding of the subject under study and its biological and behavioral processes, which affect disease and human wellbeing, including such areas as the cellular and molecular bases of diseases, genetics, immunology, neuropsychology and neurobiology.'

We used internal systems to gather data on principal investigators, including their home institution (which enabled us to consider total institutional NIMH R01 support during the study period), terminal degree, their academic rank, whether they were early-stage investigators (within 10 years of receiving their last terminal degree or have completed their medical residency within 10 years), their total prior NIH funding (in terms of dollars and in terms of project-years of support) and the number of NIH peer-review study sections they had previously served on.

Bibliometric measures and outcomes

We obtained grant-based publication data from NIH's Scientific Publication Information Retrieval and Evaluation Systems (SPIRES; https://era.nih.gov/nih_and_grantor_agencies/other/spires.cfm) and PubMed. Paper-based citation data were obtained from the Thomson Reuters' InCites database (http://researchanalytics.thomsonreuters.com/incites/). Critics have noted the weaknesses of measuring raw citation counts, as citation behaviors vary widely

across fields (for example, cardiovascular medicine versus psychiatry) and over time.[8] Therefore, we worked with Thomson Reuters to use their InCites database[9] to retrieve for each paper a citation percentile value, which was based on how frequently a paper is cited compared with contemporaneously published papers of the same type (for example, research article, letter) in the same field or category.[8] Hence, a biochemistry paper published in 2005 would only be compared with other biochemistry papers published in 2005. Subject categories were taken from a set of 252 empirically derived categories from the Web of Science; some articles were assigned to more than one subject category, in which case we used the best percentile value. For articles published in multi-disciplinary journals, Thomson Reuters focuses on the cited items within each publication to determine subject category. NIMH papers included 155 of the 252 categories; the most common categories were neurosciences (20%), psychiatry (14%), clinical psychology (6%), pharmacology (5%) and experimental psychology (4%). To assure maximum accuracy, we not only furnished Thomson Reuters with PMID numbers, we also sent a detailed EndNote library to allow for manual data cleansing.

We considered a paper to be 'Top-10%' if it had a percentile value of 0-10, meaning that it was cited more frequently than at least 90% of other same-field contemporary papers. For each paper, we calculated a 'normalized citation impact' as (100-citation percentile)/100. Thus, if a paper was the most highly cited paper in its field and in its year of publication, it would have a citation percentile value of 0, yielding a normalized citation impact of 1. If a paper was never cited, it would have a citation percentile of 100 and a normalized citation impact of 0. If a paper was cited more often than 75% of its same-field contemporaries, it would have a citation percentile of 25 and a normalized citation impact of 0.75. Of note, the InCites database yielded a percentile value for 95.3% of the papers identified by SPIRES and PubMed.

Our primary productivity end points were grant-based normalized citation impact per $million spent and number of top-10% papers per $million spent. We noted whether papers cited more than one grant, in which case we 'penalized' the bibliometric outcome for each grant to avoid superfluous counting.[3] Thus, if a paper cited two grants and had a normalized citation impact of 0.80, each grant was credited with a normalized citation impact of 0.40.

We used the InCites database to collect publication and citation data for each principal investigator's NIMH-funded papers published in the 5 years before their grant was reviewed. In this way, we could assess whether prior productivity could predict a grant's productivity. However, this left us without the ability to consider all of an investigator's prior work, so we used Scopus data to identify all publications issued in the 5 years before grant review for a random sample of 100 grants. We chose to perform this secondary analysis on a random sample because of the extensive work needed to disambiguate author names. For each paper, we also used Scopus data to obtain citation counts and annual citation rates.

