.

Equality for Granted:

What Happens when Discrimination in Academia Becomes Salient?

Lena Hensvik[♦] and Peter Nilsson

This version March, 2026[♣]

Abstract

We document the individual, organizational, and field-wide impacts following a public disclosure of substantial male premiums in the competence assessments of recent biomedical PhDs applying for an important individual grant from the Swedish NIH. Following the disclosure, three key changes occurred: (i) male-only review committees were rapidly phased out; (ii) individual reviewers adjusted their decision-making; and (iii) the average male premium was eliminated. We follow applicants' publications, promotions, and earnings for up to 18 years after application. After the disclosure, the share of female grantees rose and, consistent with improved selection on research potential, the long-run output gap between funded and unfunded applicants widened. However, the disclosure of bias not only improved fairness and allocative efficiency; but also triggered coordinated responses with broader implications for womens' health. Female enrollment in biomedical PhD programs rose by 10 pp relative to other fields, which in turn increased female health-focused research by 20 percent, without crowding out attention to men's health.

JEL: J16, I23

[♦] Hensvik: Uppsala University (Department of Economics), UCLS & CEPR, email: lena.hensvik@nek.uu.se. Nilsson (Corresponding author): Institute for International Studies (IIES), Stockholm University, Linnaeus University, CESifo & CEPR, email: peter.nilsson@iies.su.se

[♣] Comments and suggestions by seminar participants in Uppsala, IIES, Bergen, Royal Holloway and Queen Mary, Helsinki GSE, Tinbergen, Imperial College, Linnaeus University, SOFI, and EEA in Milan, the human capital and inequality workshop (Exeter), Neuchatel, Society of Labor Economists 2024 meeting in Portland, Autonoma, UPF-BSE, G-Gaps in Warsaw, Gothenburg, NUS (Singapore), Zürich, CEPR Paris Symposium, and Mannheim were all very helpful. Ulfhild Westin and Fabian Sinn provided fantastic research assistance, as did Mattias and Märta Almgren, and Anton Arbman Hansing. Funding from VR grant 2019-03177 gratefully acknowledged.

1. Introduction

"Of course no-one has intentionally been biased against women, but it could be a result of traditions" - The Chairman of the Swedish Medical Research Council, March 1995.

"[Wold and Wennerås' findings initiated a debate] of immense importance in getting what was previously more or less regarded as women's whining accepted as a legitimate problem." (SOU 1995:110)

The allocation of opportunity in science is commonly justified in the language of merit: scarce resources are supposed to flow to the most capable and promising. But in practice, the institutions that govern early scientific careers must take high-stake decisions using imperfect signals. Like employees in many other occupations, the assessments of early career academics past, current, and expected future performance involves considerable uncertainty. In such situations, decision makers may base their evaluations, at least in part, on stereotypes, past-experiences, or "traditions". Shortcuts based on easily observable traits can introduce bias in the evaluation of the competence of minority groups. When such evaluations occur at critical career junctures in jobs with a tournament structure (Lazear and Rosen, 1981; 1991), even small differences in perceived productivity can significantly affect the allocation of resources (Cornell and Welch, 1996).[1]

Documenting and disseminating evidence of evaluator bias may have a debiasing effect by directly influencing evaluators' decisions (Boring and Philippe, 2017; Pope, Price and Wolfers, 2018; Alesina et al., 2024).[2] But to what extent publishing findings of gender bias in academia (e.g. Wennerås and Wold, 1997; Sarsons, 2017; Wu, 2018; Card et al., 2020; Dupas et al., 2021; Koffi, 2021) can bring about meaningful behavioral or organizational change remains an open question. Even less is known about if and how much reveling bias at critical junctures influences researchers' careers and aggregate research output.[3] In grant-making contexts, if productive female researchers are overlooked in favour of less productive males, reducing bias should improve the selection of grantees and the overall research output of the funded researchers. However, aggregate output could also fall if the pre-disclosure bias reflects accurate expectations about group-level differences in future performance (Bohren et al., 2019).

[1] Biased assessments at these pivotal stages can generate substantial disparities in subsequent performance and careers across groups. Such disparities may, in turn, motivate and perpetuate unequal treatment in subsequent decisions (Coate and Loury, 1993; Altonji and Blank, 1999), ultimately reducing productivity both within occupations and in the broader economy.

[2] Following the widespread media attention of Price and Wolfers (2007) findings of racial bias among NBA referees, individual referees changed their decision-making process and the bias was eliminated (Pope et al., 2018). Informing middle school teachers about their implicit bias change assessments of immigrant students (Alesina et al., 2024), while informing undergraduate students about biases in teacher evaluations reduces male students' discrimination (Boring and Philippe, 2017).

[3] In a recent exception, Quach and Yu (2025) document the female supply-side responses to the dissemination of evidence for gender differences in recognition for group work in tenure evaluations by Sarsons (2017).

Such differences may arise from, e.g., the unequal division of childcare responsibilities that disproportionately hinder the long-term careers of female researchers.[4]

In this paper we document the short- and long-run individual, organizational, and field-wide impacts following a public disclosure of substantial gender bias in the competence assessments of young biomedical PhDs applying for a prestigious individual grant. The individual grant provided by the Swedish Medical Research Council (MFR) – the Swedish equivalent of the National Institutes of Health (NIH) – provided full funding for four-year research positions. To be eligible, applicants had to apply within five years of completing their PhD, submit a CV, a list of publications, and a research proposal. Evaluations followed a model adapted from the NIH. Individual reviewers first assigned scores to (i) the applicant's scientific competence, (ii) the relevance, and (iii) the methodological quality of the research proposal. Final funding decisions were based on the applicant's ranking, and a collective discussion, within review committees, resulting in a committee-specific priority list.

