The Effect of Mentor Gender on the Evaluation of Protégés

Marc J. Lerchenmueller*
University of Mannheim, L5, 4, 68161 Mannheim, Germany, marc.lerchenmueller@uni-mannheim.de

Leo Schmallenbach
University of Mannheim, L5, 4, 68161 Mannheim, Germany, lschmall@uni-mannheim.de

Karin Hoisl
University of Mannheim, L5, 4, 68161 Mannheim, Germany; Copenhagen Business School, Department of Strategy and Innovation, Kilevej 14A, DK-2000 Frederiksberg; and Max Planck Institute for Innovation and Competition, Marstallplatz 1, 80539 Munich, Germany, hoisl@bwl.uni-mannheim.de

Abstract

The scientific community calls on senior women to help mentoring the next generation of scientists. Yet research indicates that particularly senior women often get undervalued in academia, from hitting glass ceilings to receiving less recognition for accomplishments than male peers. We examine whether these gender differences in evaluations of senior scientists, who are poised to serve as mentors, also affect the evaluation of protégés and their work. Identifying mentors of 4,556 scientists with competitive early career funding from the U.S. National Institutes of Health (NIH), we document a citation discount of 10% on the average paper published by women- relative to men-mentored protégés. Using data on both the publications of the mentored protégés as well as the citing articles, we distinguish supply-side (i.e., scientists' offerings) from demand-side (i.e., actions by the scientific community) explanations. Supply-side factors appear to account for about 40% of the citation discount. The remainder is explained by men citing protégés of women less often than protégés of men. This gender-biased treatment on the demand-side particularly afflicts work in the most impactful research areas that draw the most citations. Although both men and women mentors spur protégés to producing their very best work, the science community gives women-mentored protégés less recognition for it. These findings raise concerns about an unbiased discourse on the best scientific contributions and about systemic limitations to women serving as mentors as a means to closing gender gaps in science.

* Corresponding author

Introduction

Protégés are generally unknown quantities and just embark on an academic career. The lack of a scientific track record gives rise to information asymmetries when the science community seeks to evaluate the work by "newcomers" (Kram 1983, Podolny 1993, Li et al. 2019). Mentors, on the contrary, are usually established scientists with a rich publishing record and their standing in the community provides social collateral that can positively affect the evaluation of protégés' work. Instead of evaluating the protégés' work in the mist of uncertainty that is attached to a fledgling track record, the community can look to the mentor and his or her reputation to draw inferences (Merton 1968, Simcoe and Waguespack 2011)[1].

The gender of mentors becomes salient in this context. This is because of compelling evidence that women who have climbed to the senior ranks of the academe, who are poised to serve as mentors, still get undervalued relative to male peers in the science social system. For example, women are less often deemed worthy of joining prestigious editorial boards, get paid less, and receive less recognition for their accomplishments (Rossiter 1993, Amrein et al. 2011, Jena et al. 2015, Hoisl and Mariani 2017, Ma et al. 2019, Fang et al. 2020). In short, prevailing conditions in science may decrease the extent to which women relative to men mentors can pledge social collateral for their protégés.

Combining the importance of mentors for the evaluation of protégés with the notion of senior women getting undervalued in science, a baseline hypothesis emerges: *Women relative to men mentors negatively (or at least less positively) impact the evaluation of protégés' work by the scientific community*. In a first step, we will test this baseline hypothesis. However, we cannot stop at this point. To reflect on the desirability and promise of potential countermeasures, we need to better understand why any evaluation discount arises. Hence, in a second step, we investigate the reasons for a possible gendered aspect to mentorship.

For our investigation of a possible evaluation discount experienced by women-mentored protégés, we build on the literature on gender gaps in professional labor markets. This literature broadly distinguishes between *supply-side* effects and *demand-side* effects giving rise to gendered outcomes (e.g., Kulis et al. 2002, Ding et al. 2013, Fernandez-Mateo and Fernandez 2016). Supply-side logics focus on women self-selecting into less advantageous positions, in turn, depressing their outcomes. In the context of science, women might choose not to compete for publication spots in top journals or to conduct research on niche topics that receive less recognition, for example. Whatever the root cause, supply-side effects may well impact mentored protégés, both directly through coauthorships with the mentor as well as indirectly, by imprinting a certain research strategy.

[1] The study by Simcoe and Waguespack (2011) provided the first evidence that status advantages can transfer to coauthors in the context of science. Our study differs in examining the transmission of gendered evaluations from mentors to protégés, even if the mentor may not coauthor a focal publication.

Demand-side effects, meanwhile, pertain to actions by the evaluators of women's work and also how these evaluations get shaped in the prevailing social context (Reskin 2005, Fernandez-Mateo and Kaplan 2018, Abraham 2020). One possibility is that women mentors' occupation of lower rungs in the science social hierarchy relative to men mentors (Zuckerman et al. 1991) spills over and dims the esteem of women-mentored protégés. This may translate in women-mentored protégés facing differential access to competitive research areas, for example, despite trying. Another possibility is that the scientific community holds biased evaluations of work by women (Botelho and Abraham 2017, Lerchenmueller and Sorenson 2018). Work of comparable content and quality may still receive less credit. Again, irrespective of the root cause, demand-side effects may well impact mentored protégés.

To test our baseline hypothesis and to analyze these different mechanistic explanations, we use a sample of 4,556 highly qualified postdoctoral students who received mentored early career funding from the U.S. National Institutes of Health (NIH). More specifically, we identify recipients of the prestigious Kirschstein National Research Service Award, the so called "F32 program". This setting offers several conceptual and empirical advantages, especially the fact that established senior scientists must formally mentor an elite cohort of early career scientists during a formative career stage.

An ideal, albeit theoretical, design for our analysis would employ a two-step random allocation. First, mentors would be randomly assigned a gender. Second, F32 recipients[2] would be randomly allocated to women versus men mentors. The first step ensures that mentors do not differ on task-relevant characteristics but only on gender. Random allocation of F32 recipients mitigates concerns that protégés of different quality may be systematically matched to mentors of different gender. This randomized design is for obvious reasons not feasible.

The next best empirical design satisfies two conditions. First, the mentors are observationally equivalent on pertinent quality dimensions and only differ in terms of gender. Second, protégés are of similar quality. To fulfill the first condition, we employ a matched sample approach using detailed information on the mentors' scientific track records. To meet the second condition, we select the F32 program as our empirical setting for its highly competitive nature. Our data support the assumption that successful contestants for the F32 grant are of comparable quality. We further draw on newly constructed forward citation data to assess how protégés' work may be differentially evaluated by the science community depending on their mentors' gender.

We find support for our baseline hypothesis, i.e., protégés of women relative to men mentors receive 10 percent fewer citations, on average. After accounting for observational differences in mentor characteristics, like historic publication and funding records, the citation discount drops to seven percent fewer citations, underlining the importance of the formative mentor relationship. Adding characteristics about the protégés themselves and their actual research, like choice of research topics, does not

[2] In the following, we use the terms F32 recipient, early career scientist, and protégé synonymously.

materially influence the citation discount attached to female mentors (about one percentage point). In all, supply-side factors could account for about 40% of the observed citation disadvantage for women-mentored protégés. In particular, controlling for the *expected impact* of a focal paper – the average citations received by all papers of the same (or highly similar) content published in the same year in the life sciences – does not explain the citation disadvantage. Protégés of women mentors thus do not chose to publish in less promising areas of research but receive less citations for comparable work.

To trace the source of this gender difference in citations for comparable work, we shift our attention to the demand-side. We find that the citation discount associated with women mentors originates from a relative lack of citations on work with the highest *expected impact*. There is no discernable effect on citations when the underlying work deals with less impactful topics. Pursuing the question further of who does not cite, we identify that subsequent publications authored by men cite the focal research at a 10–13 percent lower rate[3]. Women authors of subsequent publications cite women-mentored work at a somewhat higher rate, but that benefit cannot compensate for the former discount as men still outnumber women in academic science. Taken together, we interpret these findings as a demand-side bias afflicting the evaluation of women-mentored research, particularly the most impactful science.

Our contribution is at least twofold. We contribute to the literature on mentorship by first documenting an evaluative discount for protégés mentored by women versus men, while also distinguishing demand-side and supply-side explanations for the observed discount. Second, we forward our theoretical understanding of how collaboration dynamics affect the impact early career scientists may (or may not) unfold on scientific discourses, thereby enhancing the literature on the gender gap in science.

Theoretical Background

Mentorship in academic science

The importance of mentorship in the scientific community is undisputed and actively promoted as, for example, in a recent Editorial of Nature (2019) titled "Science Needs Mentors". The concept of a mentor can be traced back to the story about the Trojan War in Homer's Odyssey, where Athena the wisdom goddess acted as a protector, educator, and guide of a young prince, while Odysseus, his father, fought in the Trojan War. Today, mentors in academia are considered advisors and guides who pave the way for their protégés to enter and prosper in the academic world.

[3] This discount afflicts the evaluation of protégés irrespective of the gender of the protégés (see Appendix 1).

Mentors can fulfill various important functions to the benefit of their protégés[4]. They provide resources in support of high-quality research and career progress, such as research funding and the necessary infrastructure (Kram 1988, Scandura 1992, Allen et al. 2004). In addition to money, mentors can provide their protégés with a better understanding of how the publication and grant application processes work (Etzkowitz et al. 2000, Preston 2004). Mentors can also introduce protégés to potential collaborators. Empirical studies have repeatedly associated these functions with different measures of productivity gains for protégés (see Johnson (2007) for a literature review and Eby et al. (2008) for a multidisciplinary meta-analysis). Results show that mentorship (as opposed to no mentorship) is positively related to, for example, number of conference presentations, research output (number of publications), and research funding (Johnson 2007, Malmgren et al. 2010). Mentors are even credited for having a positive influence on the likelihood of protégés staying in academia (Dohm and Cummings 2002, 2003), an outcome tightly coupled to productivity. Mentors may also affect their protégés' progress via advancing their political and social skills. Comparing academics with similar levels of performance, studies have suggested that mentored early career scientists more often advance to higher ranks in the academe and earn more money, presumably attributable to better networking or visibility flowing from more intense mentorship (e.g., Kirchmeyer 2005).

In addition to impacting productivity, mentors can also serve in a certification function once protégés' work is published and enters the realm of evaluation by the community. A relationship with a high-quality mentor can send a signal to the research community about the abilities of the otherwise difficult to evaluate "newcomer", alleviating information asymmetries. Tellingly, early career scientists who publish with top senior scientists often have a competitive advantage over young scientists without such coauthors in terms of getting noticed and cited (Li et al. 2019). A similar, though subtly different effect emerges with potential status transfer from the mentor to the protégé. The status of the mentor can serve as a pledge of social collateral that elevates the protégé's standing in the scientific community, irrespective of coauthoring (Ridgeway 1991, Podolny 2001, Azoulay et al. 2014, Reschke et al. 2018). Protégés often have little-to-no status as a virtue of being early in their careers and, hence, depend disproportionally on the status of their mentors. Conceivably, such status transfer also improves the reception of independent work produced by the protégé.

Compared to the rich evidence on productivity gains that exists in meta-analytic form, there is scarce evidence on the effect of mentors on the evaluation of their protégés. Notable exceptions at least point to the relevance of this certifying mechanism, showing that an eminent mentor, such as an individual participating in the federal advisory system or with honorary degrees, is found to be

[4] For an extended review of the literature on the effects of mentorship see Appendix 2. All findings relevant to our argument are included in the main text.

positively related to the protégé getting a tenure-track job and succeeding in such a position (Reskin 1979, Waldinger 2010).

Taken together, existing work has mainly focused on productivity differentials of mentored versus non-mentored young scientists. The impact of different types of mentors, i.e., women versus men mentors, has not yet been systematically analyzed.

