I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12762

Money for Something: Braided Funding and the Structure and Output of Research Groups

Russell J. Funk
Britta Glennon
Julia Lane
Raviv Murciano-Goroff
Matthew B. Ross

NOVEMBER 2019

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12762

Money for Something: Braided Funding and the Structure and Output of Research Groups

Russell J. Funk
University of Minnesota

Britta Glennon
University of Pennsylvania

Julia Lane
New York University and IZA

Raviv Murciano-Goroff
Boston University

Matthew B. Ross
New York University

NOVEMBER 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Money for Something: Braided Funding and the Structure and Output of Research Groups*

In 2017, the federal government invested over $40 billion on university research; another $16 billion came from private sector sources. The expectation is that these investments will bear varied fruits, including outputs like more economic growth, more scientific advances, the training and development of future scientists, and a more diverse pipeline of STEM researchers; an expectation that is supported by the work of recent Nobel Laureate in Economics, Paul Romer. Yet volatility in federal funding, highlighted by a 35 day federal shutdown in early 2019, has resulted in an increased interest on the part of scientists in finding other sources of funding. Understanding the effect of such different funding streams on research outputs is thus of more than academic importance, particularly because there are likely to be tradeoffs, both in terms of the structure of research and in terms of research outputs. For example, federal funding is often intended to affect the structure of research, with explicit goals of training the next generation of scientists and promoting diversity; those goals are less salient for non-federal funding. On the output side, federally funded research may be more likely to emphasize producing purely scientific outputs, like publications, rather than commercial outputs, like patents. The contribution of this paper is to use new data to examine how different sources of financial support – which we refer to as "braided" funding – affect both the structure of scientific research and the subsequent outputs.

JEL Classification: O3, M5, H1

Keywords: UMETRICS, team science, research impact, science policy, research outputs

Corresponding author:
Julia Lane
Wagner School of Public Policy
New York University
The Puck Building
295 Lafayette Street
New York, NY 10012-9604
USA

E-mail: julia.lane@nyu.edu

* This research was supported by the National Center for Science and Engineering Statistics; NSF SciSIP Awards 1064220, 1262447 and 1932689; NSF Education and Human Resources DGE Awards 1348691, 1547507, 1348701, 1535399, 1535370; NSF NCSES award 1423706; NIHP01AG039347; and the Ewing Marion Kaufman and Alfred P. Sloan Foundations. Our research agenda draws on work with many coauthors, particularly Bruce Weinberg and Jason Owen Smith. Our work was enabled, in part, using resources provided by Institute for Research on Innovation and Science (IRIS), and was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. We are thankful for the high-performance computing resources of the Boston University Shared Computing Cluster.

1. Introduction

In 2017, the federal government invested over $40 billion on university research; another $16 billion came from private sector sources (*1*). The expectation is that these investments will bear varied fruits, including outputs like more economic growth, more scientific advances, the training and development of future scientists, and a more diverse pipeline of STEM researchers(*2*, *3*); an expectation that is supported by the work of recent Nobel Laureate in Economics, Paul Romer (*4*). Yet volatility in federal funding, highlighted by a 35 day federal shutdown in early 2019, has resulted in an increased interest on the part of scientists in finding other sources of funding. Understanding the effect of such different funding streams on research outputs is thus of more than academic importance, particularly because there are likely to be tradeoffs, both in terms of the structure of research and in terms of research outputs. For example, federal funding is often intended to affect the structure of research, with explicit goals of training the next generation of scientists and promoting diversity; those goals are less salient for non-federal funding. On the output side, federally funded research may be more likely to emphasize producing purely scientific outputs, like publications, rather than commercial outputs, like patents. The contribution of this paper is to use new data to examine how different sources of financial support—which we refer to as "braided" funding – affect both the structure of scientific research and the subsequent outputs.

There has been a rich literature on the results of different sources of funding since at least the seminal work of Rosenberg and Nelson(*5*). Much of this literature suggests that privately funded research is more likely to be applied than basic. (*6–9*). Some theoretical work has hypothesized that funding from public and private sources might influence scientific inquiries in laboratories in different directions (*10*). This prediction is supported by empirical evidence that suggests that having funding from different sources changes the research outputs pursued by research groups (*11–15*). The evidence, however, is mixed – some literature argues that funding from the public sector complements private sector funding by supporting the production of more scientific output within a single research agenda; others find public and private funding for research to be substitutes, enabling scientists to pursue new and different avenues of research (*9*, *16*, *17*).

While there has been an interest in understanding the output tradeoff, the literature has been largely silent as to the mechanisms whereby the tradeoff occurs. Team size is thought to be important, but that finding is based on an examination of individual contributions to outputs such as publications and patents rather than studying the structure of research groups themselves (*18*). This paper uses a new dataset that provides detailed information about both funding and the structure of the key unit of analysis, research teams. It includes longitudinal information about all sources of funding, both federal and non-federal, received by research teams, the way in which that funding was spent over time, and the composition of each team. That new dataset, UMETRICS data from the Institute for Research on Innovation and Science (IRIS) consists of highly granular transaction information from 26 major U.S. research institutions(*20*), representing about a third of federally supported university R&D activity. Depending on the university, they go back as far as the first quarter of 2000. The data include monthly information about how funding is spent, and on what it is spent. Information on multiple sources of funding over time, as well as the structure of researchers' teams—particularly the postdoctoral and graduate workforce – is also included: the data consist of information about 300,000 unique federal and non-federal awards including monthly wage payments to 540,000 individuals, including 54,000 faculty, 100,00 graduate students, and 38,000 postdocs.

The data has several interesting features particularly useful for addressing the research question. First, while a majority of the research groups are initially set up with portfolios that are highly specialized and completely funded by federal agencies, there is a substantial minority whose portfolios are initially more diversified, with braided funding from federal and non-federal sources. Second, the portfolio of funding and expenditures changes over time, even within the same research groups. Thus it is possible to examine the links between research group expenditures and workforce composition across groups. Third, links have been made with output information, such as patent and publication activity, at the level of the research group, and thus it is possible to observe the way in which research groups adjust over time, both in terms of their workforce composition and outputs.

These new data permit us to make three contributions to the literature. First, we develop a new method, based on advances in network science (*21*), for identifying and characterizing research groups in large-scale bibliometrics data. Although the research group has historically been important in sociological and economic studies of scientific production(*22*, *23*), recent large-scale analyses have typically focused on much bigger (e.g., scientific fields,(*24*, *25*) or much smaller (e.g., individual papers or researchers,(*26*, *27*) units of analyses. Part of the reason for the lack of attention stems from methodological constraints; prior work has typically relied on organizational directories or records (*19*, *28*) to identify research groups, and therefore has typically been constrained to the study of a single institution. Our approach offers a route for overcoming these limitations, thereby "bringing the research group back in" to the study of scientific production. Second, we directly examine how funding affects the initial structure of the research group and how changes in federal and non-federal funding streams affect the workforce composition at the level of the research group over time. Previous work[1] has largely focused on one source of funding, such as NIH, or has been cross-sectional in nature. Third, we provide direct links between research group composition and multiple output measures.

We find marked differences in the structure of groups that are completely federally funded and those that have braided sources of funds. Research groups that are able to spend from non-federal sources have a lower proportion of staff who are faculty. The share of faculty that is female is also lower.

Changes in the sources of funds over time also affect the way in which groups adjust their workforce composition. Increases in non-federal expenditures are associated with lower utilization rates of graduate students and postdocs. The adjustment is much more pronounced for those with diversified portfolios. In terms of gender diversity, increases in non-federal funding are generally associated with teams having fewer female graduate students and postdocs. The adjustments are greater for those with initially diversified expenditure portfolios.

There are clear differences in outputs as well.Groups with diversified portfolios that include non-federal funding are more likely to patent and less likely to publish than are the more specialized, completely federally-funded groups. Those that do patent are more likely to produce disruptive inventions. Research groups led by more female faculty are less likely to generate a disruptive patent or a highly cited patent, but more likely to publish a scientific paper than those with fewer female faculty. Differences across groups with different levels of braided funding are weaker, with the exception of their likelihood of publishing. Research groups with initially higher levels of non-federal funding are even less likely to publish in response to increases in non-federal funding. There do not appear to be significant differences in terms of receiving ongoing funding.

2. Background and Hypotheses

The enormous growth in team science is typically seen as a response to the increasingly complex nature of scientific work, which requires the pooling of both resources and knowledge (*29*, *30*). In operational terms, scientific research teams are often led by one or more senior scientists who jointly apply for funding and manage a group of postdocs, graduate students, staff scientists, and administrators.

This change in the nature of scientific production means that the historical focus on characteristics of the individual researchers alone may no longer be appropriate. In particular, science is now largely conducted

[1] With some notable exceptions(*12*)

as a collaborative activity, with both the number of team authored papers and the size of those teams growing dramatically over the past few decades (*30*). There is also mounting evidence that formal properties of organizations—building layout and departmental boundaries for example—may have profound effects on researchers' exposure to new knowledge and therefore their capacity for transformative insights (*28*, *31*). Collectively, these observations suggest that the *organizational contexts*—the informal (i.e., collaborations) and formal structure—within which research takes place are becoming increasingly important for understanding the link between scientific funding and transformative research activity. (*32*, *33*).

Expansion in organizational and managerial scale requires increased funding. Principal investigators often need to bring in multiple grants or contracts simultaneously in order to support costly operations. We call this phenomenon of merging and leveraging multiple streams of funding from different sources ``braided'' funding. The consequences of such arrangements for team structure are likely to be important because in the United States, different funding sources often have different strictures applied to workforce composition. This is particularly true when the source of funding is the federal government; other sources of funding typically have fewer requirements on workforce composition but may have different output expectations.