Data analysis

We used a machine learning technique, Breiman random forests,[10] to investigate the relative importance of percentile ranking grant with citation impact. Random forests provide unbiased estimates of relative variable importance without making strict parametric

modeling assumptions. 'Variable importance' is based on permutation importance, which assesses the prediction accuracy of a prediction model before and after randomly permuting a given predictor. Larger values of importance indicate stronger predictors, and values around zero suggests that the variable does not predict for grant productivity or impact. Random forests not only have an added advantage of being adept at handling large numbers of covariates, but they also handle non-linear associations, complex interactions, non-normal distributions and multi-collinearity in a robust and unbiased way.[11] We used R statistical packages for all analyses, including the Harrell's rms and Hmisc packages for descriptive analyses, Wickam's ggplot2 (ref 12) for descriptive visualizations (with logarithmic transformations of outcomes due to skewed distributions) and Ishwaran's and Kogalur's randomForestSRC for random forest analyses.[13] A de-identified version of the main analysis data set and accompanying R code will be available to researchers upon request.

Results

Table 1 shows characteristics of grants and their principal investigators according to peer-review priority score-based percentile ranking. Grants with worse percentile rankings tended to receive lower budgets, to go to institutions that received less overall NIMH R01 funding, and to be focused more on animal research. Investigators of these grants were less productive in terms of NIMH-based normalized citation impact in the 5 years before their grant was reviewed, were less likely to have served on multiple NIH study sections, tended to have non-tenured positions and were recipients of less prior NIH grant funding.

After accounting for multiple grant acknowledgments per paper, the 1755 grants yielded 12,332 papers, with a total normalized citation impact of 8186. As expected, normalized citation impact increased with total award dollars (Supplementary Figure 1, top panel) with diminishing marginal returns (Supplementary Figure 1, middle panel). Normalized citation impact also increased with longer project durations (Supplementary Figure 1, bottom panel).

Table 2 shows descriptive bibliometric outcomes according to grant percentile ranking. By percentile ranking categories, we found no clear pattern linking grant percentile ranking to number of papers published, normalized citation impact, number of top-10% papers or for any of these outcomes per $million spent. Figure 1 shows the associations of per-grant normalized citation impact and top-10% papers according to grant percentile ranking and type of research (basic or applied); there was no association with grant percentile ranking, but basic research projects yielded more favorable outcomes. We noted similar patterns when using the raw grant priority score instead of the percentile ranking (Supplementary Figure 2).

Figures 2 and 3 show the main results from the random forest regression models. The most important predictors of citation impact per $million spent were longer project duration, lower total grant award (consistent with diminishing marginal returns), less prior principal investigator (PI) NIH grant funding, more prior project-years of funding, fewer mean number of grants cited per paper, type of research (basic or applied), greater prior normalized NIMH citation impact and fewer mean number of authors per paper. All other variables, including grant percentile ranking, were unimportant predictors. In a

supplementary multivariable linear regression analysis, project duration and total grant award also emerged as independent predictors ($P < 0.0001$ for both), whereas prior normalized NIMH citation impact was a moderately strong predictor ($P = 0.01$).

Figure 4 shows the main results of a random forest regression on a random sample of 100 grants in which we obtained prior publication data that were not limited to NIMH grants and that underwent manual name disambiguation. The three most important predictors of normalized citation impact per $million were a lower total grant award, longer project duration and greater annual citation rates for articles published by PI's in the 5 years before grant review.

Discussion

In a large cohort of *de novo* investigator-initiated R01 grants funded by NIMH, we found no association between peer-review-based grant percentile score and subsequent field-normalized citation impact. Our results are consistent with previous findings from two other NIH institutes, National Institute of General Medical Sciences[5] and NHLBI[3,4], and from a unit at the National Science Foundation.[6] Furthermore, consistent with previous findings from NHLBI[4] and from NIH-wide analyses[14], we found that productive applicants had better grant percentile rankings; consistent with previous NHLBI findings[4], we found a moderate association between prior PI productivity and grant citation impact. We also found that basic science grants yielded greater citation impact than applied science grants, a finding supportive of NIMH's increasing focus on basic brain science.[15]

