Research Policy 47 (2018) 1007–1017

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

The gender gap in early career transitions in the life sciences☆

Marc J. Lerchenmueller*, Olav Sorenson

Yale School of Management, Yale University, 165 Whitney Avenue, New Haven, CT, 06511, United States

ARTICLE INFO

JEL classification:
O3 Research and Development
J44 Professional Labor Markets
J71 Labor Discrimination

Keywords:
Science careers
Gender gap
Productivity paradox
Differential returns
National Institutes of Health

ABSTRACT

We examined the extent to which and why early career transitions have led to women being underrepresented among faculty in the life sciences. We followed the careers of 6,336 scientists from the post-doctoral fellowship stage to becoming a principal investigator (PI) – a critical transition in the academic life sciences. Using a unique dataset that connects individuals' National Institutes of Health funding histories to their publication records, we found that a large portion of the overall gender gap in the life sciences emerges at this transition. Women become PIs at a 20% lower rate than men. Differences in "productivity" (publication records) can explain about 60% of this differential. The remaining portion appears to stem from gender differences in the returns to similar publication records, with women receiving less credit for their citations.

1. Introduction

Despite a narrowing of the gender gap, women remain underrepresented in the science, technology, engineering, and mathematics (STEM) academic labor force. According to the National Science Foundation, women earn about half of the doctoral degrees in science, yet represent a mere 22% of the faculty at the full professor level at Research I institutions in the United States (NSF, 2015). This continuing gap, in part, reflects the fact that many of today's senior faculty received their degrees thirty or more years ago. But that fact alone cannot account for this gap. Thirty years ago, women already accounted for more than 30% of doctoral degrees earned in the life sciences (Hill et al., 2010).

In attempting to explain this gap, a large body of research has documented that women produce less measurable output than men. Women, for example, publish fewer papers (Cole and Zuckerman, 1984; Long, 1992; Xie and Shauman, 1998), the papers that they publish appear in less prominent journals (Brooks et al., 2014; Lerchenmüller et al., 2018) and receive fewer citations (Larivière et al., 2013; King et al., 2016), and women receive the prestigious first and last authorships on co-authored articles less often (West et al., 2013; Filardo et al., 2016). Although these differences in publication records may themselves stem from factors such as discrimination, disparity in the time spent on childcare, or insufficient mentoring, to the extent that these elements of the research record factor into hiring, promotion, and funding decisions, one would expect fewer women to attain and retain faculty positions. But, even when men and women have equivalent research records, a parallel literature, based primarily on audit studies, suggests that hiring and promotion committees still prefer men over women (Steinpreis et al., 1999; Moss-Racusin et al., 2012).

We extend this literature on the gender gap in STEM faculty by examining the extent to which disparate publication records versus differential returns to similar records account for a critical early career transition in the life sciences, from being a lab member to being a principal investigator (PI). Because researchers in the academic life sciences require substantial resources – equipment and personnel – for their research, acquiring these grants has effectively become a precursor to being viable for tenure at a research-oriented university (Jena et al., 2015).

This shift to analyzing the correlates of a critical career transition – as opposed to identifying cross-sectional differences between men and women in their publication records – forwards our theoretical understanding of the underrepresentation of women in STEM in at least two respects. Most importantly, it examines whether differential publication records could actually account for the gender gap. Most prior studies have not been capable of disentangling cause from effect. The gender gap at the faculty level might arise from women publishing fewer or less prominent papers (Xie and Shauman, 1998). But the direction of

☆ The Initiative on Leadership and Organization at the Yale School of Management provided generous financial support. Marc Lerchenmueller also received funding through a grant from the German Research Foundation (DFG; LE 3426/1-1). We thank Meshna Koren, from Elsevier, for facilitating our access to the Scopus API for extracting comprehensive citation information, and Amandine Ody-Brasier and seminar participants at Copenhagen Business School, the European School of Management and Technology (ESMT), and the Yale School of Management for helpful comments on earlier versions of this paper. The usual disclaimer applies.

* Corresponding author.
E-mail addresses: marc.lerchenmueller@yale.edu (M.J. Lerchenmueller), olav.sorenson@yale.edu (O. Sorenson).

https://doi.org/10.1016/j.respol.2018.02.009
Received 10 February 2017; Received in revised form 6 February 2018; Accepted 13 February 2018
Available online 16 April 2018

causality could run in the reverse direction: Women might have less impressive publication records because they have not had the time and resources for research that come from being senior faculty at research-oriented institutions (Merton, 1968).

Second, our approach allows us to isolate whether – and if so, where – men and women receive differential returns to their publication records. Although audit studies suggest that these differential returns exist (e.g., Moss-Racusin et al., 2012), because those studies, by design, hold constant all elements of the publication record, they cannot determine whether women receive less credit for some specific element of their research portfolios or whether the individuals evaluating applicants simply have a preference for candidates of a particular gender among those with equal qualifications.

Our analysis focuses on a set of similar men and women – those who had received a postdoctoral (F32) training grant from the National Institutes of Health (NIH). We examined the rates at which men and women funded by those grants transitioned to being independent researchers, becoming a PI on an NIH R01 grant, and the extent to which their publishing records could account for those transitions. We first document that the transition to being a PI on an R01 grant can explain a substantial share of the gender gap in the life sciences. Women experienced 20% lower rates of transition than men. We then explored what factors might account for this disparity. Adjusting flexibly for differences in publication records could explain about 60% of this gender gap. But even women with similar publication records received R01 grants at lower rates than men. We then examined the extent to which women might receive less credit for their publication records (differential returns). These differential returns, particularly in the extent to which women benefited from citations, could account for the remainder of the gap.

In addition to the theoretical implications of the results, our study also contributes empirically to the literature on the gender gap in STEM in at least two additional respects. First, most of the prior studies on gender differences in productivity have analyzed samples of scientists who received their doctoral degrees in the 1970s or earlier. We update these findings by studying a sample of scientists who received their degrees in the 1980s, 1990s, and 2000s, a period during which the gap between the numbers of men and women enrolled in doctoral programs in the life sciences closed (Hill et al., 2010).

Second, prior research has focused on the average differences in publication records and on the linear effects of those differences on pay or promotion. But many of the returns in science come from being in the right-hand tail, to being unusually productive or producing research of particular importance, to being perceived as a star (Merton, 1968). We therefore introduce an empirical approach that allows us to capture heterogeneity in the returns to the research record across the distribution of the various dimensions of that record. Doing so can explain a substantial amount of additional variance. But the gender gap in the transition to being a PI remains even allowing for these non-linearities.

2. Career transitions

In trying to understand why women remain underrepresented in STEM fields, researchers have commonly characterized the process as being similar to a pipeline with an almost continuous series of leaks (e.g., Berryman, 1983; Etzkowitz et al., 2000; Lautenberger et al., 2014). Although this view has been criticized as being overly linear and insufficiently sensitive to the importance of social context outside of school or the workplace (Xie and Shauman, 2003), research in this vein has usefully documented the fact that the proportion of women in STEM fields declines through the college years, during graduate school, and as one considers ever more senior positions in these fields (Berryman, 1983; Shen, 2013; Lautenberger et al., 2014). Recent research suggests that the gender gap in the pipeline emerges even before college, as high school students begin to form their career ambitions and expectations (Morgan et al., 2013; Legewie and DePrete, 2014).

However, this pipeline view obscures the fact that most of the loss of women appears to occur within a short segment of the career, and one relatively far down the line. Consider the academic life sciences, the largest among the STEM fields: Women have reached near parity in both of the primary paths for entry, having a medical degree or a doctorate in a life sciences field (Lautenberger et al., 2014; Shen, 2013). They still appear almost equally represented in residency and post-doctoral training positions in research laboratories (Lautenberger et al., 2014; NPA, 2011). Yet, women hold only 40% of assistant professorships and no more than 30% of associate professorships in the life sciences (Jena et al., 2015). Their underrepresentation in the field emerges in the space of only two to ten years out of a career of forty or more. Returning to the pipeline analogy, it is less that the pipe drips continuously along the way and more that it is gushing at one or two of the joints between segments.

Given this fact, we see value in shifting the focus of analysis to understanding these critical career transitions where the gap widens most rapidly – in this case, on the transition to becoming an independent researcher in the life sciences. Individuals who complete a relevant graduate degree – a medical degree (MD) or a doctorate (PhD) – first move into a junior faculty position, either directly or following post-doctoral training. Because of the increasingly expensive nature of research in the life sciences, junior faculty must then find a means of funding their research. That usually means winning a major grant. Those who fail to do so have low odds of securing long-term (tenured) academic positions.

