Investigating the US Biomedical Workforce:

Gender, Field of Training, and Retention

Published in *Science and Public Policy*, scz039, https://doi.org/10.1093/scipol/scz039
Published online: 13 August 2019

Anne E. Winkler, Department of Economics, University of Missouri-St. Louis, St. Louis, MO 63121, USA; Sharon G. Levin†, Department of Economics, University of Missouri-St. Louis, St. Louis, MO, 63121, USA; Michael T. Allison, Department of Economics University of Missouri-St. Louis, St. Louis, MO 63121, USA.

Corresponding author: Anne E. Winkler, 314-516-5563 (phone), 341-516-5352 (fax),

awinkler@umsl.edu (email).

† Sharon Levin passed away on 8/21/17. She was the original PI on the National Institutes of Health grant that funded this research.

Abstract: The biomedical research workforce plays a crucial role in fostering economic growth and improving public health through discoveries and innovations. This study fills a knowledge gap by providing a comprehensive portrait of this workforce and retention within it. A distinguishing feature is that we use an occupation-based definition which allows us to look "backward" to field of training and assess the extent to which it has grown more interdisciplinary, and how this differs by gender. The analysis is conducted using restricted-use SESTAT data, the most comprehensive dataset on the scientific workforce in the United States, for the years 1993, 2003, and 2010. Among the findings, we identify differences in interdisciplinarity in training by gender, and these differences have widened. In the retention analysis, which focuses on the 7-year period, 2003 to 2010, we find that retention is negatively and significantly associated with interdisciplinary training for women, but not for men.

Funding: The work was supported by National Institutes of Health grant U01-GM-112599-02. Restricted data were obtained through the grant from the National Science Foundation. The analysis and conclusions are the authors' alone and do not reflect the views of the National Institutes of Health or the National Science Foundation.

Acknowledgements: Jamie Vergano provided expert research assistance. The authors are grateful for comments on earlier drafts from Donna Ginther, Misty Heggeness, Peter Henderson, Paula Stephan, and National Institutes of Health staff. This study also benefitted from feedback received from presentations at APPAM, the Midwest Economics Conference, Washington University and University of Missouri-Columbia.

1. Introduction

Understanding the US biomedical labor force is crucial because research discoveries and innovations directly foster economic growth and affect public health and well-being, both in the United States and abroad.[1] Despite the importance of this sector, much remains to be learned about the research workforce as discussed by National Institutes of Health (NIH, 2012) and Heggeness et al. (2016, 2017). In this study, we have two objectives: (1) examine the composition of the biomedical workforce using a broad definition to capture the totality of the research enterprise; and (2) analyze the factors responsible for the retention (or lack thereof) of scientific talent within this workforce. In doing so, we focus particular attention on the relationship between gender, field of training, and retention in the field. The analysis is conducted using restricted-use data from the Scientists and Engineers Statistical Data System (SESTAT), the most comprehensive dataset on the scientific workforce in the United States, for the years 1993, 2003, and 2010.

To get as broad a picture as possible of the biomedical workforce for 1993-2010, we employ an occupation-based definition, rather than using field of study (e.g. NIH, 2012). One advantage to this approach is that it captures individuals who are actively engaged in biomedical research but not necessarily trained in biomedical science. An added advantage of an occupation-

[1] NIH (2012) has recognized that in order to maintain a bright and productive scientific workforce, it must "attract and retain the best and most diverse scientists, engineers and physicians from around the world to conduct biomedical research as well as increase the number of domestic students from diverse backgrounds who excel in science and become a part of the STEM workforce."

based approach is that we are able to look backward to field of training and thereby investigate the extent to which the biomedical research enterprise has grown increasingly interdisciplinary. This question is of interest in light of an increased emphasis on interdisciplinary research as a means to address complex problems (National Academies of Science, 2015; and Stephan, 2012).[2] What is not yet well-understood is what the precise disciplinary background of the biomedical research workforce looks like. In investigating this question, we examine those with Bachelor's, Master's degrees, and MDs, in addition to the narrower group of PhD-trained scientists who are more frequently studied. Further, we broaden the scope beyond academia; our data show that 50% of the biomedical workforce is employed in government and industry.

In addition to providing a broad portrait of the biomedical workforce, we exploit the longitudinal nature of the restricted-use SESTAT data to look at the 7-year retention rate of those identified as biomedical researchers in 2003. This dimension is important in light of concerns raised about retention in biomedical research as a whole (NIH, 2012) and women's retention in STEM fields (e.g. Preston, 2004; Stephan and Levin, 2005; and Hunt, 2016).

2. Related Research

2.1 Biomedical Workforce: Description and Trends

NIH, which was formally established in 1930, plays a key role in providing funding for the training and support of the biomedical research workforce. A particularly significant development was the doubling of NIH funding (in nominal terms) over the period 1998 to 2003

[2] A related issue is whether interdisciplinary research has, in fact, enhanced scientific knowledge. For an investigation, see for instance, Wang et al. (2015).

(NIH 2012), though NIH funding then declined in real terms over the next decade (Johnson and Seka, 2018). The doubling led to a substantial rise in the number of new biomedical PhDs (NIH, 2012). Several other developments also occurred which bear on the size and composition of the biomedical workforce and the sectors in which researchers work. Kahn and Ginther (2017) and NIH (2012) detail the considerable rise in the number of biomedical postdocs since the late 1980s, including a rising share of those who obtained biomed PhDs abroad.[3] In addition, NIH (2012) points to a growing number of biomed-trained PhDs employed in non-academic, non-research positions. The end of mandatory retirement in 1994 is also a relevant factor, in that tenured incumbents are remaining longer in their academic positions, reducing the number of entry-level permanent slots. Numerous researchers (Alberts et al, 2014; Blau & Weinberg, 2017; Daniels 2015; NIH 2012; Stephan 2012 and Kahn and Ginther, 2017) have raised concerns about the confluence of these factors on the well-being of the biomedical field and its research workforce.

Most attention on the biomedical workforce has focused on those who received biomedical PhDs (NIH, 2012, Kahn and Ginther, 2017), but some work has used an occupation-based definition, as employed in this study. For instance, Heggeness et al. (2016, 2017) examined those employed in biomedical occupations (though not necessarily doing research, a key restriction examined here) using data from the US Census and the American Community

[3] Postdocs are an important component of the biomedical workforce but as discussed at length by Kahn and Ginther (2017), there is no single source that includes all postdocs, not even the NSF's Survey of Graduate Students and Postdoctorates in Science and Engineering. One group that is especially difficult to capture are postdocs who earned PhDs abroad.

Survey (ACS). For those with a PhD, they calculated the size of this labor force to be just over 75,000 in 2010; the 2010 figure for all individuals in the biomedical labor force, regardless of educational attainment, was around 220,000. For the PhD component of the workforce, they also conducted a more in-depth analysis of changes in the demographic composition over the period 1990 to 2014. Among the trends cited, women's share of the PhD workforce increased (although still just slightly above 40% in 2014), and the share of foreign-born PhDs rose to over 50%.[4] While informative, the analysis was not restricted to those engaged in research and did not investigate field of training.

2.2 Research on Retention within STEM

A fairly extensive body of research has focused on women's progress within STEM fields (e.g. Preston, 2004; Xie & Shauman, 2008; Levin and Stephan, 2005; Kahn & Ginther, 2015; and Hunt, 2016) and within academic STEM positions (Ginther and Kahn, 2009; and Xu, 2008). The vast majority of this research focuses on those trained in science or a particular subfield within science. A consistent pattern staring with the early work by Preston is that women trained in science exit from academia and from the labor force as a whole at higher rates than men. To our knowledge this is the first study to examine retention within the biomedical occupation. Here we review the most closely-related studies on retention that were recently conducted and used SESTAT.

[4] In related work, Gibbs et al. (2014) looked at the biomedical workforce based on a survey of 1,500 US citizens and permanent residents who completed their PhD in the Biomedical Sciences between 2007-2012. Using SESTAT, we are able to provide a national portrait.

Levin & Stephan (2005) examined retention in the IT workforce (defined as those who are computer analysts, computer engineers, computer programmers, information scientists) between 1993 and 1999 using restricted SESTAT data. As in the study at hand, individuals employed in the field may have a formal degree in the field, though others might not. (All have at least a bachelor's degree since the data are from SESTAT). They found that women exit at higher rates, even after controlling for other factors, and these exits are more often exits out of the labor force rather than a switch to another occupation. They point to family roles as the explanation.

Using restricted SESTAT data, Kahn and Ginther (2015) examined gender differences in early career trajectories of cohorts of bachelor of science in engineer (BSE) graduates who received their degrees between 1985 and 2009. Outcomes examined included whether the BSE graduate was engaged in engineering (or not), and whether out of the labor force (or not). They found a higher rate of exits out of the labor force for women versus men, but did not find any changes in the gender retention rate across cohorts. The key explanation for the gender difference, regardless of cohort, is associated with women's greater exit out of the labor force entirely. Family-related factors are identified as a likely explanation.

Ginther and Rosenbloom (2018) examined gender differences in retention in Computer Science/IT occupations for the bachelors' and masters' degree populations using the NSCG and NSCRG components of SESTAT. Their analysis focused on those trained in CS/IT and those in the occupation itself. They found that over one-half of all those in the occupation have training in non-CS/IT fields based on a question that asks how closely their occupation is related to their degree. They also examined retention rates within the occupation for those trained in CS/IT and for all those in the occupation, regardless of training. They found that women are more likely to exit the CS/IT occupation, especially if they have a young child or an employed spouse. Notably, these are exits out of the labor force altogether, not exits to other occupations.

