



Published in final edited form as:

Circ Res. 2015 July 17; 117(3): 239–243. doi:10.1161/CIRCRESAHA.115.306830.

Predicting Productivity Returns on Investment: Thirty Years of Peer Review, Grant Funding, and Publication of Highly Cited Papers at the National Heart, Lung, and Blood Institute

Michael S. Lauer, Narasimhan S. Danthi, Jonathan Kaltman, and Colin Wu

Division of Cardiovascular Sciences, NHLBI

Abstract

There are conflicting data regarding the ability of peer review percentile rankings to predict grant productivity, as measured through publications and citations. To understand the nature of these apparent conflicting findings, we analyzed bibliometric outcomes of 6873 de novo cardiovascular R01 grants funded by the National Heart, Lung, and Blood Institute between 1980 and 2011. Our outcomes focus on “Top-10%” papers, meaning papers that were cited more often than 90% of other papers on the same topic, of the same type (e.g. article, editorial), and published in the same year. The 6873 grants yielded 62,468 papers, of which 13,507 (or 22%) were Top-10% papers. There was a modest association between better grant percentile ranking and number of top-10% papers. However, discrimination was poor (area under ROC 0.52, 95% CI 0.51–0.53). Furthermore, better percentile ranking was also associated with higher annual and total inflation-adjusted grant budgets. There was no association between grant percentile ranking and grant outcome as assessed by number of top-10% papers per \$million spent. Hence, the seemingly conflicting findings regarding peer review percentile ranking of grants and subsequent productivity largely reflect differing questions and outcomes. Taken together, these findings raise questions about how best NIH should use peer review assessments to make complex funding decisions.

Keywords

Peer review; research funding; National Institutes of Health; bibliometrics

A just-published analysis by Li and Agha of nearly 30 years of NIH R01 grants showed associations between better percentile rankings and bibliometric outcomes.¹ These associations persisted even after accounting for a number of potential confounding variables, including prior investigator track record and institutional funding. These associations also appear to be at odds with prior analyses from the National Heart, Lung, and Blood Institute

Address correspondence to: Dr. Lauer, 6701 Rockledge Drive, Room 8128, Bethesda, MD 20892, Tel: 301-435-0422, lauerm@nhlbi.nih.gov.

DISCLOSURES

The views expressed here are those of the authors and do not necessarily reflect the view of the NHLBI, the NIH, or the Federal government.

SOURCES OF FUNDING

All authors are full-time NIH employees and conducted this work as part of their official federal duties.

(NHLBI)²⁻⁴, the National Institute of General Medical Sciences (NIGMS)⁵, the National Institute of Mental Health (NIMH)⁶, and the National Science Foundation (NSF).⁷ How can we reconcile these apparent differences? Are these differences contradictory, or do they reflect questions that differ in a subtle, though important manner?

To understand the different findings, it is important to consider the differences between the Li and Agha¹ and the prior ones. The most obvious, perhaps, is that Li and Agha¹ included a much larger number of grants that were funded over many decades.⁸ But there are two other key differences: first, Li and Agha focused on raw publication and citation counts, as opposed to field normalized counts,⁹ and second Li and Agha focused on bibliometric outcomes alone, whereas some of the previous studies focused on outcomes per \$million spent.^{2-4,6}

If you were told that a person weighs 100 pounds, you would know little. If you were then told that that person is a 6-foot tall man, we might worry about cachexia. If you were told that that person is a 10-year old girl, we would worry about serious obesity. Similarly if you were told that a paper received 100 citations, you would know little. Your interpretation would change depending on whether the paper focuses on mathematics, or cell biology, or basic cardiovascular biology, or clinical cardiovascular medicine.⁹ It would also change if the paper were published one year ago or ten years ago. One recent analysis found that clinical cardiovascular papers are cited 40% more often than basic papers, and that citation rates in cardiovascular sciences have increased dramatically over time.¹⁰ Because of these marked variations in citation practice, a number of authorities⁹ identify the “percentiles approach” as the most robust citation metric.¹¹ Here, each paper is judged against other papers published in the same year and dealing with the same topic – hence a biochemistry research paper published in 2005 is compared against other biochemistry research papers published in 2005 and not against a clinical trial paper published in 2002.