One can readily see from the much lower proportion of women at the associate professor level relative to the assistant professor level that women clear these hurdles at lower rates. What might account for differences in the transition rates experienced by men versus by women? We focused on two potential disparities: differences in publication records and differences in the returns to those publication records.

2.1. The productivity paradox

In academia as in many other settings, productivity represents an important determinant not only of who gets hired but also of who gets promoted. Given the up-or-out nature of the tenure-track job ladder, moreover, it also determines who remains in academia.

Productivity in academia, particularly in the sciences, means publications. Much attention therefore has been given to gender differences in publication records, the so-called "productivity paradox" (Cole and Zuckerman, 1984). Women publish fewer articles than men (Cole and Zuckerman, 1984; Long, 1992; Stack, 2002), and place them in less prominent outlets (Brooks et al., 2014; Lerchenmüller et al., 2018). Articles written by women, moreover, receive fewer citations, an important metric used to assess the influence of scientific research (Larivière et al., 2013).

On the articles they do publish, women appear in less prestigious authorship positions (Jagsi et al., 2006; Filardo et al., 2016). In the life sciences, the first and last authorships carry particular prestige. By convention, the individual who led the research and who analyzed and wrote up the results receives the first authorship. Last authorship goes to the head of the laboratory, who often receives credit not just for funding the research but also for conceiving of it. Interior authorships, meanwhile, go to those who assisted with data collection or analysis. Although women have reached parity in their probability of appearing in the first author position (West et al., 2013), this average belies the fact that women remain less likely to receive this prime position on articles published in the most prestigious journals (Lerchenmüller et al., 2018).

Overall, the reasons for these "productivity" differences remain a puzzle. Women may suffer discrimination both in the research lab and in the publication process, with consequences for their publication records. They may also find themselves with less time for research, either because they engage in more non-research activities at work or because they must shoulder a disproportionate share of the responsibilities at home (e.g., Craig and Mullan, 2011). Women may also choose different research paths. Leahey (2007), for example, has argued and provided evidence that women specialize less than men. Since specialization can allow researchers to produce more articles and increases the odds that they receive attention from others active in the field, it could account for multiple aspects of the productivity paradox.

But the productivity paradox may also come from comparing apples to oranges, or at least shoots to plants. Most of the studies on the gender gap in productivity have examined cross sections of authors or articles, pooling individuals across all career stages. If publication and citation rates rise over time and if fewer women transition to senior positions (perhaps due to bias in the evaluation process), then the average woman would occupy an earlier career stage than the average man in the population and one would observe these gender gaps in the cross-section even if men and women at the same career stages had equivalent publication records.

However, to the extent that productivity differences between men and women do appear early in their careers, one could see how such easily quantifiable differences in publication records could lead to differential rates of hiring, grant awarding, and promotion for men and women – regardless of whether these differences emerge from discrimination, from disparities in the allocation of parenting and other responsibilities, or from differential choices in their research agendas. We nevertheless have little direct evidence regarding the extent to which differences in publication records might account for critical career transitions.

2.2. Undervalued research records

Although academia ostensibly operates as a meritocracy, at least two lines of research suggest that differences in productivity might not account for the paucity of senior women on science faculties. First, in a series of audit studies, researchers have sent out equivalent resumes or curriculum vita, altering only the names of the candidates to signal the gender of the individual. Steinpreis et al. (1999), for example, manipulated the names of applicants for an assistant professor position in psychology and found that psychologists preferred candidates with stereotypically male names over those with female names. Moss-Racusin et al. (2012) repeated this design more recently for candidates for a lab manager position and again found a preference for applicants with male names.

Although these audit studies suggest that women receive lower returns to the same research records, this evidence remains inconclusive. On the one hand, the design of these studies holds constant every element of the research record. The same pattern of results would emerge even if the individuals screening the applications had only a slight preference for candidates of a particular gender among those equally qualified. On the other hand, the results of these studies have also not been consistent. Williams and Ceci (2015), for example, using the same study design, found that faculty preferred assistant professor candidates with female names over those with male names in every field studied, except for economics.

In a second line of research, a small number of studies have examined promotion rates and found residual effects for gender even after controlling for the number of publications. Long et al. (1993), Leahey et al. (2010), and Lutter and Schröder (2016) for example, have reported gender differences in the rates of promotion to tenure among biochemists, American sociologists, and German sociologists, respectively. After adjusting for the number of publications (and sometimes other dimensions of the publication record), these studies find lower promotion rates for women relative to men. These studies, however, do not provide direct evidence for the proportion of the overall gender gap that might stem from the productivity paradox because they have either entered gender in their regressions after or simultaneous to their measures of publication records, meaning that one cannot assess the extent to which adjusting for publication records might have narrowed the gender gap.

These literatures nonetheless suggest the possibility of differential returns – that women receive less credit for equivalent publication records. These differential returns could emerge in at least a couple of ways. Evaluators may simply place less value on the articles written by women or on the citations received by them. Such a pure form of discrimination would obviously place women at a disadvantage in selection and promotion processes. Or, it may reflect a preference for similar others. Research in social psychology has found that both men and women tend to evaluate same-sex individuals more favorably for similar levels of performance than individuals of the opposite sex (Greenberg, 1978). Given that men still account for the majority of evaluators – such as editors and grant application reviewers – in the life sciences and elsewhere, both forms of discrimination seem plausible.

But one could also imagine a more subtle dynamic. Perhaps the articles and citations themselves receive the same weight regardless of the gender of the authors but the allocation of credit for those articles differs systematically across men and women. Modern science has become a team sport, with ever larger groups of scientists involved in research projects (Wuchty et al., 2007). In mixed gender research groups, readers may perceive the men on the team as having contributed more to the research than the women. Consistent with this idea, Sarsons (2017) found that men in economics benefited much more from coauthored articles than women did, in terms of their odds of being promoted to tenure, and that this disparity appeared largest for mixed-gender coauthorships.

3. NIH funding and life science careers

The NIH is the largest funder of life science research in the United States, with an annual budget of roughly $30 billion (NIH, 2016a) – more than four times that of the National Science Foundation. The NIH supports intramural (NIH executed) and extramural research, with more than 80% of its budget going to the latter through competitive grants awarded to individuals and institutions, primarily in the United States.

A useful feature of the life sciences for our research is the fact that academic research in this field operates as a “soft money” environment in the United States. Rather than being guaranteed salaries and research funds by their universities, life scientists must compete for grants to fund their own positions and to finance their resource-intensive research projects. The funding trail therefore provides a good means for assessing the relationship between publication records and career advancement in the life sciences.

We focused on the receipt of the first R01 grant. The R01, a project-based renewable research grant awarded to scientists who have demonstrated research competence in a specific area (Azoulay et al., 2011), serves as the primary funding mechanism for the NIH. Although some other programs support the research of independent investigators, no other program comes close in importance to the R01.[1] These grants account for almost half of all NIH grant dollars and they represent the primary funding source for most academic biomedical research groups in the United States. There are about 27,000 outstanding awards, with roughly 4,000 new ones approved each year. Each award provides an average of $1.7 million in support spread over three to five years (Li, 2017).

Researchers typically receive their first R01 around the age of 42. It represents an important milestone in their careers, as both a financial enabler and as an indicator of their ability to conduct research independent from a more senior scientist (Garrison and Deschamps,

[1] Since these other programs may have somewhat different selection criteria, we restricted our analysis to the receipt of an R01.

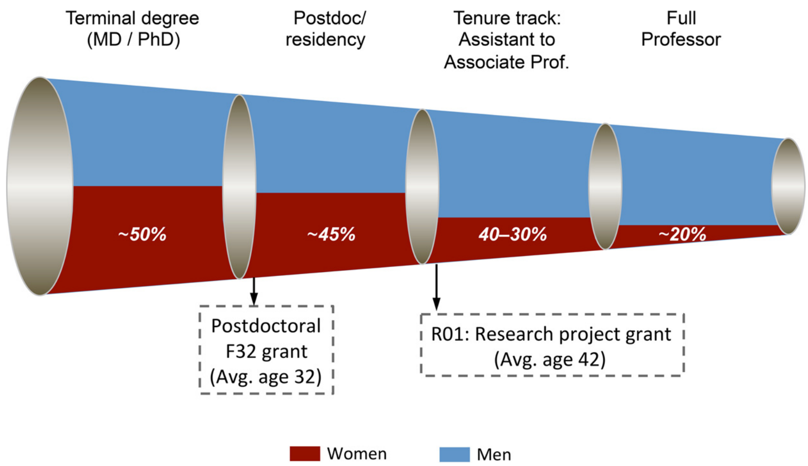

Fig. 1. Career stages, gender representation, and typical timing of first F32 and R01 award.