Another closely related paper is Hunt (2016). Hunt used cross-section NSCG data (the largest component of SESTAT data) for 2003 and 2010, the same years as studied here. Her focus was on comparing exits in engineering and in science with comparisons to male-dominated fields of economics and finance. In her study, an exit is defined differently than here: it refers to current employment that is not related to the field of highest degree (again, Bachelor's degree or more). She found a higher exit rate for women in engineering relative to science, and that the exit rate for engineering was similar to that for women in economics and finance. Contrary to other researchers, she found that the explanation was not due to family-related reasons, but rather perhaps due to lack of adequate mentoring and networks in engineering and these other male-dominated fields. Another feature of Hunt's study is that she looked at the correlation between (stated) preferred job attributes such as career advancement opportunities, desirable location, benefits and retention. These same attributes are briefly examined here.

To summarize research on retention in STEM, women tend to exit from the field of training or occupation at higher rates than men, and they more often exit the labor force entirely. Further, most, though not all studies, point to family factors as an important contributor.

3. Data and Definition of the Biomedical Workforce

3.1 Data

We use a restricted version of NSF's Scientists and Engineers Statistical Data System (SESTAT) for the years 1993, 2003, 2010. This integrated data system is a unique source of longitudinal information on the education and employment of the college-educated U.S. science and engineering workforce. These data have been previously used in related work, as described in the prior section. The SESTAT target population is defined as "individuals with a bachelor's degree or higher, educated or working in an S&E [Science & Engineering] field or occupation[5] who are age 75 or younger, noninstitutionalized, and living in the United States as of the survey reference date."[6] It is comprised of three biennial surveys: The National Survey of College Graduates (NSCG), The National Survey of Recent College Graduates (NSRCG), and the Survey of Doctorate Recipients (SDR). The NSCG is the core of the system and is drawn from individuals living in the United States at the time of the decennial population census[7] (or for the 2010 survey, the 2009 American Community Survey(ACS)) who were identified as having at least a college degree in S&E or working in S&E but not necessarily with a degree in S&E. The NSRCG supplements the NSCG data with the addition of recent college graduates. The third survey in SESTAT is the Survey of Doctoral Recipients (SDR), which is drawn from the Survey

[5] Starting with 2003, the population also includes those in S&E-related fields and occupations. For ease of writing, we refer to "S&E" for all years.
[6] Recent SDRs include a cohort of US PhDs who are employed abroad. Those individuals are excluded here since we are interested in the US-based biomedical workforce.
[7] The major change in the underlying sample for the NSCG is discussed in greater detail shortly.

of Earned Doctorates (SED) and includes the stock and inflow of scientists and engineers earning doctoral degrees in the United States. This latter survey is regarded as the "gold standard" for analysis of US-trained PhDs (NIH, 2012). Appendix A provides further details.

There are several reasons that SESTAT is a superior choice to using the SDR alone to study the scientific workforce.[8] Among these, the SDR neglects two potentially important groups: (1) those individuals whose doctorates were earned abroad and may hold postdocs as well as other research-related positions in the scientific workforce in the United States; and (2) medical doctors who may be engaged in biomedical research and yet are not included in the SDR unless they also have earned a science doctorate in the United States. Further, SESTAT permits an analysis of the biomedical workforce inclusive of those with Bachelor's or Master's degrees.[9]

During the period of analysis (1993 – 2010), there have been at least two major changes to SESTAT which merit further discussion. First, the broad categorization of occupations changed. Until the early 2000s, occupations were broadly classified as S&E and Not S&E. At that time, the NOT S&E grouping combined S&E-related occupations (and degrees) with those that might be regarded as "true" NOT S&E. S&E-Related occupations are those involving the practice or education of S&E such as secondary school math & science teachers, while "true" NOT S&E include occupations such as such as Management, Education (other), Social Services,

[8] Blau & Weinberg (2017) point to these same limitations regarding their own analysis, which principally used the SDR.

[9] Note that when we examine retention, only those individuals who are in both the 2003 and 2010 SESTAT database will be studied. Thus college and doctorate recipients since 2003 will be excluded from the analysis.

Sales & Marketing, Arts & Humanities, and Other. Starting with the 2003 survey, occupations are divided into three groups as described above: S&E, S&E-Related, and Not S&E.

The second major change concerns the 2010 redesign of the NSCG.[10] With the discontinuation of the long form of the 2010 Census, the American Community Survey (ACS) was adopted as the basis for the NSCG sample; for the 2010 NSCG, the 2009 ACS was used. The shift to the ACS also expanded coverage of non S&E occupations (and degree fields) owing to its design. In light of this change in sampling, in our discussion of 1993, 2003, and 2010 cross-sections, we take extra measures to ensure meaningful interpretation of trends in the data. This is not an issue in the longitudinal analysis because we are following the same set of individuals over time.

3.2 Definition of the Biomedical Workforce

Who should be considered members of this workforce has been a subject of discussion at NIH for many years (National Research Council, 2011). One definition is based on training in specific PhD fields as outlined in NIH report (2012, p.17).[11] Biomedical researchers might be defined narrowly as those with PhDs in "basic" biomedical fields such as biochemistry,

---

[10] The ACS addressed an additional concern about the sample frame of SESTAT raised (National Academies of Science, 2003; and National Science Foundation, n.d.); SESTAT was previously not able to incorporate immigrant S&E degree holders who entered the U.S. during the decade after the Decennial Census.

[11] Basic Biomedical is defined as Biochemistry, Bioinformatics, Biological Sciences, Biomedical Engineering, Biophysics, Biotechnology, Cell Biology, Developmental Biology/Embryology, Endocrinology, Genetics, Immunology, Microbiology, Molecular Biology, Neurosciences, Nutritional Science, Parasitology, Pharmacology, Pharmaceutical Chemistry, Physiology, Toxicology, Veterinary Medicine, and Zoology.

bioinformatics, and biomedical sciences, or they might be defined more broadly to include other life sciences or even behavioral sciences. Using field of training as a starting point is valuable in that it provides an estimate of the extent to which those trained in the S&E field remain in the field. In studies that take this approach (e.g. Preston, 2004, Hunt 2016, Ginther and Kahn 2018), an exit is defined as holding a position that is not related to the degree field or being out of the labor force entirely). In this study, we start with those who define themselves in a biomedical occupation (explained precisely shortly) and define an exit as leaving the occupation. This approach is valuable to the extent that it allows us to look back and examine the extent to which the occupation is becoming more interdisciplinary in terms of field of training.[12] What we do not observe, however, are exits of those trained in bioscience who are not employed within the biomedical occupation (either because they are employed elsewhere or not in the labor force).

In using an occupation-based definition, the next issue is what occupations to include. Heggeness et al. (2016, 2017) restricted their analysis to biomedical scientists, while in other descriptive work, Mason et al. (2017) further included biomedical engineers, statisticians, and natural science managers. NIH (2017) went so far as to include researchers and non-researchers alike (e.g. those doing science policy, science regulation and science communication}.

Given the policy importance of biomedical research discoveries, this study focuses on the narrow definition of biomedical scientists (same as Heggeness et al., 2016, 2017), with the added restriction of those who are actively engaged in research, either as a primary or secondary

[12] Between 1999 and 2003, some SESTAT occupational definitions changed, but this change does not affect the retention analysis, which focuses on the 2003 to 2010 period.

activity. Further we limit the analysis to those employed full-time defined as 35 or more hours per week. Table 1 provides population sizes for this workforce. It also shows the specific six-digit occupations that fall under Biomedical (code 22) in SESTAT: Biochemistry and Biophysics, Biological Sciences, Medical Sciences, and Other Biomedical. The table also places the Biomedical occupation within the broader Life Science occupation which further includes Postsecondary Teachers—Natural Sciences, Agricultural and Food, and Forestry. (Field of training is categorized in SESTAT using a similar taxonomy.[13]) Table 1 also enumerates two other categories separately who might be included in an expanded definition of the biomedical research workforce: bioengineers engaged in research plus post-secondary bioscience teachers engaged in research.

The analytic sample is those whose highest degree is at least a bachelor's degree.[14] Most often, the highest degree attained is in in the field of biomedical science, but it does not have to be. As discussed earlier, to be included in SESTAT, the individual has to have earned a bachelor's in S&E *or* be employed in S&E. Note that we use the terms training and (formal) education interchangeably in this work. In addition to looking at those with bachelor's degrees

[13] There is one slight wording difference. The field of study called Biological Science has the same SESTAT code (22) as the occupational field Biological and Medical Scientists (Biomedical). We refer to Bioscience as "within-field" training for the Biomedical occupation.
[14] A related workforce is the broader Biomedical and Behavioral Science (BMBS) workforce which includes Anthropology, Audiology/Speech Pathology, Demography/Population Studies, Sociology and Psychology. This group is not studied here because it is typically defined by PhD field of study (National Research Council, 2011), while in the definition employed here, the research workforce does not need to have a PhD.

or more, we also present findings for those with a PhD/MD only.[15] This is for comparison purposes with our broader educational definition and because this group is of interest in and of itself. As discussed earlier, there is a substantial literature on career outcomes of PhDs, especially those in academia. We also briefly look at postdocs given their important role in the biomedical enterprise.