In November 1994, two microbiologists, Christine Wennerås and Agnes Wold (WW), noted that despite a near gender balance in the applicant pool, less than 20 percent of the grants were awarded to women. Surprised by this statistic, WW requested access to the reviewers' evaluation summary statements, linked them to citation records, and documented striking gender disparities in grant success rates, even after conditioning on citation counts, the stated relevance of the research proposal, or the applicant's independence.[5] Their initial findings were published in an op-ed in Sweden's leading daily in January 1995 (Dagens Nyheter, 1995a). WW´s findings received widespread media attention in Sweden (see Figure 1) and later on in international outlets such as the *New York Times*, *the Economist*[6], and triggered a heated debate within the biomedical community and among policymakers. While initially dismissing the WWs findings (DN, 1995b), in March 1995, the MFR pivoted and promised to increase "the share of females on the review committees […] successively from the current 10 percent to a more equal level" (DN, 1995c).

In support of their initial conclusion that "MFR reviewers could not evaluate the applicants in a gender-neutral way," WW published additional findings in *Nature* (Wennerås and Wold, 1997) after gaining access to the applicants' competence assessment scores. WW showed that, to receive the same competence score as a male candidate, a female applicant needed to be 2.6 times more productive as the average male peer—equivalent to publishing three additional articles in *Nature* or *Science*, or twenty additional publications in leading field journals. For most of the recent PhD graduates, this gap constituted an outsized hurdle.

[4] it is not clear that competence assessments based solely on observable quality measures, such as past performance, will increase allocative efficiency.
[5] See Appendix Table A3 for a reproduction and translation of the key table from the *Dagens Nyheter* debate article (DN, 1995a).
[6] NYT (1997); The Economist (1997).

In this paper we use the public disclosure of gender bias at the MFR as a natural experiment to examine whether – and to what extent – exposing such bias can lead to meaningful change in academia. To this end, we collected data on individual peer-reviewer assessments and linked them to applicants' pre- and post-application publications, and career outcomes from administrative data. The data cover all applicants during a ten-year window (1990-1999) around the publication of Wennerås and Wold's findings, which occurred between the 1994 and 1995 application rounds. In addition, we augment the analysis with linked information on all reviewers' background characteristics and publication histories, as well as individual data on all PhD students enrolled in Swedish medical universities, their demographics, and research interests.

We split the presentation of our results into three parts:

In the first part, we revisit, replicate, and extend Wennerås and Wold's original findings. The original study used a single cross-section. We document strikingly similar gender bias also in the preceding years, that are robust to inclusion of an extensive list of applicant and reviewer specific controls, from publication records and register data. When we apply the same test for male bias to the application round immediately following WW's findings, we find no evidence of bias. This first new finding suggests that publishing evidence of bias can reduce evaluator bias in academic grant-making organizations.

Revisiting and extending the baseline findings of WW also allows for a clearer understanding of the origins of the assessment bias, the changes induced by the revelation and how our results generalize to other settings. We provide several insights:

First, as the gender bias disappeared, the merit basis of the competence assessments, if anything, increased. This suggests that female candidates were not simply awarded higher scores in the post-disclosure period independent of their quality. Rather, reviewers placed greater weight on applicants' publication record than they had prior to the disclosure.

Second, the pre-disclosure male assessment premium is large and stable throughout most of the applicant competence distribution - except in the very top, where the estimated bias is smaller and insignificant. This suggests that there is less reviewer discretion when assessing top candidates – a star is a star.

Third, we use reviewer-level data to shed light on the changes in the organization and individual reviewer decision-making.[7] In the pre-disclosure period, a higher share of female reviewers on the review boards, and a better screening ability are negatively correlated with reviewer bias. Following the disclosure, the MFR systematically changed the gender composition of the review boards. Within two years, male-only review committees vanished

[7] Individual peer-review data are not traditionally made public in Sweden or elsewhere (Carson, Graff Zivin, Schrader, 2023).

(55% at baseline). The share of female-authored reviews increased from 10 to 45% within five years. However, the overall scientific competence of the review boards did not change. These patterns are consistent with the stated policy goal to increase the share of females on the review boards and a desire to specifically recruit competent female reviewers. However, the increased female representation cannot alone explain the elimination of the male competence premium. Except in the immediate year following WW, we find no clear indication that male and female reviewers were assessing male and female applicants significantly differently.[8]

Fourth, to better understand the mechanisms at work beyond the stated policy goals, we examine changes in reviewers' assessment behavior. Analyzing changes in inter-reviewer agreement, we find that review committees exhibited greater agreement in their assessments of female applicants before WW than afterward, when agreement rates converged to those for male applicants. Additionally, we find the reviewers that stopped serving after WW were no more biased than the reviewers that continued serving. Moreover, the retained reviewers changed their decision making-process dramatically, and became, on average, unbiased.

In summary, the results in the first part of the paper suggests that gender bias is a malleable trait, and that changed reviewer behavior played an important role for the demise of the bias. However, it is not clear whether these changes only improved fairness, or whether they also led to an increase in aggregate research output.

In the second part of the paper, we therefore provide a test of how the revelation of bias affected the core objective of the medical research council: to support research with the maximum international impact.[9] More specifically, we test for changes in the long-run research output gap across grantees vs. non-grantees before vs. after WW. An increased gap indicates an improvement in the selection on true-research potential. A key problem with a simple pre-post comparison of the grantee/non-grantee gap is that there could be many other unobserved factors influencing long-run research output that vary over time and that are not

[8] Bagues et al. (2017) use randomized selected evaluators in Spain and Italy for the period 2002-2014, and find that female evaluators are not more favorable towards female applicant (consistent with our findings). Although the difference in timing (early 1990s vs 2002, 2012), cultural setting, and career stages (post-doc vs associate/full professor promotions) and that our research question is different from Bagues et al., which may clearly matter for the differences in findings, - we don't find the unexpected effect of males becoming more hostile towards female applicants when more women are on the committees in our setting.