2014). As part of its charge to develop the biomedical research workforce, the NIH has a longstanding commitment to identifying and supporting promising young scientists on the way to independence. Since the 1970s, the NIH has sought to identify New Investigators, applicants who have not previously received an R01. The NIH segregates their applications into a separate pool, reviewing them relative to other early career scientists. NIH policy, moreover, requires the agency to award grants to new and experienced principal investigators at comparable rates (NIH, 2011).

One of the difficulties in almost any research on transitions involves the definition of the set at risk.[2] Defining the population at risk too narrowly precludes the researcher from gaining insight into crucial intermediate stages in the process. Defining it too broadly increases the odds that individuals differ in meaningful ways not captured in the covariates. Consider, for example, the transition to tenured faculty. At one extreme, one might want to follow all who completed a relevant doctoral program as being at risk. But doing so would then include many who had no interest in an academic career. At the other extreme, one might limit the sample to assistant professors. That restriction, however, would exclude all of those who had pursued academic positions but failed to obtain one.

To ensure that our sample included a relatively homogeneous group of individuals while still encompassing the crucial period when much of the gender gap emerges, we focused our analysis on individuals selected by the NIH to receive an F32 postdoctoral fellowship award. Established in 1974, the Kirchstein National Research Service Award (NRSA) Fellowship program (or F grant mechanism) represents the *only* means by which the NIH directly supports the basic preparation of individuals for careers in biomedical research (Mantovani et al., 2006). The F32 grant, by far the most common of these grants, targets scientists in their early postdoctoral years, with the average recipient being about 32 years old. The fellowship offers up to three consecutive years of mentored research support, with an average annual grant size of about $50,000 (Jacob and Lefgren, 2011).[3] The support packages include a stipend, tuition support, and an allowance to defray other miscellaneous costs related to research training (Mantovani et al., 2006). Fig. 1 situates the F32 postdoctoral fellowship award and the R01 mechanism within a typical life science career and reports the approximate proportion of women at each career stage (NPA, 2011 Lautenberger et al., 2014; Jena et al., 2015).

The F32 fellowship identifies individuals likely to pursue scientific careers. F32 grant recipients have a demonstrated interest in and commitment to pursuing an academic career. About two-thirds of F32 fellows remain employed in academia eight years after the completion of the fellowship. To ensure further that our analysis does not include individuals who have no or limited interest in pursuing an academic career, we exclude from the risk set F32 recipients who did not produce a single publication during their F32 fellowship period (roughly four out of ten F32 recipients). A gender gap in this cohort of committed and nationally-competitive individuals should therefore not reflect differences in the careers that men and women would prefer to pursue.

4. Gender gap in funding

We used a January 2016 download from the NIH *ExPORTER* database to track scientists. The database includes the names of funded scientists and a unique identifier (ID) assigned by the NIH to each scientist. Because grant applicants must use their assigned ID in all subsequent NIH grant applications, with failure to do so punishable by disqualification and potentially by federal law, these identifiers have extremely high fidelity across grants. The database, which covers the period running from 1985 to 2015, records grant budget periods, areas of research inquiry, and publications citing the grant, as well as other information.

We used the forenames of the funded scientists to infer their gender, using the *Genderize* database (Lerchenmueller, 2016). *Genderize* associates a name with the probability of being a man or a woman based on the occurrence of that name in a number of official sources, such as the Social Security Administration records, and in social media sources that verify the gender of the users (Wais, 2015).[4] For example, the database designates the forename "Chris" as male with 93% confidence, based on 8,631 verified records. For our analyses, we only included cases where the algorithm assigned a 90% or greater probability to the individual being of a specific gender. Using this confidence threshold, we could assign a gender to 88% of the F32 recipients in our download of the NIH *ExPORTER* database.

At the postdoctoral level, men and women differ somewhat in the financial support that they receive from the NIH, with men accounting for roughly 60% of these awards. Fig. 2 depicts the percentage shares of women receiving F32 awards and R01 grants since 1985. Although women still receive fewer F32 awards than men, their proportion has climbed in lockstep with the proportion of women obtaining terminal

[2] The NIH does not release data on unsuccessful grant applications and, even if it did, defining the risk set as those who had applied might prove too narrow.

[3] F32 fellows have a strong incentive to complete at least two years of supported training, as the NIH Revitalization Act of 1993 specifies that recipients must reimburse any support if they leave the fellowship prior to the completion of two years.

[4] *Genderize* has an advantage over other databases primarily in its scope, incorporating name data on 216,286 unique names from 79 countries and 89 languages. Comparisons of *Genderize* to other automated algorithms have found that it provides the most accurate gender assignments (Coding News, 2015).

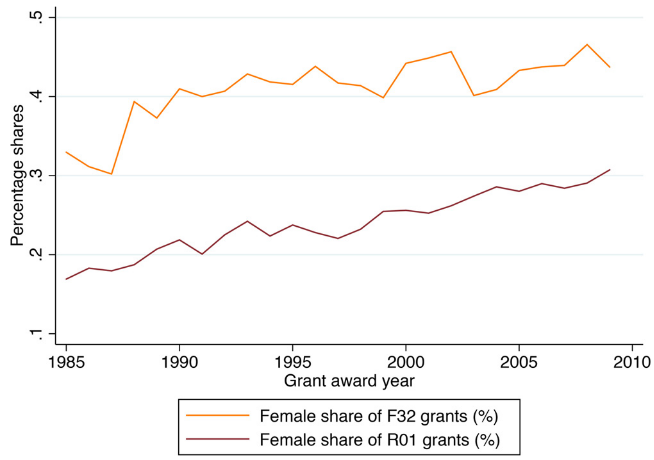

Fig. 2. Women's representation across NIH grant programs 1985–2009.

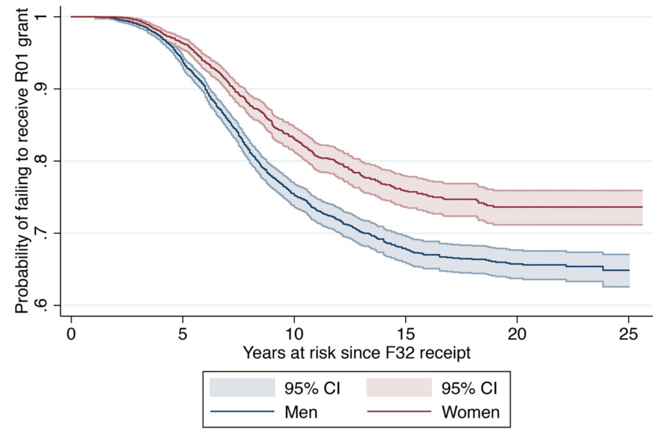

Fig. 3. Kaplan–Meier survival rate estimates.

Table 1
Transition matrix for F32s (1985–2005).

	Men(%)	Women(%)
Prior grant		
No	98	96
Yes	2	4
Prior type		
F30-31	90	92
Other	10	8
Post grant		
No	60	68
Yes	40	32
Post type		
R01	65	61
R[a]	26	28
Other	9	11
N	8,140	5,487

[a] One of 23 other R-mechanism grants (excluding R01 mechanism).
Post grants considered up to fiscal year 2015.

degrees in the life sciences. The remaining differential has largely been a function of the fact that somewhat fewer women apply for these grants. Conditional on application, the success rates do not differ by gender (Pohlhaus et al., 2011). Men and women also receive roughly the same levels of funding, with average award amounts differing by a mere $353 (not statistically significant).

But men and women receiving F32 awards have much different trajectories following this post-doctoral training. Fig. 2, which depicts the proportions of F32 and R01 awards going to women over time, reveals a substantial gender gap in R01 awards, which lags that in post-doctoral awards by at least 20 years. Table 1 reports the grant transition matrix for F32 recipients by gender, focusing on those who received their awards before 2006 (to allow for at least ten years to observe transitions). The R01 award represents the most common source for future funding, irrespective of gender. Note also that the other Research Program Grants, which account for many of the other transitions, often serve as intermediate awards on the path toward receiving an R01. Overall, men secure follow-on funding at an eight percentage point higher rate than women. The gender gap in funding levels stems almost entirely from this disparity in transition rates, as men and women receive awards of roughly equivalent sizes (with R01s for men having annual budgets of only $2,412 more than for women, a difference of less than half of one percent).