4. Findings

4.1 Cross-Section Analysis

Figure 1 provides population figures (based on weights provided in SESTAT) on the biomedical workforce for the three years of study: 1993, 2003, and 2010. It puts the full-time biomedical research workforce in the context of the biomedical and life science occupations. It shows, for instance, that about 70 percent of those in biomedical occupation are employed full-time and engaged in research (termed BIOMED here). As a point of comparison, for the broader life science occupation, about 60 percent of the workforce is employed full-time and doing research (termed LIFE here). Figure 1 also shows a considerable increase in those employed in LIFE with a bachelor's degree or more, from 188,474 in 1993 to 368,312 in 2010, a 95% increase. Even more striking, yet in line with the NIH "doubling" discussed earlier, BIOMED increased 128%, over the same period, from 126,605 to 289,147.

Table 2 provides a demographic breakdown of those employed in BIOMED. Depending on the year considered, 41 to 46% of the full-time research biomedical workforce has a

[15] Technically, this definition also includes JDs, Dentists, and other professional non-Master's degrees.

PhD/MD, which means that a considerable fraction of those employed in this sector do not. This justifies broadening analyses beyond the PhD/MD sector. Moreover, nearly 60 percent of those with a bachelor's degree or more and nearly 50 percent of those with a PhD were employed outside of academia (and research institutions), Indeed, employment in this sector has shrunk over the 1993 to 2010 period, as shown in the table. This table brings home the point that analyses of academic PhDs provide a narrow view of the workforce as a whole.

While we seek to provide a broad portrait of biomedical workers, SESTAT does provide some information for 2003 on an important component of the academic sector: postdocs.[16] Of the academic biomedical research scientists included in SESTAT, 37% of these individuals hold postdoc positions. The data further show that women in academia are over-represented among postdocs (40.7%) as compared to their percent in academia (36%).

Table 2 points to other notable patterns. For one, as has been observed elsewhere, the full-time biomedical work force engaged in research in the United States has become increasingly dependent upon foreign-born scientists. For instance, the US-born population engaged in the full-time biomedical research enterprise declined from 76.5 to 63.2 % from 1993 to 2010 for all those with at least a Bachelor's degree, and declined from 64.6 to 42.8% for those

[16] Our ability to meaningfully examine this group is limited by what is available in SESTAT. Kahn and Ginther (2017) had to piece together three surveys to as fully as possible identify US-born biomedical postdocs. For 2003 only, SESTAT provides information from all 3 surveys (variable name ACADPDOC) for those whose principal position is postdoc within a postsecondary institution during the survey week of October 1, 2003. The figure on postdocs are reported in the text only.

with a PhD/MD. Similarly, the white non-Hispanic workforce declined from 77 to 65.9 % (for bachelor's degree and more) and from 74.8 % to 59.1 % for the PhD/MD population for the full period. There has been an increase in percentage female, from 39.6 % to 50.3% for those with a bachelor's degree or more and from 30 to 42.1% for those with a PhD/MD, though the latter figure is still well below 50 percent.

One question that has not been investigated but is important given the growing emphasis on the value of interdisciplinary work (National Academies of Science, 2015; and Stephan, 2012) is the extent to which the training of individual researchers in the biomedical workforce has grown more interdisciplinary. Individuals in the biomedical workforce may have 1) received their highest training in a discipline outside of biomedical science; or 2) they may have their highest degrees in bioscience (or life science) but have a prior degree in a "different" field. Here we focus on 1) the social sciences and 2) math, computer science, and physical sciences[17] as the "different" disciplines. The social sciences are defined in SESTAT to include economics, political science, psychology, sociology and anthropology, plus other, with 50% coming from psychology alone.[18] While it would also be interesting to look at those trained in "NOT S&E", the underlying sample size is still relatively small.[19]

[17] Ideally, we would like to look at Math & Computer Science alone, but the group size is too small to draw meaningful conclusions.
[18] While NIH (2012) includes psychology as a category in their definition of biomedical training, it is captured in social sciences in SESTAT. This broad field includes clinical, social, industrial psychology, and so on.
[19] In fact, in the analysis of retention, there are no individuals with training in "Not S&E" in 2003 who remain in the sample in 2010.

Table 3 provides insights into trends in interdisciplinary training over the period 1993 to 2010. The percentage of the full biomedical research workforce with training in bioscience declined from 70.1 % in 1993, to 68.3% in 2003, and then further to 58.2% in 2010. Table 3 also shows a modest rise in the fraction with social science training among those with a bachelor's degree or more (from 2.6% in 1993 to 3.7% in 2010) and a slight rise in the fraction with math/computer science/physical sciences training among those with a PhD/MD (from 7.5% in 1993 to 8.1% in 2010). However, an important caveat is that these trends partly reflect changes in the underlying SESTAT sample. (As discussed in the prior section, the 2010 SESTAT survey includes a greater fraction of researchers with Not S&E training than the 2003 survey.) Notably, even if the Not S&E grouping is excluded and the percentages are recalculated (figures not shown), these trends persist. For additional sensitivity testing, Table 3 reports the ratio of those trained in social science relative to bioscience and in math/comp sci/physical sciences relative to those trained in bioscience. These ratios show similar, albeit still modest, trends.

Far more striking are differences by gender. For one, a greater share of women have social science training than men in each year examined. Further, for the bachelor's or more group, Table 3 points to a rise in female biomedical researchers trained in social sciences as a share of all female S&E workers as well as a rise in the ratio of females trained in social science relative to bioscience. By way of example, the latter ratio increased from .04 in 1993 to .10 in 2010. Notably there was little change in this ratio for those with a PhD/MD only. At the same time, there was a considerable increase in the ratio of males trained in math/computer science/physical science relative to bioscience for both the bachelor's or more and PhD/MD groups. For instance, for the bachelor's or more group, it rose from .12 in 1993 to .16 in 2010. To sum up, interdisciplinarity in field of training rose, and it rose differently for women and men.

The analysis in Table 3, as is the case for most research (e.g. Hunt, 2016), focuses on highest degree earned. It is quite possible that individuals who have a highest earned degree in social science, for instance, might have previously earned a life science degree of some kind. We investigate this a bit in Table 4.[20] These figures indicate that a non-negligible percentage of PhDs, 12.4 to 14.8%, depending on the year, have a prior LIFE degree. For those with a bachelor's or more in the social sciences, the figures are lower at 4.1 to 9.2%, depending on the year. We also looked at the interdisciplinary background of those who have LIFE as their highest degree. A relatively small percent, just 1.6 to 3.7%, have prior training in the social sciences.

4.2 Longitudinal Analysis

In this part of the study, we investigate retention of full-time biomedical researchers. We take advantage of the restricted nature of the SESTAT data, which enable us to look at retention of individuals between 2003 and 2010, a 7-year period. An additional advantage of this analysis is that it obviates any concerns previously raised about survey differences in the cross-section

[20] The figures provided in Table 4 are slight underestimates. In the SDR survey in particular, a small percentage of cases (under 3%) do not provide information on prior degrees earned apart from the highest degree.

data. Appendix Table B provides population sizes and Table 5 reports findings on retention rates stratified by gender.

First, we focus our attention on the 2003 BIOMED workforce with a bachelor's degree or more. If we take a very restrictive definition of stayer—those who remained full-time engaged in biomedical research—the figures in Table 5 indicate that 39.4% of the 2003 biomedical workforce were retained. There is also a distinct gender difference: 35% of women were retained compared to 42% of men. Similar to patterns identified in prior research for other S&E sectors, a much larger fraction of women than men took a position outside of S&E (13.2% versus 10.7%) or were not employed (10% versus 8.3%). Turning to the PhD/MD group, the overall retention rate was somewhat higher at 43.6%, as might be expected given the greater time investment. Notably, PhD women were retained at a *higher* rate than PhD men (46.6% versus 42.2%). In a latter section, we investigate whether gender differences in retention are maintained after adjusting for controls.

Table 6 puts retention rates for BIOMED in the context of retention rates for the broader biomedical occupation (not restricted to research and/or full-time) and for other S&E occupations. This comparison is instructive in that it shows that retention rates are substantially lower for those in a biomedical occupation, whether narrowly or more broadly defined. For instance, for those with a bachelor's degree or more, retention rates are 39.4 to 43.2% for the narrower to broader definitions of biomedical as compared to rates of 50.8% for other life sciences, 56% for engineers, and rates as high as 66.4% for those in S&E-related occupations.

For the PhD groups as well, retention rates are substantially lower for biomedical (narrowly and broadly defined) as compared to other S&E occupations.

Table 6 also provides comparisons of S&E occupational retention rates by gender. These figures show that among the PhD group, retention rates in biomedical are higher for women relative to men even when the broader definition of the occupation is used. This pattern is not universal for all STEM disciplines in the PhD group: retention rates for female PhD engineers are substantially below rates for their male counterparts. This difference, although just one example, points to the importance of studying specific S&E occupations rather than treating them as group.

The remainder of the paper returns to the focus on the biomedical workforce. Table 7 provides key demographic characteristics of BIOMED stayers (those who remain full-time in biomedical research between 2003 and 2010) and movers (the remainder) for both the full educational group and the PhD/MD subgroup. Among the patterns, as would be expected, those who are US-born are more likely to move out of BIOMED, presumably because they have the greatest flexibility (especially compared to those on temporary visas) to change sectors and occupations.