[9] To provide some context, male applicants have, on average, a much stronger publication record at application. Following the application and conditional on pre-application research output, the gender gap in research performance widens considerably. Eight years after application there is a gender gap in citations and impact factor of 60 percent and 45 percent, respectively. The substantial difference in research output is followed by a large gender gap in promotions and earnings observed in administrative data. The gender gap in the promotion to full professor peaks at 10 percentage points, 14 years after application. Given that the reviewers were all well-established and active researchers at Swedish universities—presumably aware of average gender differences in research output—it is natural to ask whether the observed disparities in treatment between male and female applicants reflect an implicit or explicit response to these differences. In other words, could the biased assessments be rationalized as a form of statistical discrimination based on group-level performance expectations?

directly caused by the changes within the MFR.[10] To address this concern, we exploit variation in the expected intensity of the impact of the WW disclosure across review committees. Specifically, we use the committee-specific male bias (pre-WW) as a measure of the expected impact in a difference-in-differences design and provide estimates of the impact on long-run research output (number of publications, impact, and citations).

We first show that the MFR did not allocate additional or fewer resources to the most biased committees following the disclosure of bias. Instead funding was reallocated to female applicants to a much higher extent.[11] In terms of long-run research output, we find that relative to non-grantees, the post-reform grantees outperform the pre-reform grantees if they were assigned to a review committee with a higher pre-WW bias. The effect is sizable: eliminating the average male bias increases the long-run research output gap by 0.23 standard deviations. The output effect is robust to a number of specification checks and is especially clear during the four years covered by the grant. But even 11 years after application there is no indication of a significant *decrease* in research output.

The long-run research output results are important since they suggest that the elimination of the bias did not generate a gender-equity/research output trade-off. On the contrary, the disclosure of bias led to an improved selection on true research potential, and as a result a higher allocative efficiency in the use of the public funds channeled through the MFR.[12] That the research-output gap between funded and unfunded applicants increased after WW is moreover difficult to reconcile with model of accurate statistical discrimination motives (as we show in the sequel).

In the third part, we examine how the revelation of the MFR bias impacted the research community more broadly. Following WW, a new policy goal was added: "to increase the number of *female researchers*" (MFR, 1997, emphasis in original). An underlying motive besides fairness concerns, is that an increased representation of women could lead to a greater attention to womens' health.

Using population wide data on PhD-enrollment, we show that over the four consecutive years directly after WW, the share of females enrolling in medical PhD programs increases by 10 pp relative to other programs. The variance in the female enrollment share across medical

[10] For example, during the late 1990s there were a number of initiatives that aimed at increasing female researchers at all stages of the academic career ladder. See e.g. Vetenskapsrådet (2015) for an excellent summary of policy initiatives during our observation period (in Swedish).

[11] This finding echoes earlier findings by Sandström and Hällsten (2008) who in a pre vs. post comparison find that the aggregate gender gap in grant probabilities is reduced after WW. We show that this aggregate pattern is not caused by changes in applicant quality, holds for accounting for applicant characteristics, and for using across review committee variation in pre-WW bias Sandström and Hällsten (2008) do not focus on the post-doc grants (as WW did), which is an important distinction since the uncertainty in applicant quality decreases significantly with seniority/publication history. Moreover, in their single cross-section data for 280 applicants in 2004 – 10 years after WW - and can therefore not investigate when the changes occur. In addition, they do not have access to the competence scores (only final ratings which is a composite of the project quality, method, and competence scores), and can therefore not adjust for quality differences. They do not investigate effects on research output, do not use individual reviewer data or look at changes in the composition of the review committees. In summary their focus, findings, and analysis are distinctly different from ours.

[12] Lundberg and Startz (1983) study how discrimination affects allocative efficiency in the labor market.

universities decreases, relative to other fields. These findings indicate that the revelation of bias at the MFR triggered coordinated efforts to increase female enrollment rates in biomedical research programs.

We then examine if the compositional changes affected the direction of medical research. There is indeed a strong correlation between PhD student gender and research focus in the cross-section: female PhD students are significantly more likely to publish work related to women's health, particularly when they are first authors.[13] However, whether increasing the representation of women in research leads to greater attention to female-focused health outcomes remains an open and important question—especially given the gendered nature of the outcome and the fact that women comprise half of the population.

To shed light on the causal question, we use the sharp increase in female PhD enrollment following WW in an instrumental variable framework – using the pre-WW share of female graduates across university-fields as an instrument for the changes in female enrollment. Our analysis indicates that a 10 percentage-point increase in female enrollment increases female health-focused research by around 20 percent. Our findings complement, extends, and are of similar magnitude as the recent findings from Truffa and Wong (2025), who focused on the extensive margin enrollment following the transition of US university from male-only to coeducated undergraduate programs. That our findings lines up with the Truffa and Wong findings is important for external validity and policy relevance of both studies, especially given that most university programs today are coeducational.

The rest of this paper is structured as follows: Sections 2 and 3 provide background and describe the events leading up to the changes in the organization of the reviewer committees and their work. Section 4 describes the data. Section 5 presents the results and Section 6 concludes.