Fig. 3 provides a sense of when these differences emerge, depicting Kaplan–Meier estimates of the cumulative transition probabilities for men (blue) and women (red).[5] Over 60% of the men and 70% of the women did not transition to an R01 grant during our observation window. Women who received grants also received them later in their careers, on average. At every point in time after the fifth year from the F32 receipt, a smaller proportion of women than of men have received their first R01. For example, 10 years after the receipt, 25% of the men had received an R01 grant compared to only 17% of the women. These unconditional transition rates, however, do not say anything about why the gap widens.

5. Correlates of the gender gap

To assess the extent to which publication records versus differential returns to those records might account for the gender gap, we connected the grant data to article-level data from the *PubMed* database (for further details of the sample construction, see Appendix A). In total, our sample for estimation comprised 6,336 F32 recipients (60% male and 40% female).

Our data set consists of one observation per publication per person. The F32 budget start date served as the beginning of the time at risk and we considered the R01 budget start date the time of the transition event. If a scientist did not receive an R01 grant by the end of 2009, we considered the person right-censored. Our data set includes 74,188 publication-person observations (11.7 publications per person on average), covering 68,834 person-years (10.9 years on average).

We created a number of variables to capture different dimensions of publication records. We included a (logged) count of the number of *ln (articles)* on which individuals had been listed as an author. Because the norms in the life sciences assign the first author position to the person responsible for leading the execution and reporting of the research, one might expect that first authors would receive more credit for any publications. We therefore calculated a *Percent (first author)* variable to capture this effect.[6]

We also incorporated measures of the importance of these publications. Having been a coauthor on a publication in a leading journal may count more than having an article in a less prominent outlet. We therefore computed the proportions of publications appearing in journals with 5-year journal impact factors (JIF) of over five and up to ten (*Percent (JIF 5-10)*) and exceeding ten *Percent (JIF > 10)*. The first category includes a number of important field journals, while the latter encompasses the most prominent journals in science and medicine. The proportion of publications in journals with JIFs of five or less served as

[5] Since we have a continuous clock, we used the asymptotic variance estimate to derive confidence intervals for $\hat{S}(t)$ (Kalbfleisch and Prentice, 2002).

[6] We used proportions for these variables rather than raw counts to reduce collinearity with the number of publications.

the baseline category.

Our models also included the average number of citations, according to *Scopus* (a database maintained by Elsevier), received by all articles published by the individual up to that point in time. Citation counts have often been used as an article-level metric of research quality but even if other factors influence them they clearly capture the attention received by the research. Our models included the (logged) average number of citations received by all articles published by the individual (*ln (avg. citations)*).[7] To ensure that our data reflect the number of citations received up to a particular point in time, this variable has been calculated at the time of each publication.

These dimensions of the research record align well with the criteria by which the NIH claims to evaluate proposals. The agency suggests that evaluators should use five criteria: (1) Does the applicant have a record of accomplishments in the field? (2) Will the scientist convert funding into research output? (3) How significant is the proposed research? (4) How innovative is it? And, (5) does the researcher operate in an environment that will support the research? The final two criteria seem least connected to these measures of publication records, though the prominence of the journal and the number of citations received often reflect some combination of quality, significance, and innovativeness. Note also that our models control for the institution's aggregate success in securing R01 grants, adjusting explicitly for the institutional environment (the fifth criterion).

5.1. Estimation

To understand better the factors underlying gender differences in the transition to principal investigator, we turned to parametric survival analysis. These models have the advantage of exploiting information on both the occurrence and the timing of events.

We began by estimating a non-parametric baseline hazard rate, without any covariates:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t | T > t)}{\Delta t}, \tag{1}$$

where T represents a random variable for the time of R01 receipt and t denotes the amount of time that has passed since individual i has received the F32 award. We estimated this function using kernal smoothing, averaging values of the function over a moving window.[8]

Fig. 4 displays these unconditional hazard rates for men and for women. The rate at which these scientists received their first R01 awards peaked at about eight years after their F32 awards. Interestingly, the hazard rate of receiving an R01 grant rose more steeply for men than for women. Irrespective of gender, however, the hazard rate had a non-monotonic relationship with t, reflecting the fact that the transition to R01 becomes increasingly less likely if it does not happen within eight years of the F32. Beyond 15 years, the hazard rate falls almost to zero. We therefore limited our observation window to the 15 years following the receipt of an F32.

Given the non-monotonic shape of the hazard function, we chose a log-logistic form of time dependence.[9] We estimated time to first R01 grant with an accelerated failure-time (AFT) model. One can interpret the exponentiated coefficients of these models as time ratios, with values in excess of one indicating a delay in the arrival of the event and values below one reflecting an acceleration of the arrival rate.

The models also include a number of control variables. We account for other sources of funding with a count of the number of non-R01 grants received from the NIH since the receipt of the F32 (*interim grants*). We also included an indicator variable for whether the individual had received an NIH grant prior to the F32 (*prior grants*) and a count of the number of *grant extensions*.[10] We also accounted for the potential effect of the number of articles published prior to F32 receipt (*ln (prior articles)*).

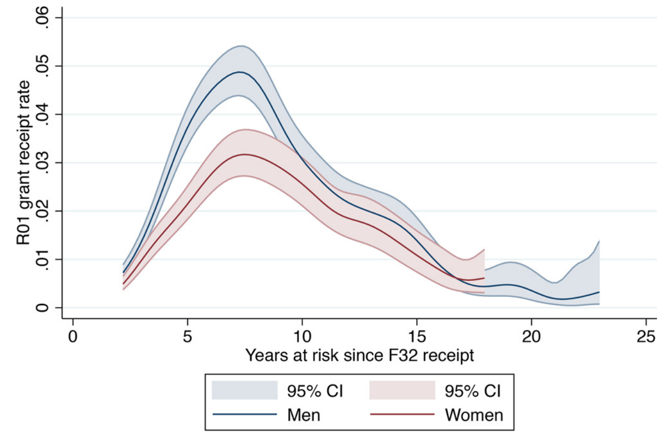

Fig. 4. Baseline hazard rate estimates.

Because the composition of the research teams of which individuals have been a member may influence the allocation of credit, we calculated a covariate for the average number of coauthors on an article (*avg. team size*). Each coauthor probably receives less credit for research produced by a larger team. We also computed the *percent female authors* across all publications as an additional control variable.

Past research outside the life sciences has found that academics with more specialized research agendas appear to enjoy greater success (e.g., Leahey, 2006). We therefore included a measure of specialization, based on medical subject header (MeSH) terms (keywords). We calculated a Herfindahl-index measure of specialization, summing the squared proportions of MeSH terms associated with an individual's articles. Because this measure correlates strongly with the number of articles, we regressed this measure on the (logged) number of articles and used the residuals from that regression as our measure of *specialization*.

We also accounted for the strength of the institutional environment by including the institution's percentile rank in terms of its number of NIH R01 awards. Finally, we included fixed effects for area of research inquiry (using the two-digit letter code for the supporting NIH Institute/Center embedded in the F32 grant number) and for grant vintage (dividing F32 grants into five year cycles, 1985–2009) to control for field and period effects, such as changes in the NIH budget.

Table 2 provides descriptive statistics for the variables used in our models. All independent variables and the control variables for interim grants, grant extensions, specialization, average team size, and percent female authors update at the time of publishing an article. All other control variables remain constant for an individual over time. Interestingly, although men and women differ in terms of their numbers of publications, they appear quite similar on almost every other dimension of their publication records.

6. Results

Table 3 reports the results of the log-logistic regressions of time to first R01 grant. The first model, including only a covariate for gender,

[7] Using the average instead of the sum helps us to distinguish publication quantity from article-level attention.

[8] In principle, one could estimate $h(t)$ by differentiating the cumulative hazard function with respect to t, using the Kaplan–Meier estimate of $S(t)$. But the Kaplan–Meier estimator creates a step function for $S(t)$. One therefore cannot differentiate it directly.

[9] Comparisons of log-logistic models to estimates using other forms of time dependence revealed that the log-logistic models fit the data better based on the Akaike Information Criterion (AIC).

[10] The NIH data do not record family events, such as child birth; however, NIH policies allow the extension of fellowship and career development grants when the grantee has family responsibilities that delay the research (NIH, 2016b).

Table 2
Descriptive statistics for F32-transition models.