Table 8 provides insight into the disciplinary training of stayers. Not surprisingly, the sector continues to be dominated by those who earned their highest degree in bioscience (71%) or in the broader field of life science (78%). However, what is most interesting is that we again see a gender difference in interdisciplinary training. Among stayers with a bachelor's degree or more, a larger fraction of women (7.4%) have social science training as compared to their male counterparts (1.8%). This gender difference in interdisciplinary training is not present among the PhD/MD group. In fact, a somewhat greater fraction of men hold terminal social science degrees (2.7%) compared to women (1.8%), though both figures are quite low. In contrast, a much larger share of men have math/computer science/physical science training, a pattern that holds for both the bachelor's or more and narrower PhD/MD group. For instance, of those with a bachelor's or more, 5.7% of men have math/computer/physical science training as compared with just 2.7% of women (the comparable figures for the PhD/MD group are 6.1% and 2.6%).

4.3 Linear Probability Models of Retention

We estimate simple linear probability models to examine correlates with retention while controlling for other factors.[21] For those employed as full-time researchers in biomedical (BIOMED) in 2003, we investigate retention outcomes for 7 years later (in 2010): (1) remain in BIOMED (full-time researcher) or not; (2) remain in a biomedical occupation broadly defined (part- or full-time, research or non-research) or not; and (3) NOT EMPLOYED versus the alternative of remaining employed in any other sector, including biomedical science.

The correlates are gender (female), race/ethnicity dummies (with non-white Hispanic as the omitted category), highest degree obtained dummies (with terminal Bachelor's degree as the omitted category), field-of-study dummies (same set as in Table 3, with bioscience as the omitted category), immigrant status (with US-born as the omitted category), a dummy for foreign-born

[21] An advantage of using this specification in estimating descriptive regressions is that the coefficients provide direct magnitudes of the relationships estimated.

and foreign-educated, sector of employment (with academia as the omitted category), potential experience (years since highest degree) and potential experience squared, age, and family variables indicating married and presence of child under age 6. All correlates are measured as of 2003 except for age and potential experience. The inclusion of age along with experience is important to explicitly control for possible cohort effects. The correlates chosen are based on a review of the literature cited earlier.[22]

Table 9 presents OLS results pooled for women and men for the bachelor's degree or more and PhD/MD only groups. These models include a simple gender dummy (female) plus interactions of gender with married and presence of children under age 6. The gender dummy is statistically insignificant in the retention models for the bachelor's or more group. However, this pooled specification does not allow (most of) the covariates to differ by gender, thereby masking potentially statistically significant gender differences. For the PhD/MD group, the coefficient on the gender dummy is positive and statistically significant (10% level) in one retention model--the "remain in biomedical" specification, broadly consistent with the earlier descriptive results. However, the same aforementioned limitation applies; this model largely assumes retention behavior is the same for women and men.

Next, Table 10 presents results from retention models estimated separately for women and men for the bachelor's or more group, thereby allowing all coefficients and the intercept to differ by gender. This specification is analytically equivalent (in terms of coefficients and

[22] The field of "Not S&E" is not included in the models because no in with this field in the 2003 BIOMED sector was observed in the sample in 2010.

magnitudes) to estimating a fully interactive dummy variable model (a model with a gender dummy, other explanatory variables, plus the dummy multiplied by each explanatory variable).[23] The specification in Table 10 is presented because it has fewer coefficients. However, the advantage of the interactive model is that the coefficients provide direct information on the statistical significance of gender differences; the text discussion incorporates this additional information.

In the models presented in Table 10, we find a significant gender differences in retention in the biomedical occupation, consistent with findings for other STEM fields. We reach this conclusion using a Chow test; we reject the null hypothesis that men's and women's retention behavior is the same at the 6% and 5% significance levels, respectively, in both BIOMED (full-time research) and the broader biomedical occupation. For the PhD/MD subgroup, we reject this same null in both retention models (5% level or better).

Next we turn to field of training and retention. Both Tables 9 and 10 include a set of dummies with bioscience (within-field training) as the omitted category. One would expect that retention would be strongest (or at least not weaker) for those trained within the same field (bioscence) because of the intellectual connection between education and occupation and the longer period over which professional networks might develop. Indeed, Table 9 shows that nonbioscence training is never positively associated with retention.[24] We see that retention rates

[23] These results plus the PhD/MD analyses mentioned in the text are available upon request.
[24] For sensitivity testing, we also estimated a simple specification with a single dummy of trained in bioscience versus nont-trained in bioscience. Retention in the full-time biomedical workforce

are significantly lower for those trained in math/computer sci/physical sci, engineering, and SE-related (and find no significant difference for social science and ag and environmental life sciences).

Table 10 further probes the relationship between field of training and retention by gender. Retention in BIOMED is negatively and significantly related to social science training for women (18 percentage point lower retention rate relative to within-field training), but for men there is no significant difference. For the other interdisciplinary field--math/computer science/physical sciences, we again see the same gender pattern, with women less likely to be retained with this type of training. In both cases, the gender difference in retention by field-of-training is statistically significant. It is worth noting that this pattern of findings (including a significant gender difference in retention) is obtained for ag and environmental life science training too. For men, retention is significantly lower if trained in engineering relative to bioscience. For this field, however, additional analysis indicates that the gender difference in retention is not statistically significant.

Finally, we turn to other findings of potential interest. Contrary to most, but not all other work reviewed, family variables are never found to be statistically significant, not even for women. For the full sample of those with bachelor's or more and for women, there is a statistically significant positive relationship between staying in BIOMED and holding a PhD. Also, across virtually all models, there is a significant relationship between retention and years

---

is positively and significantly correlated with within-field training, consistent with the results found in the more nuanced field-of-training specifications presented in Tables 9 and 10.

since degree (although at a decreasing rate). Another strong finding is that retention is almost always positive and statistically related to holding a position in government or business (relative to academia). Greater exits from academia are to be expected to the extent that many positions within this sector are non-permanent (such as postdocs). Some models point to lower rates of retention in the biomedical occupation for those who are not US-born, but these results are not robust across models. Finally, there is some evidence, again not robust, of a statistically significant relationship between minority status and exits from BIOMED.

The SESTAT data also include variables that reflect a respondent's valuation (e.g. very important, not important) regarding a set of job attributes including advancement, benefits, salary, location, challenge, responsibility, independence, and contribution to society. These are the same variables Hunt (2016) incorporated into her study. We do not find any evidence that individuals' valuation of these measured job attributes is significantly correlated with retention in BIOMED (results not reported here).

5. Discussion and Conclusions

This study expands current knowledge of the biomedical workforce by employing an occupation-based definition, which permits a look back at field of training, as well as a look forward at retention within the occupation itself. The data analyzed are restricted-use SESTAT data for 1993, 2003, and 2010. While the vast majority—around 70%-—of the full-time research biomedical workforce (bachelor's degree or more) are trained in the same field or the slightly broader field of life sciences, the remaining 30% received their highest degree in a different field. There is also evidence that the occupation has grown slightly more interdisciplinary since the mid 1990s as reflected by the decline in the share of those holding bioscience (or life science) degrees. From 1993 to 2010, we see a rise in the share of those with social science degrees, especially among women with a bachelor's degree or more, and a rise in the share of those with math/computer science/physical science degrees for both those with a bachelor's or more and the narrower PhD/MD group.

For full-time biomedical researchers in 2003, we investigate retention seven years later, in 2010. In the descriptive results, for the bachelor's or more group, we find that women are retained at lower rates than men, but these results are not maintained in simple models when we adjust for other factors and include a gender dummy. This does not mean that gender does not matter to retention. We next estimate models stratified by gender and, indeed, find evidence of an overall statistically significant gender difference in retention behavior. We undertake the same set of analyses for the narrower PhD/MD subgroup. In the descriptive results we find that women with PhDs are retained at higher rates than their male counterparts. Once we control for other factors, this finding is maintained if we define retention broadly to include retained in part-time and/or non-research biomedical positions, though not if we use a stricter definition of retained in a full-time research biomedical position. We then estimate separate models for women and men. As with the bachelor's or more group, we again identify a statistically significant overall gender difference in retention behavior. We would argue that a particular advantage of stratifying the analysis by gender is that this approach reveals statistically significant factors associated with retention that vary by gender, such as field-of-training, that would not be captured in a simpler pooled model (e.g. Stephan and Levin 2005).

A particularly interesting finding, one which is new to the STEM literature, is that retention rates are (considerably) lower in the biomedical field as a whole and lowest among biomedical researchers, as compared to other life science occupations, engineering, and other S&E occupations. These comparatively low retention rates for the biomedical workforce should be of substantial policy interest to NIH given that it is a major funder of the biomedical workforce and has expressed explicit concerns about retention (NIH, 2012).

Low retention rates in the biomedical occupation raise the question as to what specific factor(s) are significantly connected with exits from this occupation. In the regression analysis, we find that women (but not men) are significantly less likely to be retained if they are trained in the social sciences, and, moreover, the gender difference is statistically significant. We obtain this same set of results for math/computer science/physical science. These findings, taken together, suggest that interdisciplinary training works against retention of women in the biomedical occupation. The explanation cannot simply be that interdisciplinary training is more valuable in other S&E sectors because we do not see this pattern for men. For completeness, we would also point out that we found this same pattern of results (significantly low retention rates for women and a significant gender difference) regarding training in the field of agricultural and environmental life sciences.

