Federal Funding of Doctoral Recipients: Results from new linked survey and transaction data[1]

Wan-Ying Chang, National Science Foundation
Wei Cheng, Ohio State University
Julia Lane, New York University
Bruce Weinberg, Ohio State University

[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; and the Ewing Marion Kaufman and Alfred P. Sloan Foundations. Data were generously provided by the Committee on Institutional Cooperation and its member institutions. We thank Cameron Conrad, Ahmad Emad, Evgeny Klochikhin, Josh Tokle and Rosa Castro Zarzur for research support, Greg Carr, Marietta Harrison, David Mayo, Mark Sweet, Jeff Van Horn and Stephanie Willis for help with data issues, and Jay Walsh, Roy Weiss, and Carol Whitacre for their continuing support. The research agenda draws on work with many coauthors.

1. Introduction

"Instability in Federal funding .. affects the career trajectories of doctoral students, postdoctoral workers, and researchers whose employment is often supported by federally funded research grants. Recognizing that strong, steady increases in Federal R&D funding may not always be feasible, it is important for Federal agencies, Congress, and research institutions to consider how best to mitigate the adverse effects of budget conditions on career paths" -- National Science Board, p26 (1)

Funding of research is critically important because it affects the flow of new, doctorally qualified scientists into the workforce through several channels. It sets priorities for fields of research graduate students work in. It supports the people working in those fields. It may expand the pipeline of labor into the workforce by providing financial support for students who might otherwise not be able to afford an education (2–4). While many reports have called for new and better sources of data to better understand the processes whereby funding affects these channels (5, 6), the complexity of the funding structure in many countries has made it difficult to collect such data.

This paper provides new insights into how survey data can be combined with administrative records to examine the ways in which funding affects workforce decisions. Although the context is that of US federal funding, the approach can be applied to other countries, which also have multiple sources of funds for students.

It provides an approach to filling the substantial data gaps (7), which have led to calls for new data collection. It particularly points to the value of administrative records, which are often a cheaper and more extensive way of collecting information than survey data alone (8). One goal of this paper is to provide an early demonstration of the power of combining administrative and survey data for science policy. Importantly, the linked survey and administrative data approach is not limited to the U.S. Indeed, there are parallel efforts at various stages in Europe and Australia. We see the potential of such an approach as being at least as great in Europe, given the importance of research funding from multiple levels (e.g. the European Commission and individual member states).

In particular, we demonstrate how multiple funding channels can be studied by bringing together two disparate datasets: UMETRICS, which is generated from university payroll and financial records, and the Survey of Earned Doctorates (SED), which is one of the most important US survey datasets about the doctoral workforce, and which has a counterpart in most European countries.

The results presented here cannot be derived from the separate datasets. The SED provides information about research fields, the characteristics of the men and women in graduate school, and the ways in which they are supported; the link to the UMETRICS data provides the connection to the rich complexity of research funding. The SED in turn provides information about the types of jobs that individuals get immediately after graduation.

We find that research funding is indeed rich and complex. No individual agency, looking at its own data, can document how many doctoral recipients are supported by federal funds. Even the dominant supporter of doctoral research, the National Science Foundation (NSF), provides funds to just under one in five doctoral recipients. Even if agencies were able to provide separate (and accurate) counts of how many doctoral recipients were supported by their funding, the combined results would substantially overestimate how many total individuals were supported and understate the amount of support they receive, because although almost 40% of doctoral recipients are supported by only one agency, almost 7% are supported by 2 and some are supported by 3 or more. In other words, it is virtually impossible to rely on agency databases to characterize the pipeline of doctoral recipients.

We also find large differences in the number and mix of disciplines supported by different agencies. NSF supports more graduate students per dollar spent than other federal agencies. Not surprisingly, NIH heavily supports biology, health, and psychology PhDs, while NSF heavily supports PhDs in engineering, the physical sciences, mathematics, and computer science. Other agencies, which include USDA, support PhDs in agriculture.

Federal funding also affects *who* does research. Of the women in our sample, 39% are federally funded, compared to 52% of men. There are also substantial differences across agencies – a larger share of doctoral recipients supported by NIH are women (50%), African American (2.6%) and Hispanic (4.2%), compared to NSF, the Department of Defense (DOD) or the Department of Energy (DOE).

Finally, federal funding is highly correlated with the pipeline of researchers going into different fields, particularly R&D fields, and the decision to pursue postdoctoral fellowships.

As with any research, caveats apply. The analysis presented here draws on UMETRICS data from seven leading research universities. The universities may not be representative of the research university population at large (indeed, it is difficult to imagine any sample that would be fully representative), but they mix public and private schools; most, but not all, have medical schools and engineering schools. All are members of the American Association of Universities, and hence are among the most research intensive universities. The analysis here focuses only on federal funding for graduate students in the two years prior to receiving their doctoral degree; the same analysis could be done on non-federal sources of funds, as UMETRICS data is expanding to include that information for partner institutions. Lastly, the challenges of establishing a causal link between federal funding and career outcomes are well documented (9). The current paper is purely descriptive and establishes a set of basic facts. We expect more causal subsequent research, as the new Institute for Research on Innovation and Science (IRIS) adds more universities and longer time spans to its existing data infrastructure. Even beyond its important description of the support for doctoral research training, this research powerfully demonstrates that it is possible to combine two sources of data, with no additional burden on survey respondents, to provide much richer insights into the patterns of federal funding and its contribution to supporting the development of an important part of the research workforce.

2. Data description

The paper combines data from two sources: the Survey of Earned Doctorates (SED) from 2011 and 2012 and new transaction data from the UMETRICS program.

Survey of Earned Doctorates

The National Center for Science and Engineering Statistics (NCSES) produces a number of restricted data sets that are essential for research and reporting on U.S. STEM graduates and the workforce. The Doctoral Records File (DRF) is a particularly important data source in this regard. This file contains data on all earned doctorates granted by regionally accredited United States universities, in all fields, from 1920 to the present. The SED, which began in 1957, forms the backbone of the DRF. The SED is a survey administered to all the doctorate recipients in the United States. The overall survey response rate exceeds 90%, and since basic information not provided by the respondent is filled in by the department, the survey is essentially a census of doctoral recipients. Between 2011 and 2012, 8931 doctoral recipients were reported by the seven universities in the SED.