Mentor gender and the evaluation of protégés

There is a stubborn prejudice that women are worse scientists than men (e.g., AlShebli et al. 2020). Although research has made strides to undermine this view, we feel compelled to begin our theoretical evaluation of gendered mentorship effects with this notion. Women have been shown to be underrepresented in jobs that are perceived to require intellectual ability, such as science and engineering (Leslie et al. 2015, Meyer et al. 2015, Cimpian and Leslie 2017). Even though this gender gap is often attributed to a lower intellectual ability of women compared to men, such a difference has no empirical grounding. On the contrary, intellectual ability has been shown to be gender neutral (Valian 2007, pp. 32-34). However, stereotypes exist that associate brilliance to men more than women (Bian et al. 2017, 2018, Jaxon et al. 2019, Rivera and Tilcsik 2019). Other reasons for gender differences in scientific outcomes may be a lower level of qualification due to a lower level of education of women compared to men. However, women have gained the same access to the education system as men over time and a career in academic science has always required a university degree and a doctorate. Consequently, educational differences cannot explain possible gender differences between mentors who cleared these hurdles (Weichselbaumer and Winter-Ebmer 2005). Considering existing evidence, differences in innate aptitude seem unlikely between men and women mentors.

If not aptitude, perhaps women and men mentors differ in the way they contribute their abilities to the furthering of their protégés. Intensity or quality of supervision is hard to observe. Most informative work in this regard stems from smaller scale studies in psychology, attesting that women mentors tend to establish less power-asymmetric mentor-protégé relationships and cultivate a broader flow of information (Schwiebert et al. 1999). Men are reported to provide more career mentoring to their protégés, whereas women provide more psychosocial mentoring (Allen et al. 2004). Protégés of women can learn strategies for coping with the challenges that arise from stereotype-related barriers to growth and career advancement (Ragins and Mcfarlin 1990). If at all, the literature thus tends to indicate that women and men have different foci or take different roles when mentoring their protégés. But that does not mean that supervision quality is gendered. To that end, we also find no indication for potential differences in foci materially impacting protégés' productivity per se (Pezzoni et al. 2016, Gaule and Piacentini 2018).

The preceding arguments do not rule out, however, that mentor gender plays a role for the evaluation of protégés' work. As the literature does not hold noteworthy evidence on gender differences in ability or willingness of mentors, what remains from a *supply-side* perspective is that women mentors prefer to pursue different research with their protégés relative to men mentors. One of the major concerns for the evaluation of protégés is that women mentors concentrate their research in niche fields (Leahey 2007, Leahey et al. 2010), possibly to avoid competition (Gneezy et al. 2003, Croson and Gneezy 2009). Yet these fields may be less relevant to the research community and trigger a citation discount. While the literature has documented this dynamic for women in science more generally, it is unclear to what extent this holds for women mentors who have successfully climbed to the upper ranks of science. Another possibility is, that women tend to work in areas that are particularly relevant to women. Koning et al. (2019), for instance, show that patents listing female inventors are more likely to focus on female diseases. The underpinning research is presumably more relevant for other women than for men. Since there are fewer women than men in science (i.e., the interested community is smaller), such a gender-based research focus would almost automatically lead to fewer citations. In case protégés work in the same or closely related research areas as their mentors, a citation discount for women- relative to men-mentored protégés is plausible.

The impact of protégés may also be influenced by *demand-side* effects, i.e., actions and perceptions of the research community towards the mentor. There is rich evidence that senior women get undervalued in science. Women get less often invited to prestigious editorial boards (Amrein et al. 2011), hit glass ceilings in up-or-out careers (e.g., Tesch et al. 1995), earn less (across career steps) (Jagsi et al. 2012, Hoisl and Mariani 2017), and get less credit for their accomplishments (e.g., Ma et al. 2019). One may sum these disadvantages to an evaluative discount faced by senior women in the science community. If the lower social rank of senior women relative to male peers spills over and also dims the esteem of women-mentored protégés (Podolny 1993, Ridgeway 2001, Correll 2004, Reschke et al. 2018), this may affect the reception of protégés' work in multiple ways, from facing social closure in research circles (de Vaan and Wang 2020) to differential access to publishing avenues. As such, observing work of women- versus men-mentored protégés on less competitive topics or journals can both result from supply- and demand-side effects. Although, again, it seems unlikely that women mentors who have succeeded in science lead their protégés down a less competitive path, we need to disentangle these explanations to the extent possible. In that vein, if we consider comparable work of women- versus men-mentored protégés, as indicated by, for example, targeting questions of similar importance, the scientific community may still hold gender biased evaluations (Botelho and Abraham 2017, Lerchenmueller and Sorenson 2018)[5]. As women compared to men mentors can likely pledge

[5] For a detailed discussion of how status can decouple from quality and how social influence can lead to nominal characteristics, such as gender, depressing social evaluations despite repeated and unambiguous demonstrations of quality see Lynn et al. (2009).

less social collateral to inform the communities' evaluation, women-mentored protégés may suffer a citation discount as a result (Merton 1968, Simcoe and Waguespack 2011).

In the following, we will employ a research design that identifies mentors and protégés that begin the mentorship journey from a comparable starting point, and we investigate to what extent supply-side or demand-side effects explain the impact of early career scientists who are mentored by a woman versus a man.

EMPIRICAL APPROACH

Our empirical setting is a mentored, early career fellowship in the academic life sciences that is sponsored by the U.S. National Institutes of Health (NIH). The NIH is the largest funder of scientific research globally with an annual budget of approximately USD 30 billion. At any given time, roughly 80% of U.S. life science laboratories receive funding from the NIH (Li 2017). As part of its mission, the NIH also commits to developing the next generation of scientists. In particular, the agency administers the congressionally mandated NRSA program[6] that invests roughly USD 750 million per year to further capabilities of the young research workforce. The program's central funding mechanism is the so called "F32 grant", offering up to three consecutive years of mentored research support with an average annual grant size of about USD 50,000 (Lerchenmueller and Sorenson 2018).

The F32 grant is a very suitable setting to study the effects of formal mentorship among a cohort of highly qualified protégés. Very important for our research design is the fact that having a dedicated mentor is a precondition for the F32 fellowship. As part of the program, young scholars are required to join a new research environment both in terms of content and people. The often freshly minted PhDs ought to become acquainted with new methods and research questions to broaden their scientific repertoire. The NIH therefore generally expects F32 recipients to leave their terminal degree-granting institution. The mentor then assumes responsibility from the application stage in co-devising research the early career scientist will conduct as F32 fellow. During the fellowship, the mentor guides the protégée in novel techniques for addressing research questions that often fall within, or close to, the mentor's area of expertise. Since life science research projects often span several years from inception to publication, mentors commonly collaborate with their protégées for years after the fellowship[7]. Finally, life science research is resource intensive, needing laboratory personnel and equipment. Hence, the protégée's impact is also dependent on the mentor's ability to support the outlined research project (NIH 2021b).

[6] Ruth L. Kirschstein National Research Service Award (NRSA) for high potential early career scientists.
[7] This dynamic is also documented in our data on F32 recipients' publication histories (see Appendix 3a).

Analytic strategy

To be able to answer our research question, we have to make sure that mentors do not differ on task-relevant characteristics but only on gender and that early career scientists of different quality do not systematically match to men and women mentors. Since a random allocation of gender to mentors and protégés to mentors is for obvious reasons not feasible, we will use our unique context and rich data to address endogeneity concerns in the best possible way.

To ensure that we analyze men and women mentors of comparable quality, we match mentors on central characteristics that are also evaluated by the NIH during the F32 grant approval process[8]. To that end, we use granular data on mentors' scientific track records for the ten years preceding the focal F32 grant application. To ensure that F32 recipients are of similar quality insofar that the likelihood of better early career scientists systematically matching to a mentor of certain gender is negligible, we make use of the competitive nature of the F32 program. The F32 grant stipulates eligibility criteria to producing a set of homogeneously qualified awardees, limiting the impact of unobservable characteristics on selection. To begin with, all applicants must be either U.S. citizens or permanent residents and hold a research relevant terminal degree (usually MD, PhD or both). The F32 application then gets reviewed by an NIH Scientific Review Group (SRG), which, in essence, are self-governed bodies of expert scientists providing review services to the individual thematic Institutes that make up the NIH. Five dimensions get evaluated by the SRGs: (1) observable characteristics of the applicant (i.e., research record), (2) observable characteristics of the mentor (i.e., research record), (3) applicants' research potential (e.g., reference letters and statement of purpose), (4) research environment during the fellowship (i.e., host institutions' research infrastructure), and (5) the proposed research[9] (e.g., fit with mentor's expertise). The SRG ranks all dimensions from 100 (best) to 500 (worst)[10] and sends the average score to the relevant Institute for a final decision. With a mostly meritocratic allocation of funds based on the SRG review, the set of recipients is obviously even more homogeneous in terms of quality than the broader applicant pool was to begin with[11].

Still, we consider that the final decision by the Institutes may include an unobservable component alongside the evaluation of observables. The individual Institutes may also appraise applications that are in principle worthy of funding, considering, for instance, topical priorities or aspects of workforce diversity. We are concerned that early career scientists and mentors may receive F32 funding correlated with their gender, although they may score somewhat lower on observables.

[8] See Heggeness et al. (2018) for a detailed outline of the F32 program criteria and process. All F32 program features pertinent to our empirical design are described in detail in our main text.
[9] We only analyze grant recipients and can therefore assume that proposals meet a certain quality threshold.
[10] The scoring scale changed in 2009. We constrain our observational window of early career scientists' records from 1985 to 2009.
[11] The broader applicant pool already refers to the top 60% of applications only because the Institutes usually discard the bottom 40% of applications (Heggeness et al. 2018). In other words, the pool considered for eventual funding represents already a pre-selection in a very competitive program.

Recent research based on administrative NIH data suggests that this "error component" varies across Institutes and time and that, on average, one out of eight F32 grants falls out of strict scoring (Heggeness et al. 2018). Consequently, our design would be unaffected by these unobservable determinants on almost 90% of the observations (note that we analyze only F32 recipients not applicants). Nonetheless, we incorporate F32 recipients' gender, alongside their detailed research records, into our matching and as control variables.

An evaluation of mentors as well as their protégés in terms of research record (quality and quantity) makes us confident that our approach is effective in mitigating both sorting dynamics and in analyzing comparable mentors and protégés at time of F32 grant receipt (Figure 1). Before matching, comparing men and women who served as mentors shows that men mentors published more papers on average and in higher impact journals[12] and received more major grants (R01s)[13] within 10 years prior to mentoring the F32 recipient. This is to be expected, at least to a certain degree, because the cohort of senior scientists that could serve as mentors for F32 fellowships since 1985 invariably reflects a generational gender gap, pooling relatively less experienced women with more experienced men[14]. After matching on these pertinent variables plus the fiscal year of the F32 grant (to capture cohort effects) and plus the gender of the F32 recipient, we obtain a set of comparable women and men mentors. Meanwhile, in line with our expectations for the chosen empirical setting, the protégés are already very similar across mentor gender before entering the F32 program. Protégés were affiliated with institutions of similar standing[15] and published a similar share of papers in high impact journals. Only in terms of productivity do protégés with women mentors seem to have had a small advantage over those with men mentors. Yet, this difference is arguably small (less than half a paper) and would suggest, if anything, that protégés who end up with women as mentors were slightly more productive before receiving the F32 grant.

[Figure 1 about here]

The homogeneity among mentors of different gender and their protégés is further confirmed in a logit regression estimating the likelihood of the mentor being a woman (Appendix 4). After matching, neither mentor nor F32 recipient characteristics predict the mentor's gender (Kodde and Palm 1986). The one significant (positive) predictor of the mentor being a woman is the number of papers F32 recipients published before the grant. We control for this difference in F32 recipient pre-grant productivity, favoring women mentors, in all our analyses.

---

[12] We categorize journals with an impact factor exceeding 10 as high impact journals. This categorization captures both the top field journals as well as top general science journals.
[13] The R01 grant is the NIH's central funding mechanism for principal investigators (i.e., laboratory leaders).
[14] Women accounted for 12% of doctoral degrees in the biomedical sciences in the 1960s and around 30% in the 1980s, invariably constraining the pool of very experienced female mentors for our F32 cohorts (AAUW 2015).
[15] Approximated by the institutions' percentile ranking in terms of receiving major NIH grants (i.e., R grants).

To be clear, any design short of the theoretical benchmark of two-step random allocation of gender to mentors and early career scientists to mentors leaves room for counterfactuals. We submit that our design – matching on and controlling for mentor characteristics as well as analyzing a homogeneous set of early career scientists – reduces the likelihood that any gendered mentorship effect stems from quality differences among mentors or protégés matched to mentors. In summary, we are confident that the setting and matching design we choose allow us to estimate how equally qualified mentors of different gender affect the evaluation of promising early career scientists.