The most important predictors of grant citation impact were project duration and total award amount. Long project duration predicted greater citation impact, a finding that is intuitive as the longer a grant remains active, the greater the number of publications that can be generated, which increases the potential visibility of the research proposed and allows time for a promising line of scientific inquiry to mature. The association between total grant award and citation impact followed a more nuanced pattern (Supplementary Figure 1), with diminishing marginal returns as funding increased.[5] Of note, we found parallel findings when focusing on prior PI activity; grant citation impact increased as prior PI project-years increased but also as prior PI total funding decreased (Figure 3).

Although we find no association between percentile rank and downstream grant outputs, findings presented here do not suggest peer review does not have any utility for Institutes, but rather underscores the importance of further work in this area. Some may argue that advances in science involve serendipitous events. Each grant proposal only represents the potential for highly impactful work and grant reviewers can only evaluate each grant's scientific promise—not what will actually materialize. Others may argue that peer review distills out the least promising work from being discussed during a review meeting, leaving proposals with almost equal probability of becoming successes.[6]

We acknowledge a number of limitations. First, we only used bibliometric indicators to assess grant productivity and impact. We did not consider changes in existing treatment practices, reductions in disease burden, development of novel drugs and therapeutics, and

openings in new avenues of research. Second, we restricted our analyses to data from NIMH. Thus, findings cannot be generalized to all possible areas, agencies, institutes, or centers funding neuroscience or to NIH overall. Third, we only focused on grants that were actually funded, and thus may be limited by looking at a narrow window of the best projects. Others have considered funded and unfunded projects and have found varying degrees of predictive validity of peer review scores.[2,7,16]

Despite these limitations, we confirm previous analyses performed at other federal funding agencies that showed no clear monotonic association between grant peer review percentile ranking and subsequent grant productivity and citation impact. As NIH confront increased fiscal uncertainty each year, the commensurate need to make informed funding decisions using information available at the time of application and review will only intensify, making the findings presented here both timely and important.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Demicheli V, Di Pietrantonj C. Peer review for improving the quality of grant applications. Cochrane Database Syst Rev. 2007:MR000003. [PubMed: 17443627]

2. Bornmann L, Daniel HD. Reliability, fairness, and predictive validity of committee peer review. BIF Futur. 2004; 19:7–19.

3. Danthi N, Wu CO, Shi P, Lauer M. Percentile ranking and citation impact of a large cohort of National Heart, Lung, and Blood Institute-funded cardiovascular R01 grants. Circ Res. 2014; 114:600–606. [PubMed: 24406983]

4. Kaltman JR, Evans FJ, Danthi NS, Wu CO, DiMichele DM, Lauer MS. Prior publication productivity, grant percentile ranking, and topic-normalized citation impact of NHLBI cardiovascular R01 grants. Circ Res. 2014; 115:617–624. [PubMed: 25214575]

5. Berg JM. Science policy: well-funded investigators should receive extra scrutiny. Nature. 2012; 489:203. [PubMed: 22972279]

6. Scheiner SM, Bouchie LM. The predictive power of NSF reviewers and panels. Front Ecol Environ. 2013; 11:406–407.

7. Gallo SA, Carpenter AS, Irwin D, McPartland CD, Travis J, Reynders S, et al. The validation of peer review through research impact measures and the implications for funding strategies. PLoS One. 2014; 9:e106474. [PubMed: 25184367]

8. Bornmann L, Marx W. How good is research really? Measuring the citation impact of publications with percentiles increases correct assessments and fair comparisons. EMBO Rep. 2013; 14:226–230. [PubMed: 23399654]

9. Pendlebury D White Paper: Using Biobliometrics in Evaluating Research. available at http://wokinfo.com/media/mtrp/UsingBibliometricsinEval_WP.pdf2008

10. Breiman L. Statistical modeling: the two cultures. Stat Sci. 2001; 16:199–215.

11. Gorodeski EZ, Ishwaran H, Kogalur UB, Blackstone EH, Hsich E, Zhang ZM, et al. Use of hundreds of electrocardiographic biomarkers for prediction of mortality in postmenopausal women the Women's Health Initiative. Circ Qual Outcomes. 2011; 4:521–532.