The SED collects information on the doctoral recipient's educational history, demographic characteristics, and post-graduation plans. The SED contains, inter alia, the following information: the doctorate's name, birth year, country of birth, race, sex, academic institution of the doctorate, sources of financial support during graduate school and sources and type of financial support for postdoctoral study or research.

However, there is no question in the SED that provides comprehensive information about federal research funding. Two questions provide indirect information: One question (A5) asks, "Please indicate whether each of the following was a source of financial support during graduate school: fellowship/scholarship; grant; teaching assistantship; research assistantship; other assistantship; traineeship; internship; loans; personal savings; personal earnings; spouse's, partner's or family's earnings; employer reimbursement; foreign support." Another question (A6) asks, "Which two sources in question A5 provided the most support". The answers to these questions are suggestive of the source of support but not definitive (setting aside reporting issues). For instance, people supported on research assistantships are more likely to be supported by federal funds, but might be supported by state, local, corporate, foundation, or institutional funds.

UMETRICS data

The UMETRICS data captures information from payroll records of who is paid from federal research funds. It builds upon the STAR METRICS project initiated in 2009 by the US Federal agencies, with the engagement of the Federal Demonstration Partnership (FDP).(10) The goal of the original project was to build an open source, two-layered, continually evolving data platform that could be used to (1) provide policymakers with a better understanding of the process of research and (2) provide the research community with a common data infrastructure that connected research funding with research outcomes (11).

The UMETRICS project, which extends the federal STAR METRICS effort, currently contains record level information on wage payments made from federal grants to doctoral students and other university personnel. Currently, 49 universities, accounting for more than 40% of federal R&D expenditures, are committed to participating, but the project is expected to become a national program with the recent establishment of the Institute for Research on Innovation and Science (IRIS) at the University of Michigan. The UMETRICS initiative was incubated by the Big 10 Academic Alliance (B10AA, formerly the Committee on Institutional Cooperation), a consortium of large, mostly public, research universities mainly in the Midwest. Most of our data come from B10AA institutions. To maximize the coverage and the number of schools represented, our primary focus is on students completing doctorates in 2011 or 2012, looking back to the last two years in their program. It is worth noting that not all individuals working on research grants are included in the UMETRICS data. The employee transactional data includes only people/employees who were paid on research grants. Individuals paid as work study trainees and those on honoraria are not covered. Coverage of non-federal grants ranges widely from university to university. Coverage of federal grants is generally complete, with the exception of one university that did not supply classified DOD research.

Although four files are provided by the university, the key file of interest in this project is the employee file. Briefly, for each federally funded project, the file contains all payroll charges for all pay periods (period start date to period end date) with links to both the federal award id (unique award 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, their FTE status and the proportion of earnings allocated to the award. (11, 12)

The base sample was selected from the Survey of Earned Doctorates. During the period 2011 and 2012, 8,931 individuals graduated with doctoral degrees from the seven universities. Of these 8,931 doctoral recipients, 4,930 or 55% were matched to the UMETRICS funding data of these seven universities[2]. 4116 were federally funded in the two years before graduating.

[2] The match was based on the academic institution, name, year and month of birth, and conducted under strict confidentiality restrictions. Protecting respondent confidentiality, while minimizing respondent burden, is imperative if administrative records are to be used to enrich current NCSES products and expand possibilities for research and reporting for NCSES. In the case of the Survey of Earned Doctorates (SED), the data are restricted use. The license under which the data were accessed was agreed to by the American Institutes for Research for contract work allowed by the National Science Foundation in order to improve understanding of the potential value of administrative records to improve the SED. All data were secured in a dedicated project server that employs limited access controls by both using strong passwords for user access and by blocking incoming/outgoing traffic from any other network. Only limited access was allowed, by AIR researchers under contract to NCSES. Users are permitted access only from a desktop physically located onsite at AIR. Additional security measures are in place per NSF RUD requirements.

3. Federal funding

Table 1 summarizes the sources of federal funding as reported by the respondents to the Survey of Earned Doctorates. Almost 90% of the SED respondents from these universities cite one or more of four sources as their support (response to question A5): research assistantship (68%), fellowship/scholarship (64%), teaching assistantship (53%), and grant (28%). The links to UMETRICS now permit administrative information to be added about the source of funding. It is clear that federal funding is an extremely important source of support for doctoral recipients in these research intensive universities – almost half (46%) of doctoral recipients were federally funded in the two years prior to receiving their doctoral degree. Almost 56% of those who reported being supported on a research assistantship received at least some federal support. About 44% of those supported on a fellowship or a scholarship, 46% of those on a teaching assistantship, and 49% of those on a grant, also received some federal support