More specifically, the largest federal grant making agencies are the National Institutes of Health (NIH), the National Science Foundation (NSF), the Department of Defense (DoD), and the Department of Energy (DoE). Restrictions on expenditures from these sources are typically because a major motivation for governmental research funding is to encourage the training and development of future scientists. For example, grants made by the NIH and the NSF explicitly ask applicants to incorporate the mentoring and training of postdocs and graduate students in their grants, and to document how they will do so. There is also an explicit focus on building a diverse workforce, particularly by including women on grant teams, although sometimes with mixed results(*2*, *3*, *34*). The emphasis of federal funding on training the next generation of scientists and on building a diverse workforce leads to the first two hypotheses:

H1. Research groups with a higher share of resources from non-federal sources will employ fewer trainees, namely postdocs and graduate students, but more professional/technical research staff.

H2. Research groups with a higher share of resources from non-federal sources will employ fewer women.

It is worth noting that these hypotheses can be interpreted in two ways. One is that research groups that specialize in federally funded research will organize themselves in different ways than those that receive "braided" funding, or funding from both federal and non-federal sources. The second is that research groups, regardless of how they are organized, respond differently to federal funding than they do to non-federal funding. We will test both variants.

There are also differences in the types of expected outputs. While it is difficult to generalize across these very heterogeneous federal agencies, most use a peer review process and each grant is independently evaluated. By contrast, philanthropic and industry research often involves longer term relationships between program managers and researchers, and is tied more closely to the specific goals of the firm or foundation. Federally funded research is generally meant to encourage basic, more risky, work, which can be shared to support the public good aspects of science (*35*). In contrast, collaboration with industry partners may inhibit publishing of scientific findings in academic journals, as the sponsors seek to capture value from their investment (*36*, *37*). The stereotypical example of this type of partnership is a pharmaceutical company paying bio-medical researchers to evaluate the efficacy of a new drug (*38*); such work is likely to be more focused on producing short-term results that are applicable and profitable (*39*).

The empirical evidence is mixed. Some work suggests that research funded by corporate backers is more likely to emphasize commercializable results, change the priorities of research fields(*14*), result in fewer academic papers (*15*) and shift priorities towards producing patents (*12*, *40*). Others find a positive correlation between commercial sponsorship and academic productivity. (*6*, *41*). The latter set of findings, however, are cross-sectional in nature, and are likely to reflect industry selection of higher productivity researchers. This leads to the following hypotheses about industry-sponsored research groups:

H3. Research groups with a higher share of non-federal funding are more likely to patent but are less likely to publish scientific papers.

As before, there are two possible interpretations of this hypothesis. One is that research groups that specialize in federally funded research are more likely to publish and less likely to patent than groups who receive "braided" funding or funding from both federal and non-federal sources. The second is that research groups, regardless of how they are organized, respond differently to output incentives depending on whether the source of funds is federal or non-federal.

It is an open question as to whether non-federal funding increases the impact of patents produced by research groups in terms of creating departures from existing streams of technology. Some have argued that industry explicitly funds university researchers to stimulate high impact research, others have emphasized the narrowness of industry focus(*42*). Funk and Owen-Smith show that for universities, increases in federal funding are associated with higher impact patents(*43*). Papers of scientists backed by the non-federal Howard Hughes Medical Institute are on average more innovative than those written by federally sponsored scientists (*11*); patents developed from corporate-sponsored research at the nine campuses and three national laboratories operated by the University of California are more frequently licensed and more frequently cited than patents derived from purely federally supported research (*44*).

Quite separately, no study of which we are aware examines the effect of changes in the amount of non-federal funding on the impact of patents for a given research group. Our hypothesis is that non-federal funding is likely to be supportive of stimulating research that departs from existing streams of technology, and hence is more likely to be disruptive:

H4 The patents produced by research groups with a higher share of federal funding are likely to be more disruptive.

Publications and patents measures are useful indicators of scientific productivity, but we can also draw on the firm survival literature to create an alternative measure - research group survival, as measured by the receipt of subsequent grants (*45*). The logic is as follows. Funding agencies, whether federal or non federal, typically explicitly ask principal investigators requesting additional funding to describe the results of their prior funding. Scientific panels and peer reviewers are then asked to evaluate the quality and quantity of that output in making a decision to support the investigator again. Failure to receive follow-on funding can be seen as an indicator of peers' assessment of low quality contribution from previous grants. We are not aware of any literature that has examined the relationship, and as a consequence, have no prior hypotheses about the direction of the effects of funding; the analysis reported below is for exploratory purposes only.

3. Data Construction

This section provides a description of the analytical dataset. We begin with a description of the underlying data used in the analysis. We then provide details about the construction of the unit of analysis - research groups - from the underlying raw administrative data, and the measurement of the levels and changes in the key inputs - personnel within the research groups. We subsequently discuss the measures of research output and productivity linked to each research group over time.

The underlying analytical data used come from the UMETRICS 2018Q4a data release (*46*, *47*). These raw data include longitudinal administrative, personnel, and accounting records on the expenditures associated with grants held by researchers at 26 major U.S. research institutions(*20*). The main file of interest in this project is the employee file. Briefly, for each funded project, the file contains all payroll charges for all pay periods (period start date to period end date) with links to both the award id (each incoming grant or contract, regardless of source, has a tracking number) and the internal university id number (recipient account number). Also available from the payroll records are the employee's internal de-identified employee number, the occupational classification[2] and the proportion of earnings allocated to the award. These data permit the capture of all collaborations longitudinally as well as the network connections generated by the project.

3.1 Construction of the unit of analysis: research groups

As described in section 2, an important contribution of this paper is its focus on the research team rather than on the individual researcher, reflecting the change in the nature of scientific production toward collaborative teamwork. This contribution requires developing measures of research teams, since UMETRICS data, like most other data, does not identify research groups explicitly.

Research groups are defined to be collections of senior researchers who jointly manage research funding and their associated personnel. Identifying such groups is challenging because in addition to dense connections within teams, researchers frequently have sparse connections to other teams and research groups. The empirical approach builds on prior methods in the network science literature(*48*) used to detect communities with large network data.

Briefly, the approach categorizes the collaborations of senior researchers using a network community detection algorithm. The algorithm utilized to find research groups efficiently removes conflating sparse connections between researchers and identifies densely connected research groups in a transparent and straightforward manner, described in the next few paragraphs.

The network is constructed based on faculty members who are paid on grants in the UMETRICS data. It only includes employees who are faculty members in the network and whose only role across all UMETRICS grants is as a faculty member.[3] Furthermore, in order to focus on research-active teams in the sciences, the faculty members in the sample must have been funded by a NIH, NSF, or Department of Energy grant at some point in the data.[4] This results in 31,063 faculty members.

The next task is to identify all grants that paid this group of faculty in the UMETRICS data. Trivial grants are excluded. We required grants to have at least $1,000 in total spending and at least three months of positive spending. Center grants, which support large numbers of faculty who often do not collaborate as a unified group on scientific research, are also dropped. Thus, for inclusion, grants must have fewer than five distinct faculty members as employees. Grants that are primarily targeted at training individuals are also eliminated - notably, grants from the Department of Education - as many of these are fellowships that support the education of individuals rather than a specific research group's scientific work. The resulting subset consists of 111,284 research grants.[5]

---

[2] Individuals are classified on the basis of their last occupation observed in our data.

[3] This excludes graduate students or postdocs who serve as PIs.

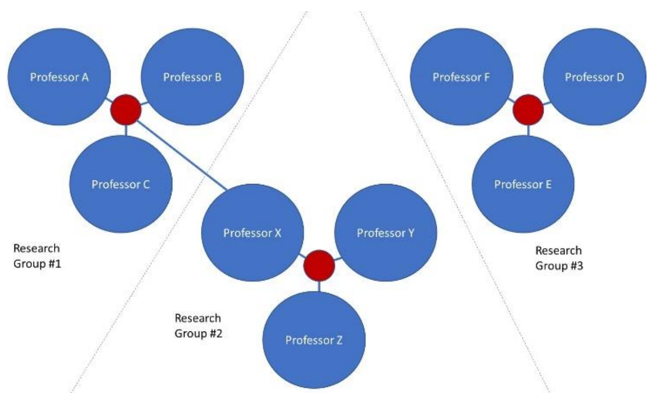
[4] Of all faculty members at the institutions being studied only 2.9% received external funding exclusively from non-federal sources, while 63.4% of faculty received funding from NIH or NSF.

[5] The grants that pass the above described criteria are used for identifying research groups by focusing on the grants in which researchers are most likely to interact, collaborate, and coordinate shared resources. These restrictions are only applied for the purpose of identifying the research groups. Our analysis of the organization and research outputs of the groups examines all funding that supported a research group from all grants visible in the data.

The next step is to combine faculty members and the research grants that paid them into a network, where each faculty member is treated as a node. An edge is formed between faculty member nodes if both researchers have been paid on the same grant (within the grants matching the above criteria).

The algorithm for identifying research groups proceeds by dividing up this network of faculty members. There are 9,113 discrete sub-components of the faculty network. If a sub-component contains more than 10 distinct faculty members (there are 212 such groups), that sub-component is further split based on a procedure inspired by Girvan and Newman (*48*) In particular, one edge at a time is sequentially removed from the sub-component based on ranking the edges by their betweenness centrality. The betweenness centrality is computed based on a weighted measure using the number of shared grants as weights. The sequential removal of edges terminates when all of the new discrete sub-components have fewer than 10 faculty members. Figure 1 provides a graphical example – faculty can be clustered in one research group around a single grant (like Professor D, E and F in Research Group 3) or can be associated with multiple grants (like Professor X in Research Groups 1 and 2). Sub-components are filtered to include those for which complete accounting and personnel data exist.[6] These derived sub-components are defined as *research groups*. The analysis focusses on active research groups by using the groups with at least five years of accounting data and with at least $100,000 in average annual spending – there are 4,790 such groups.

