I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11711

# Money for Something: The Links between Research Funding and Innovation

Britta Glennon
Julia Lane
Ridhima Sodhi

JULY 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11711

# Money for Something: The Links between Research Funding and Innovation

**Britta Glennon**
*Carnegie Mellon University*

**Julia Lane**
*New York University and IZA*

**Ridhima Sodhi**
*New York University*

JULY 2018



**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: The Links between Research Funding and Innovation[1]

Federal research funding to universities is often based on a desire to stimulate innovation – so that they spend taxpayer money for "something". There is growing understanding of the need to change the structure of research funding in order to do so; less is known about the effectiveness of different organizational structures. Yet, as Jones has pointed out, increasing the efficiency with which we transfer knowledge from one generation to the next could have important implications for innovation and productivity growth. In this paper we use new data to examine how the main organizational structure used to train the next generation of scientists and inventors – teams funded by research grants – leads to innovative activity as measured by patents.

**JEL Classification:** O30, O31, O38

**Keywords:** UMETRICS, innovation, patents, research policy, teams

**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

1 This research was supported by the National Center for Science and Engineering Statistics. NSF SciSIP Awards 1064220 and 1262447; NSF Education and Human Resources DGE Awards 1348691, 1547507, 1348701, 1535399, 1535370; NSF NCSES award 1423706; NIHP01AG039347; the US Patent and Trademark Office and the Ewing Marion Kaufman and Alfred P. Sloan Foundations. All UMETRICS results have been reviewed to ensure that no confidential information is disclosed. The research agenda draws on work with many coauthors, but particularly Bruce Weinberg and Jason Owen Smith who provided very useful comments on this paper. We are extremely grateful to Amanda Murphy, Andrew Toole, Sanjay Arora and Christina Jones for their helpful suggestions and guidance. All errors, of course, are our own.

## 1. Introduction

Federal research funding to universities is often based on a desire to stimulate innovation – so that they spend taxpayer money for "something" (*1*). There is growing understanding of the need to change the structure of research funding in order to do so; less is known about the effectiveness of different organizational structures. Yet, as Jones has pointed out, increasing the efficiency with which we transfer knowledge from one generation to the next could have important implications for innovation and productivity growth(*2*). In this paper we use new data to examine how the main organizational structure used to train the next generation of scientists and inventors – teams funded by research grants – leads to innovative activity as measured by patents.

There are some intriguing findings in the literature, derived from analyses of patent and publication data, that suggests both that the structure of science is changing and the results of these changes are mixed. There are increasing numbers of authors of publications and numbers of inventors on patents, suggesting that productive scientific research may require larger teams (*3*). Multi-institutional collaboration is also increasing, as evidenced by the increasing numbers of institutional affiliations of journal coauthors(*4*); however, research suggests that such collaborations may reduce rather than increase productivity(*5*). Diversity in teams is often cited as a concern (*6*, *7*), yet management researchers suggest that functional and educational diversity is more important than race and gender (*8*).

As interesting as the current research is, however, it results from analyzing the outcomes without understanding the process leading to these outcomes. As Whittington has noted, little is known about the structure of teams before patenting, and teams that do not patent at all(*9*). As the famous New Yorker cartoon would have it, we can now be more explicit on the "step 2" – the activities of research teams – that lie between the initial step of observing research funding and the final step of observing a miracle occurring in the form of publications and patents (*10*). In this paper, we use new data to examine the entire process, from research funding to resulting patenting, and the features of the teams that shape this process. We can, for the first time, use administrative data to describe differences in such organizational features as size and diversity between those research teams that produce work that result in patents and those that do not. We are able to directly measure whether or not multi-institution collaboration occurs prior to patenting, since the administrative data capture whether or not teams from different universities are funded on the same research grant. We are also able to describe the differences – in terms of both diversity and inventor collaboration – between patenting research teams and non-patenting research teams. Finally, we can identify which individuals on patenting research teams are named inventors on the resulting patents and who is not. This last piece has important implications for any analysis that describes teams based on inventor characteristics alone, which may contain systematic bias.

The work is made possible for two reasons. The first is the availability of a new source of information about the ties between university research funding and patents. Patents are often seen as credible proxies for the innovative activity of scientists(*11–13*); indeed, funding agencies require the reporting of patents as outputs of research funding. The second is the availability of new data about research teams derived from the administrative records of universities themselves and matched to a variety of other datasets (*14*)

Our descriptive results are striking. We find that higher levels of research funding are associated with a higher propensity to patent. But funding does not magically result in patenting; the core link to understanding the pathways to innovation is understanding how the funding affects the structure of teams that are likely to patent, and this shows up in our regression results, where funding is only one of the economically and statistically significant explanatory variables in predicting the propensity to patent. In particular, regressions using our new data suggests that larger teams, with more postdocs and graduate students, are more likely to patent.

We also find marked differences in the propensity to patent in teams funded by different agencies. Unsurprisingly, although teams with more postdocs and graduate students are more likely to patent, faculty are more likely than postdocs and graduate students to be identified as inventors. In terms of diversity, teams with female representation are less likely to patent. Female faculty, graduate students and postdocs are systematically less likely than males to be identified as inventors. These findings suggest that analyses of team structure and diversity using the inventor teams listed on patents may mischaracterize the true nature of the team performing the underlying research. Findings such as those by Whittington and Smith-Doerr, that women are less involved in patenting than men may simply reflect the lower probability of women being named as inventors(*15*). Although the analysis is not causal in nature, the results are strongly significant and robust to a variety of different specifications. Further analysis, particularly about the differences in the structure and composition of collaboration networks across universities, is warranted using the new rich data sources which are now becoming available to researchers in the field(*16*).