	Men		Women	
	Mean	SD	Mean	SD
Publication record				
Ln (articles)	2.05	0.88	1.84	0.85
Pct (first author)	0.32	0.23	0.34	0.25
Pct (JIF [0-5])	0.66	0.29	0.65	0.30
Pct (JIF (5-10])	0.18	0.21	0.20	0.23
Pct (JIF > 10)	0.16	0.22	0.15	0.22
Ln (avg. citations)[a]	4.25	0.79	4.26	0.77
Control variables				
Interim grants	0.54	0.82	0.46	0.81
Grant extensions	0.14	0.35	0.16	0.37
Prior grants	0.03	0.17	0.04	0.20
Specialization (resid.)	0.00	0.01	0.00	0.01
Avg. team size	5.11	2.62	5.22	3.91
Pct female authors	0.22	0.23	0.53	0.28
Status host institution	96.73	8.97	95.99	10.88
Ln (prior articles)	1.25	0.86	1.16	0.80
N	3,822		2,514	

[a] Based on 3,817 male and 2,512 female F32 grant holders with citations to their work.

indicates that women, on average, have a 20% slower rate of transitioning to the receipt of an R01. Adjusting for the various control variables reduces this gap by roughly 25% (to 15%).

Models 3 through 6 then examine the extent to which differences in publication records might account for the remaining gap. Model 3 accounts only for the number of articles. Not surprisingly, publications have a large effect on the expected time to receiving a first R01. A doubling in the number of articles reduces the expected time to R01 by roughly 20%. Accounting for this effect, moreover, reduced the unexplained gender gap to 12%. Model 4 then introduces the various other dimensions of the publication record. Interestingly, the proportion first authorships and the proportion of publications in prestigious journals have little influence on the expected time to first R01. Citations do, however, have a large effect: a doubling in the number of citations per article also reduces the expected time to receiving an R01 by about 22%. The residual gender gap, however, remained fairly stable at about 12%, because men and women did not differ meaningfully on these dimensions.

6.1. Functional form

An unstated assumption in much of the past literature not just on academic productivity but also on the effects of productivity on success

Table 3
Log-logistic regression of time to first R01 – publication records.

	(1) Sex only model	(2) Add controls	(3) Add productivity	(4) Add quality metrics	(5) Add funct. form	(6) Add complements
Sex	1.20**	1.15**	1.12*	1.12*	1.08*	1.08*
	(0.04)	(0.04)	(0.06)	(0.05)	(0.04)	(0.04)
Publication record						
Ln (articles)			0.71**	0.70**	0.98	1.00
			(0.03)	(0.03)	(0.05)	(0.05)
Pct (first author)				0.87†	0.64**	0.52**
				(0.07)	(0.09)	(0.08)
Pct (JIF 5-10)				1.05	1.04	1.00
				(0.10)	(0.17)	(0.17)
Pct (JIF > 10)				0.84†	0.83	0.85
				(0.08)	(0.11)	(0.11)
Ln (avg. citations)				0.69**	0.77**	0.77**
				(0.02)	(0.03)	(0.03)
Control variables						
Interim grants		0.62**	0.56**	0.60**	0.68**	0.69**
		(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Grant extensions		0.97	0.95	0.98	0.98	0.98
		(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
Prior grants		0.75**	0.70**	0.72**	0.77**	0.75**
		(0.06)	(0.07)	(0.07)	(0.06)	(0.06)
Specialization (resid.)		0.09**	0.04**	0.03**	0.16*	0.19†
		(0.05)	(0.03)	(0.02)	(0.14)	(0.16)
Avg. team size		1.00	1.02†	1.05**	1.04**	1.04**
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pct female authors		1.02	1.02	1.00	0.99	0.99
		(0.06)	(0.08)	(0.08)	(0.06)	(0.06)
Status host institution		1.00	1.00	1.00	1.00	1.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Ln (prior articles)		0.84**	0.85**	0.85**	0.89**	0.89**
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Research field fixed effects (19)	NO	YES	YES	YES	YES	YES
Grant vintage fixed effects (4)	NO	YES	YES	YES	YES	YES
Functional form fixed effects (15)	NO	NO	NO	NO	YES	YES
Complements fixed effects (45)	NO	NO	NO	NO	NO	YES
Log-likelihood	−3,170	−2,530	−2,469	−2,332	−2,294	−2,244
Observations	68,776	68,776	68,776	68,614	68,614	68,614

† Significant level: 10%.
* Significant level: 5%.
** Significant level: 1%.

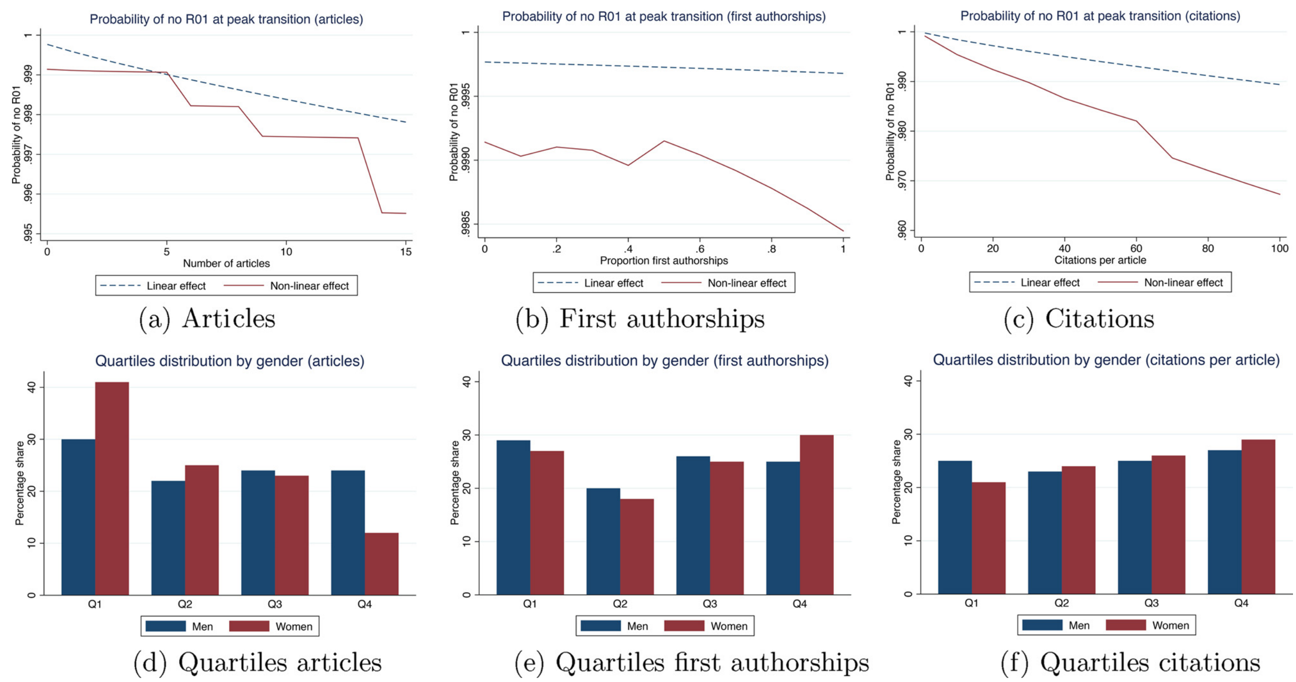

Fig. 5. Non-linear vs. linear effects on probability of transition to R01 at year eight of follow-up.

in other contexts has been that output has a linear or log-linear relationship to outcomes. For example, researchers will include the count of articles or the logged count as an independent variable as a means of adjusting for research output. But the relationship may prove more complex for a variety of reasons. On the one hand, the first publication might have inordinate importance, as a sort of proof of concept. Or, publications might have increasing marginal returns to the extent that large numbers of them lead to the individual being perceived as a star (Merton, 1968). On the other hand, many evaluations in academia compare individuals to "peers" – this implicit competition might mean that where one falls on the distribution (relative output) matters more than absolute output.

To relax these functional form assumptions, we calculated time-varying distributions of our five measures of the publication record and created vectors of indicator variables to reflect the quartile of the distribution into which the scientist fell at any particular point in time. In total, we included 15 variables to capture this distributional information, three quartile indicators for each of the five dimensions of the publication record.