Data and methods

To examine our research question, we require detailed research records of life scientists. Because no existing data set offers the needed depth of information, we merged six databases – PubMed, Clarivate's Journal Citation Report, Scopus, genderize, Author-ity, and the NIH ExPorter. PubMed is the standard and most complex bibliographic reference for the life sciences. To date, the database contains over 30 million scientific publications associated with about 100 million individual authorships. We parsed the PubMed XML database to obtain for each publication record, amongst other information, the publishing journal, time of publication, article content, and author names and order on the author byline. We supplemented journal impact factors (JIFs) for the journals in which the research was published from the 2018 Journal Citation Report[16], using unique ISSN numbers as a crosswalk. We furthermore added downstream citations for all articles from Scopus by Elsevier. Scopus and PubMed data can be merged via a common unique article identifier (the PubMed ID, or PMID). Scopus records detailed information on the citing articles, including author names and positions on the byline, journals and issue. We supplemented this information with probabilistic gender designations of authors using the genderize.io database[17] and employed this newly enriched data to analyze how downstream citation patterns differ for work by women- vs men-mentored F32 recipients. Finally, we drew on the Author-ity database to determine whether two individuals of the same (or very similar) name authoring two articles actually represent the same or different individuals. Author-ity uses author names, affiliations, coauthor networks, and scientific focus to disambiguate authors listed across articles (Torvik et al. 2005). Research based on an external gold standard of scientist IDs maintained by the NIH has shown that the Author-ity algorithm achieves more than 99% accuracy across, among other aspects, author productivity and gender (Lerchenmueller and Sorenson 2016). These data allow constructing detailed longitudinal research records of both mentors and protégés[18].

[16] We tested bivariate correlations of various journal impact factors across the years included in our analyses and obtained correlation coefficients in excess of 0.90, indicating little temporal variance in the scaling of the metric. We assigned journals without a listed impact to the lowest impact category, as the Journal Citation Report requires a minimum impact threshold for inclusion in the index.
[17] Genderize is a commercial service that uses a variety of information, such as social media records, to assign a probability that an individual with a particular forename is a man or a woman (Appendix 5 has further details).
[18] Appendix 6 presents a diagram summarizing the steps of sample construction.

We next used the NIH ExPORTER database to identify the focal mentors and protégés. The NIH records all funded projects with a unique grant ID in this database. All funded scientists are obliged to acknowledge the grant ID in resulting publications, with failure to do so punishable by disqualification and potentially by federal law. NIH grant IDs in acknowledgement sections of papers have an extremely high fidelity. We mined the ExPORTER for F32 recipients and identified the publications that stem from F32 grant funding. These publications enabled us to identify the F32 recipient's mentor in an unusually precise way. Generally, mentorship is an interpersonal interaction that evades formalism and is hard to observe and trace empirically. Besides the F32 program, we exploited a long-standing norm in the authorship order of academic articles in the life sciences to identify mentors: the senior investigator receives the last author position on the article byline (e.g., Levitt 2010). We took the first article with three or more coauthors[19] that acknowledged the F32 grant ID and determined the individual who served as the last author as the F32 recipient's mentor.

To probe the validity of our mentor identification, we examined coauthorship patterns of F32 recipients with their mentors for our full sample. Within the first ten years after F32 grant receipt, they coauthored five papers, on average. Within the first five years after grant receipt, the mentor even appeared as a coauthor on more than every second publication of the F32 recipient, on average (see Appendix 3a). Additionally, we conducted a manual search of a random draw of 100 mentor-protégé dyads, and we verified a connection beyond the first paper that acknowledged the F32 grant ID for over 95% of the dyads. Examples for such connections include referencing the relationship in academic CVs or presentations of F32 recipients' profiles on the mentors' websites (see also Appendix 3b).

Sample

Starting with 4,808 F32 recipients identified via the outlined steps, we applied a number of exclusion criteria in service of more accurate effect estimation. First, we removed 248 F32 recipients who appeared unlikely "early career stage" at the time of F32 grant receipt. This applies, for example, to F32 recipients who published more than 10 years or more than 10 papers prior to receiving the grant. Second, we removed publications where we do not expect the F32 mentorship to play a significant role. This includes papers published before F32 grant receipt, published 10 years after the F32 grant receipt, or published by more than 10 researchers. Lastly, we needed to exclude four F32 recipients whose publication data were incomplete (e.g., no information on forward citations to the focal article). The final sample comprises 58,921 journal articles published by 4,556 F32 recipients with 921 (20.2%) of recipients being women-mentored and 3,635 being men-mentored (79.8%).

[19] Requiring three or more coauthors increases the likelihood of senior authorship on original research because, in the life sciences, articles with less than three authors more often fall outside of original research (e.g., editorials). Our set of mentors remains fairly stable if we relax that condition or impose additional conditions (e.g., repeat coauthoring of mentor and protégé on papers acknowledging the F32 grant).

To trace how the impact of these early career researchers varies with mentors' gender, our analyses focus on the publications of the F32 recipients. To address endogeneity concerns induced by potential gender differences in mentor quality, we used coarsened exact matching (CEM)[20] to pair publications of women-mentored F32 recipients with publications of men-mentored F32 recipients based on mentor characteristics observed at the time of grant receipt. We first coarsened the matching variables, then defined strata based on the linear combination of coarsened variables and assigned our observations relating to women and men mentors to these strata. Only treated observations, i.e., publications of F32 recipients with a woman as mentor that are assigned to a stratum with at least one publication of F32 recipients with a man as mentor (control) were kept for analysis. Finally, we assigned weights to these observations to ensure balance within and across strata[21], and we include these weights in our regressions.

Our matching variables comprise (i) quartiles of mentor productivity ten years before grant receipt, (ii) quartiles for the share of these publications appearing in high-impact journals (i.e., with an impact factor exceeding 10), (iii) quartiles for the number of major (i.e., R01) grants the mentor received before the F32 grant fiscal year, (iv) the gender of the F32 recipient, and (v) five-year brackets for the fiscal year, in which the F32 grant was awarded. The first two variables ensure that we compare women and men serving as mentors with a similar publication track record as indicator of scientific ability. The third matching variable controls for resources available to the mentors (e.g., laboratory infrastructure) that also benefit their protégés. Matching on the gender of the F32 recipient levels out potential homophily tendencies in the mentor-protégé allocation[22] and demographic considerations in grant allocation by the Institutes. The controls for the time periods account for the fact that women mentors together with the number of women in academia have increased over the last decades. The matched sample contains 46,577 papers published by 915 F32 recipients with a woman and 2,937 F32 recipients with a man serving as mentor.

Description of the variables

Table 1 provides definitions and descriptive statistics of the key variables.

[Table 1 about here]

[20] Research suggests that CEM has several advantages over other techniques that also match on observable characteristics, for example, reducing model dependence (for a detailed review, see Iacus et al. (2012)). We execute matching without replacement. Our results based on CEM can be reproduced using propensity score matching as an alternative matching technique (see Appendix 12b).
[21] For a detailed explanation of weights see King (2012).
[22] The literature broadly distinguishes between choice and induced homophily. In our context, homophily is more likely to stem from choice, given the absence of organizational or geographical constraints (F32 recipients are expected to leave their PhD granting institution before starting their postdoc) in the mentor-protégé allocation (Kleinbaum et al. 2013).

Dependent variable. The focal dependent variable of our analysis is the number of forward citations a paper receives. Citations serve as the institutionalized metric for the impact of academic work (Merton 1988) and play a central role in hiring, pay, and promotion decisions. Citations further serve to locate scientists' relative rank in the academic social order (Azoulay et al. 2014), which allows gauging the influence of mentors' standing on the communities' evaluation of protégés. We exploit up to date information on paper-level citation counts, adding publication year fixed effects to control for different time spans during which individual papers could accrue citations. On average, the papers published by protégés received 84 citations (min=0; max=12,928). The underlying distribution is right-skewed, directing our subsequent modelling choices.

Independent variable. Our main independent variable is binary and takes on the value of one if the protégés' mentor is a woman and zero otherwise. Roughly one of five articles (18%) in our dataset was published by protégés who are mentored by a woman.

Control variables. The control variables can be grouped into three categories, respectively capturing characteristics of the mentor, the F32 grant and recipient, and the focal research. The control variables for the quality of the mentor are the same as the matching variables, i.e., the mentor's number of publications ten years prior to the F32 grant fiscal year, the share of these publications in high impact journals (journal impact factor > 10), and the number of prior R01 grants. While the weights obtained from the CEM procedure ensure balance with respect to these matching variables within and across strata based on the coarsened categories, adding the continuous variables to the regression accounts for any residual imbalance[23]. On average, mentors had 40 publications with about 15% placed in high impact journals. Lending further credence to our identification approach, mentors in our matched dataset received four R01 grants, on average, before serving as a mentor on the F32 fellowship. We also use fiscal year fixed effects to control for the year in which the F32 grant was awarded to account for the increasing representation of women across academic cohorts. This trend should affect the likelihood of observing a woman as mentor and the perception of women in the scientific community. Fiscal years vary between 1985 and 2005, with the median year corresponding to 1995.

At the level of the F32 grant and the F32 recipients, we first control for the gender of the F32 recipients (one if female and zero otherwise). Although we cannot detect homophily in protégé-mentor pairing in our matched dataset, adding protégé gender still accounts for the small fraction of F32 grants[24] that may possibly be allocated by the Institutes based on gender equality considerations. Overall, 34% of the F32 recipients in our sample were women. We control for variance in the research infrastructure of the host institutions by including institutions' percentile ranking in terms of major NIH funding (R

[23] See Blackwell et al. (2009) for a detailed discussion, especially pages 537-538. Of note, because CEM strictly bounds the level of model dependence, the model specification itself is much less consequential.

[24] Again, about 10% of F32 grants appear to be allocated not strictly by the scored pay line, according to NIH administrative data (Heggeness et al. 2018).

grants). The academic institutions in our sample are, on average, amongst the top 5% of this ranking. We also include a dummy variable scoring one when the F32 grant was extended (zero otherwise). Grant extensions under the F32 fellowship are generally only permissive for parental leave[25] and, besides the actual leave days, this variable would capture the broader ramifications of a family extension on a protégé's work life, potentially more relevant for female than male protégés (Thebaud 2015). Overall, about one in ten F32 grants were extended.

Although the pool of F32 recipients is fairly homogenous by design, we still add several variables capturing the pre-F32 quality of the protégés. Controlling for any observable differences in protégés' publications prior to receiving the F32 grant reduces the likelihood of selection influencing our estimation. On average, F32 recipients published three papers before receiving the grant with 7% appearing in high impact journals. To account for emerging productivity differences of the protégés during the mentorship period, we include a count variable for the publications starting from the first publication after F32 grant receipt and a variable indicating the number of papers published in the year before the focal paper's publication year.

Next, we account for the characteristics of the underlying research to ensure that we compare similar work, both in terms of content and quality. We include the number of authors and the share of women coauthors on the focal paper. Women-mentored F32 recipients may be more likely to coauthor with other women (Holman and Morandin 2019) and, due to potentially smaller networks of their women mentors, in smaller author teams (Woehler et al. 2021). Both could result in fewer citations to protégés' work. Author team size may also capture effective division of labor, presumably improving the odds of producing good research that attracts citations (Wuchty et al. 2007). The average article is written by five authors (solo authorships are the minimum and 10 authors the maximum). The share of women coauthors (excluding the mentor if the mentor was among the coauthors) was 28% on the average paper (min=0%; max=100%). We control for the quality of the underlying work via the impact factor (JIF) of the publishing journal. The average JIF of the publications in our sample was 5.9 and varies between 0 (i.e., the journal is not listed in the Clarivate Journal Citation Report) and 70.7 (the New England Journal of Medicine). We add fixed effects for the publication type (e.g., article or letter) since some types might attract more citations than others[26] and mentors of different gender may pursue different publishing formats. Next, we control for the presence of the mentor in the author team to account for the possibility that women or men are more likely to engage in direct coauthoring when mentoring.