2. Background

Biomedical research

Freeman, Solomon, Rosenbaum, Marincola, Weinstein (2001) describes research in the biosciences as fitting a tournament economic structure. A tournament offers participants the chance of winning a big prize. It fosters intense competition by amplifying small differences in productivity into large differences in recognition and reward. Tournament victory may result from being marginally better than competitors, e.g., completing a key experiment a week earlier.

[13] The finding is consistent with previous work documenting that female majority in research teams is associated with a higher degree of female health focused research (Koning et al., 2021).

The structure of biomedical research in Sweden is similar to other countries. The Principal Investigator (PI) is responsible for the direction of the research and the lab, while graduate students and post-docs work long hours and are motivated by a hope for an independent research career and producing impactful findings, rather than direct monetary incentives.

Transitioning from the junior position to more senior positions with their own research group to a large extent relies on the ability to attract funding, which in turn is partly based on past productivity as measured by the number and quality of the papers produced. Funding from within universities is also partly reliant on measures of research output. For example, the Karolinska Institute (KI), Sweden's leading biomedical research institution, distributes research funds across their departments based on the department's researchers' impact factor.[14] The impact factor reflects journal quality by measuring the average number of citations a paper receives within the first two years after the date of publication.

The Medical Research Council (Medicinska Forskningsrådet, MFR)

Several public and private funding agencies provide opportunities for funding of research. The largest medical research funder at the time was Medicinska Forskningsrådet (MFR). The purpose of the MFR, the Swedish equivalent of the National Institutes of Health, was to fund and promote research in the areas of medicine, pharmacology, odontology and veterinary medicine. MFR's primary objective was to "support basic research of high international quality" (MFR, 1994).

Entirely state-funded, the MFR had a national responsibility to promote independent basic research. The MFR funds were assigned to specific research projects and research positions where the individual was the applicant. Each year, the MFR funded project grants, research positions, as well as equipment and funds to visiting scientists.[15] The applications for the MFR grants and positions were evaluated and graded by field-specific review committees, which included the deputy council members, as well as other established and active researchers.

The Research Fellowship and the Peer Review Process

The stage after the PhD dissertation is seen as a crucial stage in a researcher's career, and in order to improve young researchers' career opportunities, MFR provided individual funding targeting young promising researchers at the start of their postgraduate career. The grants intended to facilitate the transition between being a dependent graduate student and an independent researcher by allowing them to continue their research and qualify for more senior

[14] KI is consistently ranked among the world's best medical schools and home of the Nobel assembly for Medicine and Physiology.
[15] Within the MFR, 60% of the budget went to project grants, and 20% went to research positions.

positions. The grant fully financed research for four years.[16] The research fellowship position was the most common position funded by the MFR. The applicants were required to have completed their PhD in the past five years to be eligible. Upon application, the researchers were required to send in a CV, bibliography, and a research proposal.

The peer review process for assessing applicants took place in review committees. The application is assigned to one of eleven committees[17], each consisting of five regular members. The individual reviewers' first independently grade the applicant/applications in three dimensions:

(i) *(R)esearch topic* - to what extent the project can contribute to medical research in an innovative and significant way,

(ii) *(M)ethodology* - to what extent the suggested methodology is well prepared, thought-through and suitable to the research question, and

(iii) *Scientific (C)ompetence* - the applicant's experience, productivity, skills, and independence as a researcher.

The reviewer (*i*) submits her *R*, *M* and *C* scores for applicant (*j*) to the MFR. The subscores are multiplied to generate a reviewer-applicant specific Total score ($T_{ij} = R_{ij} \times M_{ij} \times C_{ij}$). In other words, the subscores' are viewed as complements. The MFR calculate a final applicant score (and thereby a within-committee rank) based on the average total applicant score, $\sum_j^N T_{ij} / N_j$, where *N* is the number of reviewers assessing candidate (*j*). In almost all cases five reviewers submit their R, M, C, scores for each applicant. In cases of reported conflict-of-interest (e.g. the reviewer and applicant are co-authors) a deputy reviewer steps in and assess that particular applicant.

Next, the committee meets for a full-day session, during which each application is discussed individually. Based on the within-committee ranking and the joint discussion, the review committee agrees on a within-committee recommendation (high, medium, low). Typically, each committee designates one to three applicants as 'high priority,' as noted in the summary statements. Final decisions on grant allocations are made by a separate committee composed of the review committee chairpersons.[18]

[16] Under special and strict circumstances, the grant period could be extended to six years.

[17] Physiology I, Physiology II, Chemistry I, Chemistry II, Psychiatry, Microbiology, Epidemiology/ Public health, Medicine, Surgery, Morphology, Odontology. In 1998 the number of committees increase to 14 and some of them change names. We describe how we deal with these changes, which are important for our long-run research output analysis, below.

[18] The gender gap we document below originate from the assessments made by the review committees, and not from the committee of chair persons. Conditional on the within-committee rank, there is no indication of gender bias in the allocation of the grants (not reported).

Our empirical analyses focus on the competence score for two main reasons: First, internal-MFR documents reveal that for the personal grants "the personal competence is the main assessment criterion, closely followed by the project", and that competence is based on the applicant's publication history (see Appendix B). Second, the submitted applications have been deaccessioned by the National Archives and are no longer available.[19] Since we can recover the publication histories of the applicants using Pubmed/Medline, we can assess how the applicant competence is assessed conditional on, e.g., the quantity and quality of the publications at the time of application. We also report findings for the Method, Relevance, and Total score, for completeness in the Appendix. However, because the applications have been destroyed, we are unable to control for differences in the quality of the research proposal (*R* and *M* scores), and thereby also the Total Score. The Competence assessment scores from the 1994 application round is also what was analyzed by Wennerås and Wold (1997). In the following section, we summarize the findings as reported in their op-ed and academic article.