Another question is whether one measures outcome alone or outcomes in light of money spent. Every grant or contract that NHLBI dispenses incurs opportunity costs – if NHLBI chooses to fund a large, expensive trial, which means it won’t be able to fund a certain number of smaller (in terms of budget) R01 grants. If we focus on bibliometric outcomes – certainly not the only outcomes worth considering – we would not ask how many highly cited (for field and year) papers were produced, but how many were produced for every \$million spent.⁴ In other words, the outcome metric for the previous studies was not only return, but also return on investment.

To gain greater insight into the seemingly different outcomes of Li and Agha and prior reports, we now turn to examine bibliometric outcomes of 30-years of cardiovascular grants funded by the NHLBI.

Between 1980 and 2011, NHLBI funded 8125 de novo (i.e. not renewals) cardiovascular R01 grants. Of these 6873 were investigator-initiated and received a percentile ranking, while the remainder did not receive a percentile ranking mainly because they were reviewed by ad hoc panels assembled in response to specific programs. Of these 6873 percentiled grants, 1867 (27%) were successfully renewed at least once. Through 2013, these

percentiled grants generated 62,468 papers; of these, 13,507 (or 22%) were “Top-10% Papers,” meaning that they were cited more often than 90% of all other papers published in the same year and focused on the same topic. The expected value would be 10%, meaning that as a whole the portfolio performed at least twice as well as would be expected by chance.

The distribution of Top-10% papers was highly skewed, consistent with prior work showing the “heavy-tailed” nature of scientific output.¹² That is, a small proportion of scientific effort is responsible for a disproportionately large proportion of output; in common parlance, there is a “20–80” phenomenon in which 20% of the input yields 80% of the output. The median number of Top-10% papers per grant was 1 (IQR 0 – 2) with a range of 0 to 154. Because of the skewed distribution, we show all analyses after logarithmic transformation.

Figure 1 shows the association of Top-10% papers and percentile ranking. Consistent with Li and Agha¹, grants with better percentile scores generated more Top-10% papers. However, the individual points, each referring to one grant, shown in the top panel, illustrate the high degree of noise. To assess how well percentile ranking discriminated between grants more or less likely to generate Top-10% papers, we calculated the area under a receiving operator curve (AUC under a ROC, Online Figure I), and found modest discrimination that was only slightly better than chance (AUC 0.52, 95% CI 0.51–0.53).

Grants with better percentile scores had higher annual and total inflation-adjusted budgets (Figure 2, left panels), and grants with higher annual and total budgets yielded more Top-10% papers (Figure 2, right panels), though with varying marginal returns. The association between better percentile scores and higher budgets is not only a reflection of actual allocations (Figure 2, left panels) but also of requested allocations (Online Figure II, based on a separate set of grant applications submitted in 2011–2012). Sometimes, allocated budgets are lower than requested budgets, usually due to post-review negotiations between program staff and applicants.

Budget, of course, is a critical component because an NIH institute should not only be concerned about return (in this case number of Top-10% papers), but about return on investment (number of Top-10% papers per \$million spent). Grants with better percentile rankings may generate more Top-10% papers, but tend to also cost the NIH more money.

Figure 3 shows the association between Top-10% papers per \$million spent and percentile ranking. Consistent with prior reports^{2–4,6} – admittedly based on smaller samples studied over shorter periods of time – there was no association between grant percentile ranking and Top-10% papers per \$million spent. We also found no ability of percentile ranking to discriminate grants with higher and lower productivity by this metric (AUC under the ROC 0.49, 95% CI 0.47–0.50, Online Figure III).