Finally, we include two variables to account for possible and pertinent differences in the nature of the underlying scientific work. First, we seek to estimate citation differentials experienced by

[25] Any leave exceeding 15 days of sick leave are reserved for parental leave. Part-time training is generally not accepted under the fellowship (NIH 2018).
[26] A detailed overview of the different publication types is provided in Appendix 7b.

protégés only for research that was guided by comparable levels of expertise contributed by the mentors. To that end, we create a variable that captures the research proximity between the mentors' historic work and the protégés' current work, calculated as the number of MeSH (Medical Subject Headings) terms[27] on a protégé's publication that coincide with MeSH terms appearing on the mentor's publications within 10 years before mentoring the F32 recipient. The MeSH thesaurus is a controlled and hierarchically organized vocabulary, and the terms are used for indexing articles for PubMed. Importantly, the MeSH terms get externally assigned to articles by specially trained librarians of the National Library of Medicine, thus evading subjectivity by individual authors. Publications by protégés shared about seven MeSH terms (or 51%) of assigned keywords with historic articles of their mentors, on average.

We also create a variable for the expected impact of a given piece of research, again using the MeSH terms that get externally assigned by the National Library of Medicine. For each keyword that was assigned to an article published by a protégé in our sample, we determine the number of citations that keyword received across *all* articles published in the same year and indexed in PubMed. Since articles usually get assigned several keywords to accurately describe research content, we take the average number of citations associated with the keywords on the focal article published by the protégé. To norm the scaling of this variable, we rank this average relative to the corresponding distribution of all articles recorded in PubMed for the same year. This percentile ranking effectively captures the likely impact of the protégés' work, that is, the number of citations the article is expected to draw given its content[28]. Lastly, we also add publication year fixed effects to pick up time trends and differences in citation accrual periods across publications. We include a correlation matrix for all variables in Appendix 7, indicating that multicollinearity is not of concern.

RESULTS

Figure 2 compares the mean number of forward citations to publications of female- versus male-mentored F32 recipients. Within the first ten years after receiving the F32 grant, F32 recipients mentored by men receive roughly 10% more forward citations on their average paper compared to women-mentored protégés. In a two-tailed t-test this difference is statistically significant at the 1% level (we detect no difference in citations prior to F32 grant receipt). To better understand where these differences come from, we run a series of multivariate regressions to which we hierarchically add the explanatory variables introduced in the previous section.

[Figure 2 about here]

[27] For details see NIH (2021a).
[28] Appendix 8 provides examples for further illustration.

Supply-side effects

First, we focus on supply-side effects, i.e., the characteristics of the mentors, their protégés, and the chosen research fields. As our dependent variable – the number of forward citations – is a count variable and overdispersed, we estimate negative binomial regression models. An inspection of the data confirms that this model specification fits our data best (see Appendix 9). We report inverse rate ratios (exponentiated coefficients). These coefficients can be interpreted as percentage changes. A coefficient of one indicates no effect, coefficients above one indicate positive effects, and coefficients below one negative effects. Since an individual protégé generally (co)authored multiple publications in our sample, we cluster the standard errors in all our models at the F32 recipient level. On average, F32 recipients publish more than 10 papers within the first 10 years after receiving the grant.

[Table 2 about here]

The estimations of supply-side effects are summarized in Table 2. Consistent with the descriptive evidence in Figure 2, Model 1 shows that, on average, a woman mentor is associated with 10% fewer citations ($p < 0.01$) accruing to protégés' publications. This baseline model is estimated on the full pre-matching sample and includes only publication year fixed effects as controls.

The second model is based on the matched sample. Besides matching weights and fixed effects for academic cohorts, we add mentor characteristics that account for any remaining variance within strata of matched mentors. Accounting for observational differences in mentors' track record reduces the citation discount experienced by women-mentored protégés from 10% to 7% ($p < 0.05$). In particular, the share of the mentor's prior publications placed in high-impact journals – an indicator of the quality and potential impact of the research – appears to also influence forward citations to papers published by the protégés. The reduction in the effect of women mentors on forward citations between the first and the second model (30%) underlines the importance of this formative relationship for the protégé and the attention the work receives.

Model 3 additionally controls for the characteristics of the F32 recipients and their F32 fellowship. While some of these variables, such as the gender of the F32 recipient, the impact of the F32 recipient's publications prior to receiving the grant, or their productivity are significantly related to the number of citations their average paper receives, the effect of women mentors drops by one percentage point. This comparatively lower effect of protégé relative to mentor characteristics is not surprising given the homogeneity and competitiveness of the pool of F32 recipients, which was also documented in our previous analysis of F32 recipient quality by mentor gender.

Model 4 examines the effect of key article-level features on forward citations and the residual influence of women mentors on forward citations. While the female mentor dummy itself changes marginally (0.5 percentage points), the mentor matters. Coauthoring with the mentor boosts forward citations by roughly 9%, likely due to status conferral and greater visibility of the work. It also pays for the protégé to publish on topics related to the mentor's expertise, even if the mentor is not on the paper. It is noteworthy that these mentor-related effects appear largely independent of the mentor's gender as indicated by the stable female mentor dummy. Indicators for article level quality – number of authors as a proxy for expertise breadth and the impact factor of the publishing journal – are positively related to forward citations (about 5% increase for one additional author or a one unit increase in the journal impact factor), but also do not change the citation discount associated with women mentors. In other words, the stable citation discount of about 5% ($p < 0.05$) associated with women mentors appears neither related to visible mentorship style (coauthoring and topic counselling), nor to the quality of the work being produced. Accounting for article level features did, however, improve the precision of our estimates. The standard error on the female mentor dummy, for example, dropped by almost 20%.

Finally, we account for the possibility that the citation discount stems from women-mentored protégés selecting into fields that draw fewer citations (Model 5). To model that choice at a granular level, we make use of the externally assigned keywords (MeSH terms) to life science articles and control for a publication's expected impact given its content. As described above, we determined the average number of citations a given keyword on a 'protégé article' received across *all* articles recorded in PubMed. We then ranked the publication's average across all its keywords against all other publications appearing in the same year. This percentile ranking effectively captures the expected impact of the protégés' chosen research relative to all life science research indexed in PubMed. Model 5 shows that there is a sizeable and monotonic return to pursuing research that is poised to attract citations. Relative to pursuing research in the bottom impact quartile (base category), there is roughly a 40% to 50% to 75% increase in citations as one selects topics in the second, third, and fourth quartile of the expected impact distribution, respectively. More importantly, accounting for the choice of research area does not impact the citation discount associated with female mentors (5%, $p < 0.05$). Women- relative to men-mentored protégés appear to not systematically sort into research that differs in terms of drawing citations.

Indeed, when inspecting the representation of protégés across quartiles of expected research impact, we observe that the share of women-mentored protégés is fairly stable around 20% (Figure 3), if anything, slightly increasing with the expected impact of the research.[29]

[Figure 3 about here]

[29] In the same vein, we examined whether women-mentored protégés may publish in journals with smaller or broader audiences as indicated by journal impact and, again, do not find differences (see Appendix 10).

Demand-side effects

Although we do not discern a statistical difference in protégés' selection of impactful research areas, their work may be differentially received across these research communities depending on whether the mentor is a man or a woman. Returning to our last regression specification (Table 2, Model 5), we interact the female mentor dummy with our quartile dummies capturing the expected impact of the research. The interaction terms effectively probe whether the citation discount associated with women mentors (5% on average) differs across research areas poised to attract less versus more citations.

Figure 4 shows that the protégés of women mentors experience a citation discount only on research expected to be of high impact, that are poised to draw citations. The citation discount in the upper two quartiles is about 7% and statistically distinguishes from a null effect (red line). By contrast, we cannot discern a citation discount in the bottom two quartiles. If anything, there is a positive trend in citations for women-mentored protégés in less impactful research areas[30]. Importantly, the citation discount in the upper quartiles of the expected impact distribution is unrelated to protégés' abilities, the abilities of their mentors, or the content of their research as we account for these counterfactuals. In other words, this demand-side citation discount exists net off considered supply-side effects.

[Figure 4 about here]

To trace the origin of this citation discount afflicting women-mentored protégés – to address the question of "who does not cite?" – we split the forward citations by gender of the first and last authors of the citing article. Again, first and last authors are credited with crafting the article and with identifying and situating the research topic in the broader scientific discourse, respectively. We therefore assume that these authors hold the greatest sway over what preceding work is cited. Table 3 first shows the effects for research expected to be of high impact. Men in first and last author positions cite women-mentored protégés about 10%–13% less ($p < 0.01$) on work poised to attract citations (Models 1 and 3)[31]. The relative citation advantage gained by women-mentored protégés when the last author of the citing article is also a woman (Model 4) cannot compensate the overall citation discount since the majority of citing authors are men (see Figure 5). By contrast, men in first and last author positions do not cite women-mentored work less when the work is of lower expected impact (Models 5 and 7). Together with the stratified citation discount presented in Figure 4, we interpret these findings as a demand-side bias against the most impactful work of protégés who are mentored by a woman.

[Table 3 and Figure 5 about here]

---

[30] The wider confidence intervals for the lower ranked research areas prevent drawing definitive conclusions due to the fewer number of papers published in these less impactful areas. We will return to this finding in the discussion in connection to our empirical setting of the competitive F32 program.

[31] Although the standard errors are wider in the bottom quartiles due to the fewer number of papers, the point estimates indicate a substantively different citing pattern by men in less versus more impactful research areas. The increased rate at which women cite women-mentored protégés in bottom quartiles is also seen in Figure 4.

Additional Analyses and Robustness Checks

We conduct a series of additional analyses and robustness checks summarized in the Appendix. First, we probe whether the identified citation discount associated with women mentors also stratifies with the gender of the protégé. Although this is not germane to our research question, any stratification effect would inform who is most afflicted by the demand-side bias on women-mentored work. Looking at the locus of the citation discount detected in the previous analysis – protégé research of expected high impact and downstream citation rates by men versus women – we observe a negative gradient in the citation discount: fewer citations for female versus male mentors and fewer citations for female versus male protégés (Appendix 1a and 1b). In other words, the citation discount associated with women mentors appears to equally apply to both male and female F32 recipients. Yet, as the accuracy with which we can run this post-hoc analysis precludes drawing definitive conclusions, we do encourage further research into this question.

Next, we investigate how the matching in service of comparing equally qualified mentors (that only differ with respect to gender) alters the representation of mentors and their protégés in our sample relative to all F32 program participants we started with (Appendix 11). We run a logit regression at the level of the mentor-protégé dyad with a "final sample" dummy as dependent variable and mentor and protégé characteristics as independent variables. Women mentors are overrepresented in the final sample with a 34x higher odds of inclusion (marginalizing the relevance of the other coefficients). This pronounced oversampling of women mentors is expected given the high imbalance towards men in the mentor set and mentor gender being our matching variable. Moreover, the most prolific mentors in our data are men who lack observationally comparable women mentors for matching. Again, women accounted for 12% of doctoral degrees in the biomedical sciences in the 1960s and no more than 30% in the 1980s, invariably constraining the pool of very experienced women mentors for our F32 cohorts starting in 1985 (AAUW 2015). There may also be an element of the "glass ceiling" effect in this data, whereby senior women face disadvantages in access to funding that helps penetrating the most prolific ranks in the resource intensive life sciences. Irrespective of the precise driving forces, matching out prolific and highly visible men mentors yields conservative estimates of the identified citation discount attached to women mentors in our main analysis.

We conclude the additional analyses with a series of robustness checks, summarized in Appendix 12. First, we verify that the supply-side dynamics established in Table 2 also hold on the subsample of papers for which detailed information on forward citations is available. This applies to all publications since 1996. The results obtained in Appendix 12a are very similar both in terms of effect size and significance levels to the main effects presented in Table 2.

In a second robustness check, we employ a different matching technique, namely propensity score matching with four nearest neighbors (mimicking the ratio obtained from the CEM approach). As shown in Appendix 12b, the results remain robust under this matching specification. Additional checks with a different number of nearest neighbors yield the same results and are available from the authors. In other words, the results of our main analysis are not an artifact of the employed coarsened exact matching approach.