## 2. Background

There is a rich organizational literature that examines the processes that lead to innovation. In the case of scientific research, a great deal of attention has been paid to the size of scientific teams, the importance of diversity and the effect of cross institutional collaboration.

As Jones points out innovation is the result of a learning process which is increasingly done in teams(*2*), and increasingly done at greater scientific scale(*17*, *18*). The phenomenon of greater project size is likely to have substantive effects on the size and scope of scientific networks – and for the transmission of knowledge. Research has demonstrated that networks influence the extent to which individuals can access and use information and resources (*19*) as well as their expectations and preferences (*20*, *21*). Network position also shapes both individuals' and groups' ability to innovate (*22*, *23*). In particular, when individuals hold network positions that connect them to groups with different knowledge bases they are well positioned to innovate (*24*). Certainly, team-authored paper are characterized by higher citations – either because they are more novel or are have greater network reach.(*3*)

In terms of the structure of science, changing team size means that, as Walsh and colleagues have pointed out, science is taking on the characteristics of a small factory rather than a workbench(*25*). Simply put, principal investigators have the opportunity to structure their operations differently. Some could choose to staff a project with all senior investigators; others primarily with postdoctoral fellows; other with graduate students. We know little about how these different structures affect the production of knowledge; the most extensive work of which we are aware examines the size of author networks, rather than the entire small "factory". (*25*). However, because social capital plays a significant role in the development of human capital(*26*), it is likely that different structures do lead to different innovation potential.

Another way in which scientific structure has changed is through cross university collaborations. These have become more common both because of funding imperatives and technological advances(*3*). However, large scale collaborations have substantial drawbacks as well. Careful analysis of NSF grants (as well as widespread anecdotal evidence) suggests that the coordination costs of university collaborations can cancel out the advantages. (*5*)

The effect of diversity in both demographics and disciplines is likely to be mixed. The work of Lee, Walsh and Wang in studying research teams that published at least one Web of Science paper which suggests that scientific creativity is related to the composition and organization of research teams (*25*). However, there is evidence from the network literature that women and underrepresented minorities tend to have systematically different networks than white men in the general population(*27*). In addition, Whittington examines patent data in the life sciences and finds that women are less likely to collaborate in positions that yield innovative benefits.

It is important to build towards a better understanding of these process in the context of research funding. In the absence of such understanding, research funders are engaging in ad hoc decision making. For example, in one of the most cited attempts to examine the link between research funding and productivity, the then director of NIGMS, Jeremy Berg, plotted average publications, and their impact factor, of grant recipients and found that that researcher productivity began to diminished as grant size exceeded $600,000 to $750,000. (*28*) Lorsch,[2] Berg's successor as director of NIGMS, built on this research and a much earlier 1985 piece by Bruce Alberts[3], which outlined inefficiencies that can arise as labs become larger, to argue that inefficiencies are created when research funds are heavily concentrated among researchers rather than distributed more widely to the research community. The methodological challenges well identified by Bergstrom(*29*) have not apparently deterred NIH from formulating policies based on their analysis.

# 3. Data

## 3.1 Patent data

The first source of data is the August 2017 release of the PatentsView database4 which contains bibliographic information on all granted patent applications to the United States Patent and Trademark Office (USPTO) through August 2017. The PatentsView platform longitudinally links inventors, assignees, locations and patenting activity. It was established in 2012 and sources its data from USPTO bulk data on published patent applications (2001-present) and granted patents (1976-present). As of August 2017, it contains 6.3 million granted patents, of which there are 5.7 million granted utility patents. In this analysis, we only consider granted utility patents.

Patents filed with the USPTO can contain a field labeled "government interest" if the US government has any interests or rights in that patent. This interest typically arises for one of the two following reasons: (1) the inventor is an employee of a US government, (2) the research that produced the patent was supported by a government-sponsored grant or through a government contract. Our analysis dataset focuses exclusively on the 2nd type; we are interested in understanding how grants or contracts generate patents and how grant-funded patents are different, not how government employee inventors are different. We make this distinction by restricting our sample to patents with a grant or contract award number. It is, of course, possible that there is "over-claiming" of the link between grants and patents. This possibility is discussed in the context of life sciences research by Azoulay et al. (*30*), who make the convincing argument that scientists often use grant funding to subsidize their entire research agenda, and outputs should still be counted as related products.

[2] The argument is summarized in Lorsch http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4436771/. For an IBiology presentation by Lorsch see http://www.ibiology.org/ibiomagazine/jon-lorsch-lab-size-is-bigger-better.html.
[3] Alberts BM. Limits to growth: in biology, small science is good science. Cell. 1985;41337–338
[4] www.patentsview.org

The government interest field previously has been difficult to analyze because it contains complex textual information that is difficult to extract. For instance, the field might say “The invention described herein was made by an employee of the United States Government and may be manufactured and used by or for the Government for governmental purposes without the payment of any royalties thereon or therefor” or “The U.S. government has rights in this invention pursuant to Grant No. GI42298 awarded by the National Science Foundation”. The second example contains two important pieces of information: (1) the Federal agency with government interest in the patents (NSF), and (2) the Federal contract award number associated with the patent (Gl42298). The PatentsView database used the Stanford-maintained Named Entity Recognition library and information retrieval techniques5 to parse the raw government interest text into the two important pieces of information described above: the government organization name and contract or grant number. This information not only allows us to understand how government interest patents might differ from private patents, but it also allows us to link patent data to the UMETRICS data using Federal award numbers.