3. The Disclosure of Bias by Wennerås and Wold

In November 1994, Christine Wennerås and Agnes Wold (WW), noticed that despite the almost gender-balanced applicant pool only 20 percent of the MFR post-doc. grants were awarded to women. Surprised by this outcome, WW gained access the evaluation summary statements, linked them to citation records, and published their initial findings in January 1995 in an op-ed in the leading Swedish daily (DN, 1995a). WW documented surprisingly large gender differences in grant probabilities by gender conditional on citations, the stated relevance of the research program, and the independence of the applicant.[20] WW's findings received widespread media attention in Sweden (see Figure 1) and later on in international outlets such as the *New York Times*, *the Economist*[21] and triggered a heated debate within the biomedical science community and among policymakers. While initially dismissing the WW findings entirely, in March 1995, the MFR pivoted and promised to increase "the share of females on the review committees […] successively from the current 10 percent to a more equal level" (DN, 1995b).

Initially, WW requested access to the so-called white sheets—protocols from the review committees detailing the assessment scores for all applicants—but were informed that these documents had been destroyed. Following an investigation, the Swedish National Archives established that these documents were public records and that the MFR did not have the authority to destroy them. During this process, the National Archives received an

[19] The only free text information of the stated quality of the research program that is still available is summary statement from the review committee (typically 1/3 to 1/2 page) that summarizes the program and the assessments of the applicant.
[20] See Appendix Table A3 for the reproduction and translation of the Key table from the *Dagens Nyheter* debate article (DN, 1995a).
[21] NYT (1997); The Economist (1997).

anonymous phone call from within the MFR, prompting a site visit. During this visit, they discovered printouts of the white sheets in a box located in the basement.

When WW gained access to the competence assessment scores from the white-sheets of the 1994 applicants they published additional findings in *Nature* (Wennerås and Wold, 1997). WW showed that, to receive the same competence score as a male candidate, a female applicant needed to be 2.6 times as productive as the average male peer—equivalent to publishing three additional articles in *Nature* or *Science*, or twenty additional publications in leading field journals. For most newly minted PhD graduates, this constituted an outsized hurdle.

4. Data and Descriptive Findings

The natural experiment that occurred when WW's initial findings were published between two application rounds helps us to shed light on whether and to what extent revealing gender bias in grant decisions can bring about meaningful change. Our goal is to provide a comprehensive view of how WW impacted the researchers, the reviewers, the organization, the broader medical field, and the direction of research. To do so, we link data on individual peer-reviewer assessments to the publications, administrative career records, and background characteristics of all applicants in a ten-year window around WW. We complement the applicant data with background data on all reviewers. In addition, to shed light on the broader consequences of the revelation of bias, we gathered data on all PhD students from Swedish medical universities and the students direction of research up to 17 years after PhD enrollment. This section describes the data and provides descriptive statistics for the applicants and the reviewers.

Applicant data and characteristics: We first collected data from the MRC "white-sheets" stored at the National Archives.[22] We gathered individual data on all applicants between 1990 and 1999. For each applicant we collected: application number, name, title, type of position (general or priority area), review committee, and the reviewer scores (total score, competence score, methodology score, and research topic score), and whether the application was granted funding or not. The first name of the applicant is used to define gender.[23]

To measure academic productivity before and after the application, we link all the applicants to their publications from 10 years before and to 1 to 11 years after the application date using Medline/PubMed. To resolve name disambiguation, we use the validated

[22] Archive code SE/RA/420710.

[23] In cases when it was unclear which gender a name typically belongs to, we looked through the white sheet for clues, did web-searches and if an applicant's name could be used by both females and males, the most common gender of the name was used. In some cases we also consulted native speakers to assign the correct gender to the applicant. For the period 1996 to 1999 we can can also compare actual and assigned gender of the applicant, the correspondence for this period was very high. For the sample we can link to administrative data we also assessed the actual and coded sex of the applicants – the correspondence in the linked sample was very high >99%.

"Author-ity" MEDLINE author disambiguation database, designed for this purpose (Torvik and Smalheiser, 2009; Torvik et al., 2005).[24]

We create four productivity measures using the publication data: the number of publications, the number of first-author publications, the total impact factor of the publications, and the total impact factor of first-authored publications. In biomedicine, being the first and last author have strong signal value. The first author is the researcher who contributed the most, while the last-author is typically the most senior author (e.g., the head of the lab).

Since our sample consists of early-stage researchers, we distinguish between the publications where the focal researcher was the first author or not. The quality-adjusted measures of research productivity are created by using the impact factor of the journal. Using the Journal Citation Report (JCR), which is a standard measure of impact factor of biomedical research journals, 88 percent of the articles published by the researchers in our sample can be matched to the JCR impact factor.[25] For each applicant, we create a *pre-application* measure of productivity (articles published within 10 years before the application date), and a *post-application* measure of productivity (articles published 1 to 11 years after the application date), by summarizing the impact factors for the relevant time interval. As a complement to this impact weighted quality measure of the publications, we also retrieved the cumulative number of citations for each paper from Web of Science within the same interval as the impact factors.