Figure 1: Example of Characterizing Research Groups



3.2 Sources of funding and related expenditures

The composition of research funding is derived from grant administration accounting records in the UMETRICS dataset. Universities that contribute grant-level accounting data to UMETRICS list, for each grant, both the name of the external funding source as well as the Catalog of Federal Domestic Assistance (CFDA) code for federal agency sponsors. A combination of the CFDA codes and the names of external funders are used to determine if a grant came from a federal government agency or a non-federal source. These data are then linked to the actual spending on each grant for each research group. In other words, for each research

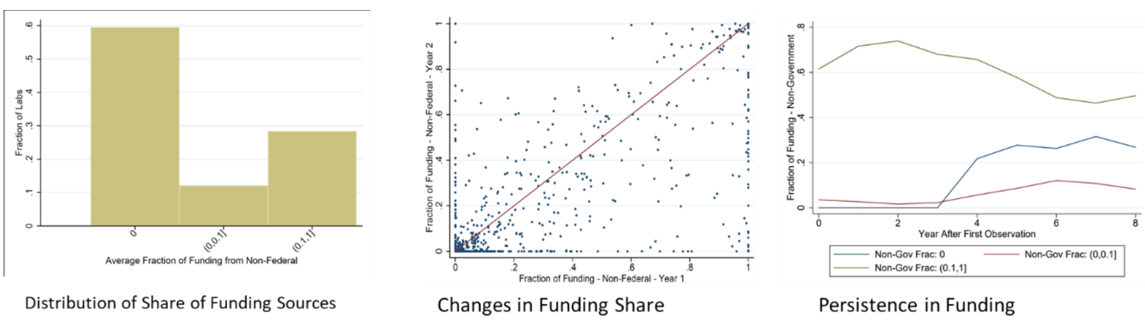
[6] In particular, we exclude research groups for which we only have information from partial years of accounting data, those with only negative (refund) vendor transactions, and those with inconsistent accounting information. For example, we remove research groups for which the sum of the vendor payments exceed the total direct expenses listed for the grant..

group in each year it is possible to construct the fraction of the group's expenditures that comes from the federal government, as well as the fraction of spending that is accounted for by non-federal sources of funds.

Recall that there were two possible interpretations of Hypotheses 1 and 2. One was that research groups will organize themselves in different ways, depending on the sources of funding. The raw data suggest that there does seem to be specialization in the way in which research groups are funded. As can be seen from the left-most panel of Figure 2, there is a dichotomy in sources of funding. Most research groups (about 70 percent) are almost entirely dependent on federal funding. However, about 20 percent have between 25 percent and 100 percent of their funding coming from non-federal sources.

The dichotomy is not random. The central panel shows persistence in funding types; research groups – represented by dots – along the 45 degree line typically had roughly the same fraction of non-federal funding in year one as in year two. Although there is some variation from year to year, there is substantial persistence, as shown by clustering around the 45 degree line. Evidence of persistence is even clearer in the third panel. When research groups are categorized into one of three groups by their levels of non-federal funding—none, low diversification (more than zero but less than 10 percent) and diversified (more than 10 percent) in the first three years—it is evident that although there is some convergence over time, those with high levels of non-government funding tend to persist at higher rates of such funding than do their counterparts.

Figure 2: Share of Non-Federal Funding for Research Groups, Levels and Persistance



The differences in the inputs and outputs of these three different groups will be investigated in Section 4.

3.3 Output measures: Patents, Disruptive Patents, Publications and Future Funding

Patent data: The first source of data is the PatentsView database, retrieved in January of 2019, which contains bibliographic information on all patents granted by the United States Patent and Trademark Office (USPTO). Established in 2012, PatentsView longitudinally links inventors, assignees, locations and patenting activity using bulk data from the USPTO on published patent applications (2001-present) and granted patents (1976-present). Data on patent inventors from PatentsView was linked to UMETRICS data on research group employees by comparing names, affiliations, and grant numbers and constructing a similarity measure based on the textual similarity of the last names, middle initials, and first names of the inventors and employees. In addition, our matching algorithm examined the university affiliation of the employee with the assignee name listed on the patent and the geographic location listed for the inventor. After comparing names and affiliations, the decision of whether or not a pair matches was based on empirical probabilities from a training dataset of known matches.[7]

Disruptive patents: The second source of data come from a recent measure, developed originally by Funk and Owen-Smith (2017), known as the CD index (*43*), which captures a new measure of "disruptive" innovations, defined as patents that create departures from existing streams of technology. The index uses networks of citations to describe the degree to which ideas (embedded in papers or patents) consolidate or destabilize the scientific or technological status quo. More specifically, the CD index characterizes a focal patent by examining how it influences the subsequent use of the papers or patents on which it builds. Consistent with theories of scientific and technological change, papers and patents that increase the use of their predecessors (i.e., by leading them to garner more citations from future work) are given larger negative values on the CD index (indicating greater consolidation), while papers and patents that decrease the use of their predecessors (i.e., garnering them fewer citations from future work) are given larger positive values (indicating greater destabilization). Thus, the CD index captures the "direction" of the effect that a focal paper or patent has on scientific or technological change.

Publication data: The third source is publication data. The PubMed database contains bibliometric records for more than 30 million scientific publications with a particular emphasis on the biomedical literature. If a scientific publication acknowledges support from a government grant, the associated PubMed database record records the grant number cited. Grant numbers cited in PubMed records are matched with the grant numbers present in the UMETRICS data. For each research group, we examine if any of the grants exclusively associated with a researcher in that research group is cited by a scientific publication in the PubMed database[8].

Future funding: This measure is constructed as a binary outcome representing if a research group started expending funds from a grant previously not used by that group. Observing a research group using funds from a new grant would provide evidence that the research group applied for and received additional sponsorship for its research work.

3.4 Summary descriptions

The final sample of 4,790 research groups, when observed over time, results in 24,683 observations. Table 1a provides summary information about the structure of the 4,790 research groups over all periods. The average research group has just over four faculty members exclusively working with that group per year, and spends a total of about $400,000 per year. About one third of the faculty is female, and about two thirds of the research groups have at least one patent.

Figure 2 suggested that there is substantial persistence in the sources of funding to research groups and that potentially there are potentially in the way in which they are structured. In order to examine these differences, the research groups are split into three categories based on their first three years of life – those that are specialized, with no non-federal research expenditures, those with low levels of diversification (less than 10 percent) and those that are diversified with more than 10 percent of their expenditures from federal sources.

[7] The dataset of known linkages between NIH grants and patents was used to compare the names of the employees in UMETRICS with the names listed on patents. This allowed the development of probabilities for variations in names and affiliations referring to the same individual.

[8] Future work will include matches to other publication datasets, since there is an inherent bias towards life sciences in this output measure

There appear to be some differences in the way research groups are organized according to the source of funding. Research groups with more diversified expenditure portfolios tend to be larger and have more funding, including from government sources. Those research groups also have fewer faculty and more postdoctoral fellows and graduate students than do those funded completely by federal funding. There are fewer women faculty, but the share of women postdocs and graduate students is not noticeably different.

Table 1b provides summary information about the unbalanced annual panel of 24,683 observations now also including output measures, such as patents (as well as disruptive patents), publications (as well as citations) and future funding. The differences in the way research groups are organized is still evident. There are fewer faculty in those research groups that have non-federal funding, and the share of faculty that is female is also lower. Industry funded groups are more likely to patent and less likely to publish than are purely federally funded groups; their likelihood of receiving ongoing funding is very similar.

Table 1a: Descriptive Statistics of the Attributes of Research Groups

	All		Initial Share Non-Fed in Years 1-3					
			0%		(0%,10%]		(10%,100%]	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Distinct Grants (per year)	3.83	4.40	3.52	4.11	6.27	5.37	4.33	5.04
Distinct Faculty (per year)	2.29	2.37	2.25	2.38	3.03	2.30	2.02	2.25
Observed Years	5.24	1.93	5.15	1.81	4.93	2.05	6.16	2.44
Total Spending	$400,493	654,385	$385,984	$650,015	$501,511	$641,497	$435,265	$687,784
Workforce Composition (Fraction of Distinct Research Group Exclusive Employees)								
Faculty	0.35	0.23	0.36	0.26	0.31	0.19	0.31	0.21
Postdocs and Graduate Students	0.30	0.26	0.29	0.22	0.32	0.24	0.36	0.27
Fraction of Distinct Exclusive Employees by Role who are Female								
All Roles	0.42	0.27	0.42	0.27	0.43	0.25	0.4	0.29
Faculty	0.33	0.38	0.33	0.37	0.35	0.36	0.3	0.39
Postdocs and Graduate Students	0.39	0.33	0.38	0.32	0.41	0.32	0.39	0.33
N research groups	4,709		3,816		367		526	

Note: The above table shows the mean and standard deviations of the attributes of research groups. An observation in the above table is a research group. The variable "Distinct Grants (per year)" is defined as the average number of distinct grant numbers from which employees of the research group are paid within each year that a research group is observed in the data. The variable "Distinct Faculty (per year)" is defined as the average number of distinct faculty members paid by grants exclusively associated with the research group within each year that a research group is observed in the data. "Observed Years" represents the total number of calendar years for which a research group is observed within our data. "Total Spending" is the average total direct expenditures of a research group within each year charged to grants exclusively associated with the research group. The variables related to the workforce composition of the research group display the fraction of all distinct employees exclusively working for the research group who are listed as having each occupational title. For example, the "Faculty" variable is defined as the fraction of the distinct employees of a research group who are listed as holding the occupational title of "Faculty." The variables related to the gender of the employees within a research group are defined as the fraction of employees within a given occupation working for a research group who are female averaged over the years in which we observe the research group. For example, the variable "Faculty" is defined as the fraction of employees with the occupation 'Faculty' who are female. The left-most columns in the above table display the averages over all research groups. The subsequent columns show the averages separately for research groups according to fraction of funding that the research group received during its first three years observed in the data derived from non-federal sources.