## 3.2 UMETRICS data

The second part of the analysis relies on new data infrastructure that can be used to examine differences between research teams that patent and those that do not. This new data captures information on all individuals funded on all research grants at a large sample of major universities and is now available at the Institute for Research on Innovation and Science (IRIS) (*14*). These data can also be combined with the patent data (through the grant identification number) to capture information on the characteristics of teams associated with grants that lead to patents, as well as those associated with grants that do not. Since the data include sub-award information, it is possible to capture whether there is cross-university collaboration[6]. Even more interestingly, the IRIS data enable the identification of which individuals on research-funded teams that patent are credited with being an inventor.

This work was based on the UMETRICS 2017Q4a[7] dataset for research. This dataset contains annual data from 26 IRIS member universities including coverage between 2001 and 2017 (this coverage varies by institution). The core files include university financial and personnel administrative data pertaining to sponsored project expenditures at IRIS member universities during a given year. IRIS core files are based on administrative data drawn directly from sponsored projects, procurement, and human resources data

[5] Further information about the process can be found in the complete Government Interest Processing Report. http://www.patentsview.org/data/GovtInterestReport-July2016.pdf
[6] The degree of collaboration, however, will be understated for those agencies that allow parallel collaborative submissions from different universities.
[7] The Institute for Research on Innovation & Science (IRIS). Documentation for UMETRICS 2017Q4a Dataset: The Second Annual Data Release. Ann Arbor, MI: IRIS [distributor], 2018-04-10, doi:10.21987/R7R94D

systems on each IRIS member university's campus. Individual campus files are de-identified, cleaned and aggregated by IRIS to produce these core files. The 2018 release includes transactions from about 300,000 unique federal and non-federal awards including wage payments to 480,000 individuals. In addition, about 13,000 unique organizations / institutions received sub-awards from IRIS member universities transferring their prime awards. Approximately 22,000 unique prime awards were used as the funding source to transfer sub-awards to sub-recipients. Vendor and sub-award payment[8] total $27.2 billion. [9]

We matched the PatentsView data with UMETRICS data by taking the award number identified in the government interest field, together with information on whether the assignees include a UMETRICS university. The analytical subset was defined to include only information on employees or expenses up through 2013, to permit sufficient time for a patent to be submitted. Analysis was limited to observations with information on both employees (employee table) & expenses (from any of vendor/sub-award/award tables). Any grant with a total expenditure less than $2000 was removed[10]. The final subset included a total of 89,527 *(*53,974 federal*)* unique grant (or award) numbers spread across 19 universities, 35 federal agencies and 255,548 employees. Out of this subset, we were able to match 1,188 distinct awards with Patentsview.

# 4. Descriptive analysis

## 4.1 Research-funded patenting activity

The new data show that research-funded patenting activity has clearly increased over time both in terms of the numbers of patents associated with federally funded grants and in relative size. The former

[8] Throughout the documentation, 'grant' and 'award' are interchangeably used as is often the case—e.g., NIH also uses these terms somewhat interchangeably, indicating "award conditions and information for NIH grants." To be precise, an award is defined as "(f)inancial assistance that provides support or stimulation to accomplish a public purpose. Awards include grants and other agreements in the form of money or property in lieu of money, by the federal government to an eligible recipient. The term does not include: technical assistance, which provides services instead of money; other assistance in the form of loans, loan guarantees, interest subsidies, or insurance; direct payments of any kind to individuals; and contracts which are required to be entered into and administered under federal procurement laws and regulations, " according to grant terminology at: https://www.grants.gov/web/grants/learn-grants/grant-terminology.html

[9] Boston University, Emory University, Indiana University, Michigan State University, New York University, Northwestern University, Ohio State University, Pennsylvania State University, Princeton University, Purdue University, Rutgers University, Stony Brook University, University of Arizona, University of California - San Diego, University of Cincinnati, University of Colorado, Boulder, University of Hawaii, University of Illinois at Urbana-Champaign, University of Iowa, University of Kansas, University of Michigan, University of Missouri, University of Pennsylvania, University of Pittsburgh, University of Virginia, University of Wisconsin - Madison

[10] Total expenditure here refers to the net aggregate of expenditure observed for the grant across the three sources: vendor, award and subaward tables over the total lifetime of grant observed within UMETRICS database (2018Q4A data release) until and including the year 2013.

have risen from almost zero in 1976 to almost 5,000 a year by 2012, and the latter from below 0.5% to 2.5% of all patenting activity (Figure 1a). Even if the benchmark year is moved to a full 10 years after the passage of Bayh-Dole, the increase is still consistently greater than patent growth overall – an almost five-fold increase in the level of grant-funded patenting, compared to 2.5 fold increase in all USPTO patents.

The increase has not come at the expense of traditional measures of quality. As Figure 1b shows, the trend is consistent when adjusted for a traditional measure of quality – forward citations- and hence does not seem to be a result of the spurious patenting in response to funding incentives noted in other countries (*31*).

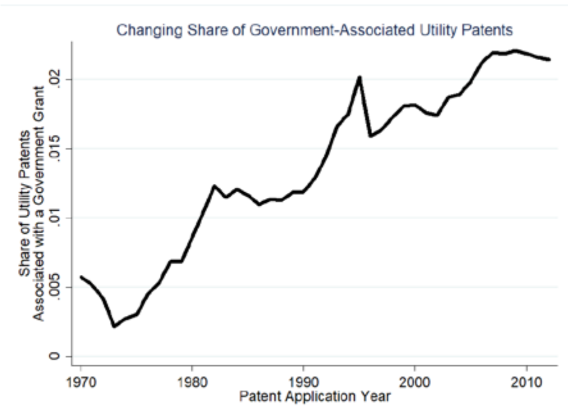



Figure 1a

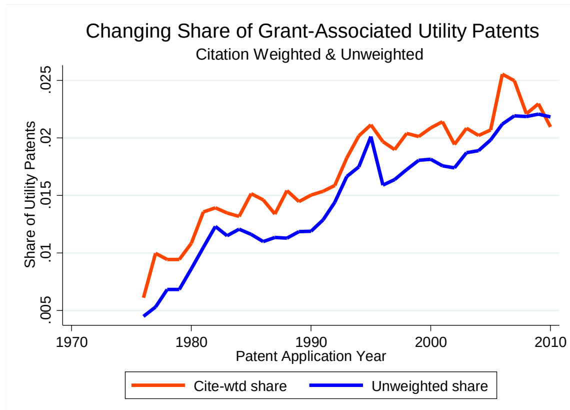



Figure 1b.