12. Wickham, H. ggplot2: elegant graphics for data analysis. 2009. http://had.co.nz/ggplot2/book

13. Ishwaran, H., Kogalur, UB. Random Forests for Survival, Regression and Classification (RF-SRC). 2015. http://cran.r-project.org/web/packages/randomForestSRC/

14. Jacob BA, Lefgren L. The impact of research grant funding on scientific productivity. J Pub Econ. 2011; 95:1168–1177. [PubMed: 21857758]

15. Carey, B. Blazing trails in brain science. New York Times; 2014. http://www.nytimes.com/2014/02/04/science/blazing-trails-in-brain-science.html?_r=0 [accessed on 22 March 2015]

16. Clavería LE, Guallar E, Camí J, Conde J, Pastor R, Ricoy JR, et al. Does peer review predict the performance of research projects in health sciences? Scientometrics. 2000; 47:11–23.
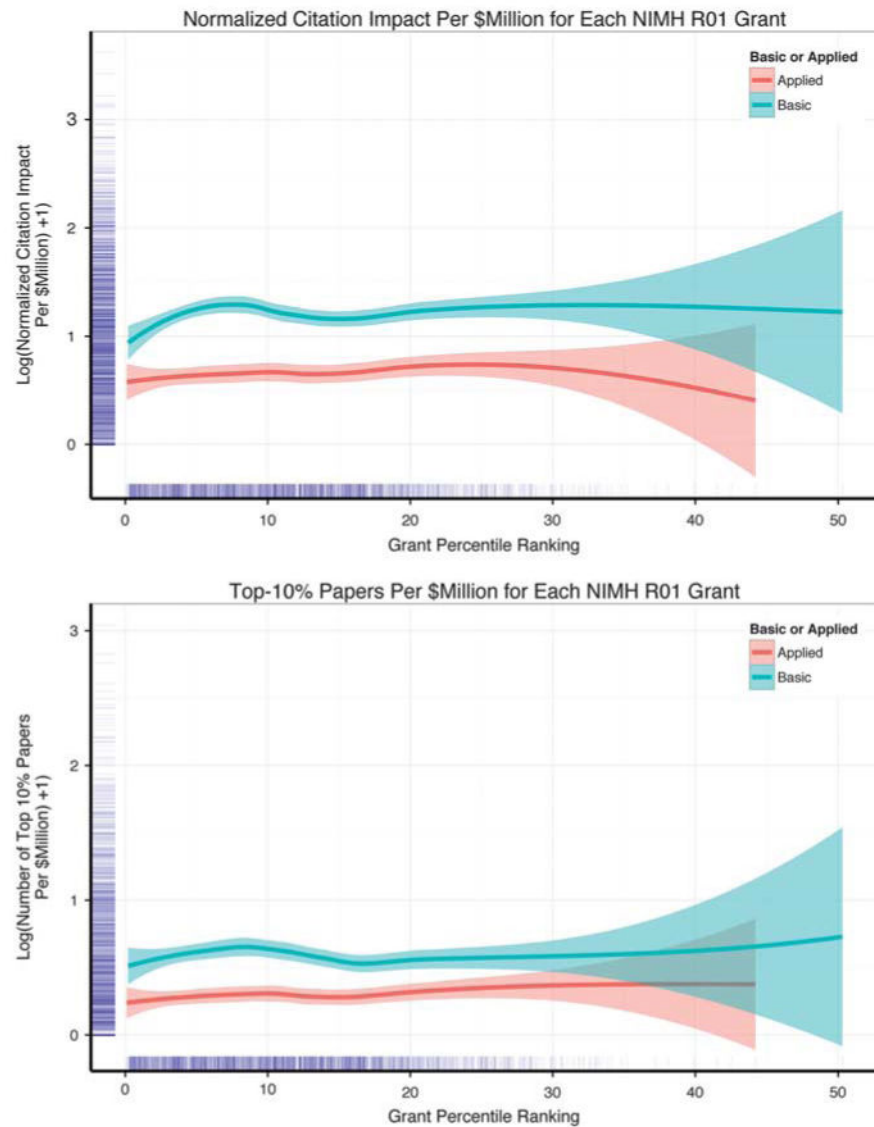
Figure 1.
Scatter and rug plots relating grant citation impact per $million spent and grant percentile ranking. The top panel shows results for normalized citation impact of each grant, calculated as the sum of the normalized citation impact of each paper. Each paper's normalized citation impact equals (100–InCites