Table 1b: Descriptive Statistics of Observations in Sample (Research Groups in a Year)

	All		Initial Share Non-Fed in Years 1-3					
			Specialized (0%)		Low Diversification (0%,10%]		Diversified (10%,100%]	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Distinct Grants	3.63	4.47	3.34	4.15	5.93	5.53	4.12	5.21
Distinct Faculty	2.17	2.59	2.16	2.64	2.76	2.36	1.87	2.34
Total Spending	$379,237	710,264	$365,881	715,805	$475,033	637,586	$406,663	710,385
Workforce Composition (Fraction of Distinct Research Group Exclusive Employees)								
Faculty	0.33	0.25	0.34	0.26	0.30	0.22	0.29	0.24
Postdocs and Graduate Students	0.32	0.29	0.31	0.29	0.33	0.27	0.37	0.30
Fraction of Distinct Exclusive Employees by Role who are Female								
All Roles	0.42	0.29	0.42	0.29	0.41	0.28	0.41	0.31
Faculty	0.31	0.39	0.32	0.39	0.32	0.37	0.27	0.39
Postdocs and Graduate Students	0.38	0.36	0.37	0.36	0.39	0.34	0.39	0.36
Patented (binary)	0.32	0.47	0.32	0.47	0.25	0.43	0.36	0.48
Publication (binary)	0.60	0.49	0.61	0.49	0.72	0.45	0.45	0.5
Disruptive Patent (binary)	0.12	0.32	0.13	0.33	0.04	0.20	0.10	0.30
Forward Citations	0.29	1.71	0.30	1.77	0.12	0.99	0.30	1.64
New Grant (binary)	0.62	0.49	0.60	0.49	0.80	0.40	0.62	0.49
N research groups	24,683		19,634		1,810		3,239	

Note: The above table shows the mean and standard deviations of the attributes of observations in our sample, a panel dataset of research groups observed each year. An observation in the above table is a research group in a year. The variable "Distinct Grants (per year)" is defined as the number of distinct grant numbers from which employees of the research group are paid. The variable "Distinct Faculty (per year)" is defined as the number of distinct faculty members paid by grants exclusively associated with the research group. "Total Spending" is the total direct expenditures charged to grants exclusively associated with the research group. The variables related to the workforce composition of the research group display the fraction of all distinct employees exclusively working for the research group who are listed as having each occupational title. For example, the "Faculty" variable is defined as the fraction of the distinct employees of a research group who are listed as holding the occupational title of "Faculty." The variables related to the gender of the employees within a research group are defined as the fraction of employees within a given occupation working for a research group who are female. For example, the variable "Faculty" is defined as the fraction of employees with the occupation 'Faculty' who are female. The left-most columns in the above table display the averages over all research groups. The subsequent columns show the averages separately for research groups according to fraction of funding that the research group received during its first three years observed in the data derived from non-federal sources.

4. Analytical Results

This section presents the empirical results. The first part describes the links between spending from braided funding sources and the workforce composition of the research group. The second part examines the relationship with gender diversity. The section concludes with an analysis of the portfolio of outputs, also at the research group level.

4.1 The link between funding sources and research group composition

The first set of results examine the link between changes in a research group's share of non-federal expenditures and their workforce composition (H1). As discussed, prior research has largely studied either cross-sectional variation in funding or used individuals, as opposed to research groups, as the unit of analysis. Here, the richness of our data permits analysis at the research group level and an examination of the relationship between changes in expenditures from non-federal research funding sources, both between and within units, over time.

The following model is estimated:

$$s_{i,t} = \alpha_{u(i),t} + \beta\ NonFed_{i,t} + \gamma ln\left(Spending_{i,t}\right) + \epsilon_{i,t} \quad (1)$$

In the above equation, the dependent variable $s_{i,t}$ is the share of the distinct employees working in research group *i* in year *t* who have a particular occupational title.

The model is estimated separately for the occupational titles: "Faculty," "Grad/PostdocsPostdoc," and "Research Staff." The key independent variable, *NonFed,* is the fraction of total spending that came from non-federal sources. The natural log of total spending, *Spenditng*$_{i,t}$ is included as a scale control measure. The baseline specification also includes institution by year fixed effects, represented by $\alpha_{u(i),t}$, where $u(i)$, is the university where the research group is based. The expectation is that when the share of expenditures from non-federal sources increases within a research group, the group reduces the share of graduate students and postdocs – that the estimate of β will be negative in the regression where the share of postdocs and graduate students is the dependent variable. There is no prior expectation on the coefficient for regressions on faculty or research staff.

The results of estimating Equation (1) are provided in Table 2a, where the dependent variable is the share of total distinct individuals from each occupation grouping who are paid by a given research group in a year. The first panel - columns 1-3 - presents results where the dependent variable is the share of faculty in a research group; the second panel - columns 4-6 - presents results where the dependent variable is the share of graduate students and postdocs, and the third panel - columns 7-9 - presents results where the dependent variable is the share of research staff.

The first column in each panel, which includes neither university nor time fixed effects, shows across-group correlations and finds that research groups with higher levels of non-federal spending have fewer faculty; the result holds across research groups, even when institution and year effects are included. The last column in each panel adds the individual research group controls, and as such describes within research group adjustments to increases in non-federal spending. The fixed-effects specification, in column (3), suggests that a ten percentage point increase in the share of expenditures from non-federal sources is associated with an increase in the share of faculty of 0.3 percentage points, relative to the unconditional mean of 31 percent this is a 0.8 percent change. The same ten percent increase was also found to be associated with a 0.6 percentage point decline in the share of graduate students and postdocs: a 0.7 percent change relative to the unconditional mean of 27 percent. These results support the first hypothesis that increases in non-federal expenditures are associated with lower utilization rates of trainee workers. Note that the results estimated from pooled cross-sectional variation (columns 1-2, 4-5, and 7-8) are different in sign from those estimated using within research group variation—adjustments across groups are very different than adjustments within groups.

Table 2a: The Link between Research Group Composition and Expenditures from Non-Federal Sources

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Faculty			Grad/Postdocs			Research Staff		
Frac. Expenditures- Non-Fed	-0.089***	-0.061***	0.025**	0.027	0.000	-0.055***	0.02	0.01	0.01
	(0.01)	(0.007)	(0.01)	(0.021)	(0.017)	(0.01)	(0.017)	(0.012)	(0.01)
Ln(Total Expenditures)	0.000	-0.010***	-0.014***	-0.020***	-0.011***	0.018***	0.039***	0.033***	0.021***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Research Group FE	No	No	Yes	No	No	Yes	No	No	Yes
Institution x Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
N	24,620								
Dep. Mean	.31			.27			.29		

Notes: This table presents the results of a regression of the share employment on share non-federal expenditures. The unit of observation in these regressions is a research group by year. The dependent variables are calculated as the share of distinct employees. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

The second set of analyses in Table 2b permits the analysis of adjustments across groups in more detail, postdoc. These analyses estimatee equation (1) on three separate subgroups of research group based on their initial proportion of research expenditures. The first group has specialized initial portfolios, and spends no money from non-federal sources in their first three years of existence. The second group has some diversification, but less than 10 percent of their expenditures are from non-federal sources. The third group is more diversified, with between 10 and 100 percent of their expenditures derived from non-federal sources. As shown below, the results for faculty (columns 1-2) indicate that a ten percentage point increase in non-federal spending is associated with a 0.7 to 1.1 percentage point decrease in the share of faculty for the specialized group, but a 1.4 to 1.9 percentage point increase and a 0.8 to 1 percentage point increase for the two diversified groups.postdocThe results on the share of employment in graduate students and postdocs (columns 3-4) were driven by the most diversified group, where a ten percentage point increase in non-federal spending had a significant positive effect on the share of employment in the range of 0.5 to 0.6 percentage points.

Table 2b: The Link between Research Group Composition and Diversification

Dep: Share Employment (Distinct %)	(1)	(2)	(3)	(4)	(5)	(6)
	Faculty		Grad/Postdocs		Research Staff	
Frac. Spending - Non-Fed	-0.109***	-0.077***	0.004	-0.012	0.038	0.014
	(0.013)	(0.009)	(0.025)	(0.02)	(0.025)	(0.016)
Ln(Total Spending)	-0.002	-0.009***	-0.018***	-0.011***	0.037***	0.033***
	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)
1[Low Diversification]	-0.055***	-0.031***	0.031*	-0.004	0.047***	0.060***
	(0.01)	(0.012)	(0.017)	(0.014)	(0.012)	(0.01)
1[Diversified	-0.067***	-0.053***	0.107***	0.073***	-0.014	0.001
	(0.013)	(0.014)	(0.017)	(0.017)	(0.01)	(0.01)
1[Low Diversification] * Frac. Spending - Non-Fed	0.189***	0.141***	-0.105**	-0.061	-0.080*	-0.074**
	(0.051)	(0.049)	(0.049)	(0.047)	(0.041)	(0.033)
1[Diversified] * Frac. Spending - Non-Fed	0.103***	0.078***	-0.056**	-0.053**	-0.027	-0.001
	(0.02)	(0.019)	(0.024)	(0.023)	(0.023)	(0.016)
Institution x Year FE	No	Yes	No	Yes	No	Yes
N	24620	24620	24620	24620	24620	24620
Dep. Mean	0.31	0.31	0.27	0.27	0.29	0.29

Notes: This table presents the results of regression the share employment on share non-federal expenditures on share non-federal spending broken out for three sub-groups defined by share of federal funding in years 1-3. The unit of observation in these regressions is a research group by year. The dependent variables are calculated as the share of distinct employees. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Each column uses observations from research groups based on the share of spending associated with non-federal sources during the first three years observed in the data. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

4.2 The link between funding sources and research group diversity

The data also permit an examination of the link between the research group's spending from non-federal funding and team gender diversity (H2). Equation (1) is modified by replacing the dependent variable with the share of distinct employees of a research group, within occupation and year, who are female. Following Hypothesis 2, the expectation is that $\beta < 0$; namely, the greater the share of spending from non-federal funding, the less likely there will be a female member on the team. Table 3a shows the results using the same approach as in Table 2a.