In sum, government-funded patents are increasing over time, and that increase has not come at the expense of patent quality as measured by the standard metric of patent citations.

## 4.2 Grant funded teams and patenting activity

Our analysis of the 89,000 grants at UMETRICS universities shows that just over 1% of awards lead to a patent that cites the grant as contributing to the invention. There are substantial differences in patenting rates across by different funders – ranging from over 5% from the Defense Department to less than .4% for other federal agencies.

| Table 1: Distribution of Awards by Agency & Patenting Activity | | | | |
|---|---|---|---|---|
| | Led to Patent | Did Not Lead to Patent | Total | Patent% |
| Dept of Defense | 205 | 3,696 | 3,901 | 5.30% |
| Dept of Energy | 70 | 1,862 | 1,932 | 3.60% |
| NSF | 348 | 11,963 | 12,311 | 2.80% |
| NIH | 377 | 22,985 | 23,362 | 1.60% |
| Non Federal | 163 | 35,852 | 36,015 | 0.50% |
| Other federal | 51 | 12,445 | 12,496 | 0.40% |
| **Total number of grants** | **1,188** | **88,339** | **89,527** | **1.34%** |

Numbers in table are count of unique awards. The rows/columns do not sum up into totals since a particular patent can be associated with multiple agencies. **Source:** UMETRICS 2017Q4A* (2018 release); PatentsView (August 2017) **Data:** Analytical subset of C3:G12 federal and non-federal grants who have a minimum net expenditure of $2,000 over the life of the grant as observed within UMetrics database; including information up to 2013.

Table 2 provides an overview of the characteristics of teams by whether or not the grant resulted in a patent. The differences are consistent with the literature. Teams on grants that result in patents are about twice as likely to work with other universities, have budgets that are almost three times larger and team sizes that are almost twice as large[11]. The structure of the teams is quite different as well. The number of faculty is greater on patenting teams, and probably by virtue of the scale, more likely to have a female represented in the senior team. Interestingly, however, the proportion of post-doctoral fellows and graduate students is substantially higher in a patenting team than in a team that doesn't patent.

[11] The difference in relative between the budget size and the team size is due to the fact that patenting teams spend more on other inputs, such as equipment and travel. UMETRICS data includes information (at the detailed object code level) about those expenditures; a parallel research thread is examining the production function of this research.

| Table 2: Grant funded Teams by Patenting Activity | | |
|---|---|---|
| | No Patent | Patent |
| Proportion with inter-university collaboration* | 14.6% | 27.6% |
| Average yearly expenditure | $172,222 | $496,257 |
| Average annual team size | 5.1 | 8.8 |
| Average proportion of post-docs and grad students on team | 30% | 40% |
| Average count of faculty | 1.2 | 2.4 |
| Average share of grants with at-least one senior female employee** | 41.9% | 50.8% |
| Count of grants | 52,925 | 1,049 |

Notes: (1) A grant is said to have an inter-university collaboration if it has a sub-award with another university; (2) A female who is a faculty/clinician/graduate/postgraduate is referred to as a senior female employee; (3) Numbers in table are count of unique awards. The rows/columns do not sum up into totals since a particular award can be associated with multiple agencies, and can be patented with one and not with another.
**Source:** UMETRICS 2017Q4A* (2018 release); PatentsView (August 2017); **Data:** Analytical subset of 53,974 federal grants that have a minimum net expenditure of $2,000 over the life of the grant as observed within UMETRICS database; including information until and for the year 2013.

As noted in the previous section, the data also permit an examination of who gets named as inventors. Of the almost 16,000 individuals funded on patenting teams, about 12% are named as inventors. Almost one in three faculty working on a grant leading to a patent are named as inventors, while just over one in ten postdocs or graduate students. Those in other categories have a one in twenty chance of being named. While these results are consistent with prior expectations, they do indicate that prior work that describes teams based on inventor characteristics alone is likely to understate the size and structure of the team in systematic ways.

| Table 3: Who gets named as inventors on patenting teams | | | |
|---|---|---|---|
| | **Inventors** | **Total Employees** | **Inventor Rate** |
| Faculty | 848 | 2,958 | 28.7% |
| Grad or Postdoc | 776 | 6,017 | 12.9% |
| Other | 387 | 6,948 | 5.6% |
| Total | 1,976 | 15,786 | 12.5% |

Source: UMETRICS 2017Q4A* (2018 release); PatentsView (August 2017). Data: Analytical subset of 15,786 employees working on federal grants (up to 2013).

# 5. Analytical Results

We estimate two separate models. The first estimates the likelihood that a grant funded research team does work that results in a patent. The second estimates the likelihood that an individual on the research team is named as an inventor.