In the third and fourth robustness check we probe alternative model specifications to the negative binomial distribution that also account for the skewness of forward citations. Since a few publications typically attract most citations (Lotka 1926), the observed citation discount attached to women mentors might be driven by a few highly cited publications that may systematically be more often produced by men mentors and their protégés. To test for this possibility, we first rerun the demand-side analysis in OLS regressions with logged citations as dependent variables. Logging the dependent variables consolidates the entire distribution and reduces the lever of extremely large values. As shown in Appendix 12c our results remain robust with this model specification. In a second check addressing the same question, we adjust the highest values of forward citations to a pre-determined maximum given the underlying distribution and rerun the negative binomial regression models with the adjusted values[32]. In this model specification our results remain robust, too (Appendix 12d). Of note, these alternative specifications and the obtained stable results also reflect the fact that the chosen matching procedure is fairly immune to model specification (Blackwell et al. 2009).

In a last robustness check, we alter the period considered relevant for the influence of mentorship from ten to five years, only including publications that appeared within five years after F32 grant receipt (Appendix 12e). Again, the results of this robustness check echo our main analyses.

CONCLUSION

We started with the question whether the gender of the mentor affects the evaluation of protégés by the scientific community. The question has smoldered in the literature because, on the one hand, there is ample evidence for the importance of mentorship in advancing protégés whilst, on the other hand, there is a deep literature on senior women getting undervalued in science. To get a better understanding of whether the undervaluation of senior women who serve as mentors also affects the evaluation of protégés and their work, we investigated the citation records of over 4,500 highly qualified postdocs with a mentored research fellowship of the NIH. We find that the work of women- relative to men-mentored protégés draws 10 percent fewer citations on the average paper, an order of magnitude that has been shown to tangibly affect career progress (Lerchenmueller and Sorenson 2018).

[32] We employ the "nb_adjust" Stata package (Enzmann 2015) and stick to the default options, which define an outlier as any value for which the expected frequency is less than 0.5.

Our study contributes to the literature on mentorship and furthers our understanding of the gender gap in science. There are rich strands of literature on the benefits of mentorship versus no mentorship. This research presents evidence that mentors fulfill several important functions to the benefit of their protégés, from hard factors (e.g., funding) to soft factors (e.g., psychosocial support). These factors have been shown to improve the productivity of protégés (see Eby et al. (2008) and Allen et al. (2004) for meta-analyses). Recent research in science specifically indicates, that working closely with a senior scientist is highly beneficial in terms of publishing and receiving recognition (Li et al. 2019, Ma et al. 2019). Again, this research uses publications by junior scientists with no senior collaborators as a control group and does not differentiate the effects for mentors of different gender. In fact, coauthoring with a senior scholar does not equate to being mentored (see AlShebli et al. 2020). What evidence exists on gender effects in mentorship suggests that women mentors may accelerate careers, particularly of young women (Blau et al. 2010). Persistent calls for women mentors to serve as role models and mentoring the next generation of scientists may therefore come as no surprise (Noe 1988, Tong and Kram 2013, Gaule and Piacentini 2018).

Building from this rich literature on mentorship, our study does not question the importance of mentorship but asks whether the associated benefits differ if the mentor is a woman versus a man. Indeed, we find that mentorship is beneficial for the evaluation of protégés' work, indicated by, for example, an increase in citations when the mentor coauthors a paper. But, our theoretical vantage point of senior women being undervalued in science, which may constrain their ability to pledge social collateral for the evaluation of their protégés, enhances our understanding of gendered mentorship effects. It further allows us testing the relative merits of demand- versus supply-side explanations giving rise to an evaluative discount for protégés.

Our results also shed light on hitherto concealed dynamics of gender gaps in science. We complement the existing literature by offering a closer look at how gender biases do not only affect the focal actor (here: the mentor), but also appear to transcend to collaborators (here: the protégé). About 40% of the citation discount experienced by women-mentored protégés were attributed to supply-side factors, chiefly to gender differences in mentor characteristics captured via our matching design. It is important to note that our research did not focus on examining how these gender differences in mentors' publication or funding records emerged. Some of these differences are likely also explained by gender biases faced by women mentors when climbing the science career ladder, considering the evidence available on the gender gap in science. We were nonetheless concerned with examining the evaluation of highly qualified protégés who start out with mentors of comparable observational quality but of different gender. One of the most important, yet troubling findings is that protégés' association with a mentor – a relationship almost uniformly perceived as advantageous and formative for the protégé in the literature – may (inadvertently) hurt protégés' evaluation and visibility early on in their careers. The identified citation discount applies independent of the protégés' gender, underpinning the importance of addressing gender bias to the benefit for all. Our sample of F32 scientists contains individuals at an early stage of their career and committed to pursuing an academic career long-term. This early window is usually critical for getting the attention that allows winning independent grants (Bol et al. 2018) and for opening the door to a tenure track career (Jena et al. 2015). With these results, we also add to a line of research that draws our attention to career barriers that women face early (e.g., Lerchenmueller and Sorenson 2018, Sterling and Fernandez 2018), relative to the deep literature on women hitting a "glass ceiling" when having already climbed the career ladder to more senior ranks.

Given that women-mentored protégés experience an undue demand-side bias on their most promising work, research of high promise for advancing science, it seems desirable to reflect on possible countermeasures and policy implications. Our findings may sensitize women-mentored protégés, their mentors, and women in science more generally to actively championing their work in service of increasing its visibility. Scholars have started to examine how women can positively influence their career progress through presenting their accomplishments more actively (Bikard and Fernandez-Mateo 2018, Exley and Kessler 2019, Lerchenmueller et al. 2019), for example. Future research appears warranted that examines ensuing benefits and drawbacks for women when more forcefully pursuing credit where credit is due. Again, it is worth emphasizing that the identified citation discount afflicts the most impactful work of protégés, work that can legitimately be championed and that deserves credit. In addition to the affected actors, the institutions of science can consider measures that may contribute to change. For example, it seems feasible that publishing journals devote more editorial coverage that highlights women's work. Some journals have already gone that route by publishing dedicated issues on seminal work by female researchers once a year. But even prior to publication, conference or seminar organizers can draw the attention of the science community to great work by women, for instance, by inviting women keynote speakers, inviting women on visible research panels and so forth. Overall, our study calls for measures that legitimately raise attention to women's research so that the probability of getting cited meets the quality of the research.

As with most empirical undertakings, our study involves trade-offs and limitations. First, mentorship is hard to observe in practice and hence hard to verify, particularly for field research that seeks evidence beyond controlled laboratory experiments or smaller scale case studies (for a detailed discussion of this challenge, see AlShebli et al. (2020)). Part of the reason why we chose the F32 program as our empirical setting was to explicitly address challenges of mentorship identification. Our key identifying assumption for mentorship is that the senior author on the first paper that acknowledges the mentored F32 funding (and that is by construction coauthored by the F32 recipient) acts as mentor. The stringent F32 grant criteria, requiring a sponsoring mentor by design, and the fact that authorship norms in the life sciences reserve the last author position to the sponsoring senior scholar, lend credence to our approach. We document frequent and stable collaboration between mentors and protégés.

Moreover, a manual search for a random draw of 100 mentor-protégé dyads verified a connection for almost all cases.

Second, we cannot quantify the degree to which either the F32 recipient has been able to influence mentor allocation, or the mentor been able to cherry pick protégés. This two-sided matching dynamic is a notorious challenge in research on tie formation (Azoulay et al. 2017), and we do not claim that our approach is immune to this challenge. But we submit that our empirical setting mitigates sorting concerns to the extent possible because this two-sided matching problem becomes pronounced with quality heterogeneity in the sets of individuals who seek to match. Our data shows that mentors stratify in terms of quality and that this stratification is also influenced by gender. Although the selective F32 program likely shrinks heterogeneity also for mentors relative to the population of senior scientists (e.g., mentors in our set have already earned four major R01 grants on average), there is still a concern that better F32 recipients want to match to men mentors, potentially via status preferential attachment or triadic closure (de Vaan and Wang 2020). Turning the tables, mentors also want to pick the best talent but the unusually homogenous group of F32 contestants makes it a lot harder to know who to pick. Borrowing a queuing imagery (Podolny 2001), mentors line up (the best come first) to pick protégés but with a limited ability to tell the wheat from the chaff, being first in line is of limited use for the picking. More serendipitous factors (e.g., current location, conference meetings) likely play an important role for the matching in this context (Boudreau et al. 2017), though the parties involved might be reluctant to acknowledge that in hindsight. If anything, our data indicate that women mentors get paired with slightly better protégés in terms of productivity, after matching out any preferential attachment based on gender.

Finally, we do not observe mentoring quality, i.e., the extent to which the mentor puts time and resources into supervising the protégé. The fact that the mentors are reputable scientists and that the mentor is a decisive factor for a successful F32 fellowship makes us confident that the overall effort the mentors dedicate to their protégés is high. We also have no grounds to surmise that mentorship effort would systematically differ by gender. The benefits of our approach in that regard notwithstanding, focusing on F32 recipients potentially reduces generalizability of our findings. We identify a citation discount for women-mentored F32 recipients, especially on topics that are likely to attract citations. Comparing the work of the competitive F32 cohorts to the broader life science research enterprise, our protégés indeed publish disproportionally on topics poised to attract the attention of the research community. Again, it is troublesome that the citation discount materializes on work that likely matters the most for science advancement. We would nonetheless welcome research that extends our findings to other cohorts and contexts.

In conclusion, we provide the first evidence of an evaluative discount afflicting promising protégés based on the gender of their mentors, however inadvertent the discount may be. Since these effects appear pronounced early on in scientists' careers, the identified dynamics also conceivably contribute to the gender gap in the life sciences and, potentially, in other professional labor markets.

REFERENCES

AAUW. (2015) Why so few? Women in Science, Technology, Engineering, and Mathematics. Accessed February 8, 2021, https://time.com/wp-content/uploads/2015/05/why-so-few-women-in-science-technology-engineering-and-mathematics.pdf.

Abraham M (2020) Gender-role Incongruity and Audience-based Gender Bias: An Examination of Networking among Entrepreneurs. *Administrative Science Quarterly* 65(1):151-180.

Allen TD, Eby LT, Poteet ML, Lentz E, Lima L (2004) Career benefits associated with mentoring for protegee: a meta-analysis. *Journal of Applied Psychology* 89(1):127-36.

AlShebli B, Makovi K, Rahwan T (2020) [RETRACTED ARTICLE] The association between early career informal mentorship in academic collaborations and junior author performance. *Nature Communications* 11(1):5855.

Amrein K, Langmann A, Fahrleitner-Pammer A, Pieber TR, Zollner-Schwetz I (2011) Women underrepresented on editorial boards of 60 major medical journals. *Gender Medicine* 8(6):378-87.

Azoulay P, Liu CC, Stuart TE (2017) Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs. *American Journal of Sociology* 122(4):1223-1271.

Azoulay P, Stuart TE, Wang YB (2014) Matthew: Effect or Fable? *Management Science* 60(1):92-109.

Bian L, Leslie SJ, Cimpian A (2017) Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355(6323):389-391.

--- (2018) Evidence of Bias Against Girls and Women in Contexts That Emphasize Intellectual Ability. *American Psychologist* 73(9):1139-1153.

Bikard M, Fernandez-Mateo I (2018) Gender differences in publishing and patenting rates: Does scientific style matter? Working Paper, London Business School.

Blackwell M, Iacus S, King G, Porro G (2009) cem: Coarsened exact matching in Stata. *The Stata Journal* 9(4):524-546.

Blau FD, Currie JM, Croson RTA, Ginther DK (2010) Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial. *American Economic Review* 100(2):348-352.

Bol T, de Vaan M, van de Rijt A (2018) The Matthew effect in science funding. *Proceedings of the National Academy of Sciences of the United States of America* 115(19):4887-4890.

Botelho TL, Abraham M (2017) Pursuing Quality: How Search Costs and Uncertainty Magnify Gender-based Double Standards in a Multistage Evaluation Process. *Administrative Science Quarterly* 62(4):698-730.

Boudreau KJ, Brady T, Ganguli I, Gaule P, Guinan E, Hollenberg A, Lakhani KR (2017) A Field Experiment on Search Costs and the Formation of Scientific Collaborations. *Review of Economics and Statistics* 99(4):565-576.