Model 5 includes these quartile fixed effects in the models. Their inclusion improved the model fit ($p < 0.01$). As one might expect, these quartile indicators largely absorbed the effect of the logged number of articles. Surprisingly, however, their inclusion *increased* the predictive power associated with the linear term for the proportion of first authorships. This effect emerged because all of the action on that dimension comes from the top quartile. Fig. 5a–c displays the predicted values including the time-varying quartiles relative to the predicted values using only the continuous measures for three of the variables (article count, proportion first author, and citations).[11] As one can see, the deviation for first authorships occurs for those who have been first authors on the majority of their publications. Without the quartile fixed effects, the effect in the first three quartiles of the distribution drives the average estimate. Fig. 5d–f depicts the number of men and women in each of these quartiles. Overall, adjusting flexibly for functional form reduced the residual gender gap by another third to 8%.

6.2. Complementarity

Regression estimates of promotion also typically assume that the various components of productivity have additive effects on the outcome. For example, being a first author on an article in *Science* should have the same effect as being a first author in a less prominent outlet and being an interior coauthor on an article in *Science*. But even casual observation of how decisions get made in academia suggests that these various measures may interact in important ways. First authorships, for example, may prove particularly valuable if they occur on publications in highly visible outlets (JIF > 10). To allow for these interactions, we created a set of indicator variables for every possible combination of the quartile variables, a total of 45 additional terms. Although these terms proved jointly significant ($p < 0.01$), suggesting that the various dimensions of publication records do interact in important ways in determining grant receipt, their inclusion neither narrowed nor widened the unexplained gender gap.

6.3. Differential returns

All of the models thus far have assumed that men and women benefit equally from their publication records. But, as noted above, women may receive less credit for the same output – in essence, they may receive differential returns to their publication records.

To assess this possibility of differential returns, we first estimated gender-specific models of the time to first R01. Doing so effectively interacts gender with every other variable in the model. Table 4 reports the results of these models in the first two columns. Note that we excluded all fixed effects from these models so that any differences in the returns would only appear in reported coefficients.[12] Although all of the point estimates appear somewhat different for men and women, the 90% confidence intervals around these estimates substantively overlap in all cases except for one, citations. While a doubling in the number of

[11] To allow for a visual comparison of these time-varying effects, we computed the predicted probabilities of survival (i.e., not receiving a R01 grant) at eight years after the receipt of the F32 (the peak transition year). Kinks in the lines illustrate the effect of shifting into the next quartile of the respective distributions.

[12] The functional forms of the relationships of the publication measures to time of grant receipt did not vary across men and women ($prob > \chi^2 = 0.44$).

Table 4
Log-logistic regression of time to first R01 - differential returns.

	(7) Men	(8) Women	(9) Pooled
Sex			0.67[†]
			(0.15)
Publication record			
Ln (articles)	0.74[**]	0.67[**]	1.01
	(0.03)	(0.05)	(0.05)
Pct (first author)	0.87	1.01	0.51[**]
	(0.09)	(0.16)	(0.08)
Pct (JIF 5-10)	0.98	1.13	0.95
	(0.11)	(0.19)	(0.17)
Pct (JIF > 10)	0.81[*]	0.70[*]	0.86
	(0.08)	(0.11)	(0.11)
Ln (avg. citations)	0.67[**]	0.76[**]	0.74[**]
	(0.02)	(0.04)	(0.03)
Differential returns			
Sex × ln (articles)			0.96
			(0.04)
Sex × pct (first author)			1.05
			(0.14)
Sex × pct (JIF 5-10)			1.11
			(0.17)
Sex × pct (JIF > 10)			0.93
			(0.13)
Sex × ln (avg. citations)			1.12[**]
			(0.05)
Control variables			
Interim grants	0.60[**]	0.63[**]	0.69[**]
	(0.02)	(0.03)	(0.01)
Grant extensions	1.00	1.02	0.98
	(0.06)	(0.08)	(0.04)
Prior grants	0.82[†]	0.63[**]	0.75[**]
	(0.09)	(0.09)	(0.06)
Specialization (resid.)	0.03[**]	0.03[†]	0.21[†]
	(0.04)	(0.05)	(0.18)
Avg. team size	1.05[**]	1.06[**]	1.04[**]
	(0.01)	(0.02)	(0.01)
Pct female authors	1.08	0.91	1.00
	(0.11)	(0.11)	(0.06)
Status host institution	1.00	1.00	1.00
	(0.00)	(0.00)	(0.00)
Ln (prior articles)	0.85[**]	0.85[**]	0.89[**]
	(0.02)	(0.03)	(0.02)
Research field fixed effects (19)	NO	NO	YES
Grant vintage fixed effects (4)	NO	NO	YES
Functional form fixed effects (15)	NO	NO	YES
Complements fixed effects (45)	NO	NO	YES
Log-likelihood	−1,540	−817	−2,238
Observations	44,712	23,902	68,614

[†] Significant level: 10%.
[*] Significant level: 5%.
[**] Significant level: 1%.

citations per paper reduced the time to the first R01 grant by about 23% for men, the same increase in citations per paper only reduced the time to R01 by roughly 16% for women.

The final model then pools the estimates again, including interaction terms for gender and the various dimensions of the publication record. These models include the field, period, functional form, and complements fixed effects. Only one interaction has a significant effect: citations. A woman with the same number of average citations per publication appears to benefit about 12% less from them in terms of time to receiving her first R01. Note that women received a slightly greater proportion of their citations from first authorships (23.1%) relative to men (22.5%) in our data, so this disparity does not come from women receiving less credit for articles on which they had been interior authors.

Although the main effect of gender in this model would appear to suggest that women actually transition more rapidly to R01 than men, all else equal, note that one cannot interpret the gender coefficient in the same manner when estimated with an interaction term. The coefficient represents the time to R01 for women relative to men for those with zero citations. But scientists without any citations have almost no chance of being granted an R01. A more useful way of calibrating this coefficient calculates at what number of citations men appear advantaged relative to women. That occurs when the average citations per article exceeds roughly 2.1 ($= e^{(1/.88)/1.5}$), a level less than the fifth percentile of the citations distribution.

7. Discussion

The gender gap in academic STEM employment has attracted much attention. Research and policy agendas have been focused on two types of effects. As in other settings, there has been concern that women face a "glass ceiling" – a level beyond which they simply cannot advance. Research has also called attention to the idea of a "leaky pipeline" – that the number of women active in STEM professions declines from early education to college to post-doctoral training and at every subsequent career stage (Etzkowitz et al., 2000).

We document that a large share of this gap emerges in a relatively short period of time, as men and women move from being a member of another researcher's lab to leading their own lab. Rather than women dripping out of the STEM career pipe every centimeter along the way, they appear to pour out at one of the critical junctures. We therefore shifted the lens to focus on this period where the gap widens most rapidly.

In particular, we analyzed the rates at which men and women received their first R01 grant from the NIH. Among those who had already held post-doctoral grants and who had published, women had 20% lower transition rates to an R01. Although not a measure of promotion per se, the importance of funding in the life sciences means that an R01 has effectively become a precursor to receiving tenure at a research university (Jena et al., 2015).

Why does this gap in funding emerge? In trying to understand the underrepresentation of women in STEM and recognizing that the publication record plays a prime role in determining who gets hired, funded, and promoted in academia, past research has documented a number of dimensions on which women experience worse outcomes than men, from fewer publications overall to publishing in less prominent outlets (e.g., Cole and Zuckerman, 1984; Long et al., 1993; Stack, 2002; Lerchenmüller et al., 2018).

But these studies have not been able to connect these gender gaps in publication records to the paucity of women at the senior levels of the professorate for two reasons. First, many of them have been cross-sectional. One cannot even say then in which direction causality might run. Less impressive publication records might contribute to the underrepresentation of women in science but gender gaps in the publication record might also emerge as an artifact of comparing the records of more senior men to more junior women. Second, even those studies that have been longitudinal in their design have not structured and reported their analyses in a manner that allows one to determine what portion of the gender gap in STEM might stem from differences in the publication records of men and women.

We therefore estimated the extent to which publication records could account for the lower rates at which women received their first R01 grants. Various dimensions of the publication record, most notably the number of publications and the average number of citations received per article, can account for roughly 60% of the gender gap in the receipt of these grants. We adopted extremely flexible functional forms, allowing publication records to have non-linear and even non-monotonic effects and to have complementarities between aspects of the publication record – for example, allowing first authorship on a paper in a top-tier journal to count more than first authorship on a paper in a less prominent one. These flexible functional forms substantially improve the explanatory power of publication records. In contrast, the typical approach – assuming a linear or log-linear relationship between the number of publications and grant receipt – underestimates the importance of being near the top of the distribution, of being highly prolific. But even allowing for extremely flexible functional forms and for complementarity between dimensions of the publication record could not fully explain the gender gap in the receipt of R01 grants.