Table 1: Survey sources of federal funding

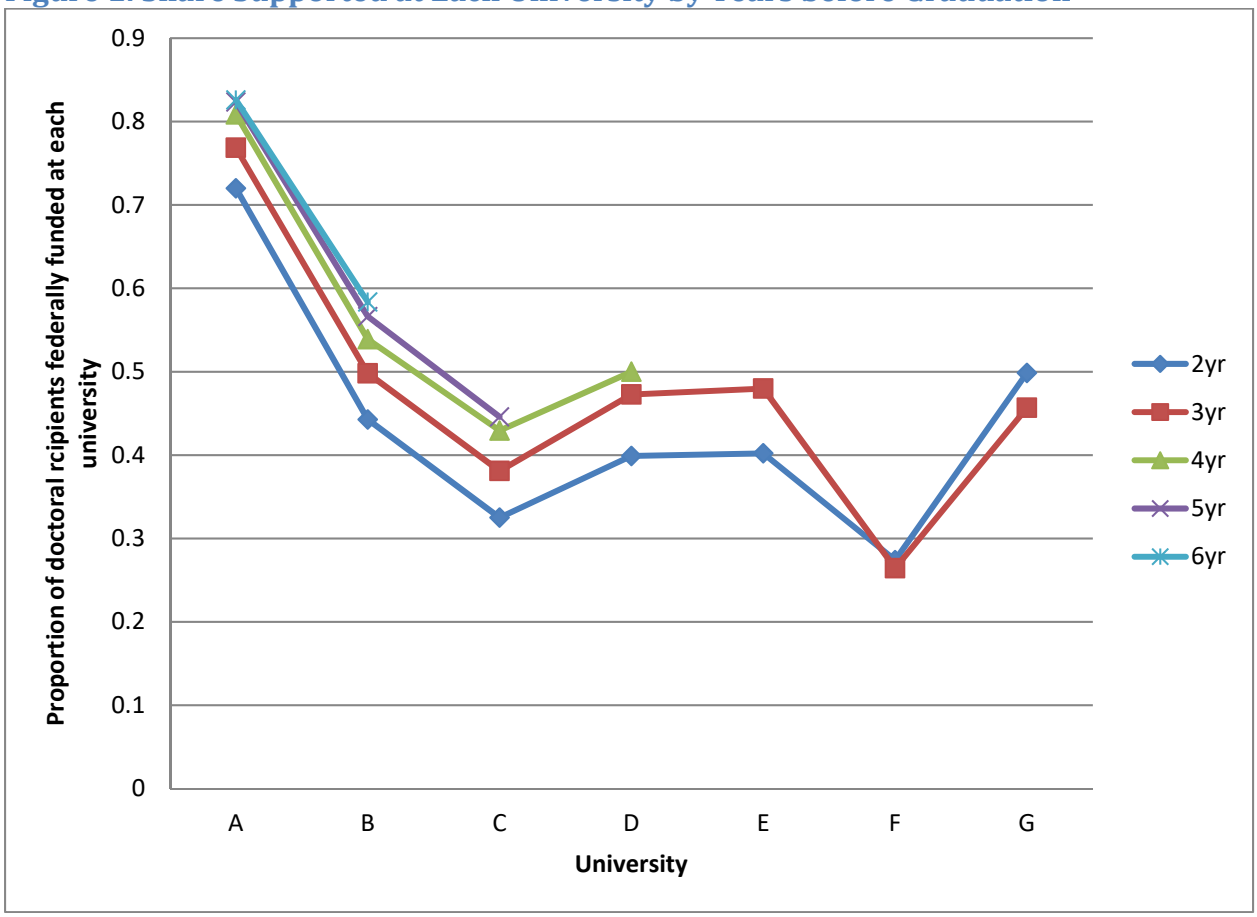
Source of Funding	SED[1]	SED-UMETRICS[2]	Federal SED-UMETRICS[3]
Research assistantship	6117	4006	3410
Fellowship, scholarship	5703	3036	2522
Teaching assistantship	4745	2613	2166
Grant	2534	1494	1239
Missing (did not respond)	2584	1084	852
Traineeship	2054	882	689
Spouse's, partner's, or family's earnings or savings	1712	663	501
Foreign (non-U.S.)	1568	541	399
Personal earnings during graduate school	338	270	231
Loans (from any source)	391	200	164
Personal savings	550	177	135
Employer reimbursement/assistance	356	163	132
Other	375	117	81
Internship, clinical residency	680	341	268
Other assistantship	5	2	1

Responses to SED Question A5: Which of the following were sources of financial support during graduate school?
*814 individuals from the matched SED-UMETRICS sample were either federally funded after their PhD completion or in a period earlier than 2 years before their PhD graduation date
[1] Individuals in SED (8,931)
[2] Individuals in the UMETRICS-SED matched sample (4,930)
[3] Individuals in the UMETRICS-SED matched sample who were federally funded* (4,116)

It is also possible to use UMETRICS data to understand the patterns of federal funding for each individual over a period of time. Federal funding is not just a one-time source of support, since the data show that, on average, doctoral recipients were supported for 10 months out of the two years prior to receiving their degree; the median duration of coverage was 8 months.

As Figure 1 shows, there are large variations in the share supported across universities, which appear to be driven by the amount of research funding as well as the fields in which the schools are concentrated. The share supported also increases by as much as 15% as the horizon over which support is calculated increases from 2 years to (up to) 5 or 6 years before graduation (two of the universities in the data provided data for more than 10 years).

Figure 1: Share Supported at Each University by Years before Graduation



Note. The figure shows the share of people supported by federal funds based on the time horizon – the number of years prior to graduation - over which support is calculated. Schools provided data for different periods, so data are available for fewer schools over longer time horizons. We used data on all people with data at a given time horizon, so the composition of people at each school varies by time horizon.

The duration of support is substantially lower in agriculture, mathematics, psychology, social science and non-Science, Engineering, or Health (SEH) fields, where the average duration of those funded is just over 8 months. Federal funding is also not monolithic: most students rely on more than one research grant for funding; the doctoral recipients were supported by 1.7 federal grants on average during the last two years prior to receiving their degree. Consistent with our intuition, the share of people supported by 2 agencies increases substantially as the window over which support is calculated increases – from 7% over a 2-year window to 15% or higher over a 5- or 6-year horizon. While few people are supported by 3 agencies, the rate

increases even more dramatically – from .6% over a 2-year horizon to 3% over a 5- or 6-year horizon.

Of course, there are many different sources of federal funding[3]; because each agency only has access to its own data, it has hitherto not been possible to capture how agency funding is interconnected at the individual student level. UMETRICS data do provide that information, and our analysis uncovers some striking interdependencies. As Table 2 shows, the NIH provided funding for roughly 40% of the federally funded cohort; NSF supported slightly more. The DOD supported just over one in ten and DOE roughly one in ten, all depending on the horizon over which support is calculated.

[3] The major agencies are the NIH, the NSF, the DOD and DOE.