Cimpian A, Leslie SJ (2017) The Brilliance Trap. *Scientific American* 317(3):60-65.

Correll SJ (2004) Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review* 69(1):93-113.

Croson R, Gneezy U (2009) Gender Differences in Preferences. *Journal of Economic Literature* 47(2):448-474.

de Vaan M, Wang D (2020) Micro-structural foundations of network inequality: Evidence from a field experiment in professional networking. *Social Networks*.

Ding WW, Murray F, Stuart TE (2013) From Bench to Board: Gender Differences in University Scientists' Participation in Corporate Scientific Advisory Boards. *Academy of Management Journal* 56(5):1443-1464.

Dohm F-A, Cummings W (2002) Research mentoring and women in clinical psychology. *Psychology of Women Quarterly* 26(2):163-167.

--- (2003) Research mentoring and men in clinical psychology. *Psychology of Men & Masculinity* 4(2):149.

Eby LT, Allen TD, Evans SC, Ng T, DuBois DL (2008) Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior* 72(2):254-267.

Editorial (2019) Science needs mentors. *Nature* 573(7772):5.

Enzmann D (2015) NB_ADJUST: Stata module to identify and adjust outliers of a variable assumed to follow a negative binomial distribution.

Etzkowitz H, Kemelgor C, Uzzi B (2000) *Athena unbound: The advancement of women in science and technology* (Cambridge University Press).

Exley CL, Kessler JB (2019) The gender gap in self-promotion. Report.

Fang R, Zhang Z, Shaw JD (2020) Gender and social network brokerage: A meta-analysis and field investigation. *Journal of Applied Psychology*.

Fernandez-Mateo I, Fernandez RM (2016) Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs. *Management Science* 62(12):3636-3655.

Fernandez-Mateo I, Kaplan S (2018) Gender and Organization Science: Introduction to a Virtual Special Issue. *Organization Science* 29(6):1229-1236.

Gaule P, Piacentini M (2018) An advisor like me? Advisor gender and post-graduate careers in science. *Research Policy* 47(4):805-813.

Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quarterly Journal of Economics* 118(3):1049-1074.

Heggeness ML, Ginther DK, Larenas MI, Carter-Johnson FD (2018) The impact of postdoctoral fellowships on a future independent career in federally funded biomedical research. Report.

Hoisl K, Mariani M (2017) It's a Man's Job: Income and the Gender Gap in Industrial Research. *Management Science* 63(3):766-790.

Holman L, Morandin C (2019) Researchers collaborate with same-gendered colleagues more often than expected across the life sciences. *Plos One* 14(4).

Iacus SM, King G, Porro G (2012) Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20(1):1-24.

Jagsi R, Griffith KA, Stewart A, Sambuco D, DeCastro R, Ubel PA (2012) Gender Differences in the Salaries of Physician Researchers. *Journal of the American Medical Association* 307(22):2410-2417.

Jaxon J, Lei RF, Shachnai R, Chestnut EK, Cimpian A (2019) The Acquisition of Gender Stereotypes about Intellectual Ability: Intersections with Race. *Journal of Social Issues* 75(4):1192-1215.

Jena AB, Khullar D, Ho O, Olenski AR, Blumenthal DM (2015) Sex Differences in Academic Rank in US Medical Schools in 2014. *Journal of the American Medical Association* 314(11):1149-58.

Johnson WB (2007) Student–Faculty Mentorship Outcomes. Allen TD, Eby LT, eds. *The Blackwell handbook of Mentoring: A Multiple Perspectives Approach* (Blackwell Publishing, Malden), 189-210.

King G. (2012) An Explanation for CEM Weights. Accessed March 23, 2021, https://docs.google.com/document/d/1xQwyLt_6EXdNpA685LjmhjO20y5pZDZYwe2qeNoI5dE/edit.

Kirchmeyer C (2005) The effects of mentoring on academic careers over time: Testing performance and political perspectives. *Human Relations* 58(5):637-660.

Kleinbaum AM, Stuart TE, Tushman ML (2013) Discretion Within Constraint: Homophily and Structure in a Formal Organization. *Organization Science* 24(5):1316-1336.

Kodde DA, Palm FC (1986) Wald Criteria for Jointly Testing Equality and Inequality Restrictions. *Econometrica* 54(5):1243-1248.

Koning R, Samila S, Ferguson J-P (2019) Female inventors and inventions. *Available at SSRN 3401889*.

Kram KE (1983) Phases of the Mentor Relationship. *Academy of Management Journal* 26(4):608-625.

--- (1988) *Mentoring at work: Developmental relationships in organizational life* (University Press of America).

Kulis S, Sicotte D, Collins S (2002) More than a pipeline problem: Labor supply constraints and gender stratification across academic science disciplines. *Research in Higher Education* 43(6):657-691.

Leahey E (2007) Not by productivity alone: How visibility and specialization contribute to academic earnings. *American Sociological Review* 72(4):533-561.

Leahey E, Keith B, Crockett J (2010) Specialization and promotion in an academic discipline. *Research in Social Stratification and Mobility* 28(2):135-155.

Lerchenmueller MJ, Sorenson O (2016) Author Disambiguation in PubMed: Evidence on the Precision and Recall of Author-ity among NIH-Funded Scientists. *Plos One* 11(7).

--- (2018) The gender gap in early career transitions in the life sciences. *Research Policy* 47(6):1007-1017.

Lerchenmueller MJ, Sorenson O, Jena AB (2019) Gender differences in how scientists present the importance of their research: observational study. *British Medical Journal* 367:l6573.

Leslie SJ, Cimpian A, Meyer M, Freeland E (2015) Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347(6219):262-265.

Levitt DG (2010) Careers of an elite cohort of U.S. basic life science postdoctoral fellows and the influence of their mentor's citation record. *BMC Medical Education* 10.

Li D (2017) Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal-Applied Economics* 9(2):60-92.

Li WH, Aste T, Caccioli F, Livan G (2019) Early coauthorship with top scientists predicts success in academic careers. *Nature Communications* 10.

Lotka AJ (1926) The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(12):317-323.

Lynn FB, Podolny JM, Tao L (2009) A Sociological (De)Construction of the Relationship between Status and Quality. *American Journal of Sociology* 115(3):755-804.

Ma YF, Oliveira DFM, Woodruff TK, Uzzi B (2019) Women who win prizes get less money and prestige. *Nature* 565(7739):287-288.

Malmgren RD, Ottino JM, Amaral LAN (2010) The role of mentorship in protege performance. *Nature* 465(7298):622-U117.

Merton RK (1968) The Matthew effect in science. The reward and communication systems of science are considered. *Science* 159(3810):56-63.

--- (1988) The Matthew Effect in Science .2. Cumulative Advantage and the Symbolism of Intellectual Property. *Isis* 79(299):606-623.

Meyer M, Cimpian A, Leslie SJ (2015) Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology* 6.

NIH. (2018) Summary of Leave, Part-Time and Extension Policies Available to Ruth L. Kirschstein National Research Service Awards (NRSA) Trainees and Fellows. Accessed March 23, 2021, https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-154.html.

---. (2021a) Welcome to Medical Subject Headings. Accessed March 23, 2021, https://www.nlm.nih.gov/mesh/meshhome.html.

---. (2021b) Ruth L. Kirschstein Postdoctoral Individual National Research Service Award F32 program. Relevant Policy Notices and Program Purpose. Accessed February 8, 2021, https://researchtraining.nih.gov/programs/fellowships/f32.

Noe RA (1988) Women and Mentoring - a Review and Research Agenda. *Academy of Management Review* 13(1):65-78.

Pezzoni M, Mairesse J, Stephan P, Lane J (2016) Gender and the Publication Output of Graduate Students: A Case Study. *PLoS One* 11(1):e0145146.

Podolny JM (1993) A Status-Based Model of Market Competition. *American Journal of Sociology* 98(4):829-872.

--- (2001) Networks as the pipes and prisms of the market. *American Journal of Sociology* 107(1):33-60.

Preston AE (2004) Plugging the leaks in the scientific workforce. *Issues in Science and Technology* 20(4):69-78.

Ragins BR, Mcfarlin DB (1990) Perceptions of Mentor Roles in Cross-Gender Mentoring Relationships. *Journal of Vocational Behavior* 37(3):321-339.

Reschke BP, Azoulay P, Stuart TE (2018) Status Spillovers: The Effect of Status-conferring Prizes on the Allocation of Attention. *Administrative Science Quarterly* 63(4):819-847.

Reskin BF (1979) Academic Sponsorship and Scientists Careers. *Sociology of Education* 52(3):129-146.

--- (2005) Including mechanisms in our models of ascriptive inequality. *Handbook of employment discrimination research*:75-97.

Ridgeway CL (1991) The Social Construction of Status Value - Gender and Other Nominal Characteristics. *Social Forces* 70(2):367-386.

--- (2001) Gender, status, and leadership. *Journal of Social Issues* 57(4):637-655.

Rivera LA, Tilcsik A (2019) Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation. *American Sociological Review* 84(2):248-274.

Rossiter MW (1993) The Matthew-Matilda Effect in Science. *Social Studies of Science* 23(2):325-341.

Scandura TA (1992) Mentorship and Career Mobility - an Empirical-Investigation. *Journal of Organizational Behavior* 13(2):169-174.

Schwiebert VL, Deck MD, Bradshaw ML, Scott P, Harper M (1999) Women as mentors. *The Journal of Humanistic Counseling* 37(4):241.

Simcoe TS, Waguespack DM (2011) Status, Quality, and Attention: What's in a (Missing) Name? *Management Science* 57(2):274-290.

Sterling AD, Fernandez RM (2018) Once in the Door: Gender, Tryouts, and the Initial Salaries of Managers. *Management Science* 64(11):5444-5460.

Tesch BJ, Wood HM, Helwig AL, Nattinger AB (1995) Promotion of women physicians in academic medicine. Glass ceiling or sticky floor? *Journal of the American Medical Association* 273(13):1022-5.

Thebaud S (2015) Business as Plan B: Institutional Foundations of Gender Inequality in Entrepreneurship across 24 Industrialized Countries. *Administrative Science Quarterly* 60(4):671-711.

Tong C, Kram KE (2013) The efficacy of mentoring-the benefits for mentees, mentors, and organizations. Passmore J, Petersen D, Freire T, eds. *The Wiley-Blackwell handbook of the psychology of coaching and mentoring* (John Wiley & Sons, Chichester), 217-242.

Torvik VI, Weeber M, Swanson DR, Smalheiser NR (2005) A probabilistic similarity metric for Medline records: A model for author name disambiguation. *Journal of the American Society for Information Science and Technology* 56(2):140-158.

Valian V (2007) Women at the Top in Science--And Elsewhere. *Why aren't more women in science?: Top researchers debate the evidence.* (American Psychological Association, Washington, DC, US), 27-37.

Waldinger F (2010) Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany. *Journal of Political Economy* 118(4):787-831.

Weichselbaumer D, Winter-Ebmer R (2005) A meta-analysis of the International gender wage gap. *Journal of Economic Surveys* 19(3):479-511.

Woehler ML, Cullen-Lester KL, Porter CM, Frear KA (2021) Whether, How, and Why Networks Influence Men's and Women's Career Success: Review and Research Agenda. *Journal of Management* 47(1):207-236.

Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036-9.

Zuckerman HE, Cole JR, Bruer JT (1991) The outer circle: Women in the scientific community. *This volume is based on papers from four symposia held at Stanford University, CA, from 1983 to 1986.* (WW Norton & Co).