As before, the first column in each panel (columns 1, 4, and 6) provides an aggregate overview of the relationship between spending from non-federal funding sources and workforce composition, while the second column controls for time, but pooling both across and within variation in spending from non-federal sources (columns 2, 5, and 7). The last column in each panel shows the within-group adjustment as spending from non-federal sources changes. The results suggest that as the share of spending from non-federal sources increases, there are fewer female graduate students and postdocs. In terms of magnitude, a ten percentage point increase in the share of spending from non-federal funding sources was associated with a 0.5 percentage point decline in female graduate students and postdocs; relative to the unconditional mean of 40 percent this is a 1.1 percent change.

Finally, as expected, there is no significant correlation regarding changes in the gender composition of faculty members when looking within research groups over time. This is intuitive since the faculty member leaders of a research group are unlikely to change over time.

Table 3a: The Link between Gender Diversity and Expenditures from Non-Federal Sources

Dep: Share Female (Distinct %)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Faculty			Grad/Postdocs			Research Staff		
Frac. Spending- Non-Fed	-0.023***	-0.008	0.012	-0.035***	-0.019	-0.045**	-0.034***	-0.004	0.041**
	(0.009)	(0.009)	(0.013)	(0.021)	(0.017)	(0.01)	(0.017)	(0.012)	(0.01)
Ln(Total Spending)	0.003*	-0.06***	-0.003***	0.015***	0.005	0.001	0.007***	-0.003*	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.002)	(0.004)
Research Group FE	No	No	Yes	No	No	Yes	No	No	Yes
Institution x Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
N	19,373			17,299			16,830		
Dep. Mean	.31			.40			.57		

Notes: The unit of observation in these regressions is a research group by year. The dependent variables are the share of distinct employees of a research group within an occupational title who are female in a given year. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

Table 3b permits the analysis of adjustments across groups in more detail; as before, it shows equation (1) for three separate subgroups based on their initial proportion of research expenditures, but with a different dependent variable. The results for faculty (columns 1-2) show that a ten percentage point increase in non-federal spending is associated with a 2.4 to 4.2 percentage point decrease in the share of female faculty in a research group but only for the high initial share groups. For the graduate student and postdoc group, the result was again concentrated in the high initial share group with a ten percentage point increase in non-federal spending associated with a decrease of 2.6 to 4.8 percentage points. The results for research staff show that a ten percentage point increase in non-federal spending was associated with a marginally precise increase of 3.6 to 6 percentage points for the medium initial share group but a decrease of 4.7 to 9.2 percentage points for the high initial share group.

Table 3b: Estimates of the Link between Gender Diversity and Expenditures from Non-Federal Sources Across Levels of Diversification

Dep: Share Female (Distinct %)	(1)	(2)	(3)	(4)	(5)	(6)
	1[Female & Faculty]		1[Female & Grad/Postdocs]		1[Female & Research Staff]	
Frac. Spending - Non-Fed	0.212***	0.074	0.066	-0.096	0.500***	0.074
	(0.075)	(0.082)	(0.074)	(0.087)	(0.177)	(0.206)
Ln(Total Spending)	0.069***	0.052***	0.071***	0.074***	0.206***	0.180***
	(0.009)	(0.008)	(0.014)	(0.012)	(0.028)	(0.025)
1[Low Diversification]	0.175*	0.198**	0.411***	0.298**	0.157	0.295**
	(0.101)	(0.096)	(0.13)	(0.132)	(0.147)	(0.135)
1[Diversified]	0.182***	0.233***	0.438***	0.396***	0.533***	0.668***
	(0.063)	(0.065)	(0.093)	(0.099)	(0.157)	(0.157)
1[Low Diversification] * Frac. Spending - Non-Fed	-0.192	-0.083	-0.1	0.105	0.364	0.595**
	(0.126)	(0.119)	(0.204)	(0.195)	(0.281)	(0.288)
1[Diversified] * Frac. Spending - Non-Fed	-0.419***	-0.241***	-0.483***	-0.264**	-0.921***	-0.473*
	(0.071)	(0.08)	(0.096)	(0.112)	(0.215)	(0.257)
Institution x Year FE	No	Yes	No	Yes	No	Yes
N	24620	24620	24620	24620	24620	24620
Dep. Mean	0.3	0.3	0.33	0.33	0.66	0.66

Notes: The unit of observation in these regressions is a research group by year. The dependent variables are the share of distinct employees of a research group within an occupational title who are female in a given year. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

4.3 The link between funding sources and research group outputs

The final hypothesis concerned the expected differences in the types of scientific output produced by a research group with different funding sources (H3). The following model is estimated:

$$output_{i,t+2} = \alpha_{u(i),t} + \beta NonFed_{i,t} + \gamma ln\left(Spending_{i,t}\right) + \delta \boldsymbol{X}_{i,t} + \epsilon_{i,t} \tag{2}$$

In this regression, as in earlier ones, *i* represents a research group and *t* represents a year. The dependent variable $output_{i,t+2}$, represents the scientific output of the research group, measured as patents (including whether or not they are disruptive), publications (including citations) as well as whether the research group recived a new grant two years after the expenditure occurred.

The regression also includes a set of time-varying variables, $\boldsymbol{X}$, that control for changes in the workforce and the diversity composition of the research groups. These covariates mirror the dependent variables of equations (1). β can be interpreted as describing how patenting and publishing changes when the composition of funding changes; X controls for workforce allocation changes that are correlated with changes in non-federal funding. Because these regressions examine the correlation of research group's workforce composition and funding with outputs occurring two years later, the analytical dataset excludes observations in the final two years to avoid truncation. In total, there are 16,032 observations in these regressions.

Table 4a shows the results of the estimation of Equation 2 excluding the research group fixed-effects and relying on both the between and within variation. Columns 1-2 displays the results of regressing a binary indicator of whether the research group patented on the fraction of expenditures associated with non-federal funding two years prior. An increase of ten percentage points in spending from non-federal sources across research groups was associated with between a 2.3 and 2.4 percentage point increase in the likelihood that the group patents. Relative to the unconditional average rate of patenting, 0.32, this represents a 7.2 to 7.5 percent higher rate of patenting. Columns 3-4 demonstrate that research groups that spend more from non-federal sources are less likely to publish a scientific article two years later - a ten percentage point increase in spending from non-federal sources across research groups was associated with a decrease in the likelihood of publishing of between 1.2 to 1.4 percentage points or 2.5 to 3.2 percent of the mean of 44 percent.

Columns 5-6 of Table 4a display the results of regressing a binary indicator of whether a patent was disruptive based on whether it measured at or above the 95th percentile based on the CD index proposed by Funk and Owen-Smith (2017).[9] An increase of ten percentage points in spending from non-federal sources across research groups was associated with a 0.64 to 0.68 percentage point increase in the likelihood of producing a disruptive patent or about 5.3 to 5.8 percent above the mean of 12 percent. Columns 7-8 regress the log of one plus the total patent citations on the fraction of spending from non-federal sources: an increase of ten percentage points across research roups was associated with a 0.2 to 0.3 percent increase in patent citations, however, these coefficients are not significantly different from zero.

[9] Note the 95th percentile is calculated based on all patents in the PatentsView data with the sample application year not relative to the subset of those patents linked to UMETRICS data.

Table 4a: Relationship between Outputs and Spending from Non Federal Sources

Dep: scientific output	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1[Patent]		1[Publish]		1[Disruptive Patent]		ln(Patent Citations)	
Frac. Spending- Non-Fed	0.225***	0.235***	-0.141***	-0.118***	0.064***	0.068***	0.028	0.034
	(0.028)	(0.026)	(0.02)	(0.019)	(0.023)	(0.023)	(0.024)	(0.023)
Ln(Total Spending)	0.093***	0.096***	0.046***	0.024***	0.041***	0.042***	0.037***	0.038***
	(0.006)	(0.007)	(0.005)	(0.004)	(0.006)	(0.006)	(0.007)	(0.008)
Fraction of Faculty		0.104***		0.215***		0.025		0.066*
		(0.044)		(0.057)		(0.022)		(0.039)
Frac. Workforce - Grad/Postdoc		0.065		-0.08		0.44		0.027
		(0.047)		(0.052)		(0.027)		(0.048)
Frac. Workforce - Research Staff		-0.072		0.425***		-0.03		0.012
		(0.045)		(0.077)		(0.027)		(0.057)
Frac Female – Faculty		-0.012		0.043**		-0.015***		-0.024***
		(0.01)		(0.016)		(0.007)		(0.009)
Frac Female - Grad/Postdoc		-0.035**		0.137***		-0.024**		-0.042
		(0.015)		(0.019)		(0.01)		(0.029)
Frac Female - Research Staff		-0.016		0.097***		0.001		-0.019
		(0.012)		(0.018)		(0.015)		(0.011)
Institution x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	16,302							
Dep. Mean	0.32	0.32	0.44	0.44	0.12	0.12	0.10	0.10

Notes: The unit of observation in these regressions is a research group by year. The dependent variables are calculated based on occurrences two years after the expenditure data. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups. Note that the unit of observation is a single research group as opposed to previous tables where the unit of observation was a research group by year.

One additional point of interest with respect to Table 4a pertains to the coefficient estimates on the share females in various occupations. In Column 2, the negative coefficients on the covariates for the share of the faculty and graduate students who are female indicates that research groups with a higher proportion of female workers are on average less likely to patent. Research groups led by more female faculty are less likely to generate a disruptive patent (column 6) or a highly cited patent (column 8). In contrast, increases in the proportion of female faculty are positively correlated with the publishing of scientific papers (column 4). At the extreme, research groups that are led by all female faculty are predicted to be 4.3 percentage points, or 9.7 percent, more likely to publish a scientific paper two years later than a research group led by an all male set of faculty.