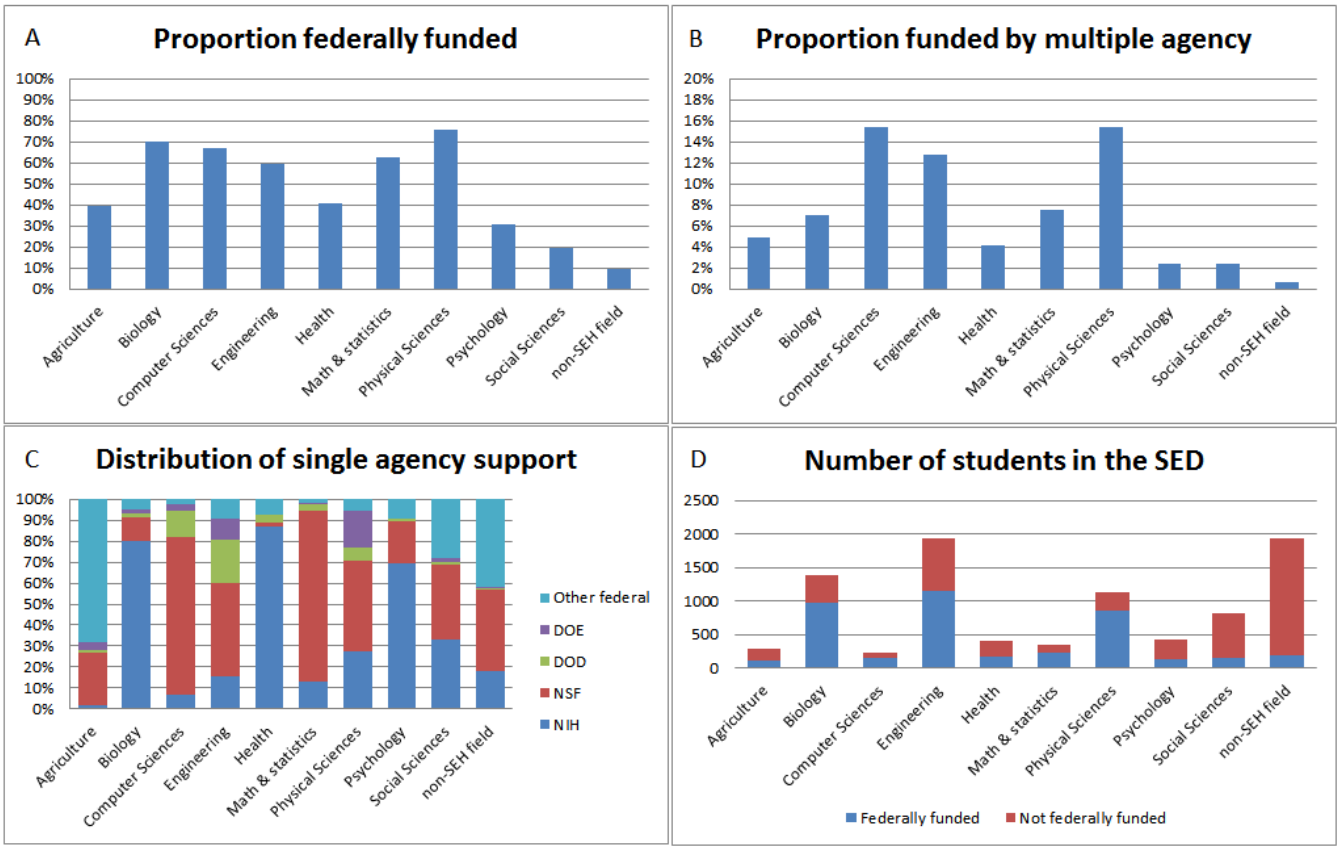
Table 2: Federal funding by SED Discipline and funding agency

SED Discipline	SED Total	Federally Funded	Ever NIH	Only NIH	Ever NSF	Only NSF	Ever DOD	Only DOD	Ever DOE	Only DOE	Ever Other Federal	Only Other Federal	Multiple Sources of Funding
Agriculture	282	111	3	2	32	24	1	1	7	4	79	66	14
Biology	1388	970	773	696	168	98	33	18	33	17	64	43	98
Computer Sciences	234	156	14	8	123	90	31	15	16	4	9	3	36
Engineering	1944	1160	196	141	607	406	326	187	154	90	140	87	249
Health	412	167	145	130	11	3	8	6	0	0	21	11	17
Math & statistics	356	222	43	26	183	158	15	6	5	2	4	3	27
Physical Sciences	1127	852	239	188	445	293	85	43	178	117	86	38	173
Psychology	423	131	92	84	29	24	4	2	0	0	16	11	10
Social Sciences	825	159	63	46	61	50	1	1	3	3	49	39	20
non-SEH field	1940	188	36	32	77	68	2	1	2	1	81	74	12
Total	8931	4116	1604	1353	1736	1214	506	280	398	238	549	375	656

Many students are supported by grants from multiple agencies, but the agency dependence differs dramatically by agency. For example, of the 506 individuals ever supported by DOD funding, 226 or almost 45% were supported by funding from another federal agency. Of the 398 individuals ever supported by DOE funding, 160 or 40% were funded from another agency. These proportions drop dramatically for NIH funding. Of the 1604 individuals ever supported by NIH funding in the two years prior to graduating, just over 15% were supported by grants from other agencies. Put another way, 84% of doctoral recipients supported by NIH had NIH as their sole source of externally-funded support. Students supported by NSF funding are slightly less dependent: 30% of those ever supported by NSF funding were also funded from other sources.

The new data also make it possible to examine the distribution of federal funding by discipline. This point is illustrated in Figure 2. There are stark differences in the proportion of doctoral recipients in each discipline who are supported by federal funding. As panel A in Figure 2 shows, well over three in four students in physical sciences, agriculture / biology, and computer / mathematical sciences are on federal grants in the two years prior to getting their degree. That number drops to less than one in five for social sciences and one in ten for humanities. Panel B shows that there are also large field differences in the probability that people receive support from multiple federal agencies. Panel C shows how the various agencies support work by field, with large, but intuitive, differences – NIH overwhelmingly supports PhDs in biology, health, and psychology; NSF dominates mathematics and computer science. Panel D gives counts of the numbers of PhD graduates overall and with federal support.

Figure 1: The incidence of federal funding by discipline



Much concern has been voiced about diversity in the research workforce(13). These linked data will enable new insights about the connection between federal funding and workforce composition. As Table 3 demonstrates, in this linked sample, there are many fewer women than men in the SED cohort -- 56% are male and 44% female. Only 39% of females are federally funded, while 52% of males are. Another way of putting this is that of the over 8924 individuals in the dataset, only 17% are federally funded females; 29% are federally funded males. Of course, these demographics are different across funding agencies. Of those doctoral recipients ever supported by NIH, half are female, compared with 29% for NSF, 20% for DOD, and 17% for DOE.