A subsequent exploration of potential differences in the returns to the same features of the publication record suggested that women benefit less from the same number of average citations per article (but that they experience similar returns to every other dimension of the research record). These differential returns can account for the remaining gender gap in funding. Although our results would appear consistent with audit studies, some of which have suggested that women with equivalent records receive less favorable outcomes than men (e.g., Moss-Racusin et al., 2012), people rarely put their average number or even their total number of citations on their resumes. Our results therefore appear to point to a novel specific dimension on which women receive lower returns than men.

Our research design does not, however, allow us to say precisely why these differential returns occur. Note that the models do control for the proportion of women coauthors (and that men and women did not differ significantly on the effects of that variable). The differential returns to citations therefore would not seem to stem from women receiving less of the credit when coauthoring with men.

One possibility is that evaluators err in their estimates of the influence of research. They may effectively overestimate the importance of prior research done by men relative to that done by women. In contrast to other elements of the publication record, NIH applications do not typically include information on the citations that applicants have received. This absence of explicit information may allow more latitude for cognitive biases – even implicit ones – to creep into evaluations. Consistent with that idea, men and women did *not* differ in their apparent returns to any of the dimensions of the publication record that appear on the grant applications. One potential remedy worth considering therefore would involve including explicit information on citations on grant applications.

Another possibility is that evaluators perceive the research done by women as less valuable than that done by men and that this bias applies most strongly to the most novel and most influential research. That idea seems consistent with some of the prior research on stereotyping which suggests that many scientists and engineers perceive science as a male occupation (Joshi, 2014). Unfortunately, however, if that explains the effect, then it becomes hard to imagine any simple policy intervention that could rectify the situation.

But differences in publication records – the number of articles and the average number of citations per article – appear even more important than differential returns in explaining the gender gap in funding. Why men and women differ on these dimensions, however, also remains an open question. In our definition of the sample, we tried to rule out gender differences in career preferences by focusing our analysis on cohorts of F32 post-doctoral grant recipients – scientists with an interest in and commitment to pursuing an academic career (Mantovani et al., 2006).

Biases may, of course, directly influence the ability of women to publish and the number of citations that they receive. If reviewers or editors perceive publications written by women as less important or of lower quality than those written by men, women might receive more rejections before finally placing a paper or go through more rounds of review. Either could slow down their rate of publication. Published research by women, moreover, may receive less recognition by others, in the form of citations.

Blind review, particularly of a form where the editors did not have information on the identities of the authors before coming to a decision, could help to limit gender differences in the journal evaluation process. When orchestras began to have musicians audition from behind screens so that the judges could not guess the gender of the musician, the gender balance of orchestras rapidly shifted from being mostly men to the majority being women (Goldin and Rouse, 2000). But if the same biases exist among readers, the consumers of research, blind review would not necessarily eliminate disparities in citation rates.

Men and women may also differ in their output because of differences in the time that they have available for research. Some of these differences likely stem from the home. Even among dual-career couples, women typically shoulder most of the burden in childcare and in maintenance of the household (Craig and Mullan, 2011). But a large share of these differences may also emerge from the workplace. Women often do more than their fair share of administration and service in academic settings.

These differences in productivity might also stem from differential access to mentoring and role models. One of the difficulties in expanding the representation of women in the life sciences and elsewhere has been the very paucity of senior women. Not only does this absence mean that junior women have fewer role models who they may consider relevant but also it means that they may not have access to senior women who can act as mentors. Mentors can play a number of important roles, from providing their junior colleagues with a better understanding of how the publication and grant application processes work to introducing them to potential collaborators and to gatekeepers in the field (Preston, 2004; Etzkowitz et al., 2000). If mentors favor those of the same gender in these processes, women may find themselves disadvantaged in this early access to tacit knowledge and social capital. Understanding better the ultimate source of these differences in publication records therefore represents an important question for future research.

Although our focus has been on early career transitions, our results may also have relevance for gender stratification at later career stages as well. Funding continues to matter at more senior levels, determining who can pursue their research agendas and therefore who can publish and receive accolades for their contributions. These factors therefore may continue to disadvantage women even if they receive an R01 and earn tenure.

Appendix A

We began the construction of our dataset by connecting F32 grant recipients to their publication records by first associating the grant holders with articles that had acknowledged these F32 grants (according to the NIH *ExPORTER* database). Those articles then served as a means of connecting the NIH data to the Authority author codes in *PubMed*, which allowed us to identify their entire publication records both prior to and after the period covered by the F32 grants. The Authority disambiguation algorithm, which has been assigned to all authors of *PubMed* articles written prior to 2009, identifies which cases of authors of different articles with the same name have a high probability of referring to the same individual. It has been found to have a greater than 99% accuracy (Lerchenmueller and Sorenson, 2016).

Our initial dataset included 7,623 F32 grants that had been acknowledged in at least one article published by the respective grantees prior to 2009 (the end of the period covered by the Authority algorithm). From this sampling frame, we could match the unique author identifiers from NIH *ExPORTER* to the unique author identifiers available in *PubMed* for 7,169 F32 recipients (94%).

We excluded scientists who had published ten years or more prior to the receipt of their F32. Given that the F32 should almost immediately follow a doctoral degree, either these records include an error or these individuals would have had to have published as a high school student or as an undergraduate.

From this set, we excluded scientists for whom we could not determine the dates of their publications for a large share of their publication records. Our longitudinal analysis required us to assign publications to specific dates so publications without these dates effectively add noise to the analysis. Specifically, we dropped from the analysis scientists whose publication portfolios included ten articles or fewer with publication dates missing for two or more of these articles (i.e., more than 20% of their articles). We also excluded scientists, with more than ten publications, who had missing dates for more than 10% of their articles. These restrictions reduced the sample to 6,549 F32 scientists (but it did not alter the overall gender distribution of the sample).

In cases where *PubMed* recorded only a month and year of publication (as opposed to a specific date within the month), we assigned a random publication day (1–30) from a uniform distribution to avoid tied spells in the time-to-event analysis. In cases where *PubMed* only recorded publication years (11% of records), we assigned a random publication day from the full year (1–365), from a uniform distribution.

Although the NIH represents the major source of third-party funding in the life sciences, some individuals may have benefited from other sources of funding that we could not observe and may therefore have entered the F32 program at a later stage in their careers or may have received a major award prior to receiving their first R01. We therefore dropped from the analysis individuals in the long tail of the distribution, with more than 50 articles (97th percentile of the distribution). Note, however, that including these cases did not substantively change the results.

References

Azoulay, Pierre, Graff Zivin, Joshua S., Manso, Gustavo, 2011. Incentives and creativity: evidence from the academic life sciences. RAND J. Econ. 42 (3), 527–554.

Berryman, Sue E., 1983. Who Will Do Science? Minority and Female Attainment of Science and Mathematics Degrees. Rockefeller Foundation, New York.

Brooks, Chris, Fenton, Evelyn M., Walker, James T., 2014. Gender and the evaluation of research. Res. Policy 43 (6), 990–1001.

Coding News, 2015. Gender Detection. Tech. Rep. Coding News (accessed 11.07.17). http://codingnews.info/post/gender-detection.html.

Cole, Jonathan R., Zuckerman, Harriet, 1984. The productivity puzzle: persistence and change in patterns of publication of men and women scientists. Adv. Motiv. Achiev. 2, 217–258.

Craig, Lyn, Mullan, Killian, 2011. How mothers and fathers share childcare. Am. Sociol. Rev. 76 (6), 834–861.

Etzkowitz, Henry, Kemelgor, Carol, Uzzi, Brian, 2000. Athena Unbound: The Advancement of Women in Science and Technology. Cambridge University Press, New York.

Filardo, Giovanni, da Graca, Briget, Sass, Danielle M., Pollock, Benjamin D., Smith, Emma B., Ashley-Marie Martinez, Melissa, 2016. Trends and comparison of female first authorship in high impact medical journals: observational study (1994–2014). BMJ 352. http://dx.doi.org/10.1136/bmj.i847.

Garrison, Howard H., Deschamps, Anne M., 2014. NIH research funding and early career physician scientists: continuing challenges in the 21st century. FASEB J. 28 (3), 1049–1058.