Thus, it appears that some of the apparent contradiction between the recent analysis of Li and Agha¹ and the prior ones^{2–4,6} is that each focused on different questions. To maximize return on investment, it may be more appropriate for NIH to focus on bibliometric outcomes per \$million spent rather than on bibliometric outcomes alone. We must acknowledge, though, that this metric has its limitations; that is, a metric like number of highly-cited

papers per \$million may not be able to fully discriminate which projects yielded greatest scientific impact. In some cases it may make sense to spend more even for a small number of highly cited papers; for example, NIH may wisely choose to spend proportionally more money to fund clinical trials that directly impact practice and improve public health.¹³ Other projects may generate new methods that are “well worth the money” because they make whole new areas of scientific inquiry possible. Citation metrics have well-known limitations; for example, papers may be cited because they are controversial rather than meritorious. Nonetheless, recent literature has found a correlation between expert opinion and citation impact.¹⁴ Furthermore, assessments of return on investment should also consider other factors, such as the type of research. We recently reported that at NIMH, basic science projects appear to yield a greater return on investment than applied projects⁶; we are planning similar analyses at NHLBI.

What can we say from all this? It appears that peer review is able, to a modest extent and with modest degrees of discrimination, to predict which grants will yield more Top-10% papers.¹ The modest degree of discrimination reflects previous reports suggesting that most grants funded by a government sponsor are to some extent chosen at random.¹⁵ At the same time, it appears that this association is closely entangled with budgets, budgets that NIH institutes must wrestle with. Because of the diminishing marginal returns seen with more expensive – and better scoring – grants, these data challenge us to question the assumption that it is best to rely primarily on the payline for making funding decisions. Is it smart investment strategy to fund all grants below the cutoff payline at close to requested budgets, but fund a tiny fraction of grants for those scoring above the payline?

In an important sense, we are not so much dealing with a peer review question, but a larger systems question⁸ – how best should NIH make funding decisions about those grants that pass the muster of peer review? NIH might ask reviewers to address explicitly their perspectives on the opportunity costs of applications, especially those that are more expensive. Instead of making funding decisions solely based on peer review rankings, decision makers could consider a number of alternate approaches that leverage, but do not solely rely upon, peer review. Some NIH institutes have already taken explicit steps to move away from strict adherence to paylines, choosing instead to fund a proportion of grants from among those that score generally well, that is within a “zone of opportunity.”¹⁶ Some thought leaders have called for NIH to pay closer attention to the distribution of funding, thereby enabling more scientists to benefit from increasingly constrained funds.⁵ Some institutes are choosing to change the focus of peer review from a system that primarily focuses on the merit of projects to one that focuses on the expected performance of people and their research teams.¹⁷ Still another approach depends on the type of research; for example, when NIH considers which clinical trials to fund it might choose to prioritize those that focus on hard clinical endpoints and to take steps to assure that peer review panels appreciate the importance of such trials.^{18,19} These alternate approaches – the “zone of opportunity,” additional scrutiny for well-funded investigators, focus on people instead of projects, and NIH-stipulation of preferred clinical trials – are only a subset of possibilities, which all beg for their own rigorous analyses to determine whether they enable NIH to make better decisions.²⁰ Some critics have argued that it is not enough for NIH to try new and different mechanisms: NIH should “turn the scientific method on itself” by conducting its

own randomized trials^{21,22}, some of which could well involve peer review and how NIH program staff respond to peer reviewers' scores and comments. In any case, the recently burgeoning literature on grant peer review promises to usher in a period in which NIH, the scientific community, and the public will engage in a rich, data-driven dialogue on how best to leverage the scientific method to improve public health.