Acknowledgements

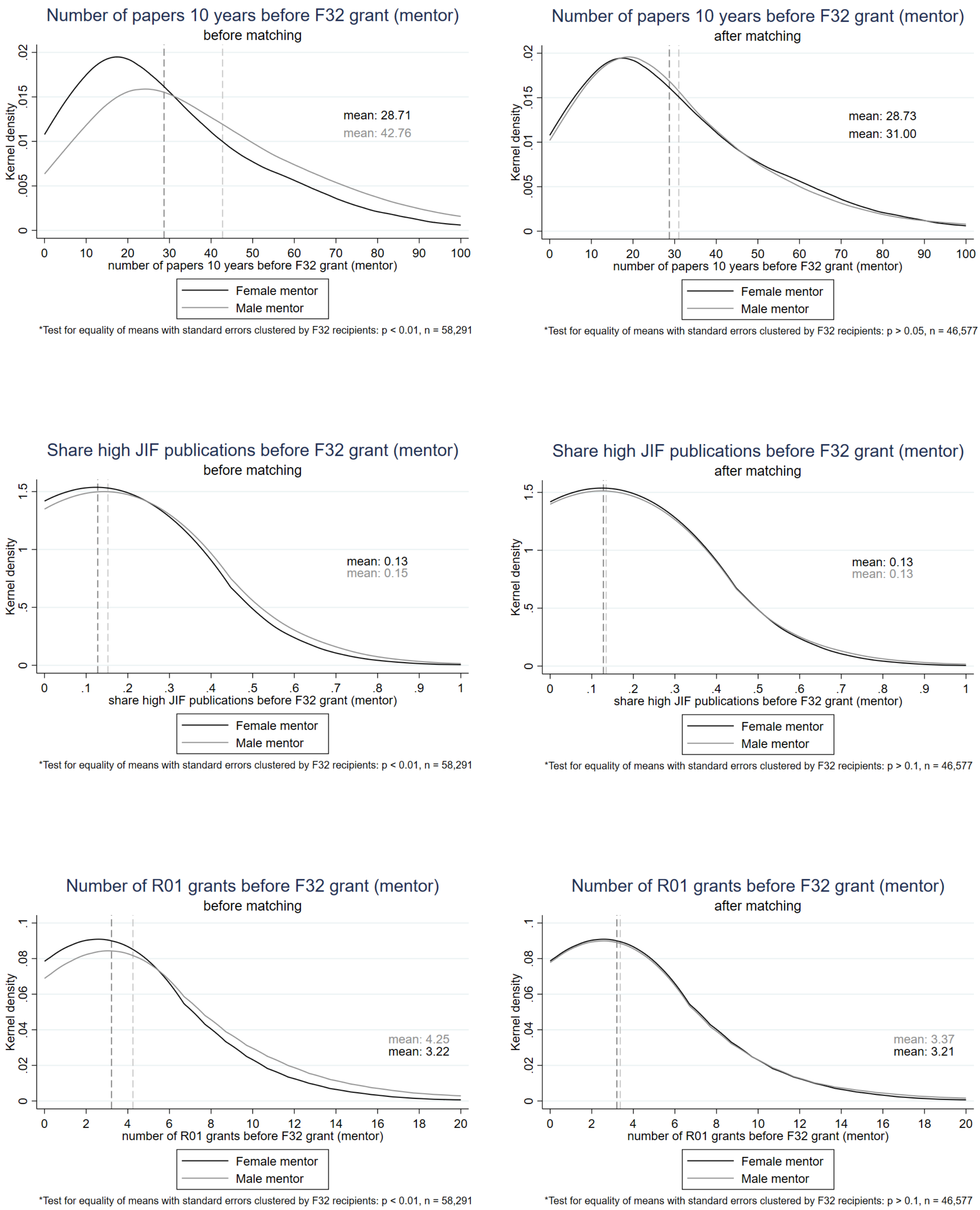
The authors acknowledge support by the German state of Baden-Wuerttemberg through bwHPC, enabling the use of high-performance cluster computing resources. The usual disclaimer applies.

Leo Schmallenbach is supported by the Joachim Herz Foundation.

TABLES AND FIGURES

Figure 1 - Pre-post matching distribution of mentor and F32 recipient characteristics

a) Mentor characteristics



b) F32 recipient characteristics

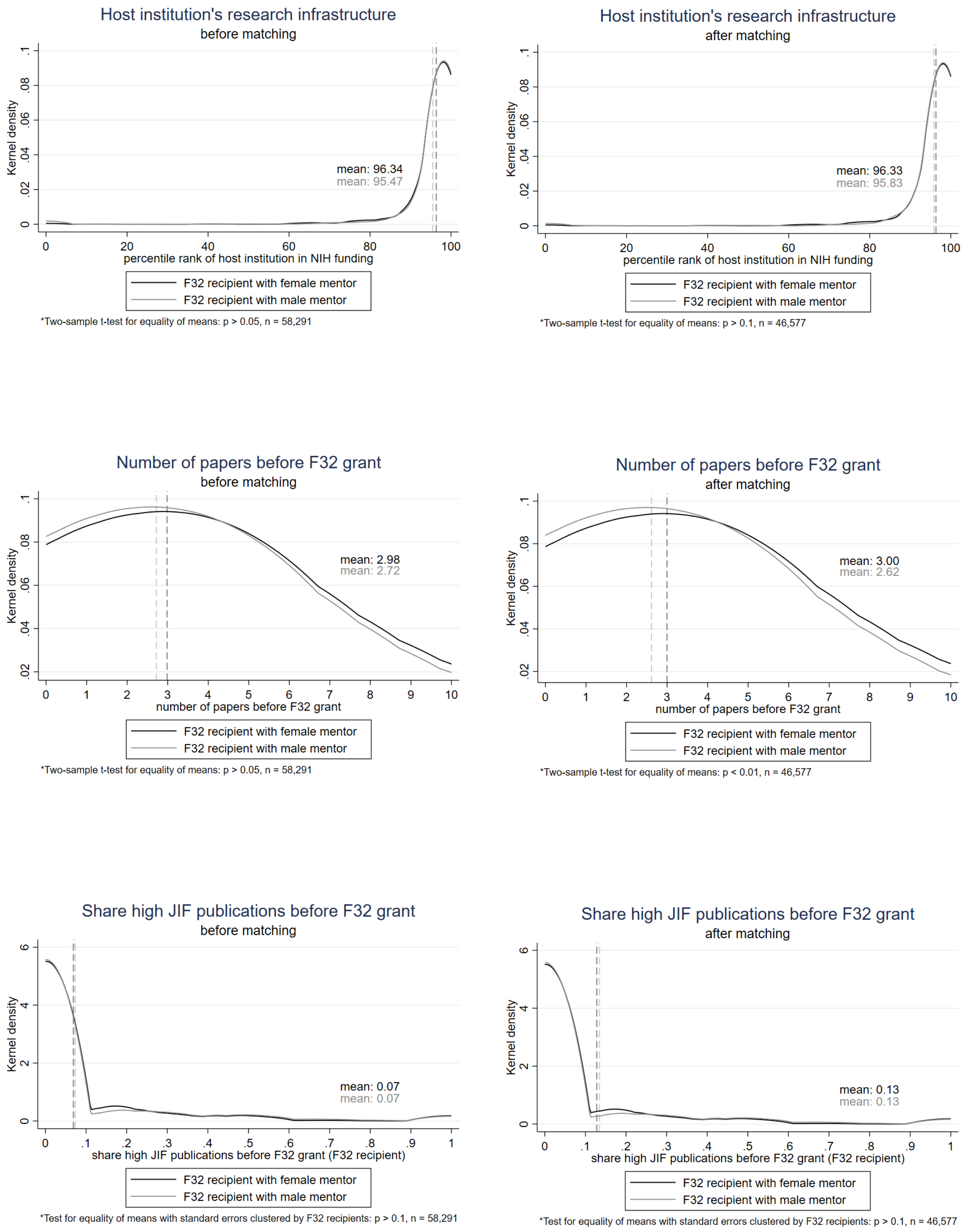


Figure 2 - Comparison of average citations

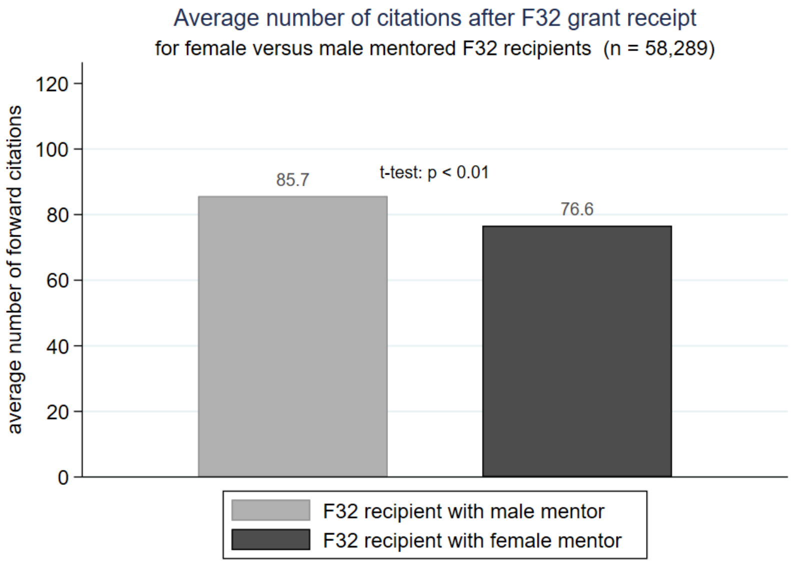
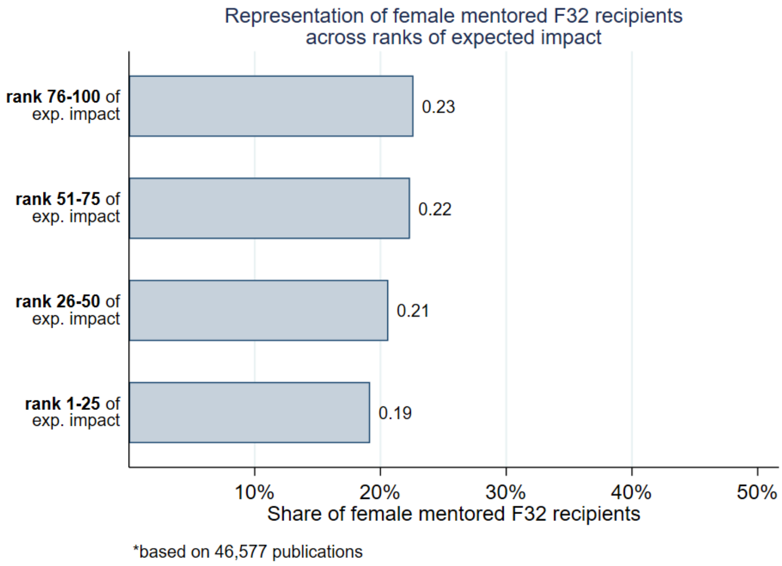
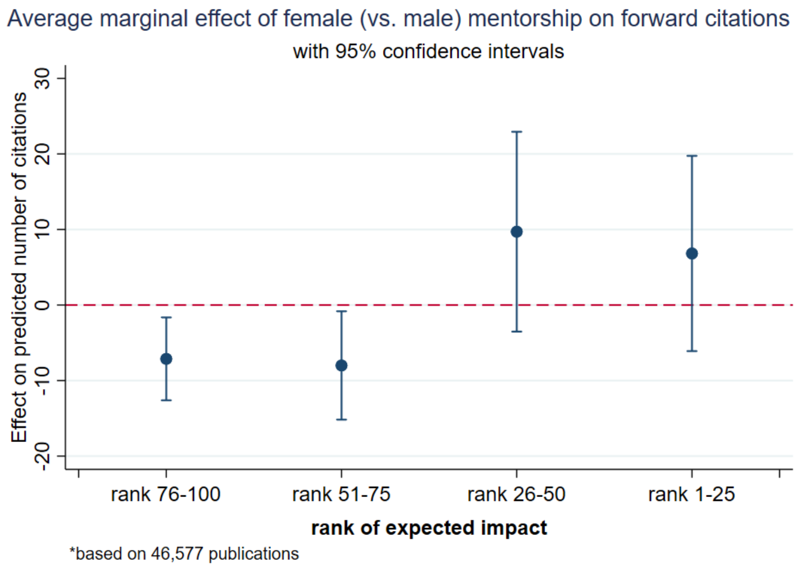


Figure 3 – Representation of female mentored F32 recipients across ranks of expected citations*



*Rank 100 referencing the highest level of expected citations.

Figure 4 – Average marginal effects of female mentorship on citations for different quartiles of expected citations*



*Rank 100 referencing the highest level of expected citations.