Table 4b permits the analysis of adjustments across groups in more detail. As before, Equation (2) focuses on three subgroups, defined according to their initial proportion of research expenditures across sources. The results of a ten percentage point increase in non-federal spending for the likelihood of patenting (columns 1-2) are mixed. They are positive for the groups characterized as low diversity in the range of 2.7 to 2.8 percentage points but negative for the more diversified groups and in the range of 0.9 to 1.7 percentage points. On the other hand, the same increase in non-federal spending was uniformly negative across the groups for the likelihood of publishing (columns 3-4) and in the range of 0.9 to 0.11 for the group with a low level of diversification, 0.09 to 1.9, and 1.1 to 1.7 percentage points for the most diversified group. There is no systematic evidence of the links between non-federal expenditures and the likelihood of producing a disruptive patent or a patents with high levels of citations.

Table 4b: Relationship between Outputs and Spending from Non Federal Sources Across Levels of Diversification

Dep: scientific output	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1[Patent]		1[Publish]		1[Disruptive Patent]		ln(Patent Citations)	
Frac. Spending - Non-Fed	0.270***	0.284***	-0.118***	-0.093***	0.103**	0.108**	0.36	0.43
	(0.034)	(0.046)	(0.023)	(0.023)	(0.041)	(0.041)	(0.044)	(0.103)
Ln(Total Spending)	0.092***	0.095***	0.045***	0.023***	0.041***	0.041***	0.037***	0.038***
	(0.007)	(0.007)	(0.005)	(0.004)	(0.006)	(0.007)	(0.017)	(0.007)
1[Low Diversification]	-0.038	-0.032	0.143***	0.095***	-0.055***	-0.056***	-0.042***	-0.041***
	(0.037)	(0.035)	(0.04)	(0.036)	(0.017)	(0.017)	(0.014)	(0.015)
1[Diversified]	-0.013	-0.02	0.154***	0.084***	0.019	-0.027*	-0.061**	-0.023
	(0.025)	(0.026)	(0.03)	(0.022)	(0.019)	(0.015)	(0.03)	(0.017)
1[Low Diversification] * Frac. Spending - Non-Fed	0.043	-0.097	-0.097	-0.190**	0.031	0.022	0.468**	0.189
	(0.10)	(0.096)	(0.096)	(0.093)	(0.102)	(0.053)	(0.179)	(0.12)
1[Diversified] * Frac. Spending - Non-Fed	-0.085	-0.169***	-0.169***	-0.109***	-0.073	-0.033	-0.009	-0.009
	(0.055)	(0.041)	(0.041)	(0.038)	(0.079)	(0.076)	(0.061)	(0.061)
Fraction of Faculty		0.108**		0.221***		0.023		0.065*
		(0.042)		(0.056)		(0.055)		(0.038)
Frac. Workforce - Grad/Postdoc		0.09		-0.04		0.051*		0.131
		(0.057)		(0.061)		(0.028)		(0.097)
Frac. Workforce - Research Staff		-0.079		0.410***		0.002		0.019
		(0.06)		(0.083)		(0.061)		(0.058)
Frac Female – Faculty		-0.012		0.044***		-0.015**		-0.024***
		(0.02)		(0.017)		(0.007)		(0.032)
Frac Female - Grad/Postdoc		-0.034***		0.137***		-0.024**		-0.016
		(0.015)		(0.018)		(0.001)		(0.046)
Frac Female - Research Staff		-0.022		0.096***		0.005		-0.018
		(0.019)		(0.017)		(0.02)		(0.053)
Institution x Year FE	yes	yes	yes	yes	yes	yes	yes	yes
N	16302							
Dep. Mean	0.32	0.32	0.44	0.44	0.12	0.12	0.1	0.1

Notes: The unit of observation in these regressions is a research group by year. The dependent variables are based on two years after the expenditure data. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups. Note that the unit of observation is a single research group as opposed to previous tables where the unit of observation was a research group by year.

Table 5 presents the results of estimating (2), using research group fixed-effects and relying exclusively on within research group variation; the effects are not significant for patents. This result may be due to data limitations. Patenting is a relatively rare event, and our sample period is relatively short. In terms of publications, the results suggest that a ten percent increase in the share of non-federal research funding is associated with a relatively small 0.85 percentage point decline in the likelihood of publishing a scientific article. Relative to the prior cross-sectional estimates, we note that our estimates with research group fixed-effects are the same in sign but much smaller in magnitude and less statistically precise. One conclusion in comparing the between and within estimates is that within and across group adjustment is similar.

Table 5: Research Group Fixed-Effects Regression of Scientific Output on Share Non-Federal Funding

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1[Patent]		1[Publish]		1[Disruptive Patent]		ln(Patent Citations)	
Frac. Spending- Non-Fed	0.024 (0.027)	0.023 (0.028)	-0.085*** (0.017)	-0.074*** (0.018)	0.003 (0.025)	0.006 (0.025)	-0.035 (0.043)	-0.029 (0.043)
Ln(Total Spending)	0.006 (0.007)	0.005 (0.007)	0.009*** (0.003)	0.004 (0.003)	0.003 (0.005)	0.001 (0.006)	0.006 (0.005)	0.009 (0.006)
Fraction of Faculty		0.03 (0.035)		-0.022 (0.027)		-0.036 (0.032)		0.07 (0.074)
Frac. Workforce - Grad/Postdoc		-0.03 (0.035)		0.003 (0.022)		0.019 (0.032)		0.085 (0.084)
Frac. Workforce - Research Staff		-0.002 (0.025)		0.025 (0.024)		0.044 (0.047)		0.089 (0.118)
Frac Female – Faculty		0.033 (0.015)		0.029* (0.016)		0.029 (0.018)		0.033 (0.021)
Frac Female - Grad/Postdoc		0.021 (0.018)		0.010 (0.013)		0.008 (0.012)		0.02 (0.021)
Frac Female - Research Staff		-0.001 (0.001)		0.005 (0.011)		0.008 (0.012)		0.02 (0.016)
Research Group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	16,032							
Dep. Mean	0.32	0.32	0.44	0.44	0.12	0.12	0.10	0.10

Notes: The unit of observation in these regressions is a research group by year. The dependent variables are from two years after the expenditure data. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups. Note that the sample size is different than in the previous samples because of truncation in the dependent variable.

Table 6 presents the results of estimating Equation (2), with new grant funding as an aggregate metric of research group success. The regression estimates in columns 1-2 show that research groups with a 10 percentage point higher share of funding coming from non-federal sources have a 4.6 to 5.4 percentage point lower chance of receiving a new grant two years later. This is equivalent to a 7.4 to 8.7 percent lower rate of receiving new funding. With the addition of fixed effects for the research group in columns 3-4, the estimated coefficinents on the share of funding from non-federal sources becomes insignificant.

The initial composition of research portfolios appears to make a difference in funding success, however. As columns 5 and 6 demonstrate, diversified research groups are more successful in getting follow on funding, although the effect of spending additional non federal dollars appears to be mixed.

Table 6: Regression of New Grant Two Years Later on Share Non-Federal Funding

	(1)	(2)	(3)	(4)	(5)	(6)
	1[New Grant]		1[New Grant]		1[New Grant]	
Frac. Funding- Non-Fed	-0.054***	-0.046**	-0.022	-0.028	-0.109***	-0.102***
	(0.02)	(0.02)	(0.037)	(0.038)	(0.021)	(0.02)
Ln(Total Funding)	0.046***	0.041***	-0.007	-0.006	0.047***	0.043***
	(0.005)	(0.005)	(0.006)	(0.007)	(0.005)	(0.004)
[LowDiversification]					0.155***	0.105***
					(0.022)	(0.017)
[Diversified]					0.143***	0.124***
					(0.029)	(0.028)
[Low Diversification] * Frac. Spending - Non-Fed					0.286***	0.224***
					(0.029)	(0.101)
[Diversified] * Frac. Spending - Non-Fed					-0.008	-0.01
					(0.036)	(0.035)
Fraction of Faculty		0.076**		-0.013		0.071***
		(0.034)		(0.045)		(0.034)
Frac. Workforce - Grad/Postdoc		-0.080*		-0.038		-0.091**
		(0.041)		(0.047)		(0.042)
Frac. Workforce - Research Staff		0.022		0.001		-0.009
		(0.043)		(0.056)		(0.039)
Frac Female - Faculty		-0.041***		0.014		-0.038***
		(0.014)		(0.026)		(0.013)
Frac Female - Grad/Postdoc		-0.024*		0.002		-0.024*
		(0.013)		(0.02)		(0.013)
Frac Female - Research Staff		-0.013		-0.021		-0.062***
		(0.008)		(0.022)		(0.014)
Research Group FE	No	No	Yes	Yes	No	No
Institution x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	16,302					
Dep. Mean	0.62	0.62	0.62	0.62	0.62	0.62

Notes: The unit of observation in these regressions is a research group by year. The dependent variables occur two years after the expenditure data. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups. Note that the unit of observation is a single research group as opposed to previous tables where the unit of observation was a research group by year.

Conclusion

This paper has used new data to examine the mechanism whereby changes in the sources of funding affect the composition and the output of research groups. Using these data, it is possible to examine the links between research funding structure and workforce composition across groups, as well as the composition of their outputs. They also allow us to examine the way in which research groups adjust in response to changes in funding sources over time, both in terms of their workforce composition and outputs.

As always, much more can be done. This study has simply contrasted federal and non-federal funding sources. The data have much more detailed information about not only the source of federal funds (for example, from NIH, NSF, DoD and DoE), but also the type of grants (for example, their size, term, and structure) and the degree of collaboration with other universities. The type of non-federal expenditures is also included in the data, notably whether the funding is from industry or philanthropic foundations, although we do not exploit this information in our own analysis. Also included is information about purchases from vendors, thereby permitting the examination of complementarities between purchases of scientific inputs and the usage of research staff.