In addition to the quantity and quality of researchers' output, we use the Medical Subject Headings (MeSH) data from PubMed. The MeSH thesaurus is a controlled and hierarchically-organized vocabulary produced by the National Library of Medicine.[26] It is used for indexing, cataloging, and searching of biomedical and health-related information. We use the MeSH data for each article to create two measures capturing the direction and novelty of applicants' research portfolio. The article novelty measure developed by Mishra and Torvik (2016) quantifies the conceptual novelty of the researcher's publications.[27] The novelty index reflects the number of papers that were published on the topic, or alternatively the years since the first publication on the topic relative to the year the applicant's paper was published. Using

[24] Our RA Ulfhild Westin did a heroic job in manually searching for all applicants in the Authority data (Author-ity Exporter). Correcting for name changes/spelling errors etc. Verifying uncertain case using internet searches, research topics/paper titles etc. We also used the SCB name-database to compare assigned vs. actual gender for the administrative data matched sample. The assigned and actual sex of the applicant was >99% correct. The Author-ity database gives us all PMIDs (article IDs) in Pubmed (since 1950) for all the applicants. At the time of data collection the Author-ity exporter contained publications up to 2011. In our later analysis of the output of all PhD students we used the full 2018 Author-ity database which covers paper up to 2018.

[25] The first digitalized JCR impact ranking is from 1997. For all publications published before 1998, we assign an impact factor using the 1997 JCR. For all publications after 1997, we use the JCR ranking in the year of publication to assign an impact factor to the publication.

[26] Each MeSH term has a unique identifier and are the indexing of new MeSH terms supplemented by a long period of intense mentorship at NLM that takes place between a new indexer and his/her reviser, who reviews and corrects all of the indexer's work.

[27] Mishra S, Torvik VI. Quantifying Conceptual Novelty in the Biomedical Literature. *D-Lib magazine : the magazine of the Digital Library Forum*. 2016;22(9-10):10.1045/september2016-mishra.

the MeSH terms, we also constructed a measure of direction of research by measuring the number of articles which focus on the medical needs of women.

To further study the career dynamics of the researchers, we gathered data on academic position and income for all the applicants using Statistics Sweden LISA database (1985-2019), University graduation register (Universitet- och Högskoleregistret - doktorander 1973-2019), and University employee register (Universitet- och Högskoleregistret – personal 2001-2019). The administrative data are comprehensive and allow us to examine several additional outcomes, and to take, e.g., university of graduation and year of graduation into account. A drawback compared to the publication data is that the administrative data only covers individuals who graduated and reside in Sweden.

Table 1 provides descriptive statistics for the applicants during the observation period. From this table, we can see that the applicant pool is almost gender balanced (57 percent male). The grant probability in a given round is 15 percent. Since it is possible to reapply within five years after PhD (40 percent reapply if not granted), the share of applicants ever granted funding is higher (27 percent).

The administrative data show that the average applicant is aged 36 and completed their degree at age 34. 60 percent of the applicants have children at the time of application, and 69 percent are born in Sweden. Information on reported conflict of interest indicates that 10 percent of the applicants have a connection (typically previous co-author) to a member of the review committee. Note that the applicants are assigned to review committees based on research topic.

In biomedicine, parts of or the entire PhD thesis is normally published or in the process of publication before submitting the thesis for examination. The average applicant published their first paper 6 years before application, and at the time of application they had published 13 papers. On average they had 93 citations, with a large dispersion across applicants (SD 159). Following Card et al. (2020) we also calculated a measure of post-application citations of pre-application papers, in order to measure unobserved quality of the papers (i.e. unobserved to us but potentially observed by careful reviewers).

The descriptive statistics in Columns (2) and (3) of Table 1 reveal several interesting differences across the male and female applicants. The clearest difference is in the publication history and grant probabilities which overall are much higher for men than women. On average, male applicants have published more papers, in higher impact journals, more often as a first authors, have a larger set of co-authors, and have more citations than their female peers. On average, a higher share of men than women receive the grant in the application year (18 percent vs 12 percent, and 32 percent vs 19 percent over time).

Finally, Statistics Sweden provided us with information on whether the applicant had changed their name using information from the name register. Name changes could be

important since the Author-ity algorithm, and hence our search for the applicants' publication histories in part relies on last-name matching.[28][29] In our sample, there is a clear difference in the share of last name changers between men (1.5 percent) and women (13 percent). Since marriage rates and divorce rates have increased over time, name changes could be a problem for analyses based on publication records, since name changes could reduce the accuracy of the matching across men's and women's publication records.[30] The name change data gives us a unique opportunity to test if excluding name-changers from the analysis matters.

Reviewer data and characteristics: The baseline reviewer data were collected from the official yearly project catalogues published by the MFR (e.g. MFR, 1997). In these publications the full names, titles, workplace, review committee, and reviewer role (chair, member, deputy) are recorded. The reviewers are well-established and research-active professors (a few are associate professors). Reviewers usually serve a six-year term. We use the first names of the reviewers to determine gender, and we collect information on last-author publication impact factor of the reviewers using the same approach as for the applicants.

Individual peer-reviewer data are traditionally not made public in Sweden or elsewhere (Carson, Graff Zivin, Schrader, 2024). WW uses the average competence assessment score across reviewers in their analysis. We also collected the individual reviewer assessments to provide insights into the mechanisms behind the changes we document below. Moreover, we were able to link individual reviewers to their assessments of each applicant by combining information from internal MFR documents.[31] This allows us to assess how individual reviewer characteristics correlate with male competence premiums.