percentile)/100, where the InCites percentile value is based on how often a paper is cited compared with other papers of the same category published in the same year. The bottom panel shows results for the number of top-10% papers per grant, that is, the number of papers with an InCites percentile value   10. The shaded areas reflect 95% confidence ranges. The curves were derived using lowess smoothing.
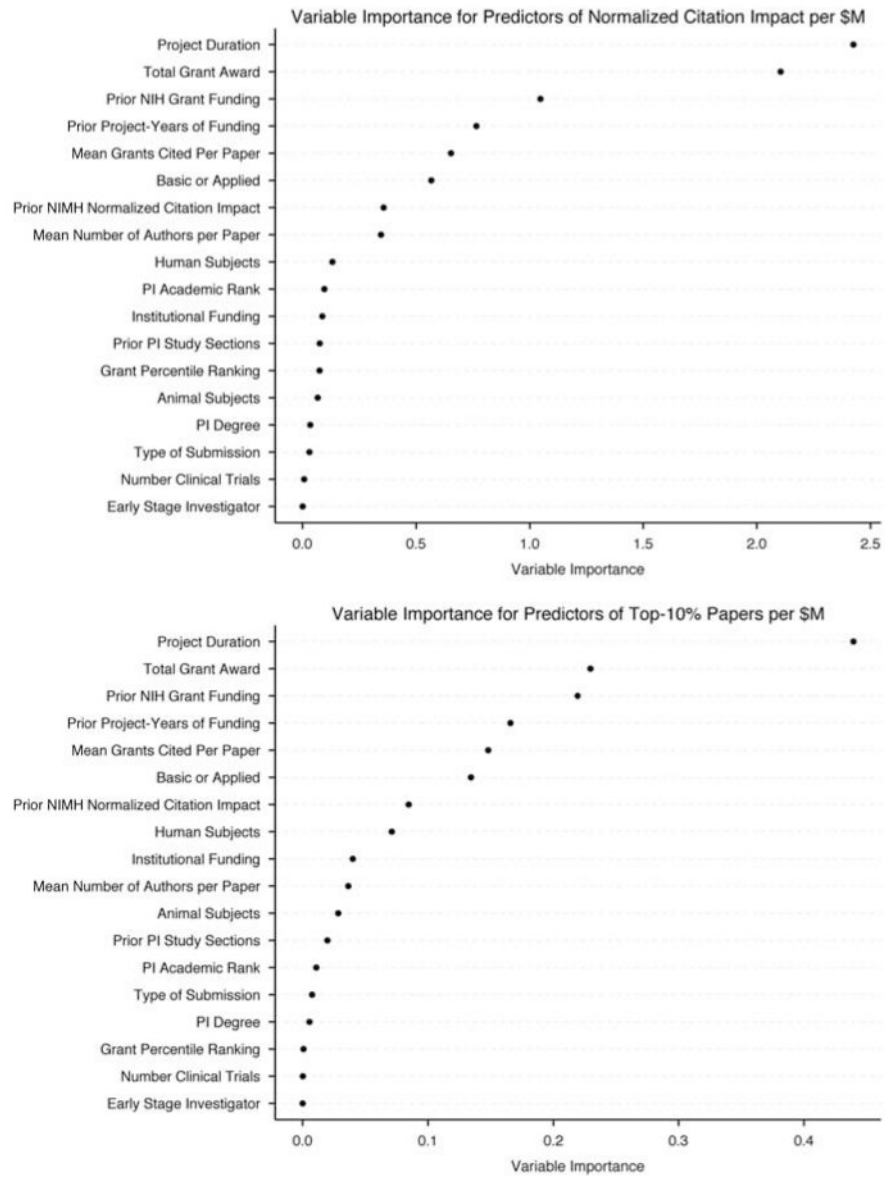
Figure 2.
Results of random forest regressions for identifying the most important predictors of normalized