In addition, the data and its use continues to expand. Over 35 major universities have joined the Institute for Research on Innovation and Science, with another 30 universities in various stages of agreement.

The research community has added linkages to thousands of dissertations, 8 million U.S. patents, tens of millions of scientific publications, and as well as public information on hundreds of thousands federal grants. There is also a growing community of more than 120 researchers from around the world that has made use of 3 annual UMETRICS data releases both at IRIS and at the Federal Statistical Research Data Center network.

References

1. National Science Foundation National Center for Science and Engineering Statistics, Higher Education Research and Development Survey. (2017).
2. National Institutes of Health, "NIH-Wide Strategic Plan Fiscal Years 2016-2020" (2015), (available at https://www.nih.gov/sites/default/files/about-nih/strategic-plan-fy2016-2020-508.pdf).
3. National Science Foundation Strategic Plan, "Investing in Science, Engineering and Education for the Nation's Future" (Washington DC, 2014).
4. P. M. Romer, Government Can Do More to Support Science and Innovation. *Wall Str. J.* (2018), (available at https://www.wsj.com/articles/government-can-do-more-to-support-science-and-innovation-1542758326?mod=e2two).
5. N. Rosenberg, R. R. Nelson, American universities and technical advance in industry. *Res. Policy*. **23**, 323–348 (1994).
6. M. Gulbrandsen, J.-C. Smeby, Industry funding and university professors' research performance. *Res. Policy*. **34**, 932–950 (2005).
7. M. Perkmann *et al.*, Academic engagement and commercialisation: A review of the literature on university–industry relations. *Res. Policy*. **42**, 423–442 (2013).
8. A. Geuna, The changing rationale for European university research funding: are there negative unintended consequences? *J. Econ. Issues*. **35**, 607–632 (2001).
9. P. A. David, B. H. Hall, A. A. Toole, Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Res. Policy*. **29**, 497–529 (2000).
10. P. Aghion, M. Dewatripont, J. C. Stein, Academic freedom, private-sector focus, and the process of innovation. *RAND J. Econ.* **39**, 617–635 (2008).
11. P. Azoulay, J. S. G. Zivin, G. Manso, Incentives and creativity: evidence from the academic life sciences. *RAND J. Econ.* **42**, 527–554 (2011).
12. C. Beaudry, S. Allaoui, Impact of public and private research funding on scientific production: The case of nanotechnology. *Res. Policy*. **41**, 1589–1606 (2012).
13. M. Coccia, G. Falavigna, A. Manello, The impact of hybrid public and market-oriented financing mechanisms on the scientific portfolio and performances of public research labs: a scientometric analysis. *Scientometrics*. **102**, 151–168 (2015).
14. A. Fabbri, A. Lai, Q. Grundy, L. A. Bero, The influence of industry sponsorship on the research agenda: a scoping review. *Am. J. Public Health*. **108**, e9–e16 (2018).
15. B. Goldfarb, The effect of government contracting on academic research: Does the source of funding affect scientific output? *Res. Policy*. **37**, 41–58 (2008).
16. P. A. David, B. H. Hall, Heart of Darkness: Modeling Public\textendash{}Private Funding Interactions inside the {{R}} & {{D}} Black Box q, 19 (2000).
17. A. A. Payne, Measuring the Effect of Federal Research Funding on Private Donations at Research Universities: Is Federal Research Funding More than a Substitute for Private Donations? *Int. Tax Public Financ.* **8**, 731–751 (2001).
18. L. Wu, D. Wang, J. A. Evans, Large Teams Have Developed Science and Technology; Small Teams Have Disrupted It (2017).
19. A. Conti, C. C. Liu, Bringing the lab back in: Personnel composition and scientific output at the MIT Department of Biology. *Res. Policy* (2015), doi:10.1016/j.respol.2015.01.001.

20. J. I. Lane, J. Owen-Smith, R. F. Rosen, B. A. Weinberg, New linked data on research investments: Scientific workforce, productivity, and public value. *Res. Policy* (2015), doi:10.1016/j.respol.2014.12.013.
21. M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks. *Phys. Rev. E*. **69**, 26113 (2004).
22. K. Knorr-Cetina, *Epistemic Cultures: How the Sciences Make Knowledge* (Harvard University Press, Cambridge, MA, 2009).
23. J. Owen-smith, Managing Laboratory Work through Skepticism: Processes of Evaluation and Control. *Am. J. Sociol.* **66**, 427–452 (2001).
24. A. Varga, Shorter distances between papers over time are due to more cross-field references and increased citation rate to higher-impact papers. *Proc. Natl. Acad. Sci.*, 201905819 (2019).
25. J. Moody, The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *Am. Sociol. Rev.* **69**, 213–238 (2004).
26. J. A. Evans, Industry induces academic science to know less about more. *Am. J. Sociol.* **116**, 389–452 (2010).
27. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. *Science*. **342**, 468–472 (2013).
28. C. Catalini, Microgeography and the direction of inventive activity. *Manage. Sci.* (2017).
29. F. G. Murciano-Goroff, Why and Wherefore of Increased Scientific Collaboration. *Nber*, 1–5 (2014).
30. S. Wuchty, B. F. Jones, B. Uzzi, The Increasing Dominance of Teams in Production of Knowledge. *Science (80-.).* **316**, 1036–1039 (2007).
31. S. Paruchuri, A. Nerkar, D. C. Hambrick, Acquisition Integration and Productivity Losses in the Technical Core: Disruption of Inventors in Acquired Companies. *Organ. Sci.* **17**, 545–562 (2006).
32. W. Powell, E. Giannella, in *The Handbook of Innovation* (Elsevier, Amsterdam, 2010).
33. K. B. Whittington, L. Smith-Doerr, Women inventors in context: Disparities in patenting across academia and industry. *Gend. Soc.* **22**, 194–218 (2008).
34. D. K. Ginther *et al.*, Race, ethnicity, and NIH research awards. *Science (80-.).* **333**, 1015–1019 (2011).
35. K. Arrow, in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, National Bureau of Economic Research, Ed. (1962).
36. P. Dasgupta, P. A. David, Toward a new economics of science. *Res. Policy*. **23**, 487–521 (1994).
37. R. R. Nelson, The Market Economy, and the Scientific Commons. *Res. Policy*. **33**, 455–471 (2004).
38. T. Bodenheimer, Clinical investigators and the pharmaceutical industry-Reply. *N. Engl. J. Med.* **343**, 511 (2000).
39. A. Muscio, D. Quaglione, G. Vallanti, Does government funding complement or substitute private research funding to universities? *Res. Policy*. **42**, 63–75 (2013).
40. A. Banal-Estañol, M. Jofre-Bonet, C. Lawson, The Double-Edged Sword of Industry Collaboration: Evidence from Engineering Academics in the UK. *Res. Policy*. **44**, 1160–1175 (2015).
41. A. Banal-Estañol, I. Macho-Stadler, D. Pérez-Castrillo, Research Output From University–Industry Collaborative Projects. *Econ. Dev. Q.* **27**, 71–81 (2013).
42. J. Owen-Smith, W. W. Powell, The expanding role of university patenting in the life sciences: assessing the importance of experience and connectivity. *Res. Policy*. **32**, 1695–1711 (2003).
43. R. J. Funk, J. Owen-Smith, A dynamic network measure of technological change. *Manage. Sci.* **63**, 791–817 (2017).
44. Technology transfer: Industry-funded academic inventions boost innovation. *Nat. News*. **507**, 297 (2014).
45. J. Haltiwanger, J. Lane, J. Spletzer, Productivity Differences Across Employers: The Role of Employer Size, Age, and Human Capital. *Am. Econ. Rev.* **89**, 94–98 (1999).
46. The Institute for Research on Innovation & Science (IRIS), Documentation for UMETRICS 2017Q4a Dataset: The Second Annual Data Release, , doi:doi:10.21987/R7R94D.
47. I. F. R. O. I. A. S. (IRIS), UMETRICS 2017Q4a Core Files (2018), (available at http://iris.isr.umich.edu/research-data/2018datarelease-corefiles/).
48. M. Girvan, M. E. J. Newman, Community structure in social and biological networks. *Proc. Natl. Acad.*

Sci. **99**, 7821–7826 (2002).

Appendix

Heterogeneity in Workforce Composition Changes

We examine differences in how non-federal funding correlates with changes in the workforce composition by estimating Equation (1) using observations from different subsets of research groups. Specifically, we examine subsets of research groups based on the fraction of their funding that came from non-federal sources during the first three years that we observe them in our panel dataset. The results are displayed in the tables below.