The results of this study have broader implications. Interdisciplinary research has been regarded as the next frontier for research innovations and discoveries. To the extent that we seek a biomedical workforce with a more diverse set of skills, low retention rates for those with interdisciplinary training pose a potentially serious impediment to achieving this goal. We see several directions for future research. One direction is to better understand how interdisciplinary teams are formed and what sort of steps might be taken to retain effective interdisciplinary team members. Another direction for further study is to look at salaries and investigate the extent to which differential monetary rewards affect retention rates and gender differences in these rates.

References

Alberts, Bruce, Marc W. Kirschner, Shirley Tilghman, and Harold Varmus. 2014. "Rescuing US Biomedical Research From its Systemic Flaws." *PNAS* 111 (16): 5773-5777. [April 22] http://www.pnas.org/content/111/16/5773.full.pdf

Blau, David M. and Bruce A. Weinberg. 2017. "Why the US Science and Engineering Workforce is Aging Rapidly." *PNAS* Early Edition.

Daniels, Ronald J. 2015. "A Generation at Risk: Young Investigators and the Future of the Biomedical Workforce." *Proc Natl Acad Sci* 112 (2):313–318.

Gibbs, Kenneth D. Jr, John McGready, Jessica C. Bennett, and Kimberly Griffin. 2014. "Biomedical Science Ph.D. Career Interest Patterns by Race/Ethnicity and Gender." PLoS ONE 9(12): 1-18.

Ginther, Donna K. and Shulamit Kahn. 2018. "The Engineering Education-Workforce Continuum." In *National Academy of Engineering: Understanding the Pathways of Engineering*. Washington: DC: National Academies Press, pp. 129-148.

Ginther, Donna K. and Joshua L. Rosenbloom. 2018. "Why Do Women Leave Computer Science and Information Technology Jobs?" Working Paper.

Heggeness, Misty L, Kearney T. Gunsalus, Jose Pacas, and Gary S. McDowell. 2016. "Preparing for the 21st Century Biomedical Research Job Market: Using Census Data to Inform Policy and Career Decision-Making." Unpublished paper.

Heggeness, Misty, Kearney T. Gunsalus, Jose Pacas and Gary S. McDowell. 2017. "The New Face of US Science." *Nature* 541: 21-23.

Hunt, Jennifer. 2016. "Why Do Women Leave Science and Engineering?" *ILR Review* 69 (1): 199-226.

Ginther, Donna K. and Shulamit Kahn. 2009. "Does Science Promote Women? Evidence from Academia 1973-2001." *Science and Engineering Careers in the United States*. Richard B. Freeman and Daniel F. Goroff (eds), Chicago, IL: University of Chicago.

Johnson, Judith A. and Kavya Seka, 2018. "NIH Funding FYI1994-FY 2019," Congressional Research Service 7-5700. Updated October 15, 2018.

Kahn, Shulamit and Donna K. Ginther. 2015. "Are Recent Cohorts of Women with Engineering Bachelors Less Likely to Stay in Engineering?" *Frontiers in Psychology* 6 : 1-15.

Kahn, Shulamit and Donna K. Ginther. 2017. "The Impact of Postdoctoral Training on Early Careers in Biomedicine." *Nature Biotechnology* 35: 90-04. https://www.nature.com/articles/nbt.3766.pdf

Mason, Julie L., Elizabeth Johnston, Sam Berndt, Katie Segal, Ming Lei, and Monathan S. Wiest. 2017. "Labor and Skills Gap Analysis of the Biomedical Research Workforce," *FASEB Journal*.

National Academies of Science. 2003. *Improving the Design of the Scientists and Engineers Statistical Data System (SESTAT)*. Washington, D.C: National Academies Press Chapter 2.

National Academies of Science. 2015. *Enhancing the Effectiveness of Team Science*. Washington, DC: National Academies Press.

National Institutes of Health. 2012. *Biomedical Medical Research Workforce Working Group Report* (NIH, Bethesda, MD). https://acd.od.nih.gov/documents/reports/Biomedical_research_wgreport.pdf [Source of definition of biomedical].

NIH (2019). "The Biomedical Research Workforce." https://researchtraining.nih.gov/dbrw/biomedical-research-workforce (accessed 5-1-19).

National Institutes of Health. 2017. "NIH Awards to Test Ways to Store, Access, Share, and Compute on Biomedical Data in the Cloud." *News Release* (November 6). https://www.nih.gov/news-events/news-releases/nih-awards-test-ways-store-access-share-compute-biomedical-data-cloud

National Research Council. 2011. *Research Training in the Biomedical, Behavioral, and Clinical Research Sciences.* Washington, DC: The National Academies Press. http://www.ncbi.nlm.nih.gov/books/NBK56994/

National Science Board. 2016. *Science and Engineering Indicators 2016.* Chapter 3, Arlington, VA: National Science Foundation.

National Science Board. 2018. *Science and Engineering Indicators 2018.* Chapter 5, "Academic Research and Develoment." Arlington, VA: National Science Foundation.

National Science Foundation. 2015. *Annual Survey, Graduate Students and Postdoctorates in S&E* (Arlington, VA).

National Science Foundation, National Center for Science and Engineering Statistics, *Scientists and Engineers Statistical Data System (SESTAT)*. 1993 and 2010.

National Science Foundation. n.d. “2000 Decade Redesign of NSF’s Science and Engineering Workforce Surveys” (unpublished).

Preston, Anne E. *Leaving Science: Occupational Exit from Scientific Careers*, NY: Russell Sage, 2004.

Stephan, Paula E. and Levin, Sharon G. 2005. “Leaving Careers in IT: Gender Differences in Retention.” *Journal of Technology Transfer* 30 (2005): 383-96.

Stephan, Paula E. 2012. *How Economics Shapes Science* (Harvard University Press, Cambridge, MA).

Wang, Jian, Bart Thijs, and Wolfgang Glänzel, 2015. Interdisciplinarity and Impact: Distinct *Effects of Variety, Balance, and Disparity* PlosOne (22 May).

Xie, Yu and Kimberlee A. Shauman 2003. *Women in Science: Career Processes* (Cambridge: Harvard University Press).

Xu, Yonghong Jade, 2008. “Gender Disparity in STEM Disciplines: A Study of Faculty Attrition and Turnover Intentions,” *Research in Higher Education*, 49 (7) :607-624.

Figure 1. Size of Biomedical Workforce using Various Occupational Definitions (Bachelor's Degree or More, in thousands)

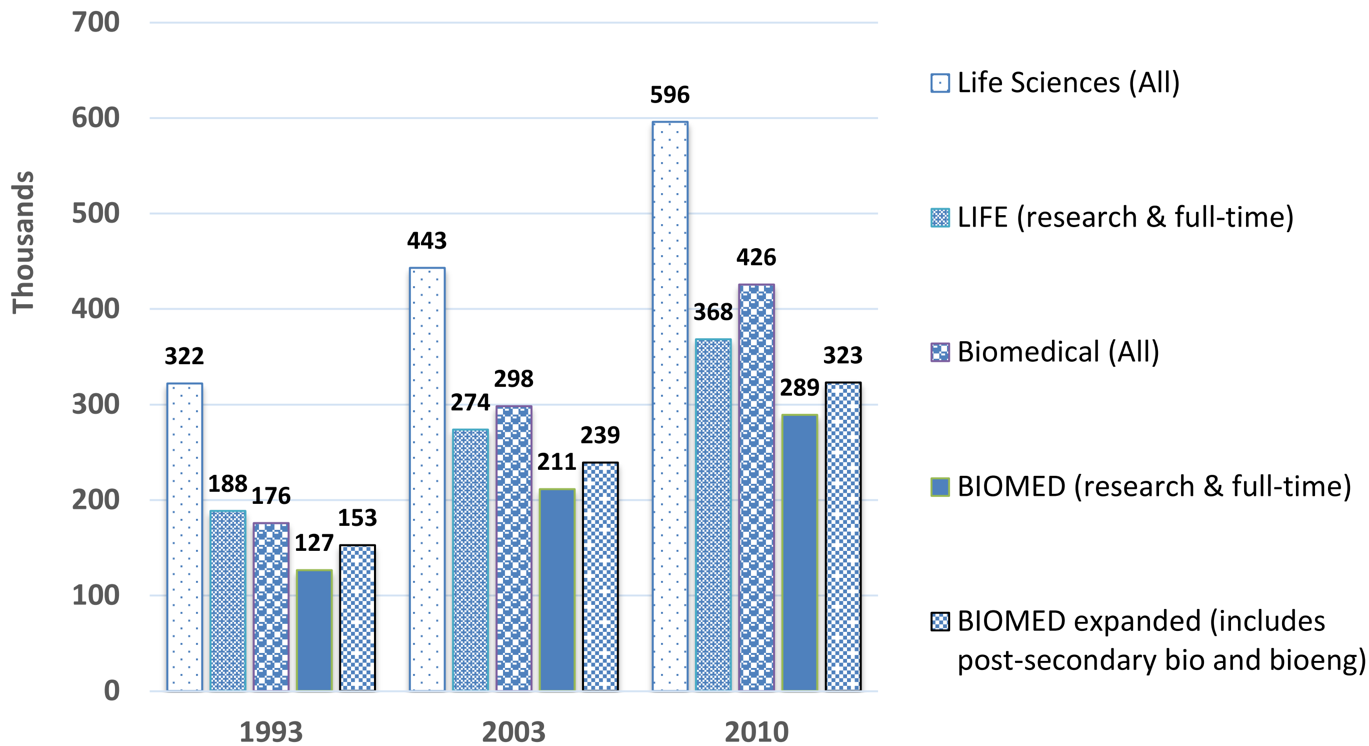

Table 1. Population Size of the Full-Time Life Science Research (LIFE) Workforce and Full-Time Biomedical Research (BIOMED) Workforce