Goldin, Claudia, Rouse, Cecilia, 2000. Orchestrating inpartiality: the impact of 'blind' auditions on female musicians. Am. Econ. Rev. 90, 715–741.

Greenberg, Jerald, 1978. Allocator-recipient similarity and the equitable division of rewards. Soc. Psychol. 41 (4), 337–341.

Hill, Catherine, Corbett, Christianne, St. Rose, Andresse, 2010. Why so few? Women in Science, Technology, Engineering, and Mathematics. Tech. Rep. American Association of University Women, Washington, DC (accessed 11.07.17). https://www.aauw.org/files/2013/02/Why-So-Few-Women-in-Science-Technology-Engineering-and-Mathematics.pdf.

Jacob, Brian A., Lefgren, Lars, 2011. The impact of NIH postdoctoral training grants on scientific productivity. Res. Policy 40 (6), 864–874.

Jagsi, Reshma, Guancial, Elizabeth A., Worobey, Cynthia Cooper, Henault, Lori E., Chang, Yuchiao, Starr, Rebecca, Tarbell, Nancy J., Hylek, Elaine M., 2006. The "gender gap" in authorship of academic medical literature – a 35-year perspective. N. Engl. J. Med. 355 (3), 281–287.

Jena, Anupam B., Khullar, Dhruv, Ho, Oliver, Olenski, Andrew R., Blumenthal, Daniel M., 2015. Sex differences in academic rank in US medical schools in 2014. JAMA 314 (11), 1149–1158.

Joshi, Aparna, 2014. By whom and when is women's expertise recognized? The interactive effects of gender and education in science and engineering teams. Adm. Sci. Q. 59 (2), 202–239.

Kalbfleisch, John D., Prentice, Ross L., 2002. The Statistical Analysis of Failure Time Data, 2nd edition. Wiley, New York.

King, Molly M., Bergstrom, Carl T., Correll, Shelley J., Jacquet, Jennifer, West, Jevin D., 2016. Men Set Their Own Cites High: Gender and Self-Citation Across Fields and Over Time. (accessed 11.07.17. arXiv:1607.00376.

Larivière, Vincent, Ni, Chaoqun, Gingras, Yves, Cronin, Blaise, Sugimoto, Cassidy R., 2013. Global gender disparities in science. Nature 504, 211–213.

Lautenberger, Diana M., Dandar, Valerie M., Raezer, Claudia L., Sloane, Rae Anne, 2014. The State of Women in Academic Medicine: The Pipeline and Pathways to Leadership. Tech. Rep. Association of American Medical Colleges, Washington, DC (accessed 11.07.17). https://members.aamc.org/eweb/upload/The%20State%20of%20Women%20in%20Academic%20Medicine%202013-2014%20FINAL.pdf.

Leahey, Erin, 2006. Gender differences in productivity – research specialization as a missing link. Gender Soc. 20 (6), 754–780.

Leahey, Erin, 2007. Not by productivity alone: how visibility and specialization contribute to academic earnings. Am. Sociol. Rev. 72 (4), 533–561.

Leahey, Erin, Keith, Bruce, Crockett, Jason, 2010. Specialization and promotion in an academic discipline. Res. Soc. Stratif. Mob. 28 (2), 135–155.

Legewie, Joscha, DePrete, Thomas A., 2014. Pathways to science and engineering bachelor's degrees for men and women. Sociol. Sci. 1, 41–48.

Lerchenmüller, C., Lerchenmueller, M.J., Sorenson, O., 2018. Long-term analysis of sex differences in prestigious authorships in cardiovascular research supported by the National Institutes of Health. Circulation 137, 880–882. http://dx.doi.org/10.1161/circulationaha.117.032325.

Lerchenmueller, Marc J., 2016. Gender Designation via Genderize API With Unlimited Request. https://figshare.com/articles/Genderize_unlimited_API_request/4563814.

Lerchenmueller, Marc J., Sorenson, Olav, 2016. Author disambiguation in PubMed: evidence on the precision and recall of Authority among NIH-funded scientists. PLOS ONE 11 (7), e0158731.

Li, Danielle, 2017. Expertise versus bias in evaluation: evidence from the NIH. Am. Econ. J.: Appl. Econ. 2 (9), 60–92.

Long, J. Scott, 1992. Measures of sex-differences in scientific productivity. Soc. Forces 71 (1), 159–178.

Long, J. Scott, Allison, Paul D., McGinnis, Robert, 1993. Rank advancement in academic careers: sex differences and the effects of productivity. Am. Sociol. Rev. 58 (5), 703–722.

Lutter, Mark, Schröder, Martin, 2016. Who becomes a tenured professor, and why? Panel data evidence from German sociology, 1980–2013. Res. Policy 45, 999–1013.

Mantovani, Richard, Look, Mary V., Wuerker, Emily, 2006. The Career Achievements of National Research Service Award Postdoctoral Trainees and Fellows: 1975–2004. Tech. Rep. National Institutes of Health, Bethesda, MD.

Merton, Robert K., 1968. The Matthew effect in science. Science 159 (3810), 56–63.

Morgan, Stephen L., Gelbgiser, Dafna, Weeden, Kim A., 2013. Feeding the pipeline: gender, occupational plans, and college major selection. Soc. Sci. Res. 42 (4), 989–1005.

Moss-Racusin, Corinne A., Dovidio, John F., Brescoll, Victoria L., Graham, Mark J., Handelsman, Jo, 2012. Science faculty's subtle gender biases favor male students. Proc. Natl. Acad. Sci. 109 (41), 16474–16479.

National Postdoctoral Association, (NPA), 2011. Postdoctoral Scholars, Gender, and the Academic Career Pipeline. Tech. Rep. National Postdoctoral Association, Rockville, MD (accessed 11.07.17). http://c.ymcdn.com/sites/www.nationalpostdoc.org/resource/resmgr/Docs/postdoc-gender-fact-sheet-20.pdf.

NIH, 2011. A History of New and Early Stage Investigator Policies and Data. Tech. Rep. National Institutes of Health (accessed 11.07.17). https://grants.nih.gov/policy/new_investigators/history.htm.

NIH, 2016a. National Institutes of Health Budget. Tech. Rep. National Institutes of Health (accessed 11.07.17). http://www.hhs.gov/about/budget/budget-in-brief/nih/index.html.

NIH, 2016b. NIH Family-Friendly Initiatives. Tech. Rep. National Institutes of Health (accessed 11.07.17). https://grants.nih.gov/grants/family_friendly.htm.

NSF, 2015. Women, Minorities, and Persons with Disabilities in Science and Engineering: 2015. Special Report NSF 15-311. Tech. Rep. National Center for Science and Engineering Statistics, Arlington, VA (accessed 11.07.17). https://www.nsf.gov/statistics/2017/nsf17310/.

Pohlhaus, Jennifer Reineke, Jiang, Hong, Wagner, Robin M., Schaffer, Walter T., Pinn, Vivian W., 2011. Sex differences in application, success, and funding rates for NIH extramural programs. Acad. Med. 86 (6), 759–767.

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

Sarsons, Heather, 2017. Recognition for group work: gender differences in academia. Am. Econ. Rev.: Papers Proc. 107 (5), 141–145.

Shen, Helen, 2013. Inequality quantified: mind the gender gap. Nature 495 (7439), 22–24.

Stack, Steven, 2002. Gender and scholarly productivity: 1970–2000. Sociol. Forum 35 (3), 285–296.

Steinpreis, Rhea E., Anders, Katie A., Ritzke, Dawn, 1999. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study. Sex Roles 41 (7), 509–528.

Wais, K., 2015. genderizeR: Gender Prediction Based on First Names. Tech. Rep. Github (accessed 11.07.17). https://github.com/kalimu/genderizeR.

West, Jevin D., Jacquet, Jennifer, King, Molly M., Correll, Shelley J., Bergstrom, Carl T., 2013. The role of gender in scholarly authorship. PLOS ONE 8 (7), e66212.

Williams, Wendy M., Ceci, Stephen J., 2015. National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. Proc. Natl. Acad. Sci. 112 (17), 5360–5365.

Wuchty, Stefan, Jones, Benjamin F., Uzzi, Brian, 2007. The increasing dominance of teams in production of knowledge. Science 316 (5827), 1036–1039.

Xie, Yu, Shauman, Kimberlee, 1998. Sex differences in research productivity: New evidence about an old puzzle. Am. Sociol. Rev. 63 (6), 847–870.

Xie, Yu, Shauman, Kimberlee, 2003. Women in Science: Career Processes and Outcomes. Harvard University Press, Cambridge.