Table A1: The relationship between expenditures from non federal funding sources and workforce composition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Faculty			Grad/Postdocs			Research Staff		
Initial Share Non-Fed in Years 1-3	0%	(0%,10%]	(10%,100%]	0%	(0%,10%]	(10%,100%]	0%	(0%,10%]	(10%,100%]
Frac. Spending- Non-Fed	-0.066***	0.061	0.009	-0.024	-0.134***	-.075***	0.024	0.005	0.015
	(0.009)	(0.042)	(0.013)	(0.021)	(0.05)	(0.016)	(0.015)	(0.039)	(0.012)
Ln(Total Spending)	-0.005**	-0.026***	-0.027***	-0.012***	-0.0136***	0.015***	0.031***	0.057***	0.025***
	(0.002)	(0.004)	(0.004)	(0.003)	(0.008)	(0.004)	(0.002)	(0.005)	(0.003)
Research Group FE	No	No	No	No	No	No	No	No	No
Institution x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	19,582	1,805	3,233	19,582	1,805	3,233	19,582	1,805	3,233
Dep. Mean	0.32	0.28	0.26	0.26	0.27	0.34	0.29	0.34	0.26

Notes: Regression of Share Employment on Share Non-Federal Funding by Share of Federal Funding in Years 1-3The unit of observation in these regressions is a research group by year. The dependent variables are calculated as the share of distinct employees. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Each column uses observations from research groups based on the share of funding that group received from non-federal sources during the first three years that we observe thenm in our data. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

Table A2: The relationship between expenditures from non federal funding sources and workforce composition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Faculty			Grad/Postdocs			Research Staff		
Initial Share Non-Fed in Years 1-3	0%	(0%,10%]	(10%,100%]	0%	(0%,10%]	(10%,100%]	0%	(0%,10%]	(10%,100%]
Frac. Spending- Non-Fed	0.000	0.004	0.046***	-0.041**	-0.105***	-.059***	0.001	0.02	0.005
	(0.01)	(0.01)	(0.01)	(0.001)	(0.018)	(0.01)	(0.017)	(0.028)	(0.01)
Ln(Total Spending)	-0.012***	-0.043***	-0.012	0.016***	0.009	0.028***	0.022***	0.025***	0.018***
	(0.003)	(0.001)	(0.001)	(0.003)	(0.007)	(0.006)	(0.002)	(0.007)	(0.003)
Research Group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	19,582	1,805	3,233	19,582	1,805	3,233	19,582	1,805	3,233
Dep. Mean	0.32	0.28	0.26	0.26	0.27	0.34	0.29	0.34	0.26

Notes: Regression of Share Employment on Share Non-Federal Funding by Share of Federal Funding in Years 1-3The unit of observation in these regressions is a research group by year. The dependent variables are calculated as the share of distinct employees. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Each column uses observations from research groups based on the share of funding that group received from non-federal sources during the first three years that we observe thenm in our data. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

Gender of Employees and Non-Federal Funding

The table below repeats the estimation done in Table 3, but uses a binary indicator of whether an employee from a given occupational grouping is a female. The results are similar but slightly larger in magnitude and more precise.

Table A.3: Regression of Female Employment on Share of Expenditures from Non-Federal Funding

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Female Faculty			Female Grad/Postdocs			Female & Research Staff		
Frac. Spending- Non-Fed	0.101**	0.05	-0.038	0.034	-0.06	-0.117**	0.349***	0.155	-0.268***
	(0.047)	(0.049)	(0.042)	(0.051)	(0.056)	(0.054)	(0.115)	(0.124)	(0.096)
Ln(Total Spending)	0.073***	0.053***	0.037***	0.071***	0.073***	0.068***	0.208***	0.175***	0.109***
	(0.009)	(0.009)	(0.005)	(0.014)	(0.012)	(0.010)	(0.027)	(0.024)	(0.014)
Research Group FE	No	No	Yes	No	No	Yes	No	No	Yes
Institution x Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
N	24620								
Dep. Mean	0.3			0.33			0.66		

Notes: The unit of observation in these regressions is a research group by year. The dependent variables are indicators of female employment in the requisite job category. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

The below table repeats the estimation done in Table 3, where the employment shares are contstructed from employees exclusively working for only one research group. The results are similar except that we find a positive coefficient for the within-group correlation with non-federal funding for research staff being female.

Table A.4: OLS Regression of the Share of Distinct Exclusive Employees by Occupation who are Female on Share Non-Federal Funding

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Faculty			Grad/Postdocs			Research Staff		
Frac. Spending Non-Fed	-0.050***	-0.040***	-0.012	-0.047***	-0.028**	-0.048*	-0.013	-0.001	0.014
	(0.01)	(0.012)	(0.011)	(0.013)	(0.013)	(0.025)	(0.015)	(0.018)	(0.023)
Ln(Total Spending)	0.008***	-0.001	0.001	0.012***	0.005	0.005	0.008***	-0.006	0.009
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.005)
Research Group FE	No	No	Yes	No	No	Yes	No	No	Yes
Institution x Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
N	24620								
Dep. Mean	.31			.38			.56		

Notes: The unit of observation in these regressions is a research group by year. The dependent variables are the share of distinct exclusive employees of a research group within an occupational title who are female in a given year. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups.

Cross-Sectional Analysis of Outputs

We also provide a cross-sectional examination of the correlation between spending from non federal funding sources and worforce configurations of research groups with outputs. In this regression, we look at how the worforce and funding of a research group during years 1-3 correlate with outputs during years 4-5. We use the same outputs as Equation (2)—the patents, publications, and new grants of a group—defined over the two specific years of 4-5.

The following table shows summary statistics for this cross-sectional dataset. In this cross-sectional dataset, an observation is a research group. The attributes of the research group are shown for years 1-3. The output variables shown are for years 4-5.

Table A5: Descriptive Statistics of Cross-Sectional Dataset

	All		Initial Share Non-Fed in Years 1-3					
			0%		(0%,10%]		(10%,100%]	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Distinct Grants (per year)	3.46	3.91	2.92	3.22	6.14	5.28	4.14	4.8
Distinct Faculty (per year)	2.12	2.26	2.02	2.2	2.99	2.36	2.02	2.3
Observed Years	5.45	1.99	5.38	1.83	4.93	2.05	6.16	2.44
Total Spending	$375,663	656,953	$342,885	642,674	$507,615	665,011	$438,390	701,693
Workforce Composition (Fraction of Distinct Research Group Exclusive Employees)								
Faculty	0.35	0.24	0.37	0.25	0.30	0.27	0.36	0.28
Postdocs and Graduate Students	0.31	0.27	0.30	0.27	0.32	0.29	0.25	0.22
Fraction of Distinct Exclusive Employees by Role who are Female								
All Roles	0.43	0.21	0.31	0.28	0.32	0.23	0.29	0.27
Faculty	0.31	0.27	0.42	0.29	0.42	0.26	0.4	0.27
Postdocs and Graduate Students	0.41	0.28	0.56	0.29	0.56	0.25	0.54	0.28
Patented (binary)	0.42	0.49	0.36	0.48	0.59	0.49	0.58	0.49
Publication (binary)	0.56	0.5	0.55	0.5	0.72	0.45	0.5	0.5
Disruptive Patent (binary)	0.2	0.4	0.17	0.37	0.37	0.48	0.26	0.44
Forward Citations	2.1	6.41	1.94	6.18	2.1	4.19	2.85	8.41
New Grant (binary)	0.92	0.27	0.91	0.29	0.99	0.12	0.95	0.22
N research groups	4,709		3,816		367		526	

Note: The above table shows the mean and standard deviations of the attributes of research groups. An observation in the above table is a research group. The variable "Distinct Grants (per year)" is defined as the average number of distinct grant numbers from which employees of the research group are paid within each year that a research group is observed in the data. The variable "Distinct Faculty (per year)" is defined as the average number of distinct faculty members paid by grants exclusively associated with the research group within each year that a research group is observed in the data. "Observed Years" represents the total number of calendar years for which a research group is observed within our data. "Total Spending" is the average total direct expenditures of a research group within each year charged to grants exclusively associated with the research group. The variables related to the workforce composition of the research group display the fraction of all distinct employees exclusively working for the research group who are listed as having each occupational title. For example, the "Faculty" variable is defined as the fraction of the distinct employees of a research group who are listed as holding the occupational title of "Faculty." The variables related to the gender of the employees within a research group are defined as the fraction of employees within a given occupation working for a research group who are female averaged over the years in which we observe the research group. For example, the variable "Faculty" is defined as the fraction of employees with the occupation 'Faculty' who are female. The left-most columns in the above table display the averages over all research groups. The subsequent columns show the averages separately for research groups according to fraction of funding that the research group received during its first three years observed in the data derived from non-federal sources.

In the below table, we show the estimated coefficients from the cross-sectional regression.

Table A6: Cross-Sectional Regression of Scientific Output on Share Non-Federal Funding

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1[Patent]		1[Publish]		1[Disruptive Patent]		ln(Patent Citations)	
Frac. Spending – Non-Fed	0.189***	0.208***	-0.191***	-0.183***	0.138***	0.148***	0.035***	0.412***
	(0.052)	(0.053)	(0.064)	(0.057)	(0.037)	(0.037)	(0.119)	(0.121)
Ln(Total Spending)	0.108***	0.117***	0.050***	0.026**	0.060***	0.064***	0.147***	0.156***
	(0.015)	(0.014)	(0.011)	(0.011)	(0.014)	(0.015)	(0.042)	(0.043)
Fraction of Faculty		0.216**		0.494***		0.112**		0.389**
		(0.092)		(0.091)		(0.052)		(0.163)
Frac. Workforce – Grad/Postdoc		0.059		0.112		0.025		0.147
		(0.085)		(0.084)		(0.047)		(0.092)
Frac. Workforce – Research Staff		0.014		0.608***		-0.029		0.012
		(0.023)		(0.121)		(0.049)		(0.091)
Frac Female – Faculty		-0.095		0.059**		0.017		-0.071
		(0.01)		(0.029)		(0.018)		(0.057)
Frac Female – Grad/Postdoc		-0.136***		0.186***		-0.049*		0.000
		(0.03)		(0.032)		(0.025)		(0.06)
Frac Female – Research Staff		-0.008		0.140***		0.004		-0.048
		(0.024)		(0.032)		(0.031)		(0.05)
Institution x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4,702							
Dep. Mean	0.42	0.42	0.56	0.56	0.20	0.20	0.42	0.42

Notes: The unit of observation in these regressions is a research group during years 1-4. The dependent variables are calculated using years 4-5. The two key variables of interest are constructed using total direct expenditure, not including overhead charges. Coefficient estimates concatenated with * represents a p-value < 0.1, ** represents a p-value < 0.05, and *** represents a p-value <0.1. Standard errors are in parenthesis and clustered at the institution by year level but robust to clustering on research groups. Note that the sample size is different than in the previous samples because of truncation in the dependent variable.