Figure 5 - Relative importance of the sources of forward citations

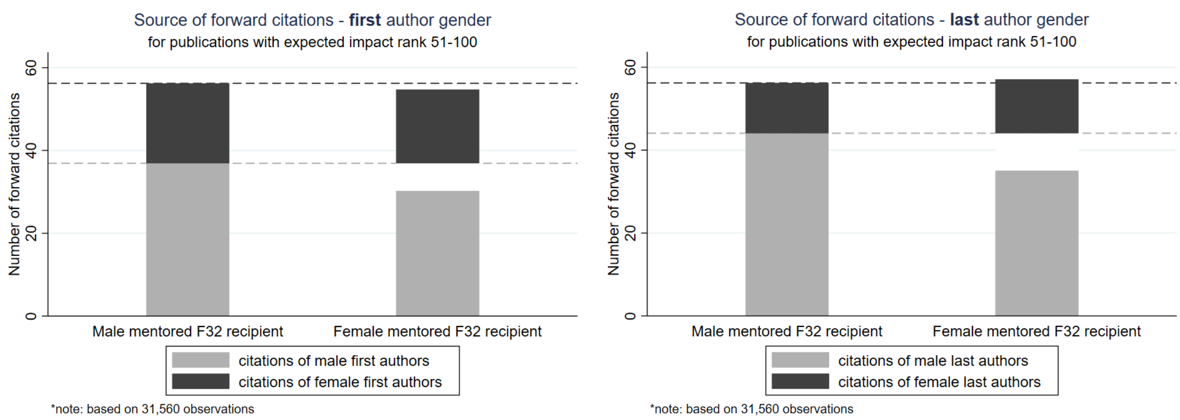


Table 1 - Descriptive statistics and variable descriptions

variable name	description	mean	sd	min	max
dependent variables					
citations	forward citations for a given paper	84.042	167.635	0	12,928
independent variable					
female mentor	dummy variable indicating gender of mentor: 1 if female; 0 otherwise	0.178	-	0	1
mentor characteristics					
number of papers 10 years before F32 grant (mentor)	number of papers mentor published within 10 year before F32 grant receipt	40.256	35.243	0	338
share high JIF pubs 10 years before F32 grant (mentor)	share of mentor publications in high impact journals (JIF > 10) 10 years before F32 grant receipt	0.148	0.151	0	1
number of prior R01 grants (mentor)	number of R01 grants issued to the mentore before F32 grant receipt	4.063	4.318	0	39
academic cohort*	fiscal year of the F32 grant award	1995	-	1985	2005
F32 characteristics (recipient and grant)					
female F32 recipient	dummy variable indicating gender of F32 grant receipient: 1 if female; 0 otherwise	0.34	-	0	1
host institution's research infrastructure	percentile rank of hosting institution in terms of NIH research funding (0 lowest rank, 100 highest rank)	95.627	12.702	0	100
F32 grant extension	dummy variable indicating F32 grant extension: 1 if grant extended; 0 otherwise	0.117	-	0	1
number of papers before F32 grant	number of papers published by F32 grant recepient before F32 grant receipt	2.767	2.557	0	10
share high JIF pubs before F32 grant	share F32 recipient publications in high impact journals (JIF > 10) before F32 grant receipt	0.072	0.199	0	1
paper count since grant receipt	count variable indicating the number of publications since F32 grant receipt	11.835	12.439	1	161
papers in last year	number of papers published by F32 recipient in previous year	2.263	2.673	0	31
research characteristics					
number of authors	number of authors on a given paper	4.647	2.22	1	10
share of female coauthors	share of female authors (w/o mentor and F32 recipient) on a given paper	0.279	0.335	0	1
journal impact factor	journal impact factor of a given paper	5.918	7.668	0	70.67
publication type**	type of publication: article, article in press, book, chapter, conference paper, editoroal, erratum, letter, note, review, short survey	-	-	-	-
mentor among coauthors	dummy variable indicating whether mentor is among coauthors: 1 if mentor is coauthor; 0 otherwise	0.421	-	0	1
research proximity	number of MeSH terms on a given paper that also appeared on one of the mentor's publication from within 10 years before F32 grant receipt	6.907	4.036	0	34
expected impact	percentile rank based on average citations associated with assigned MeSH terms of a given paper in the same year (0 lowest rank, 100 highest rank)	74.874	23.074	1	100
publication year*	publication year of a given paper	2001	-	1985	2009
observations					58,291

* median instead of mean reported
** for a detailed overview on descriptive statistics for publication types see Appendix 7b

Table 2 - Regression models of forward citations on female mentorship – supply-side

Dependent variable: forward citations	NBREG	NBREG	NBREG	NBREG	NBREG
	(1)	(2)	(3)	(4)	(5)
Independent variables:					
female mentor	0.902***	0.930**	0.940*	0.945**	0.947**
	(0.028)	(0.034)	(0.032)	(0.026)	(0.026)
mentor characteristics					
number of papers 10 years before F32 grant (mentor)		0.999	0.999*	0.999**	0.999**
		(0.001)	(0.001)	(0.001)	(0.001)
share high JIF pubs 10 years before F32 grant (mentor)		1.904***	1.837***	1.038	0.987
		(0.272)	(0.243)	(0.092)	(0.086)
number of prior R01 grants (mentor)		1.010**	1.008	1.002	0.997
		(0.005)	(0.005)	(0.004)	(0.004)
F32 characteristics (recipient and grant)					
female F32 recipient			0.929**	0.953*	0.954*
			(0.034)	(0.027)	(0.027)
host institution's research infrastructure			0.998	0.999	0.999
			(0.001)	(0.001)	(0.001)
F32 grant extension			0.974	0.986	0.980
			(0.047)	(0.042)	(0.042)
number of papers before F32 grant			1.008	1.010*	1.004
			(0.007)	(0.005)	(0.005)
share high JIF pubs before F32 grant			1.263**	1.127*	1.125
			(0.123)	(0.081)	(0.082)
paper count since grant receipt			0.997	0.997*	0.997
			(0.002)	(0.002)	(0.002)
papers in last year			1.014*	1.014**	1.018***
			(0.007)	(0.006)	(0.006)
research characteristics					
number of authors				1.049***	1.048***
				(0.006)	(0.006)
share of female coauthors				0.988	0.985
				(0.030)	(0.029)
journal impact factor				1.058***	1.057***
				(0.002)	(0.002)
mentor among coauthors				1.093***	1.069***
				(0.028)	(0.027)
research proximity				1.013***	1.012***
				(0.003)	(0.003)
1st quartile of expected impact (rank 1-25)					base
2nd quartile of expected impact (rank 26-50)					1.413***
					(0.085)
3rd quartile of expected impact (rank 51-75)					1.525***
					(0.092)
4th quartile of expected impact (rank 76-100)					1.734***
					(0.106)
publication year fixed effects	included	included	included	included	included
CEM weights		included	included	included	included
academic cohort fixed effects		included	included	included	included
publication type fixed effects				included	included
constant	53.06***	67.55***	80.52***	47.00***	29.84***
	(4.968)	(18.892)	(25.804)	(13.428)	(8.984)
observations	58,291	46,577	46,577	46,577	46,577

standard errors clustered by F32 recipients in parantheses; *p < 0.1, ** p < 0.05, *** p < 0.01; exponentiated coefficents (IRRs) reported

Table 3 – Regression models of forward citations on female mentorship – demand-side

	rank 51-100 of exepcted citations				**rank 1-50 of exepcted citations**			
	NBREG	NBREG	NBREG	NBREG	NBREG	NBREG	NBREG	NBREG
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: forward citations	*citations of* ***male first*** *authors*	*citations of* ***female first*** *authors*	*citations of* ***male last*** *authors*	*citations of* ***female last*** *authors*	*citations of* ***male first*** *authors*	*citations of* ***female first*** *authors*	*citations of* ***male last*** *authors*	*citations of* ***female last*** *authors*
Independent variables:								
female mentor	0.898***	0.992	0.870***	1.138***	1.122	1.315***	1.134	1.384***
	(0.025)	(0.030)	(0.024)	(0.036)	(0.101)	(0.125)	(0.104)	(0.131)
mentor characteristics								
number of papers 10 years before F32 grant (mentor)	1.000	1.000	1.000	1.000	0.997**	0.994***	0.996**	0.994***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
share high JIF pubs 10 years before F32 grant (mentor)	1.087	0.963	1.066	0.967	0.630	0.337***	0.534*	0.433**
	(0.094)	(0.090)	(0.094)	(0.091)	(0.222)	(0.131)	(0.198)	(0.149)
number of prior R01 grants (mentor)	0.996	0.994	0.996	0.996	1.028	1.039*	1.033*	1.027
	(0.004)	(0.005)	(0.004)	(0.005)	(0.020)	(0.022)	(0.020)	(0.023)
F32 characteristics (recipient and grant)								
female F32 recipient	0.901***	1.070**	0.903***	1.150***	0.813**	1.270***	0.847*	1.370***
	(0.025)	(0.032)	(0.025)	(0.037)	(0.070)	(0.111)	(0.073)	(0.128)
host institution's research infrastructure	0.998	0.998	0.998	0.998	0.996	0.998	0.994*	1.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
F32 grant extension	0.957	0.974	0.964	0.954	0.934	1.109	0.915	1.164
	(0.048)	(0.048)	(0.047)	(0.050)	(0.154)	(0.187)	(0.153)	(0.198)
number of papers before F32 grant	1.009*	1.011**	1.010**	1.008	0.973	0.970*	0.973*	0.963**
	(0.005)	(0.006)	(0.005)	(0.006)	(0.017)	(0.017)	(0.016)	(0.018)
share high JIF pubs before F32 grant	1.129	1.144	1.132	1.151	1.180	1.107	1.206	1.057
	(0.107)	(0.104)	(0.103)	(0.114)	(0.203)	(0.271)	(0.234)	(0.212)
paper count since grant receipt	0.998	1.000	0.998	1.001	1.001	1.004	1.002	1.001
	(0.003)	(0.003)	(0.002)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
papers in last year	1.018**	1.013	1.019**	1.008	1.005	0.978	0.993	1.000
	(0.008)	(0.010)	(0.008)	(0.012)	(0.023)	(0.024)	(0.023)	(0.025)
research characteristics								
number of authors	1.048***	1.044***	1.051***	1.032***	1.081***	1.020	1.080***	1.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.020)	(0.019)	(0.020)	(0.019)
share of female coauthors	0.841***	1.107***	0.849***	1.211***	0.876	1.714***	0.922	1.861***
	(0.025)	(0.035)	(0.025)	(0.041)	(0.120)	(0.241)	(0.129)	(0.261)
journal impact factor	1.060***	1.056***	1.060***	1.056***	1.075***	1.056***	1.073***	1.050***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.011)	(0.009)	(0.011)	(0.008)
mentor among coauthors	1.078***	1.048*	1.073***	1.048*	0.977	1.159	1.012	1.109
	(0.028)	(0.028)	(0.028)	(0.029)	(0.086)	(0.118)	(0.090)	(0.112)
research proximity	1.011***	1.024***	1.013***	1.024***	1.009	1.031**	1.014	1.026**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.011)	(0.012)	(0.011)	(0.013)
1st quartile of expected impact (rank 1-25)					base	base	base	base
2nd quartile of expected impact (rank 26-50)					1.379***	1.938***	1.410***	2.109***
					(0.119)	(0.159)	(0.119)	(0.200)
3rd quartile of expected impact (rank 51-75)	base	base	base	base				
4th quartile of expected impact (rank 76-100)	1.296***	1.235***	1.304***	1.188***				
	(0.036)	(0.040)	(0.036)	(0.041)				
publication year fixed effects	included	included	included	included	included	included	included	included
CEM weights	included	included	included	included	included	included	included	included
F32 grant fiscal year fixed effects	included	included	included	included	included	included	included	included
academic cohort fixed effects	included	included	included	included	included	included	included	included
constant	3.573***	1.238	3.645***	1.196	20.23***	4.378***	25.08***	2.035*
	(0.963)	(0.345)	(0.973)	(0.349)	(9.865)	(1.643)	(12.239)	(0.852)
observations	31,356	31,356	31,356	31,356	3,959	3,959	3,959	3,959

standard errors clustered by F32 recipients in parantheses; *p < 0.1, ** p < 0.05, *** p < 0.01; exponentiated coefficents (IRRs) reported