## 5.1 Teams and patenting

In the former case, the data (described in more detail below) permit us to characterize all research funded grants by their level of funding, the size of the team, the organizational structure (namely, the number of faculty, the proportion of postdocs, and whether there is a female on the team), whether or not the grant includes a subaward to another university and controls for the university and funding source.

$$P(patent)_g = f(funding\ _g, Size_g, Structure_g, Collaboration_g, Controls_g)\ (1)$$

The unit of observation is the grant (g), and the model is run on the full sample of grants. The funding is measured directly by the log of average annual spending on the grant. Team size is the log of the annual average of funded individuals working on the grant. Team structure is measured with three variables. The proportion of graduate students and postdocs is the annual average of graduate students and postgraduates on the team divided by team size. The faculty count is the average number of faculty. There is also a binary indicator if at least one of the research team working on the grant is identified as a female.[12] Collaboration is a binary indicator for whether or not the grant had more than two universities collaborating. This is defined here if the grant had a subaward. We include agency fixed effects and university fixed effects. Year fixed effects are included for the first year in which the award showed spending activity[13].

Table 4 shows the results of estimating this model. Columns one and two only include funding agency fixed effects, while columns three and four also include university and year fixed effects. Columns two and four simply examine team size, while columns one and three break team size down into two components: number of faculty, and proportion of postdocs and graduate students.

The first immediate takeaway is that even after controlling for the agency and the funding amount, organizational and team characteristics are strongly associated with the propensity to patent. Second, there is a great deal of heterogeneity in the patenting process, as is evident from the relatively low explanatory power. Third, while university collaboration and larger team sizes are positively associated with the propensity to patent, teams that have a female are negatively associated with the propensity to patent, although this result is dependent on specification. In particular, university collaboration is associated with a one percent higher propensity to patent, and a ten percent increase in team size is associated with a seven percent higher propensity to patent. If team size is broken down into components, we see that growth in team size in terms of faculty members or in terms of additional

[12] A binary indicator is also included if it cannot be determined whether a research team member working on the grant is a female from the gender inference algorithm.
[13] This variable was revised for various conditions such as (a) life within UMETRICS, (b) the applic ation year of the matched patent, if any.

postdocs or graduate students are both associated with a higher propensity to patent. In short, while the amount of grant funding matters, the structures of the organizations and teams involved in the research that the grant funds are also highly indicative of whether patenting is the outcome.

**Table 4: Team Characteristics and Patenting**
**Dependent Variable: Grant cited by a patent**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Constant | -0.067*** | -0.053*** | 0.879*** | 0.890*** |
| | (0.010) | (0.012) | (0.010) | (0.010) |
| Collaboration | 0.012*** | 0.012*** | 0.012*** | 0.012*** |
| | (0.003) | (0.003) | (0.002) | (0.002) |
| Ln(expenditure) | 0.006*** | 0.004** | 0.006*** | 0.005*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Proportion of grad students and postdocs | 0.028*** | | 0.029*** | |
| | (0.004) | | (0.004) | |
| IHS(count of faculty) | 0.011*** | | 0.011*** | |
| | (0.003) | | (0.002) | |
| Ln(team size) | | 0.007*** | | 0.007*** |
| | | (0.002) | | (0.001) |
| Is there a female faculty member on the team? | -0.006** | 0.002 | -0.008*** | -0.000 |
| | (0.002) | (0.001) | (0.002) | (0.001) |
| DEPT OF DEFENSE | 0.037*** | 0.044*** | 0.033*** | 0.040*** |
| | (0.008) | (0.008) | (0.008) | (0.008) |
| DEPT OF ENERGY | 0.022*** | 0.028*** | 0.018*** | 0.024*** |
| | (0.004) | (0.005) | (0.004) | (0.004) |
| NIH | 0.003 | 0.006* | -0.002 | 0.000 |
| | (0.002) | (0.003) | (0.003) | (0.003) |
| NSF | 0.017*** | 0.023*** | 0.013*** | 0.018*** |
| | (0.003) | (0.003) | (0.002) | (0.003) |
| University Fixed Effects? | No | no | yes | yes |
| Year Fixed Effects? | No | no | yes | yes |
| Observations | 54291 | 54291 | 54291 | 54291 |
| $R^2$ | 0.020 | 0.017 | 0.043 | 0.041 |
| Adjusted $R^2$ | 0.020 | 0.017 | 0.043 | 0.040 |
| Log likelihood | 30595.433 | 30520.471 | 31250.938 | 31175.460 |

Standard errors clustered at the university level in parentheses ; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; controls included for imputation

## 5.2 The likelihood of being named inventor

The award data permit one more interesting insight into the patenting/inventor process. Because the award data capture information on all individuals who are funded on all research awards in general and all awards that lead to patents in particular, we can investigate the characteristics of the team members who are named as inventors on each patent. We used name matching techniques to do this for the 12,665 unique individuals who worked on the 800 federally funded research teams that were granted a patent. 1951 of these were named inventors on their team's patent.

We estimate a simple linear probability model regression, restricted to individuals who worked on teams that were granted a patent, to examine the factors that are correlated with being named as an inventor. We use the same factors as in model 1, and we include information on the individual's position in the grant, as well as the individual's gender. This regression is at the individual level (i).

$$P(named\ inventor)_{ig(p)} = f\left(funding\ _{g(p)}, Size_{g(p)}, Structure_{g(p)}, Collaboration_{g(p)}, Individual_{ig(p)}\right)(2)$$

As with the first model, we include agency fixed effects and university fixed effects. Year fixed effects are included for the first year in which the award showed spending activity[14]. While the first regression was run on the full sample of grants, this sample is restricted to the subsample of grants that led to a patent.

The results, shown in Table 5, reveal that faculty were (unsurprisingly) much more likely to be named as an inventor (a 42% chance) than was a post doc or a graduate student. What was surprising was that female faculty were half as likely as male faculty to be named an inventor. Similarly, while postdoc/graduate students in general had about a 5% chance of being named an inventor, the effect of being a female postdoc or graduate student completely offset that chance – on net, a female postdoc/graduate student had no chance of being named an inventor in this data.