Data and methods to generate plots

Data on NHLBI cardiovascular R01 grants, including data on percentile rankings and budgets were obtained from internal tables. All budget figures were inflation-adjusted to 2000 constant dollars using the Biomedical Research and Developments Price Index (or BRDPI at <http://officeofbudget.od.nih.gov/gbiPriceIndexes.html>). We used the SPIRES system to match grant numbers to publication identifiers (PMIDs), and supplemented PMIDs with bibliographic data stored in an EndNote library. We worked with Thomson Reuters to link these publications to their InCites database, which generated for each publication a "percentile value" describing how often that publication was cited compared to similar publications on the same topic, of the same type (e.g. article, letter, editorial), and in the same year. The InCites database empirically classifies papers according to 252 distinct topics; among the 62,468 papers considered here, the most common topics were "cardiac and cardiovascular systems," "biochemistry and molecular biology," "physiology," "peripheral vascular disease," and "pharmacology and pharmacy." As we have described previously, we divided credit for papers if they cited more than one grant in the portfolio; thus if a paper was classified as a top-10% paper and it acknowledged 2 grants, each grant was credited with 0.5 top-10% papers. We generated scatter plots, loess smoothers, and 95% confidence intervals with Wickham's ggplot2 R package.²³ We calculated areas under receiver operating curves using the pROC R package²⁴ and plotted the curves with the plotROC package.²⁵

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Gary Gibbons (Director of NHLBI), Jon Lorsch (Director of NIGMS), and Richard Hodes (Director of NIA) for their helpful comments on earlier versions of this paper. We are grateful to Ebyan Addou and Sean Coody for curating grant and citation data. We also thank Donna DiMichele for helping NHLBI secure InCites data.

Nonstandard Abbreviations and Acronyms

AUC	area under the curve
CI	confidence interval
NHLBI	National Heart, Lung, and Blood Institute
NIGMS	National Institute of General Medical Sciences
NIMH	National Institute of Mental Health

NSF	National Science Foundation
ROC	Receiver operating characteristic curve

References

- Li D, Agha L. Big names or big ideas: Do peer-review panels select the best science proposals? *Science* (80-). 2015; 348:434–438.
- 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]
- Danthi NS, Wu CO, DiMichele DM, Hoots WK, Lauer MS. Citation Impact of NHLBI R01 Grants Funded Through the American Recovery and Reinvestment Act as Compared to R01 Grants Funded Through a Standard Payline. *Circ Res*. 2015; 116:784–8. [PubMed: 25722441]
- 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; 115:617–24. [PubMed: 25214575]
- Berg JM. Science policy: Well-funded investigators should receive extra scrutiny. *Nature*. 2012; 489:203. [PubMed: 22972279]
- Doyle JM, Quinn K, Bodenstern YA, Wu CO, Danthi N, Lauer MS. Association of percentile ranking with citation impact and productivity in a large cohort of de novo NIMH-funded R01 grants. *Mol Psychiatry*. 2015
- Scheiner SM, Bouchie LM. The predictive power of NSF reviewers and panels. *Front Ecol Environ*. 2013; 11:406–407.
- Mervis J. Scientific publishing. NIH's peer review stands up to scrutiny. *Science*. 2015; 348:384. [PubMed: 25908803]
- Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*. 2015; 520:429–31. [PubMed: 25903611]
- Ophthof T. Differences in citation frequency of clinical and basic science papers in cardiovascular research. *Med Biol Eng Comput*. 2011; 49:613–621. [PubMed: 21567267]
- 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]
- Press WH. Presidential address. What's so special about science (and how much should we spend on it?). *Science* (80-). 2013; 342:817–822.
- Lauer MS. Investing in clinical science: make way for (not-so-uncommon) outliers. *Ann Intern Med*. 2014; 160:651–652. [PubMed: 24798528]
- Bornmann L. Interrater reliability and convergent validity of F1000Prime peer review. *J Assoc Inf Sci Technol*. 2015:n/a–n/a.
- Graves N, Barnett AG, Clarke P. Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. *BMJ*. 2011; 343:d4797. [PubMed: 21951756]
- Pettigrew, R. A Message from the Director (October 2014): Development of a New Funding Plan: The Expanded Opportunity Zone. [Internet]. 2014. Available from: <http://www.nibib.nih.gov/about-nibib/directors-page/directors-update-oct-2014>
- Kaiser J. Funding. NIH institute considers broad shift to “people” awards. *Science*. 2014; 345:366–7. [PubMed: 25061183]
- Gordon DJ, Lauer MS. Publication of trials funded by the National Heart, Lung, and Blood Institute. *N Engl J Med*. 2014; 370:782. [PubMed: 24552337]
- Devereaux PJ, Yusuf S. When it comes to trials, do we get what we pay for? *N Engl J Med*. 2013; 369:1962–1963. [PubMed: 24224630]
- Mervis J. Peering into peer review. *Science*. 2014; 343:596–8. [PubMed: 24503830]