Table 3: Federal funding by gender

Gender	Total SED	Federally Funded	Ever NIH	Ever NSF	Ever DOD	Ever DOE	Ever Other Federal
Male	5029	2593	802	1226	406	330	307
Female	3895	1520	801	509	100	68	241
Missing	7	3	1	1	0	0	1
Total	8931	4116	1604	1736	506	398	549

We performed similar calculations by race, ethnicity, and US born individuals: NIH supports more African Americans; NIH and DOE support more domestic relative to foreign-born doctoral recipients than the other agencies. Table 4 provides a summary. Column 1 shows that without controlling for discipline, Asians are most likely to be federally funded and females are less likely to be funded. Columns 2 and 3 show that individuals with degrees in biology, computer science, and engineering are most likely to be federally funded.

Table 4: Correlates of ever federally funded

	(1)	(2)	(3)
Asian	0.129	0.041	-0.013
	(0.038)***	(0.034)	(0.019)
African American	-0.070	-0.053	-0.056
	(0.045)	(0.041)	(0.024)**
White	0.007	-0.022	-0.047
	(0.036)	(0.032)	(0.018)***
Hispanic	-0.053	-0.063	-0.032
	(0.042)	(0.037)*	(0.021)
US Born	0.007	0.029	0.011
	(0.016)	(0.014)**	(0.009)
Mother BA	0.069	0.039	0.010
	(0.024)***	(0.021)*	(0.014)
Father BA	-0.012	-0.010	0.000
	(0.017)	(0.015)	(0.010)
Both parents BA	-0.017	-0.008	0.004
	(0.029)	(0.025)	(0.016)
Female	-0.117	-0.022	-0.009

	(0.011)***	(0.010)**	(0.006)
Agriculture		0.300	-0.002
		(0.031)***	(0.019)
Biology		0.585	0.099
		(0.015)***	(0.011)***
Computer Science		0.577	0.085
		(0.033)***	(0.022)***
Engineering		0.489	0.062
		(0.015)***	(0.010)***
Health Sciences		0.305	-0.011
		(0.027)***	(0.018)
Math & Stats		0.518	0.104
		(0.028)***	(0.019)***
Physical Sciences		0.641	0.116
		(0.016)***	(0.012)***
Psychology		0.215	-0.012
		(0.025)***	(0.016)
Social Sciences		0.094	-0.022
		(0.017)***	(0.011)**
Years complete<=3			0.012
			(0.018)
Years complete =4			-0.023
			(0.013)*
Years complete =6			-0.082
			(0.019)***
Years complete =7			0.015
			(0.032)
Years complete >=8			-0.122
			(0.021)***
Years complete missing			-0.807
			(0.011)***
Constant	0.496	0.124	0.812
	(0.040)***	(0.036)***	(0.024)***
Controls for	University and graduation year	University and graduation year	University and graduation year
N	8,015	8,015	8,015
Adjusted R-squared	0.05	0.27	0.70

*** p<0.1; ** p<0.05; *** p<0.01; OLS models with marginal effects. Robust standard errors in parentheses. Reference Category: never federally funded; Male; no parental BA; Not US Born; Other race; Non-SHE field (discipline); Years complete=5.**

4. Relationship between Federal Funding and Post-Graduation Plans

An important contribution of this approach is to be able to examine the relationship between federal funding and the future career pathways of PhD recipients; the SED provides rich information about the career plans of graduating doctoral recipients. Information is available on three types of pathways: plans to take a postdoctoral position, plans to work in the private

sector, and plans to work in the R&D sector. Only about 70% of the cohort responded to the question, but of those, 44% reported that they planned to be either a postdoctoral fellow or research associate, and 71% would be working in R&D (either in industry or academia).

Plans to take a Postdoc Position: Our data indicate that federal funding is strongly correlated with the likelihood that individuals choose a postdoctoral position – individuals with federal funding are 5-22% more likely to take a postdoc position. The intensity of support (the number of months a person is supported by federal funding) is weakly positively correlated with the decision to take a postdoc; the magnitude is also quite small (.001-.003). Of course, much of the effect is related to the field of study. There is anecdotal evidence that many doctorates in the life sciences continue to a postdoc, either because additional training is necessary or because there are not many jobs available for PhDs in the life sciences. Indeed, graduating with a degree in the life sciences is highly positively correlated with the decision to take a postdoc. Once field differences are included, the relationship between federal funding and postdoc plans is weaker. The relationship with gender is not significant, but foreign-born individuals are more likely to take postdoc positions, perhaps because it provides a way to stay in the U.S.

Table 5: Correlates of Planned Postdoc (Ever funded)

	(1)	(2)	(3)	(4)
Ever Funded	0.224		0.052	
	(0.021)***		(0.028)*	
Ever Funded NIH		0.281		0.060
		(0.022)***		(0.025)**
Ever Funded NSF		0.089		0.026
		(0.020)***		(0.022)
Ever Funded DOD		0.029		0.030
		(0.030)		(0.031)
Ever Funded DOE		0.142		0.076
		(0.033)***		(0.034)**
Ever Funded Other		0.145		0.091
		(0.029)***		(0.030)***
Months funded	0.003	0.005	0.001	0.001
	(0.002)**	(0.001)***	(0.001)	(0.001)
Agriculture			0.498	0.490
			(0.037)***	(0.038)***
Biology			0.578	0.577
			(0.021)***	(0.023)***
Computer Science			0.006	0.015
			(0.039)	(0.039)
Engineering			0.153	0.156
			(0.021)***	(0.021)***
Health Sciences			0.256	0.253
			(0.034)***	(0.035)***
Math & Stats			0.280	0.292