	Bachelor's Degree or More			PhD/MD only		
	1993	**2003**	**2010**	**1993**	**2003**	**2010**
Life Sciences (LIFE)	188,474	273,643	368,312	90,714	127,618	159,306
Biomedical Science (BIOMED)	126,605	211,439	289,147	56,151	97,300	118,463
Biochemists & Biophysicists	29,686	35,011	47,713	14,267	19,258	21,656
Biological Scientists	39,380	67,431	84,370	13,192	22,170	25,711
Medical Scientists (excluding practitioners)	42,539	84,845	92,935	24,703	48,255	59,482
Other	14,999	24,152	64,129	3,989	7,616	11,614
Other Life	61,870	62,204	79,165	34,563	30,318	40,843
PostSecondary Teachers (Ag, Bio, Other Natural Sci)	34,527	28,735	35,675	27,686	22,242	32,348
PostSecondary -- Bioscience	21,750	19,179	24,864	17,545	15,681	23,167
Agricultural & Food Scientists	21,598	25,631	30,404	6,349	6,654	7,266
Forestry & Conservation	5,744	7,837	13,086	528	1,422	1,229
Bioengineers	4,445	8,668	8,890	1,080	2,701	3,112
BIOMED expanded (BIOMED + PostSecondary--Bioscience + Bioengineer)	152,800	239,286	322,901	74,776	115,682	144,742

Notes:

In SESTAT, Life Sciences is code 2 and Biomedical is code 22.

All individuals in this table are employed full time and engaged in research.

Educational level refers to highest degree attained.

Population figures based on SESTAT weights.

Table 2. Demographic Characteristics of BIOMED, 1993, 2003, 2010

	Bachelor's Degree or More			PhD/MD only		
	1993	**2003**	**2010**	**1993**	**2003**	**2010**
	(%)	(%)	(%)	(%)	(%)	(%)
PhD/MD	44.4	46.0	41.0	100.0	100.0	100.0
Female	39.6	44.5	50.3	30.0	34.4	42.1
Race/Ethnicity						
White, Non-Hispanic	77.0	68.8	65.9	74.8	62.0	59.1
Asian	na	22.7	24.6	na	30.4	32.4
Other	na	8.5	9.5	na	7.5	8.5
Employment Sector						
Academia/Research Inst.	45.7	50.8	41.3	55.9	58.0	51.1
Government	17.1	16.9	20.9	12.4	13.7	14.2
Business/Industry	37.2	32.3	37.7	31.8	28.4	34.8
Citizenship						
US-Born	76.5	67.4	63.2	64.6	50.1	42.8
Naturalized	8.6	13.6	12.1	11.9	19.1	17.4
Permanent Resident	9.0	10.4	11.9	14.8	17.5	22.4
Temporary	5.8	8.5	12.8	8.7	13.4	17.4
Total Population	126,605	211,439	289,146	56,151	97,300	118,463

Notes:

Race variables are defined differently in 1993; hence na for other groups.

Population figures obtained using SESTAT weights.

Table 3. Disciplinary Training (Highest Level) of BIOMED, by Gender, 1993, 2003, 2010

Panel A:
Bachelor's Degree or More

	1993			2003			2010		
	Female	**Male**	**Total**	**Female**	**Male**	**Total**	**Female**	**Male**	**Total**
Educational Training (% Distribution)									
Life Sciences	71.7	76.0	74.3	71.9	73.2	72.6	57.6	69.5	63.5
Biological Science	67.0	72.1	70.1	66.9	69.4	68.3	54.4	62.1	58.2
Ag & Env Life Sciences	4.7	3.8	4.2	5.1	3.9	4.4	3.2	7.4	5.3
Math, Comp Sci, Phys Sci	9.0	8.8	8.9	7.8	7.3	7.5	5.9	10.2	8.0
Social Science and Related	2.8	2.5	2.6	5.7	2.0	3.6	5.7	1.7	3.7
Engineering	0.7	1.2	1.0	0.4	1.2	0.8	2.2	3.1	2.7
S&E-Related Fields	na	na	na	11.4	12.2	11.9	12.9	11.7	12.3
Not S&E*	15.8	11.4	13.2	2.8	4.1	3.5	15.7	3.7	9.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Total Population	50,155	76,450	126,605	94,134	117,305	211,439	145,553	143,593	289,147
Educational Ratios									
Social Science/Biosci	0.04	0.04	0.04	0.08	0.03	0.05	0.10	0.03	0.06
(Math, Comp Sci, Phys Sci)/Biosci	0.13	0.12	0.13	0.12	0.10	0.11	0.11	0.16	0.14

(TABLE CONTINUES ON NEXT PAGE)

Table 3. ***Continued***

Panel B: PhD/MD only	**1993**			**2003**			**2010**		
	Female	**Male**	**Total**	**Female**	**Male**	**Total**	**Female**	**Male**	**Total**
Educational Training (% Distribution)									
Life	76.8	78.9	78.3	73.4	70.7	71.6	68.5	67.1	67.7
Biological Science	73.7	75.4	74.9	71.7	68.5	69.6	67.1	64.5	65.6
Ag & Env Life Sciences	3.2	3.5	3.4	1.7	2.2	2.0	1.5	2.7	2.2
Math, Comp Sci, Phys Sci	5.8	8.2	7.5	4.6	6.7	6.0	4.0	11.0	8.1
Social Science and Related	2.2	2.6	2.5	2.7	1.9	2.2	2.3	1.9	2.1
Engineering	0.5	0.9	0.8	0.6	1.1	0.9	2.5	2.6	2.6
S&E Related Fields	na	na	na	18.7	18.1	18.3	21.4	16.8	18.7
Not S&E*	14.7	9.4	11.0	0.0	1.5	1.0	1.1	0.6	0.8
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Total Population	16,855	39,296	56,151	33,500	63,800	97,300	49,856	68,591	118,447
Educational Ratios									
Social Science/Biosci	0.03	0.04	0.03	0.04	0.03	0.03	0.03	0.03	0.03
(Math, Comp Sci, Phys Sci)/Biosci	0.08	0.11	0.10	0.06	0.10	0.09	0.06	0.17	0.12

Notes:

*Not S&E includes Management, Education (not STEM), Social Service, Arts & Humanities and Other Not S&E.

Population figures based on SESTAT weights.

Table 4. Interdisciplinarity of Training of BIOMED, 2003 and 2010

	Percent with Soc Science as HIGHEST Degree who have a prior LIFE Science degree (%)	Percent with LIFE as HIGHEST Degree who have a prior Social Science degree (%)
2003		
All	4.1	3.1
PhD	14.8	1.6
2010		
All	9.2	3.7
PhD	12.4	3.4

Notes:

All figures are weighted to population using SESTAT weights

The denominators for each calculation are in Table 3.

LIFE is defined as sum of Biomedical plus Other Life.

Population Figures based on SESTAT weights.

Table 5. Longitudinal Analysis. BIOMED in 2003: Where are they in 2010?

	Bachelor's Degree or More			PhD/MD Only		
	All	Female	Male	All	Female	Male
OCCUPATIONS in 2010 (% Distribution)						
Still in Biomedical Field	47.7	45.9	48.9	49.6	54.3	47.3
Biomed Researchers Full-Time (STAYERS)	39.4	35.2	42.2	43.6	46.6	42.1
Biomedical Not Research Full-Time	6.4	7.9	5.4	4.1	4.7	3.8
Biomedical Not Research Part-Time	0.3	0.6	0.1	0.2	0.2	0.2
Biomedical Research Part-Time	1.6	2.2	1.2	1.7	2.8	1.2
Other Life Sciences (excluding Biomedical)	9.6	6.1	12.0	12.0	10.4	12.8
Other S&E (includes Biomed Engineering)	5.5	7.9	4.0	4.1	3.6	4.3
S&E Related	16.5	16.9	16.2	18.8	17.3	19.5
Not S&E	11.8	13.2	10.7	9.3	8.1	9.9
Not Employed	9.0	10.0	8.3	6.3	6.3	6.2
Unemployed	3.6	2.4	4.4	2.5	2.4	2.5
Not In Labor Force	5.4	7.6	3.9	3.8	3.9	3.7
Total	100.0	100.0	100.0	100.0	100.0	100.0
TOTAL Population	97,561	39,714	57,847	58,290	19,287	39,002

Notes: Stayers are those who remain full-time researchers in the biomedical field. All others still in biomedical are described as Movers.
Figures weighted using SESTAT weights.
Population figures based on SESTAT weights.

Table 6. Longitudinal Analysis. Comparison of 7-Year Retention Rates for Various Occupational Categories

	Bachelor's or More			PhD/MD Only		
	All	**Female**	**Male**	**All**	**Female**	**Male**
Focal Occupation	(%)	(%)	(%)	(%)	(%)	(%)
Biomedical, Researcher and Full-Time	39.4	35.2	42.2	43.6	46.6	42.1
Comparison Occupations						
Biomedical (not restricted to researcher or full-time)	43.2	41.6	44.5	48.3	52.7	46.0
Other Life Sciences (excluding Biomedical)	50.8	50.0	56.1	65.4	67.4	64.7
Other S&E Occupations	61.5	56.1	61.7	72.5	74.7	71.7
Math & Computer Science	59.2	51.0	61.9	66.4	69.6	65.7
Physical Science	56.3	54.0	57.1	64.9	61.5	65.4
Social Science	52.5	54.2	50.7	72.1	76.5	68.5
Engineers	56.6	57.4	56.6	66.8	57.5	67.8
S&E Related Occupations	66.4	71.1	60.3	81.0	80.4	81.4

Notes: Biomedical, Researcher and Full-Time is identical to STAYERS in Table 5.