21. Ioannidis JP. More time for research: fund people not projects. *Nature*. 2011; 477:529–531. [PubMed: 21956312]
22. Azoulay P. Turn the scientific method on ourselves. *Nature*. 2012; 484:3–5.
23. Wickham, H. *ggplot2: elegant graphics for data analysis* [Internet]. 2009. Available from: <http://had.co.nz/ggplot2/book>
24. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011; 12:77. [PubMed: 21414208]
25. Sachs MC. *plotROC: Generate Useful ROC Curve Charts for Print and Interactive Use*. 2015

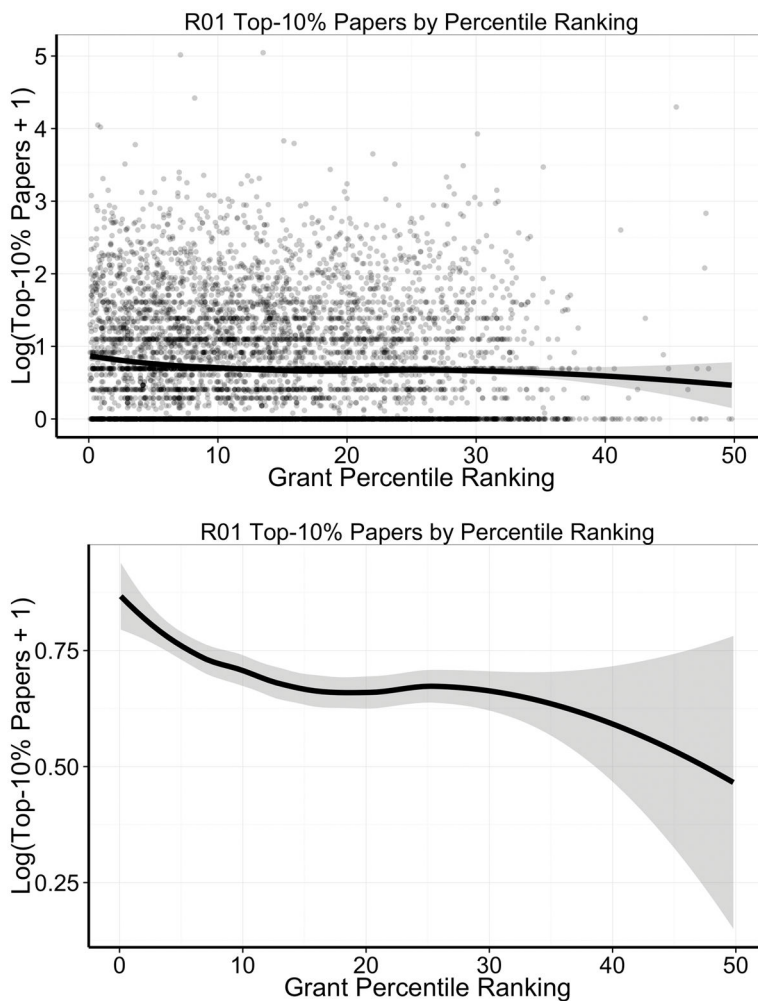


Figure 1. Loess smoother and 95% confidence ranges for the association between number of Top-10% papers generated per grant and grant percentile ranking among 6873 de novo cardiovascular NHLBI grants. Y-axis values are logarithmically transformed due to skewed distributions

The top panel shows all data in a scatter plot, whereas the bottom panel focuses on the loess smoother alone. Note the difference in scale of the Y-axis between the top and bottom panels.