			(0.036)***	(0.036)***
Physical Sciences			0.409	0.410
			(0.024)***	(0.024)***
Psychology			0.397	0.398
			(0.032)***	(0.033)***
Social Sciences			0.113	0.112
			(0.023)***	(0.023)***
Asian			-0.082	-0.079
			(0.040)**	(0.040)**
Black			-0.050	-0.047
			(0.053)	(0.053)
White			-0.082	-0.081
			(0.038)**	(0.038)**
Hispanic			-0.066	-0.067
			(0.046)	(0.046)
US Born			-0.041	-0.042
			(0.018)**	(0.018)**
Mother/BA			-0.030	-0.030
			(0.026)	(0.026)
Father/BA			0.017	0.016
			(0.019)	(0.019)
Both parents/BA			0.036	0.036
			(0.032)	(0.032)
Female			0.007	0.005
			(0.013)	(0.013)
Years complete<=3			-0.021	-0.026
			(0.030)	(0.030)
Years complete =4			0.022	0.023
			(0.021)	(0.021)
Years complete =6			0.012	0.013
			(0.027)	(0.028)
Years complete =7			-0.007	-0.005
			(0.055)	(0.055)
Years complete >=8			-0.060	-0.065
			(0.032)*	(0.032)**
Years complete missing			-0.084	-0.082
			(0.027)***	(0.024)***
Controls for	University and graduation year	University and graduation year	University and graduation year	University and graduation year
Constant	0.387	0.390	0.371	0.371
	(0.019)***	(0.019)***	(0.051)***	(0.050)***
N	5,090	5,090	5,090	5,090
Adjusted R-squared	0.09	0.10	0.24	0.24

*** p<0.1; ** p<0.05; *** p<0.01; OLS models with marginal effects. Robust standard errors in parentheses. Reference Category: never federally funded; Male; no parental BA; Not US Born; Other race; Non-SHE field (discipline); Years complete=5. Sample includes people with definite plans and non-missing responses about type of pathways: plans to take a postdoctoral position; plans to work in the private sector and plans to work in the R&D sector.**

The SED includes two interesting questions about the post-graduation employer. The first is a question about the principal employer for which respondents will be working after graduation, and whether or not the employer is in the private sector. Those respondents who had received federal funding were just under 3-6% more likely to work in the private sector than those who had never received federal funding. The second is a question about whether doctoral recipients plan to work in the R&D sector. The results here are striking. The relationship with federal funding is large and significant until field of degree is included. Once field of degree is included, only NIH funding is correlated with R&D employment plans – not surprisingly, those with degrees in engineering, computer science, and biology are most likely to work in the R&D sector[4].

Table 6: Correlates of R&D Work Placement (Ever Funded)

	(1)	(2)	(3)	(4)
Ever Funded	0.273		0.023	
	(0.017)***		(0.022)	
Ever Funded NIH		0.209		0.039
		(0.017)***		(0.018)**
Ever Funded NSF		0.163		0.000
		(0.015)***		(0.015)
Ever Funded DOD		0.170		-0.003
		(0.020)***		(0.019)
Ever Funded DOE		0.176		0.020
		(0.022)***		(0.022)
Ever Funded Other		0.095		0.018
		(0.023)***		(0.022)
Months funded	0.003	0.006	0.001	0.001
	(0.001)***	(0.001)***	(0.001)	(0.001)
Agriculture			0.473	0.474
			(0.033)***	(0.034)***
Biology			0.516	0.506
			(0.020)***	(0.021)***
Computer Science			0.398	0.406
			(0.035)***	(0.035)***
Engineering			0.473	0.477
			(0.020)***	(0.021)***
Health Sciences			0.337	0.329
			(0.033)***	(0.034)***

[4] The measurement of plans to work in the R&D sector (for those who had definite plans) was based on the response to Question B10 of the SED survey: "What will be your primary and secondary work activities? Using question B10a (answer to primary work activity) we generated the following binary outcome: R&D=1 If B10a=Research and Development; R&D=0 If B10a=Teaching; Management or Administration; Professional Services to Individuals; Other and R&D=.if B10a=99 (NO DEFINITE PLANS) or If B10a=. (individual with definite plans did not answer question

Math & Stats			0.332	0.337
			(0.032)***	(0.033)***
Physical Sciences			0.461	0.462
			(0.023)***	(0.023)***
Psychology			0.257	0.254
			(0.033)***	(0.033)***
Social Sciences			0.269	0.268
			(0.026)***	(0.026)***
Asian			0.007	0.007
			(0.038)	(0.038)
Black			-0.085	-0.085
			(0.051)*	(0.052)*
White			-0.043	-0.042
			(0.036)	(0.036)
Hispanic			-0.036	-0.036
			(0.044)	(0.044)
US Born			-0.104	-0.105
			(0.016)***	(0.016)***
Mother/BA			-0.036	-0.036
			(0.025)	(0.025)
Father/BA			0.023	0.023
			(0.017)	(0.017)
Both parents/BA			0.050	0.050
			(0.029)*	(0.029)*
Female			-0.039	-0.041
			(0.012)***	(0.012)***
Years complete<=3			-0.015	-0.015
			(0.021)	(0.021)
Years complete =4			0.002	0.001
			(0.015)	(0.015)
Years complete =6			-0.014	-0.013
			(0.021)	(0.021)
Years complete =7			0.008	0.010
			(0.038)	(0.038)
Years complete >=8			-0.065	-0.067
			(0.027)**	(0.027)**
Years complete missing			-0.114	-0.118
			(0.022)***	(0.019)***
Controls for	University and graduation year	University and graduation year	University and graduation year	University and graduation year
Constant	0.605	0.614	0.527	0.531
	(0.017)***	(0.017)***	(0.046)***	(0.046)***
N	5,090	5,090	5,090	5,090
Adjusted R-squared	0.13	0.12	0.32	0.32

*** p<0.1; ** p<0.05; *** p<0.01; OLS models with marginal effects. Robust standard errors in parentheses. Reference Category: never federally funded; Male; no parental BA; Not US Born; Other race; Non-SHE field (discipline); Years complete=5. Sample includes people with definite plans and non-missing responses about type of pathways: plans to take a postdoctoral position; plans to work in the private sector and plans to work in the R&D sector.**

5. Summary

This paper has combined a new administrative dataset (UMETRICS) with a major established survey dataset (SED) to provide new insights into the links between federal funding structure and the composition of the doctorally trained workforce. The contribution is substantial, since current surveys do not provide information about the extent or intensity of federal funding, and no individual agency, looking at its own data, can document how many doctoral recipients are supported by federal funds. We find large differences in funding patterns across federal funding agencies in terms of the number and disciplinary training of doctoral students. Federal funding is also strongly related to individual characteristics and career pathways.