Population figures are based on SESTAT weights.

Table 7. Longitudinal Analysis: Differences in Key Demographic Characteristics of Stayers and Movers, BIOMED

	Bachelor's Degree or More			PhD/MD Only		
	BIOMED in 2003 and ANY SECTOR, 2010 (ALL)	**BIOMED 2003 AND 2010 (STAYERS)**	**BIOMED 2003, but not 2010 (MOVERS)**	**BIOMED in 2003 and ANY SECTOR, 2010 (ALL)**	**BIOMED 2003 AND 2010 (STAYERS)**	**BIOMED 2003, but not 2010 (MOVERS)**
Key Demographic Characteristics (%)						
Highest Degree						
PhD/ MD	59.8	66.2	55.6	na	na	na
Less than PhD/MD	40.3	33.8	44.4	na	na	na
Gender						
Female	40.7	36.4	43.5	33.1	35.4	31.3
Male	59.3	63.6	56.5	66.9	64.6	68.7
Race/Ethnicity						
White, non-Hispanic	67.9	62.6	71.3	65.5	61.5	68.6
Asian	25.3	30.2	22.5	28.2	31.7	25.4
Other	6.6	7.2	6.2	6.3	6.8	6.0
Citizenship Status, as of 2003						
US-Born	63.5	55.8	68.5	52.3	46.4	56.9
Naturalized	16.8	19.4	15.1	20.3	23.8	17.5
Permanent Resident	13.2	15.4	11.8	17.9	18.1	17.8
Temporary	6.3	9.5	4.7	9.5	11.7	7.8
Total Population	97,561	38,398	59,163	58,290	25,406	32,884

Notes:

Stayers and Movers defined in Table 5.

Population figures based on SESTAT weights.

Table 8. Longitudinal Analysis. Female and Male BIOMED Stayers, by Field of Training

Panel A: Bachelor's Degree or More

	All	**Females**	**Males**
Field of Training (% Distribution)	**%**	**%**	**%**
Life Science	77.9	74.2	80.0
Biological Science	70.8	70.1	71.1
Ag, Environmental ,Life	7.1	4.1	8.8
Math, Comp Sci, Phys Sci	4.6	2.7	5.7
Social Science and Related	3.8	7.4	1.8
Engineering	0.4	0.3	0.4
S&E Related	11.8	14.7	10.2
Not S&E	1.5	0.7	1.9
Total	100.0	100.0	100.0
Total Population	38,398	13,977	24,421

Panel B: PhD/MD Only

	All	**Males**	**Females**
Field of Training (% Distribution)	**%**	**%**	**%**
Life Science	76.9	79.6	75.4
Biological science	74.0	77.7	72.0
Ag, Environmental Life	2.9	1.8	3.4
Math, Comp Sci, Phys Sci	4.9	2.6	6.1
Social Science and Related	2.4	1.8	2.7
Engineering	0.5	0.5	0.6
S&E Related	14.7	15.5	14.2
Not S&E	0.6	0.0	1.0
Total	100.0	100.0	100.0
Total Population	25,407	8,990	16,417

Notes: Stayers defined in Table 6.

TPopulation figures based on SESTAT weights.

Table 9. Linear Probability Models: Occupation in 2010 for those Employed in BIOMED in 2003

	Bachelor's Degree or More						PhD/MD Only					
Dependent Variable:	**BIOMED in 2010**		**Bio-medical in 2010**		**NOT EMP in 2010**		**BIO-MED in 2010**		**Bio-medical in 2010**		**NOT EMP in 2010**	
PHD	0.108	**	0.036		-0.047	*						
	(0.054)		(0.055)		(0.028)							
MA	0.109		0.074		-0.012							
	(0.071)		(0.073)		(0.037)							
PROF	0.162		0.037		-0.131	**						
	(0.108)		(0.11)		(0.056)							
FEMALE	-0.011		0.039		0.013		0.018		0.077	*	0.007	
	(0.043)		(0.044)		(0.022)		(0.046)		(0.047)		(0.023)	
ASIAN	0.043		0.043		0.02		0.036		0.033		0.019	
	(0.033)		(0.034)		(0.017)		(0.035)		(0.035)		(0.017)	
MINORITY	-0.058		-0.074	**	-0.016		-0.05		-0.076	*	-0.007	
	(0.036)		(0.037)		(0.019)		(0.038)		(0.039)		(0.019)	
AGE	-0.002		-0.001		0.006	***	-0.001		0		0.007	***
	(0.002)		(0.003)		(0.001)		(0.003)		(0.003)		(0.001)	
POT EXP	0.018	***	0.016	***	-0.026	***	0.016	***	0.008		-0.023	***
	(0.006)		(0.006)		(0.003)		(0.007)		(0.007)		(0.003)	
POT EXP SQ/100	-0.042	***	-0.035	***	0.058	***	-0.037	***	-0.021		0.052	***
	(0.012)		(0.012)		(0.006)		(0.013)		(0.013)		(0.007)	
MARRIED	-0.01		0.002		-0.02		0.006		0.031		-0.028	
	(0.037)		(0.037)		(0.019)		(0.039)		(0.039)		(0.019)	
FEMALE * MARRIED	0.041		-0.013		0.032		0.026		-0.035		0.044	
	(0.053)		(0.054)		(0.027)		(0.056)		(0.057)		(0.028)	
CHILD < 6	0.029		0.009		0.003		0.035		0.012		0.006	
	(0.035)		(0.036)		(0.018)		(0.036)		(0.037)		(0.018)	
FEMALE * CHILD < 6	-0.082		-0.061		-0.013		-0.08		-0.061		-0.034	
	(0.056)		(0.058)		(0.029)		(0.06)		(0.061)		(0.03)	
AG & ENV LIFE SCIENCES	-0.028		-0.048		-0.013		-0.012		-0.062		-0.006	
	(0.056)		(0.057)		(0.029)		(0.064)		(0.065)		(0.032)	
MATH, COMP, PHYS SCI	-0.149	***	-0.104	***	-0.011		-0.143	***	-0.101	**	-0.014	
	(0.041)		(0.042)		(0.021)		(0.045)		(0.045)		(0.022)	
SOCIAL SCI	-0.027		-0.054		-0.033		-0.039		-0.064		-0.03	
	(0.064)		(0.065)		(0.033)		(0.066)		(0.067)		(0.033)	
ENGINEERING	-0.18	*	-0.218	**	-0.048		-0.166	*	-0.199	**	-0.036	
	(0.092)		(0.094)		(0.048)		(0.098)		(0.1)		(0.049)	

(TABLE CONTINUES ON NEXT PAGE)

Table 9. *Continued*

Dependent variable:	Bachelor's Degree or More: BIOMED in 2010		Bachelor's Degree or More: Bio-medical in 2010		Bachelor's Degree or More: NOT EMP in 2010		PhD/MD Only: BIO-MED in 2010		PhD/MD Only: Bio-medical in 2010		PhD/MD Only: NOT EMP in 2010	
SE-RELATED	-0.142	***	-0.104	**	-0.014		-0.131	***	-0.107	***	-0.027	
	(0.042)		(0.043)		(0.022)		(0.041)		(0.042)		(0.021)	
PERMANENT	-0.059		-0.06		-0.015		-0.063		-0.055		-0.017	
	(0.044)		(0.045)		(0.023)		(0.045)		(0.046)		(0.023)	
TEMP STATUS	0.046		0.095		-0.064	*	0.034		0.092		-0.059	**
	(0.053)		(0.054)		(0.028)		(0.055)		(0.056)		(0.028)	
NATURALIZED	0.065	*	0.061		-0.006		0.071	*	0.072	*	-0.01	
	(0.038)		(0.039)		(0.02)		(0.04)		(0.041)		(0.02)	
FOREIGN BORN & FOREIGN ED	0.097	**	0.068		0.034		0.114	***	0.072		0.032	
	(0.044)		(0.045)		(0.023)		(0.046)		(0.047)		(0.023)	
GOVERNMENT	0.137	***	0.177	***	-0.016		0.129	***	0.172	***	-0.02	
	(0.032)		(0.033)		(0.017)		(0.034)		(0.035)		(0.017)	
BUSINESS	0.05	***	0.08	***	0.016		0.059	***	0.096	***	0.024	*
	(0.025)		(0.025)		(0.013)		(0.026)		(0.026)		(0.013)	
Intercept	0.205	***	0.311	***	0.037		0.259	***	0.334	***	-0.037	
	(0.102)		(0.104)		(0.053)		(0.112)		(0.114)		(0.056)	
Observations	2048		2048		2048		1858		1858		1858	
F-statistic	3.96		3.69		10.55		3.57		3.56		9.55	
R-squared	0.045		0.042		0.114		0.028		0.028		0.088	

Notes:
BIOMED = 1 refers to still employed full-time and doing research in biomedical;
Biomedical = 1 refers to biomedical occupation (part or full-time, research or not research);
NOT EMPLOYED = 1 if unemployed or not in labor force.
Educational fields are same as those in Table 4. "Not trained in S&E" is not included because no observations in sample.
Omitted ed field is bioscience; omitted immigration status is US-born;
Omitted sector is Academic; omitted ed level is Bachelor's.
Regressions are unweighted. Standard errors are beneath coefficients.
*, **, ***, significant at 1, 5, 10% respectively.