Figure 2 provides a first view of the stark changes in reviewer composition following WW's revelation of bias, as shown using the reviewer data. Figure 2A shows the trend in the share of new reviewers serving on the committees. In the first application round after WW, the turnover increases by around 40 percentage points compared to a normal year. Figure 2B shows that the share of "all-male" reviewer committees decreases from 55 percent to 0 within two years, while the share of female reviewers increases from 10 percent to 20 percent in the first year and continues rising throughout the observation window, reaching almost 40 percent in 1999.[32]

[28] The Author-ity model (Torvik et al. 2005) was developed for estimating the probability that two articles in MEDLINE sharing the same author name are written by the same individual. Author-ity use information on first-names, initials, last-names, author affiliation, publication topics (MeSH), journal names, email addresses etc. to match the authors across PubMed records. See e.g. Torvik and Smalheiser (2009) for further details.
[29] Historically in Sweden it was standard practice to take the husbands last-name or combine the wife's maiden name with the husbands last name and form a new last name, while the husband typically kept his last name.
[30] Similar problems are likely to occur, to some extent, in all research using names to create information on researchers' careers.
[31] Appendix B describes the method for linking reviewers to their assessments.
[32] The share of female authored reviews follows a similar path (see appendix Figure A3)

The trend in reviewer competence, as measured by their cumulative impact during the previous ten years, is shown in Figure 2C. Internal documents from the MFR, dated shortly after WW, highlight a clear intention to increase the share of female reviewers—an aim that was also publicly communicated—while simultaneously emphasizing that newly appointed female reviewers should be "competent". Figure 2C shows that despite the large changes in the gender composition of the reviewers, there is no major change in their overall scientific competence. Figure 2C also highlights that the reviewers had a high scientific productivity in the previous 10 years, with average impact factors well above 200 (corresponding to, on average, one publication in *Nature* per year during the past 10 years).

Below, we use the linked individual reviewer data to assess the extent to which observable reviewer characteristics predict higher male competence assessment premiums, to control for reviewer fixed effects, and to see how individual reviewers' competence assessments change over time.

5. Results

We present our findings in the following three sections: (i) the male competence assessment premium, (ii) the impact on long-run research output, and finally (iii) the broader impact on the medical research community. The empirical approaches we use are described as we go along.

Part 1: The Male Competence Assessment Premium Before vs. After WW

The previous section documents the promised and observed changes in the composition of the medical research council review boards. In this section, we first replicate the findings by Wennerås and Wold (1997) who used a single cross-section of applicants in 1994, and then document how the male bias changed following this event.

WW focuses on the gender gap in the average competence assessment score. To account for differences in quality, WW uses single quality measures (total impact, first-author impact, or first-author citations) with a focus on total impact. They also include information on the connection to the review board and other background applicant characteristics in their specification.[33]

Table 2, column (2) reiterates and replicates WW's key specification with the data at hand. We standardize the competence score (mean 0, std. 1) for comparability across years, but the unstandardized specification replicates the findings by WW (1997) almost perfectly (not reported). WW (1997) draw two conclusions from the results in column (2). First, conditional

[33] In their models WW account for applicant's nationality, education, field of research, postdoctoral experience, recommendation letter and the review committee assessing the applicant. Besides total impact, the three significant determinants of the competence score are applicant gender, connection to a reviewer and reviewer committee fixed effects.

on quality adjusted measures of applicant publication history (the total impact factor at the time of application), connection to the review board, and review committee fixed effects, there is a substantial male competence premium of 0.65 standard deviations.

To put this male premium into perspective, we use the estimated return for an additional unit of impact factor (0.009) on the competence score to calculate the gender gap in terms of output equivalents. To receive the same competence score, a female applicant would have needed to outperform the male applicant by 72 impact points (0.65 / 0.009). This represents approximately three extra papers in *Nature* or *Science* (with impact factors of 25 and 22 respectively at the time), or 24 extra papers in top field journals (impact factor 3). In our sample, the mean impact of the 1994 cohort is 45 impact points, indicating that a female applicant had to be 2.6 times as productive as the average male applicant to receive the same competence score (45+72)/45=2.6), identical to the estimate in WW.

Our analysis relies on the judgment made by WW about which factors biomedical researchers use to assess applicant competence at the time. Independent support for the importance of publication history in general, and the impact factor in particular, is documented in a letter from the chair of the microbiology review committee (Britta Wahren) to the council chairman. The letter, reproduced and translated in Appendix B, contains Wahren's "list of variables that [she] tries to identify for the [subscores]", and she adds "I believe most reviewers use similar variables to prioritize the candidates" […].

Table 2, column (1) provides the unadjusted gender gap in competence scores, and column (3) provides the results when using multiple measures of applicant productivity at the same time. Comparing columns (1) and (2) highlights that part of the observed gender difference in the competence score indeed reflects observable gender differences, captured by the single total impact productivity measure and connectedness to the review committee. However, adding additional measures has limited impact on the estimated bias, although it seems (consistent with Wahren's list) that the order of the author (first-author impact) is more important than the overall total impact measures, highlighting the importance of independence.

Wennerås and Wold (1997) provide results on the male premium in 1994. What happened after the disclosure of the bias? In Column (4) of Table 2 we provide the first corresponding estimate of the bias in the year immediately after the revelation. First, immediately after the revelation, the male competence assessment premium vanishes completely. Second, the merit basis of the competence score, as reflected in the estimated coefficients for total impact and first-author impact, if anything increases. This result provides the first evidence suggesting that the revelation of bias triggered immediate changes that

eliminated the male competence bias.[34] It also shows that the elimination of male bias was not simply due to favoritism towards women, since the observable productivity measures (especially the first-author impact) are, if anything, given more weight after WW than before. If female candidates were given an unmotivated premium afterwards, the merit basis of the competence score would decrease. We interpret this finding as suggestive that the assessment criteria were made stricter and/or that the reviewers changed their behavior (we return to reviewer behavior below).[35]

Figure 2D shows the evolution of applicant quality for male and female applicants as measured by their cumulative impact factor at the time of application. As can be seen from this figure, there is an increasing trend in the total impact factor, but the differences between men and women remain constant over time.[36] Figure 2D is important for the interpretation of our baseline findings since it indicates that the elimination of the male premium cannot be explained by dramatic changes in the composition of applicants. Appendix Figure A3 provides further findings indicating no dramatic changes in the research overlap with the reviewers.