Much more can be done; the analysis draws on data for only seven major research universities for which long historical data exist. As data from the new Institute for Research on Innovation and Science expands across more institutions and longer time periods, the analysis can become more generalized. It can also be extended to examine the links to non-federal sources of funds, as UMETRICS data is expanding to include that information for partner institutions. And with those links, more causal analysis can be added to the descriptive characterization provided in this paper. In the long run, the opportunity exists to design surveys around the types of data that cannot be obtained from administrative sources, letting respondents focus on topics that cannot be obtained from non-survey methods. Moreover, we are pleased that researchers in other countries are beginning to develop parallel systems.

6. References

1. National Science Board (2015) *Revisiting the STEM Workforce*.
2. National Academy of Sciences (2010) *Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5* (The National Academies Press) Available at: http://www.nap.edu/openbook.php?record_id=12999.
3. Schools CG (2007) Graduate education: The backbone of American competitiveness and innovation. *A Rep from Counc Grad Sch Advis Comm Grad Educ Am Compet Washington, DC Counc Grad Sch*.
4. Government Accountability Office (2015) *WOMEN IN STEM RESEARCH Better Data and Information Sharing Could Improve Oversight of Federal Grant-making and Title IX Compliance*.
5. National Research COuncil (2010) *Data on Federal Research and Development Investments: A Pathway to Modernization*.
6. National Science and Technology Council (2008) *The Science of Science Policy: A Federal Research Roadmap* (National Science and Technology Council,Science of Science Policy Interagency Task Group, Washington, D.C.).
7. Reedy EJ, Litan B, Teitelbaum M (2011) The Current State of Data on the Science and Engineering Workforce, Entrepreneurship, and Innovation in the United States. *The Handbook of Science of Science Policy*, eds Fealing KH, Lane J, Marburger JH, Shipp S (Stanford University Press).
8. Office of Management and Budget (20915) *Statistical Programs of the United States Government: Fiscal Year 2015* Available at: https://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/statistical-programs-2015.pdf.
9. Husbands Fealing K, Lane J, Marburger J, Shipp S (2011) *The Handbook of Science of Science Policy* (Stanford University Press).
10. Largent M, Lane J (2012) STAR METRICS and the Science of Science Policy. *Res Policy Rev*.
11. Lane JI, Owen-Smith J, Rosen RF, Weinberg BA (2015) New linked data on research investments: Scientific workforce, productivity, and public value. *Res Policy*. Available at: http://www.sciencedirect.com/science/article/pii/S0048733315000025 [Accessed February 23, 2015].
12. Zolas N, et al. (2015) Wrapping it up in a person: Examining employment and earnings outcomes for PhD recipients. *Science (80-)* 360(6266):1367–1371.
13. National Institutes of Health (2012) *Draft Report of the Advisory Committee to the Director Working Group on Diversity in the Biomedical Research Workforce*.

Appendix: Additional Tables

Table 1: Sources of funding by race and ethnicity

Race	Total SED	Federally Funded	Ever NIH	Ever NSF	Ever DOD	Ever DOE	Ever Other Federal
Asian	2729	1492	522	705	234	126	119
Black	302	99	41	35	6	9	14
Hispanic	431	165	67	60	16	16	26
Other	237	103	46	34	14	16	8
White	4671	2043	847	804	210	216	208
Missing	561	214	81	98	26	15	22
Total	8931	4116	1604	1736	506	398	549

Table 2: Sources of funding by birthplace

Place of Birth	Total SED	Federally Funded	Ever NIH	Ever NSF	Ever DOD	Ever DOE	Ever Other Federal
Foreign Born	3748	1881	610	891	283	169	254
US Born	4701	2052	919	765	203	216	275
Missing	482	183	75	80	20	13	20
Total	8931	4116	1604	1736	506	398	549

Table 3: The number of people supported by various combination of federal agencies

	NIH	NSF	DOD	DOE	Other Fed
NIH	1353	160	36	8	27
NSF		1214	140	104	83
DOD			280	17	14
DOE				238	9
Other Fed					375

Table shows support for 4,073 people with support from 1 or 2 federal agencies. Diagonal elements show the number supported by each agency alone (including multiple projects supported by that agency). Other indicates federal agency other than NIH, DOD, DOE and NSF. Off diagonal elements show the number supported by each pair of agencies. There are 43 people supported by 3 or more agencies.

Table 4: Percent of doctoral recipients with federal funding supported by various combination of federal agencies

	NIH	NSF	DOD	DOE	Other Fed
NIH	32.87%	3.89%	0.87%	0.19%	0.66%
NSF		29.49%	3.40%	2.53%	2.02%
DOD			6.80%	0.41%	0.34%
DOE				5.78%	0.22%
Other Fed					9.11%

Table shows support for 4,073 people, all with support from federal agencies. Diagonal elements show the number supported by each agency alone. “Other Fed” indicates federal agency other than NIH, DOD, DOD and NSF. Off diagonal elements show the number supported by each pair of agencies. Only 1% of people are supported by 3 or more agencies.

Table 5: The Share of Doctoral Recipients supported by Agency by gender, race, and immigrant status.