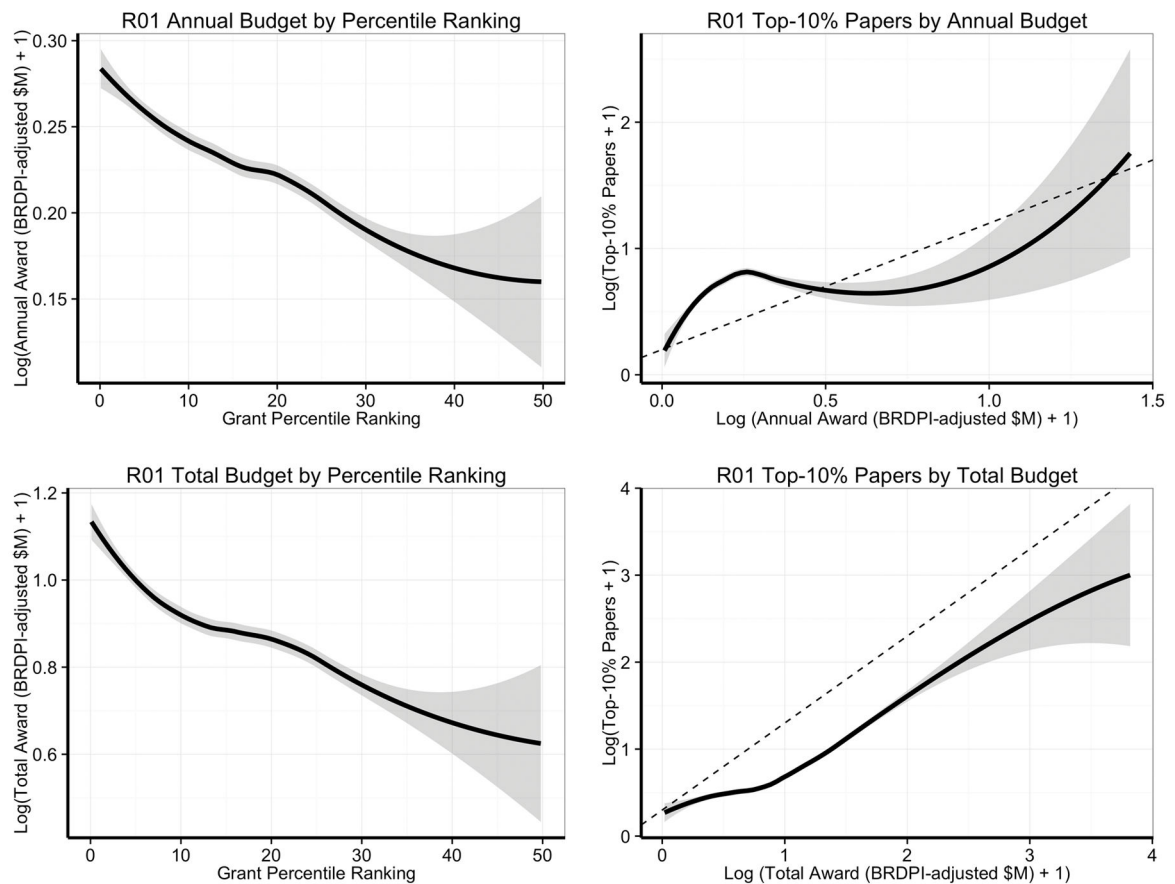


Figure 2. Loess smoother and 95% confidence ranges for the association between budgets, grant percentile ranking, and number of Top-10% papers generated per grant