citation impact per $million spent (top panel) and of number of top-10% papers per $million (bottom panel).
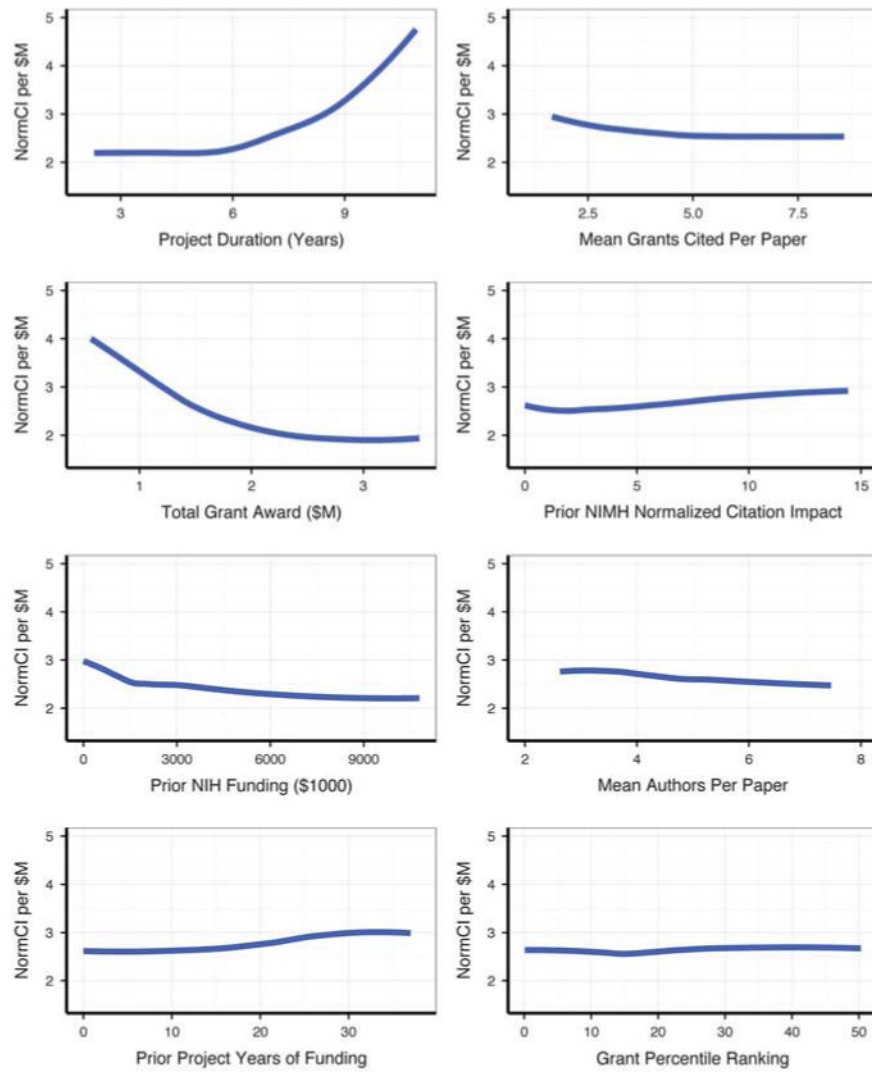
Figure 3.
Results of random forest regressions for identifying confounder- and interaction-adjusted
associations of the seven most important continuous variables and for grant percentile
ranking with normalized citation impact per $million spent. To enable meaningful
comparisons, all y axes are shown on the same scale. X axes ranges reflect 10th to 90th x-
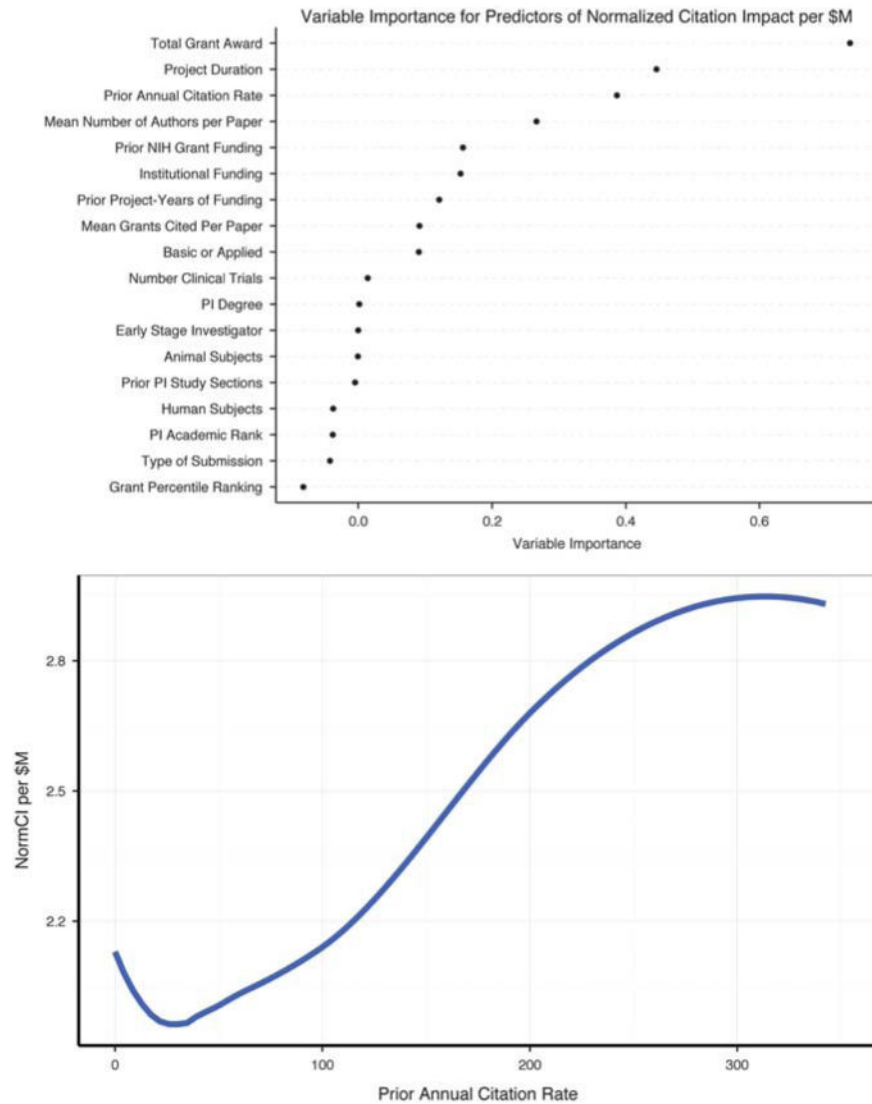tile values.