There is a legal basis underpinning differences in being named an inventor. Patent law is clear that not every person who works on a research project constitutes an inventor[15]. An individual must contribute to the "conception" of an invention, as opposed to reduction to practice[16]. Therefore, we would

[14] This variable was revised for various conditions such as (a) life within UMETRICS, (b) the application year of the matched patent, if any.

[15] We are grateful to Amanda Myers, from the USPTO, for pointing us to the relevant text.

[16] Manual of Patent Examining Procedure (MPEP) 2137.01.I. "The definition for inventorship can be simply stated: "The threshold question in determining inventorship is who conceived the invention. Unless a person contributes to the conception of the invention, he is not an inventor. … Insofar as defining an inventor is concerned, reduction to practice, per se, is irrelevant"

not expect every team member to appear as an inventor on the patent. However, this requirement leaves room for interpretation and could be implemented differently across applicant institutions or Principal Investigators. However, patent law also notes that "Difficulties arise in separating members of a team effort, where each member of the team has contributed something, into those members that actually contributed to the conception of the invention, such as the physical structure or operative steps, from those members that merely acted under the direction and supervision of the conceivers" and "there is no requirement that the inventor be the one to reduce the invention to practice so long as the reduction to practice was done on his behalf"[17], although "Each joint inventor must generally contribute to the conception of the invention."[18].

The work of sociologists is of particular interest in this context(*32–35*). Their research has found that the observed gender productivity puzzle in patenting and publications might be due to organizational structure rather than innate productivity differences by gender. Whittington and Smith-Doerr find that different network structures and hierarchies are important reasons for gender disparities(*34*).

[17] MPEP 2137.01.IV.
[18] MPEP 2137.01.V.

| Table 5: Likelihood of an individual being named an inventor | | |
|---|---|---|
| **Dependent Variable: Binary Indicator for** Named Inventor | | |
| | (1) | (2) |
| Constant | 0.777*** | 0.752*** |
| | (0.060) | (0.074) |
| Collaboration | -0.031* | 0.027 |
| | (0.013) | (0.016) |
| Is the inventor a faculty member | 0.438*** | 0.448*** |
| | (0.021) | (0.021) |
| Is the inventor a female faculty member | -0.262*** | -0.254*** |
| | (0.024) | (0.024) |
| Is the inventor a grad student/postdoc | 0.072*** | 0.075*** |
| | (0.009) | (0.010) |
| Is the inventor a female grad student/postdoc | -0.078*** | -0.074*** |
| | (0.010) | (0.010) |
| Ln(expenditure) | -0.047*** | -0.047*** |
| | (0.004) | (0.004) |
| DEPT OF DEFENSE | 0.021 | 0.033 |
| | (0.023) | (0.025) |
| DEPT OF ENERGY | 0.000 | -0.002 |
| | (0.021) | (0.022) |
| NIH | -0.066*** | -0.023 |
| | (0.020) | (0.023) |
| NSF | -0.017 | 0.012 |
| | (0.021) | (0.023) |
| Observations | 14726 | 14726 |
| $R^2$ | 0.202 | 0.217 |
| Adjusted $R^2$ | 0.202 | 0.215 |
| *AIC* | 10558.640 | 10353.547 |
| Log lik. | -5268.320 | -5132.773 |
| Standard errors clustered at the grant level in parentheses; $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$ | | |

## Conclusion

This work provides new evidence on the links between research funding and one measure of innovation (patenting). We exploit two sources of new data – information about government funding on patents linked to information about government funded research teams – to examine the ways in which research funding and the structure of teams affects their propensity to patent.

The results are striking. There are marked differences in the propensity to patent and the likelihood of researchers being named as inventors along the dimensions of team size, funding agency, and gender representation. There is evidence that larger teams and more funding generates more patents, which militates against the somewhat specious findings by NIGMS that resulted in recommendations of capping funding levels(*29*). The field specific differences are completely in line with prior results. Most interestingly, however, the gender results are very consistent with the work of sociologists who have shown that women in the academy occupy much less central positions in scientific networks, and that this affects their productivity(*34*).

As noted, however, these results are descriptive, not causal in nature. One possible way to estimate causal effects would be to implement randomized controlled trials (*36*), given that there are limited research funds to be distributed across agencies and fields, and similar approaches have been implemented in other policy areas (such as health care, training programs and education). Another approach would be to expand the network analyses to – for the first time – explain the role of team network positions in patenting behavior. It should also be possible to analyze the impact of individuals switching teams and institutions at all occupation levels.(*15*)  As Owen-Smith has pointed out, the UMETRICS offers a unique potential to examine the full dynamics of networks and characterize network positions in ways that not hitherto been possible(*16*). Other approaches, which are possible with links to Census Bureau data, are to investigate changes in team diversity and changes in agency funding structures on innovation.

In sum, this paper represents what we hope is the first step in filling the data infrastructure gap identified in the introduction – and begins to unpack the human mechanism by which research funding(money) results in innovation (something).

## References

1. D. C. Mowery, Universities in National Innovation Systems (2004).

2. B. F. Jones, The burden of knowledge and the "death of the Renaissance man": is innovation getting harder? *Rev. Econ. Stud.* **76**, 283–317 (2009).

3. S. Fortunato *et al.*, Science of science. *Science*. **359**, eaao0185 (2018).

4. V. Larivière, Y. Gingras, C. R. Sugimoto, A. Tsou, Team size matters: Collaboration and scientific impact since 1900. *J. Assoc. Inf. Sci. Technol.* **66**, 1323–1332 (2015).

5. J. N. Cummings, S. Kiesler, Coordination costs and project outcomes in multi-university collaborations. *Res. Policy*. **36**, 1620–1634 (2007).

6. D. K. Ginther *et al.*, Race, ethnicity, and NIH research awards. *Science (80-. )*. **333**, 1015–1019 (2011).

7. V. Larivière, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: Global gender disparities in science. *Nat. News*. **504**, 211 (2013).