Table 10. Linear Probability Models: Occupation in 2010 for those Employed in BIOMED in 2003 , by Gender

Bachelor's Degree or More

	Female						Male					
Dependent variable:	**BIOMED in 2010**		**Biomedical in 2010**		**NOT EMP in 2010**		**BIOMED in 2010**		**Biomedical in 2010**		**NOT EMP in 2010**	
PHD	0.182	**	0.079		-0.061		-0.011		-0.05		-0.04	
	(0.073)		(0.074)		(0.042)		(0.081)		(0.083)		(0.039)	
MA	0.168	*	0.036		-0.023		0.034		0.099		-0.009	
	(0.095)		(0.097)		(0.054)		(0.107)		(0.109)		(0.052)	
PROF	0.264	*	0.145		-0.095		0.029		-0.139		-0.137	*
	(0.162)		(0.165)		(0.093)		(0.149)		(0.152)		(0.072)	
ASIAN	0.018		0.042		0.027		0.07		0.056		0.014	
	(0.05)		(0.052)		(0.029)		(0.044)		(0.045)		(0.021)	
MINORITY	-0.107	**	-0.107	**	-0.009		0.003		-0.023		-0.025	
	(0.051)		(0.052)		(0.029)		(0.052)		(0.053)		(0.025)	
AGE	-0.001		-0.002		0.005	***	-0.002		0.00		0.007	***
	(0.003)		(0.004)		(0.002)		(0.004)		(0.004)		(0.002)	
POT EXP	0.027	***	0.027	***	-0.027	***	0.01		0.005		-0.024	***
	(0.009)		(0.009)		(0.005)		(0.008)		(0.008)		(0.004)	
POT EXP SQ/100	-0.058	***	-0.054	***	0.06	***	-0.027	*	-0.018		0.054	***
	(0.019)		(0.019)		(0.011)		(0.015)		(0.015)		(0.007)	
MARRIED	0.023		-0.025		0.016		0.01		0.022		-0.029	
	(0.038)		(0.039)		(0.022)		(0.038)		(0.038)		(0.018)	
CHILD < 6	-0.062		-0.066		-0.019		0.025		0.011		0.012	
	(0.046)		(0.047)		(0.027)		(0.037)		(0.037)		(0.018)	
AG & ENV LIFE SCI	-0.207	**	-0.245	***	0.054		0.063		0.062		-0.051	
	(0.091)		(0.092)		(0.052)		(0.071)		(0.073)		(0.034)	
MATH, COMP, PHYS SCI	-0.269	***	-0.242	***	0.006		-0.074		-0.021		-0.021	
	(0.07)		(0.071)		(0.04)		(0.051)		(0.052)		(0.025)	
SOCIAL SCI	-0.181	*	-0.196	**	-0.053		0.104		0.061		-0.017	
	(0.094)		(0.096)		(0.054)		(0.087)		(0.088)		(0.042)	

(TABLE CONTINUES ON NEXT PAGE)

Table 10. ***Continued***

	Female						Male					
Dependent variable:	**BIOMED in 2010**		**Biomedical in 2010**		**NOT EMP in 2010**		**BIOMED in 2010**		**Biomedical in 2010**		**NOT EMP in 2010**	
ENGINEERING	-0.035		0.058		-0.069		-0.225	**	-0.298	***	-0.047	
	(0.196)		(0.2)		(0.113)		(0.105)		(0.107)		(0.051)	
SE-RELATED	-0.159	***	-0.183	***	-0.009		-0.138	**	-0.033		-0.014	
	(0.059)		(0.06)		(0.034)		(0.061)		(0.062)		(0.029)	
PERMANENT	-0.044		-0.008		-0.001		-0.085		-0.115	*	-0.024	
	(0.065)		(0.067)		(0.038)		(0.059)		(0.06)		(0.028)	
TEMP STATUS	0.166	**	0.223	***	-0.099	**	-0.031		0.02		-0.049	
	(0.083)		(0.084)		(0.047)		(0.07)		(0.072)		(0.034)	
NATURALIZED	0.123	**	0.098		-0.028		0.024		0.032		0.01	
	(0.06)		(0.062)		(0.035)		(0.05)		(0.05)		(0.024)	
FOREIGN BORN & FOREIGN ED	0.101		0.136	*	-0.015		0.102	*	0.041		0.065	**
	(0.071)		(0.073)		(0.041)		(0.056)		(0.057)		(0.027)	
GOVERNMENT	0.142	***	0.197	***	-0.067	**	0.137	***	0.164	***	0.014	
	(0.05)		(0.051)		(0.029)		(0.042)		(0.043)		(0.02)	
BUSINESS	0.095	**	0.097	**	-0.013		0.03		0.069	**	0.029	*
	(0.041)		(0.041)		(0.023)		(0.031)		(0.032)		(0.015)	
Intercept	0.025		0.241		0.138	*	0.392	***	0.441	***	-0.032	
	(0.145)		(0.148)		(0.083)		(0.141)		(0.143)		(0.068)	
Observations	794		794		794		1254		1254		1254	
F-statistic	3.78		3.85		3.53		2.32		2.43		10.26	
R-squared	0.093		0.095		0.088		0.038		0.023		0.149	

BIOMED = 1 refers to still employed full-time and doing research in biomedical;
Biomedical = 1 refers to biomedical occupation (part or full-time, research or not research);
NOT EMPLOYED = 1 if unemployed or not in labor force.
Educational fields are same as those in Table 4. "Not trained in S&E" is not included because no observations in sample.
Omitted ed field is bioscience; omitted immigration status is US-born,
Omitted sector is Academic; Omitted ed level is Bachelor’s.
Regressions are unweighted.
Standard errors are beneath coefficients.
*, **, ***, significant at 1, 5, 10% respectively.

Appendix A: Description of SESTAT Data Used in Analysis			
SURVEY INCLUDED IN SESTAT	SURVEY YEARS		
	1993	2003	2010
NSCG	First year of biennial survey. Underlying sample is from the 1990 Decennial Census. All individuals living in the US during the survey week who have either a bachelor's S&E degree, an S&E occupation, or both. Includes individuals who earned degrees outside of the United States.	Underlying sample is from the 2000 Decennial Census. All individuals living in the US during the survey week who have either a S&E degree, a S&E occupation, or both. Includes individuals who earned degrees outside of the United States. Key change: Those in S&E-related fields and occupations are re-categorized from NOT S&E to S&E-related. This group is included with "S&E" above.	Includes earlier survey respondents. Major change due to discontinuation of the long form in the 2010 Census. Adds in new individuals from the 2009 American Community Survey. All individuals who are living in the United States during the survey and have either a S&E degree, a S&E occupation, or both. Includes individuals who earned degrees outside of the United States.
NSCRG	Survey conducted every two to three years. 1993 survey used to supplement NSCG. Individuals under the age of 76 who received a Bachelor's or Master's S&E degree from a US academic institution. Individuals are subsequently added to NSCG.	2001 and 2003 surveys are used to supplement NSCG. Individuals under the age of 76 who received a Bachelor's or Master's S&E degree from a US academic institution. Individuals are subsequently added to NSCG. Key change: Those in S&E-related fields and occupations are re-categorized from NOT S&E to S&E--related and included with S&E.	2006, 2008, 2010 surveys are used to supplement NSCG. Individuals under the age of 76 who received aa Bachelor's or Master's S&E degree from a US academic institution. Note: After 2010, NSCRG discontinued given change that ACS now serves as underlying survey for NSCG.
SDR	Biennial survey initiated in the 1970s. Individuals under the age of 76 who earned a doctorate degree in S&E from a US institution. Excludes foreign-earned doctorates.	Individuals under the age of 76 who earned a doctorate degree in S&E from a US institution. Excludes foreign-earned doctorates. Key change: Those in S&E-related fields and occupations are re-categorized from NOT S&E to S&E-related. This group is included with S&E.	Same definition as 2003.

Appendix Table B:
Population Sizes for Narrow and Broader Definitions of Full-time Research Biomedical (BIOMED) and Life Science (LIFE) Workforce

Cross-Sectional and Longitudinal Analyses

BIOMED	**Population**	**LIFE**	**Population**
Cross-Sectional Analysis		*Cross-Sectional Analysis*	
1993, Bachelor's Degree or More	126,605	1993, Bachelor's Degree or More	188,474
MD/PHD only	56,151	MD/PHD only	90,714
2003, Bachelor's Degree or More	211,439	2003, Bachelor's Degree or More	273,643
MD/PhD only	97,300	MD/PhD only	127,618
2010, Bachelor's Degree or More	289,147	2010, Bachelor's Degree or More	368,312
MD/PhD only	118,462	MD/PhD only	159,306
Longitudinal Analysis		*Longitudinal Analysis*	
BIOMED in 2003 and in 2010	97,561	LIFE in 2003 and in 2010	131,957
STAYERS (same occupation, still research, still full-time)	38,398	STAYERS (same occupation, still research, still full-time)	71,607
MOVERS	59,163	MOVERS	60,350

BIOMED is drawn from occupational code 22 (Biomedical) in SESTAT and is restricted here to those doing research and employed full-time.
LIFE is drawn from occupational code 2 (Life Science) in SESTAT and is restricted here to those doing research and employed full-time.
Population figures are obtained using SESTAT weights.