	Any Federal	NIH	NSF	DOD	DOE	All Other Federal
Gender						
Male	63.04%	50.03%	70.66%	80.24%	82.91%	53.33%
Female	36.96%	49.97%	29.34%	19.76%	17.09%	46.67%
All	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Race						
White	49.64%	52.81%	46.31%	41.50%	54.27%	53.20%
Asian	36.25%	32.54%	40.61%	46.25%	31.66%	30.43%
Missing	5.20%	5.05%	5.65%	5.14%	3.77%	4.09%
Hispanic	4.01%	4.18%	3.46%	3.16%	4.02%	6.65%
Black	2.41%	2.56%	2.02%	1.19%	2.26%	3.58%
Other	2.50%	2.87%	1.96%	2.77%	4.02%	2.05%
All	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Immigrant Status						
Foreign Born	45.70%	38.03%	51.32%	55.93%	42.46%	47.83%
US Born	49.85%	57.29%	44.07%	40.12%	54.27%	48.34%
Missing	4.45%	4.68%	4.61%	3.95%	3.27%	3.84%
All	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Table 6: Correlates of ever funded by agency

	Ever NSF Funded	Ever NIH Funded	Ever DOD Funded	Ever DOE Funded
Asian	0.045	0.022	0.014	-0.030
	(0.027)*	(0.027)	(0.018)	(0.018)*
African American	-0.015	-0.026	-0.020	-0.036
	(0.031)	(0.032)	(0.018)	(0.019)*
White	0.015	-0.026	-0.002	-0.026
	(0.025)	(0.025)	(0.016)	(0.017)
Hispanic	-0.023	-0.027	-0.015	-0.028
	(0.029)	(0.030)	(0.018)	(0.019)
US Born	-0.002	0.041	0.007	0.010
	(0.011)	(0.011)***	(0.007)	(0.006)*
Mother BA	-0.012	0.013	0.023	-0.010
	(0.017)	(0.016)	(0.012)*	(0.010)
Father BA	-0.003	-0.005	0.004	-0.008
	(0.013)	(0.011)	(0.008)	(0.007)
Both parents BA	0.020	-0.001	-0.031	0.014
	(0.021)	(0.019)	(0.014)**	(0.011)
Female	-0.023	0.022	-0.017	-0.024
	(0.008)***	(0.008)***	(0.005)***	(0.004)***
Agriculture	0.077	0.007	0.002	0.016
	(0.021)***	(0.008)	(0.004)	(0.009)*
Biology	0.068	0.542	0.014	0.019
	(0.010)***	(0.014)***	(0.004)***	(0.004)***
Computer Science	0.476	0.058	0.133	0.071
	(0.035)***	(0.017)***	(0.024)***	(0.019)***
Engineering	0.259	0.095	0.152	0.077
	(0.013)***	(0.010)***	(0.009)***	(0.007)***
Health Sciences	-0.017	0.338	0.014	0.005
	(0.010)	(0.025)***	(0.007)**	(0.002)***
Math & Stats	0.464	0.123	0.021	0.011
	(0.029)***	(0.020)***	(0.011)*	(0.008)
Physical Sciences	0.343	0.203	0.054	0.153
	(0.017)***	(0.014)***	(0.009)***	(0.012)***
Psychology	0.040	0.194	0.012	0.003
	(0.014)***	(0.021)***	(0.005)**	(0.002)*
Social Sciences	0.035	0.054	-0.002	0.003
	(0.011)***	(0.011)***	(0.002)	(0.002)
Constant	0.030	0.056	0.030	0.022
	(0.028)	(0.029)*	(0.018)*	(0.018)
Controls for	University and graduation year	University and graduation year	University and graduation year	University and graduation year
N	8,015	8,015	8,015	8,015
Adjusted R-squared	0.16	0.24	0.09	0.07

*** p<0.1; ** p<0.05; *** p<0.01; OLS models with marginal effects. Robust standard errors in parentheses. Reference Category: Male; no parental BA; Not US Born; Other race; Non-SHE field (discipline).**

.

Table 7: Correlates of only funded by agency

	Only NSF Funded	Only NIH Funded	Only DOD Funded	Only DOE Funded
Asian	0.047	0.016	0.000	-0.016
	(0.022)**	(0.025)	(0.015)	(0.014)
African American	0.022	-0.008	-0.012	-0.023
	(0.026)	(0.030)	(0.016)	(0.015)
White	0.026	-0.021	-0.008	-0.016
	(0.020)	(0.024)	(0.014)	(0.014)
Hispanic	0.007	-0.033	-0.018	-0.012
	(0.024)	(0.028)	(0.015)	(0.015)
US Born	-0.008	0.030	0.009	0.007
	(0.010)	(0.010)***	(0.005)*	(0.004)
Mother/BA	-0.000	0.013	0.014	-0.004
	(0.015)	(0.015)	(0.009)	(0.007)
Father/BA	-0.006	-0.011	-0.001	-0.001
	(0.011)	(0.011)	(0.006)	(0.005)
Parents/BA	0.018	0.006	-0.012	0.005
	(0.018)	(0.018)	(0.011)	(0.009)
Female	-0.020	0.023	-0.010	-0.017
	(0.007)***	(0.008)***	(0.003)***	(0.003)***
Agriculture	0.053	0.005	0.005	0.005
	(0.019)***	(0.007)	(0.004)	(0.005)
Biology	0.030	0.489	0.009	0.011
	(0.009)***	(0.014)***	(0.003)***	(0.003)***
Computer Science	0.334	0.031	0.066	0.015
	(0.034)***	(0.014)**	(0.018)***	(0.010)
Engineering	0.162	0.068	0.090	0.042
	(0.011)***	(0.009)***	(0.007)***	(0.005)***
Health Sciences	-0.033	0.305	0.012	0.003
	(0.007)***	(0.025)***	(0.006)*	(0.001)**
Math & Stats	0.399	0.073	0.006	0.002
	(0.028)***	(0.016)***	(0.007)	(0.005)
Physical Sciences	0.220	0.160	0.030	0.099
	(0.015)***	(0.013)***	(0.006)***	(0.010)***
Psychology	0.030	0.175	0.005	0.003
	(0.013)**	(0.020)***	(0.004)	(0.001)**
Social Sciences	0.026	0.035	-0.001	0.002
	(0.010)**	(0.009)***	(0.002)	(0.002)
Constant	0.011	0.051	0.020	0.010
	(0.024)	(0.027)*	(0.015)	(0.014)
Controls for	University and graduation year	University and graduation year	University and graduation year	University and graduation year
N	8,015	8,015	8,015	8,015
Adjusted R-squared	0.11	0.23	0.05	0.05

*** p<0.1; ** p<0.05; *** p<0.01; OLS models with marginal effects. Robust standard errors in parentheses. Reference Category: Male; no parental BA; Not US Born; Other race; Non-SHE field (discipline).**