8. M. P. Feldman, D. B. Audretsch, Innovation in cities:: Science-based diversity, specialization and localized competition. *Eur. Econ. Rev.* **43**, 409–429 (1999).

9. K. B. Whittington, "A tie is a tie? Gender and network positioning in life science inventor collaboration." *Res. Policy*. **47**, 511–526 (2018).

10. S. Harris, And then a miracle, (available at https://s-media-cache-ak0.pinimg.com/736x/d6/e7/54/d6e754d24aaef324c1595e68583ace7a.jpg).

11. National Science Board, "Science and Engineering Indicators 2018" (Washington DC, 2018).

12. M. J. Kalutkiewicz, R. L. Ehman, Patents as proxies: NIH hubs of innovation. *Nat. Biotechnol.* **32**, 536–537 (2014).

13. National Institutes of Health, Impact of Our Research (2018), (available at https://www.nih.gov/about-nih/what-we-do/impact-nih-research/our-society).

14. 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.

15. K. B. Whittington, in *Science and Engineering Careers in the United States: An Analysis of*

*Markets and Employment*, R. Freeman, D. Goroff, Eds. (University of Chicago Press, Chicago IL, 2009).

16. J. Owen-Smith, in *Big Data and Social Science,* I. Foster, R. Ghani, R. Jarmin, F. Kreuter, J. Lane, Eds. (Taylor & Francis, 2016).

17. N. J. Hilton, M. L. Cooke, "Enhancing the Effectiveness of Team Science" (Washington DC, 2015).

18. S. Milojević, Principles of scientific research team formation and evolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3984–9 (2014).

19. R. Burt, *Structural Holes: The Social Structure of Competition* (Cambridge University Press, Cambridge, 1992).

20. P. Marsden, N. Friedkin, Network Studies of Social Influence. *Sociol. Methods Res.* (1993).

21. J. Owen-Smith, N. C. Cotton-Nessler, H. Buhr, Network effects on organizational decision-making: Blended social mechanisms and IPO withdrawal. *Soc. Networks*. **41**, 1–17 (2015).

22. R. S. Burt, Structural Holes and Good Ideas. *Am. J. Sociol.* **110**, 349–399 (2004).

23. J. Owen-Smith, W. W. Powell, Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. *Organ. Sci.* **15**, 5–21 (2004).

24. D. Obstfeld, Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation. *Adm. Sci. Q.* **50**, 100–130 (2005).

25. Y.-N. Lee, J. P. Walsh, J. Wang, Creativity in scientific teams: Unpacking novelty and impact. *Res. Policy*. **44** (2015), pp. 684–697.

26. J. S. Coleman, Social Capital in the Creation of Human Capital. *Am. J. Sociol.* **94**, S95 (1988).

27. Birds of a Feather : Homophily in Social Networks Author ( s ): Miller McPherson , Lynn Smith-Lovin , James M . Cook Source : Annual Review of Sociology , Vol . 27 ( 2001 ), pp . 415-444 Published by : Annual Reviews Stable URL : http://www.jstor.org/stab. **27**, 415–444 (2009).

28. J. Berg, Measuring the Scientific Output and Impact of NIGMS Grants (2010), (available at https://loop.nigms.nih.gov/2010/09/measuring-the-scientific-output-and-impact-of-nigms-grants/).

29. C. Bergstrom, J. D. West, "NIH's Rule of 21" (Seattle, Washington, 2018), (available at http://callingbullshit.org/case_studies/case_study_rule_of_21_part_2.html).

30. D. Li, P. Azoulay, B. N. Sampat, The applied value of public investments in biomedical research. *Science (80-. ).* **81**, 78–81 (2017).

31. J. Vanecek, The effect of performance-based research funding on output of R&D results in the Czech Republic. *Scientometrics*. **98**, 657–681 (2014).

32. Y. Xie, K. A. Shauman, Sex differences in research productivity: New evidence about an old puzzle. *Am. Sociol. Rev.*, 847–870 (1998).

33. Y. Xie, K. A. Shauman, *Women in science* (Harvard university press, 2003).

34. K. B. Whittington, L. Smith-Doerr, Women inventors in context: Disparities in patenting across academia and industry. *Gend. Soc.* **22**, 194–218 (2008).

35. D. Tomaskovic-Devey, Women's Work: Gender Equality vs. Hierarchy in the Life Sciences. *Adm. Sci. Q.* **50**, 661–662 (2005).

36. P. Azoulay, Research efficiency: Turn the scientific method on ourselves. *Nature*. **484**, 31–32 (2012).

37. N. Wasi, A. Flaaen, Record linkage using Stata: Preprocessing, linking, and reviewing utilities. *Stata J.* **15**, 672–697 (2015).

38. W. E. Winkler, in *Bureau of the Census* (Citeseer, 2006).

39. N. Monath, A. Mccallum, Discriminative Hierarchical Coreference for Inventor Disambiguation.

40. A. Ikudo, J. Lane, B. Weinberg, "Occupational Classifications: A Machine Learning Approach" (New York University, 2018).