Budget values are shown in constant 2000 dollars (after logarithmic transformation) with inflation-adjustments according to the Biomedical Research and Development Price Index (BRDPI). The top left panel focuses on annual budgets, whereas the bottom left panel focuses on total budgets (which reflect annual budgets and project duration). The two right-sided panels show the association between the number of Top-10% papers generated per grant and budgets awarded over a grant's lifetime (including renewals when applicable). Both X- and Y-axes are logarithmically transformed. The dotted line represents a slope of 1, where a 20% increase in budget would be associated with a 20% increase in the number of top-10% papers. The top right panel focuses on annual budgets, whereas the bottom right panel focuses on total budgets (which reflect annual budgets and project duration). In the bottom right panel, the actual curve is beneath the slope-of-one line, consistent with diminishing marginal returns.

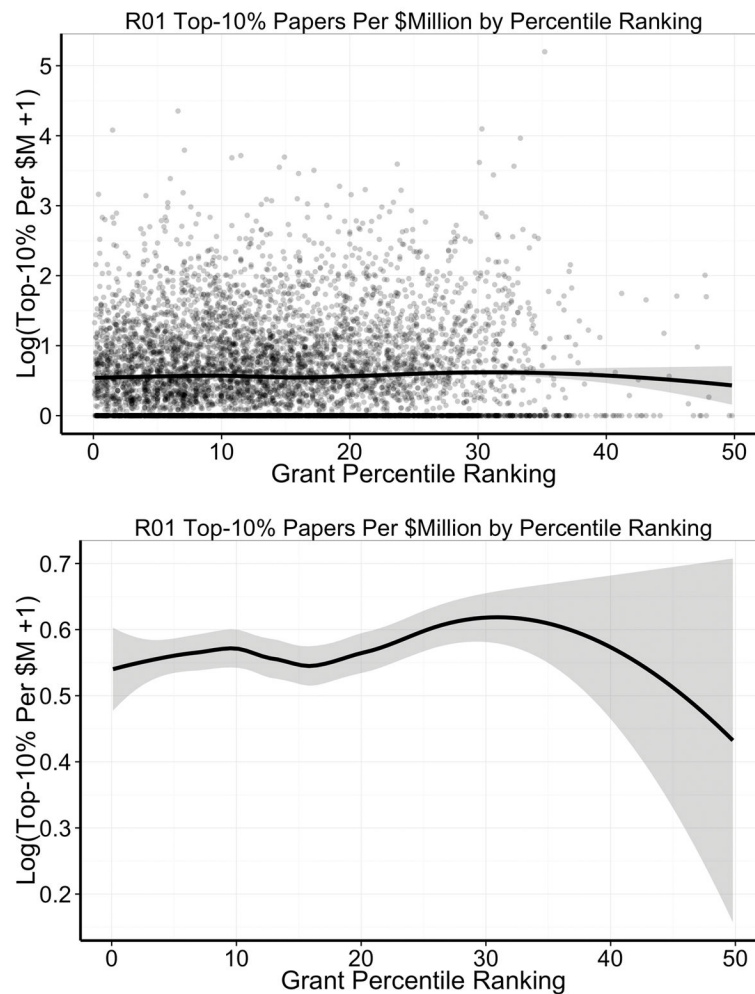


Figure 3. Loess smoother and 95% confidence ranges for the association between number of Top-10% papers generated per grant for each \$million spent and grant percentile ranking among 6873 de novo cardiovascular NHLBI grants

Y-axis values are logarithmically transformed due to skewed distributions. The top panel shows all data in a scatter plot, whereas the bottom panel focuses on the loess smoother alone. Note the difference in scale of the Y-axis between the top and bottom panels.