Figure 4.
Results of random forest regressions for identifying the most important predictors of normalized citation impact per $million for a random subset of 100 grants, in which we obtained prior publication data on all PI publications (not just those funded by NIMH) and manually disambiguated names. The top panel shows the most important predictors, whereas the bottom panel shows the confounder- and interaction-adjusted association of prior annual citation rate with normalized citation impact per $million spent.

Table 1

Grant, principal investigator and paper characteristics according to grant percentile ranking

Characteristic	Best (0–5)	Middle (5.1–15)	Worst (415)
	N = 398	N = 807	N = 550
Grant percentile ranking	1.3/2.7/3.8	7.4/10.0/12.5	16.7/19.3/23.3
Type of submission			
A0	18% (72)	26% (207)	33% (179)
A1	46% (185)	45% (362)	43% (238)
A2	35% (138)	29% (233)	24% (131)
A3	1% (3)	1% (5)	0% (2)
Priority score	124/132/140	140/150/160	160/170/183
Total grant award ($M)	1.4/1.8/2.8	1.1/1.7/2.6	1.1/1.6/2.5
Institutional funding ($M)	13.5/35.8/73.2	10.3/23.7/64.0	9.7/26.4/70.9
Project duration (years)	5.0/5.9/6.9	4.9/5.8/6.7	4.9/5.8/6.6
Basic (versus applied)	61% (243)	62% (504)	65% (358)
Human subjects	63% (251)	60% (486)	56% (306)
Animal subjects	32% (128)	36% (292)	39% (213)
Early stage investigator	2% (8)	1% (11)	2% (9)
PI degree			
MD or equivalent	21% (83)	20% (160)	18% (100)
MD/PhD	9% (36)	8% (63)	10% (54)
PhD or equivalent	70% (279)	72% (584)	72% (396)
Academic rank			
Assistant	21% (85)	25% (198)	30% (167)
Associate	24% (95)	22% (180)	21% (117)
Full	37% (147)	36% (291)	29% (162)
None	18% (71)	17% (138)	19% (104)
Prior NIMH normalized citation impact	0.84/2.87/7.66	0.75/2.61/6.28	0.52/2.10/5.69
Prior study section meetings	0/3/8	0/2/9	0/1/5
Prior NIH grant funding ($K)	146/1097/3368	43/816/3502	0/553/2063
Prior project-years of funding	2/7/18	1/6/17	0/5/12
Mean number of authors per paper	3.5/4.6/6.2	3.5/4.8/6.3	3.4/4.7/6.1
Mean grants cited per paper	2.2/3.7/5.8	2.1/3.4/5.2	2.0/3.2/4.8

Abbreviations: NIH, National Institutes of Health; NIMH, National Institute of Mental Health; PI, Principal Investigator. Continuous variables are represented as *a*/ *b/c*, with *a* the lowest quartile, *b* the median and *c* the upper quartile. Categorical variables are represented as % (*N*).

Table 2

Productivity and citation outcomes according to grant percentile ranking

Characteristic	Best (0–5) N=398	Middle (5.1–15) N =807	Worst (>15) N=550
Number of papers	1.5/4.0/7.9	1.7/3.8/8.9	2.0/4.0/8.0
Normalized citation impact	0.90/2.44/5.14	0.98/2.53/5.70	0.98/2.53/5.12
Number of top-10% papers	0.00/0.50/2.00	0.00/0.75/2.11	0.00/0.50/2.00
Number of papers per $million	0.72/1.97/4.42	0.97/2.63/5.29	1.04/2.72/5.60
Normalized citation impact per $million	0.42/1.25/2.90	0.56/1.68/3.62	0.65/1.67/3.61
Number of top-10% papers per $million	0.00/0.28/1.09	0.00/0.40/1.34	0.00/0.34/1.16

Continuous variables are represented as *a/b/c*, with *a* the lowest quartile, *b* the median and *c* the upper quartile. Categorical variables are represented as % (*N*).