## Appendix 1: Patent Data

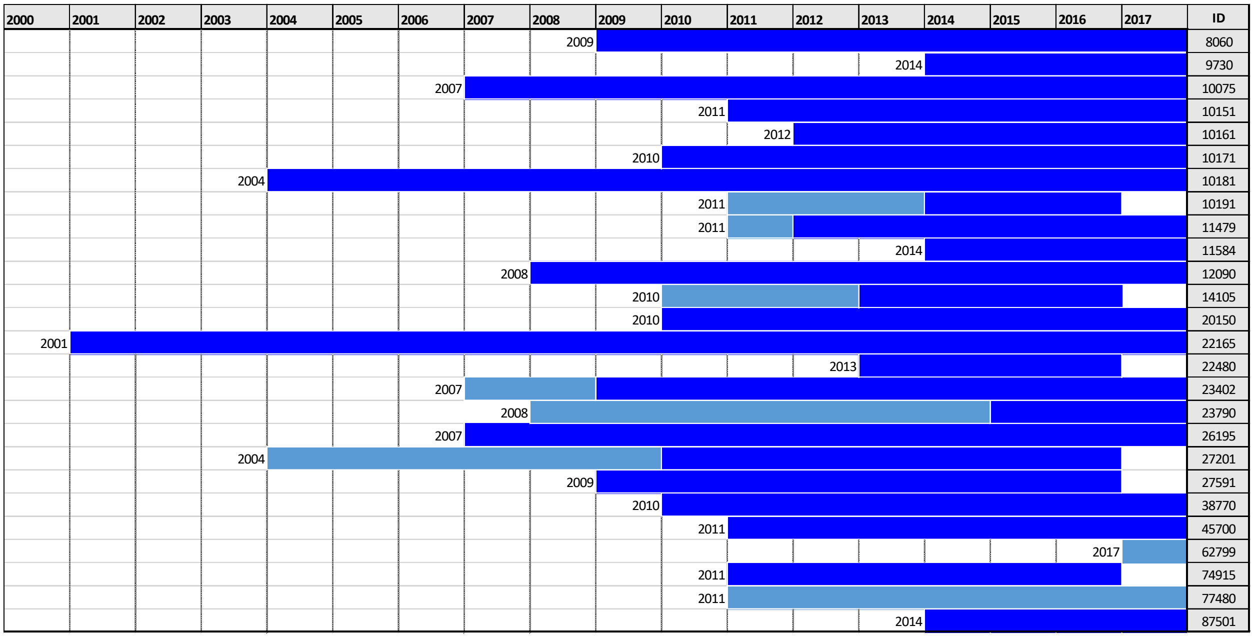
In our analysis, we also use additional information from the PatentsView data: the patent assignee, forward and backward citations, the inventors on the patent, the application and grant years, and the technology class of the patent. Company names can vary slightly in the raw patent data, so the PatentsView data uses a disambiguation process to group variations of the same company name together. This process first uses the University of Michigan's STATA Utilities to correct typos and misspellings (*37*) and then uses the Jaro-Winkler string similarity algorithm to identify similar assignee names (*38*). An inherent problem in analyzing inventor information in patent data is that there is no unique identifier for each inventor in the patent data. However, the PatentsView data uses an inventor disambiguation algorithm developed from a September 2015 Inventor Disambiguation Workshop by Andrew McCallum and Nicholas Monath to do so (*39*). This algorithm uses discriminative hierarchical conference[19] to disambiguate inventors, allowing us to trace each inventor's patenting activity over their inventive career. There are four distinct schemes for classifying patents according to their technology area, all of which are available in the PatentsView data: cooperative patent classification (CPC), World Intellectual Property (WIPO) technology fields, US patent classification (USPC), and the National Bureau of Economic Research (NBER) technology area subcategories. In our analysis we primarily use the USPC classification. Finally, we also utilize information on application and grant year of the patent. The application year is the year of application for issue of the patent. If the patent is approved, it is eventually granted or issued, and the grant year reflects that date. There is typically a lag of about 3 years between application and grant year, but there is a lot of heterogeneity in this lag. The application year is closer to the actual year of invention, so it is typically used in measuring the timing of inventive activity, but the shortcoming of its use is the problem of truncation. Because it can take several years for a patent to be granted, this means that there will be an artificial decline in the number of granted patents corresponding to a given application year in the most recent years of the sample.

## Appendix 2: UMETRICS

Figure 1 provides information about the coverage of each institution.

[19] http://www.patentsview.org/data/presentations/UMassInventorDisambiguation.pdf

Figure 1: IRIS Member universities- Temporal Coverage by Institution




Note: Institutions/Universities in the figure are masked by ID. The color coding demonstrates coverage by file. Deep blue indicates file availability for all four files (award, employee, vendor & subaward); light blue indicates that not all of the 4 files are available for a given year.

IRIS adds value to the data received by member institutions by identifying and resolving data discrepancies when possible, providing standardized occupational classification codes, as well as through various cleaning processes including name standardization. IRIS carefully masks information from the release files in order to minimize the risk of re-identifying member universities or individuals from particular data elements.

We matched the PatentsView data with UMETRICS data by taking the award number identified in the government interest field, together with information on whether the assignees include a UMETRICS university. The analytical subset was defined to include only information on employees or expenses up through 2013, to permit sufficient time for a patent to be submitted. Analysis was limited to observations with information on both employees (employee table) & expenses (from any of vendor/sub-award/award tables). Any grant with a total expenditure less than $2000 was removed[20]. The final subset included a total of 90,512 observations out of which 54,291 belong to federal agencies. This consists of 89,527

[20] Total expenditure here refers to the net aggregate of expenditure observed for the grant across the three sources: vendor, award and subaward tables over the total lifetime of grant observed within UMETRICS database (2018Q4A data release) until and including the year 2013.

*(*53,974 federal*)* unique grant (or award) numbers spread across 19 universities, 35 federal agencies and 255,548 employees. Out of this subset, we were able to match 1,188 distinct awards with Patentsview.

It is worth noting that a particular employee can be associated with multiple awards and multiple occupations. The difference can occur because of many different reasons, including career progression, classification imprecision and data entry errors (*40*) . In these cases, if an employee-award combination has more than one occupation listed, precedence was given to the highest rank: faculty>grad/postgrad>other.