Appendix Figure A4 follows DiNardo, Fortin and Lemieux (1996) and shows the distributions of the difference in the observed and predicted competence score of males before and after WW, where the prediction for males is based on the baseline model estimated only from the female applicant sample. Before WW the distribution is clearly shifted to the right – highlighting the male advantage, whereas in the year after WW the distribution is centered around zero.

Persistence, Robustness and Heterogeneity

How sensitive are these baseline results to alternative specifications, and how persistent was the bias before the revelation and how persistent is the absence of bias after the revelation? Figure 3A provides year-specific estimates throughout the observation period and also reports a number of robustness checks. The baseline specification estimates (Table 2, column 3) are plotted in black. The coefficients reveal that the bias was a persistent phenomenon in the years leading up to WW, and that it shrinks immediately after and hovers at a permanently lower level.

The second set of estimates in Figure 3A stems from a double-selection LASSO linear regression model (Belloni, Chernozhukov, and Hansen, 2014) which, on top of the baseline productivity controls, selects from a wider set of productivity controls based on the Wahren

[34] This result is in contrast to conclusions in Wold and Chrapkowska (2004) that the changes in council procedures in 1997 rendered the reviewer decisions gender neutral.

[35] Table A4 provides the results when excluding name changers (based on information in Statistics Sweden's name change data base). The results do not change when excluding name changers.

[36] Appendix Figure A2 plots the difference between male and female impact factor.

letter and the other productivity measures listed in Table 1.[37] [38] The third set of estimates uses the individual reviewer scores and accounts for reviewer fixed effects. The fourth set of estimates is based on a rank-ordered logit model, allowing reviewer scores to reflect the relative ranking of the applicants rather than the absolute quality differences. The fifth set of estimates provides Double LASSO selection estimates for the sample of applicants that we can link to the administrative data (including year of PhD, place of PhD, etc.) on top of the baseline productivity controls. All of these alternative specifications provide a very similar overall picture as our baseline specification: a high and stable bias in the pre-reform, followed by a dramatic level shift in the estimated bias immediately after the reform.

Figure 3B shows where in the competence distribution the bias is most severe, based on unconditional quantile regressions (Firpo, Fortin, Lemieux, 2009). The figure plots the baseline OLS estimate in the pre-reform period as the dashed grey line, and the corresponding quantile estimates in black. The bias in the pre-period is evident throughout most of the distribution, except at the very top (above the 80th percentile). We interpret this as an indication that there is less reviewer discretion when assessing top candidates.

Changes in reviewers' decision making

Our baseline results use the average competence assessment across reviewers in the same review committee. In this section, we focus on the individual reviewer assessments.

In Figure 3C we examine the differences in the male premiums across female and male reviewers. The black point estimates and dashed confidence bands show the estimated male premiums for male reviewers, and the white point estimates show those for female reviewers. Recall that in the pre-WW period, the share of female reviewers was only 10 percent as reflected in the wide pre-WW confidence bands. The results in Figure 3C suggest that both male and female reviewers put lower weight on the gender of the applicant in the post-period than in the pre-period. Except for the year directly after WW, there is no difference in how male and female applicants are assessed by male and female reviewers.

Since reviewers assess many different candidates, we can also estimate the reviewer-specific bias. Figure 3D plots the kernel densities of these reviewer bias estimates before and after the disclosure of bias, weighting the reviewer estimates by the inverse of the standard error. First, prior to the WW revelation it is clear that there is a wide distribution of

[37] The full double robust LASSO model variable selection list is on top of the baseline productivity controls: any NIH supported publications, the share of the total impact from the last two years before application (and square), share first author impact from the past two years (and square), share publications from the past two years (and squared), four novelty measures for each publication based on MeSH term novelty (the mean novelty in time for single MeSH, and pairwise MeSH, and the corresponding time since first MeSH novelty measures, see Mishra and Torvik (2016)), committee connectedness, the overlap of the review board research interest with the applicants research interest based on the average share of shared MeSH terms in their publications, the post applications citations of pre-application papers, total impact per author (and square), and number of unique co-authors.

[38] Table B1 provides the mapping between the Wahren list and our data.

reviewer bias, with a clear majority of reviewers exhibiting bias in favor of male applicants. Second, after WW, the distribution shifts to the left and becomes centered around zero. This result is interesting since it shows that the debiasing of the competence assessments is not due to, for example, a polarization of the reviewers, which could have generated a bimodal distribution.

In Table 3, we split the reviewers into a pre-WW group consisting of "exiting" and "retained" reviewers, and a post-WW group consisting of "retained" and "entering" reviewers. Column (1) confirms our baseline findings in the reviewer-level data: the bias changes dramatically and significantly across the two periods. Column (2) shows that the exiting and retained reviewers have similar levels of bias prior to the revelation. In other words, the reviewers that exited after WW were not dramatically different than those retained. However, as shown in column (3), the retained reviewers change their bias significantly after the revelation. In the post-period, they are no longer biased, and their assessments are not significantly different from the entering reviewers. Within a year, the retained reviewers' decision-making process changed dramatically, and, on average, was no longer biased in either direction.

In Appendix Table A1 we also assess how reviewer and review board characteristics correlate with the male competence premium in the pre-period. Using the linked reviewer data, we show that a higher screening ability (scientific competence and career exposure to female PhD students), and a higher share of females on the review board are correlated with a significantly lower male competence assessment premium. These findings are consistent with that individual specific information reduces the reliance group averages for the assessments of applicants' quality. Note that the findings in Appendix Table A1 are correlates and should not be interpreted causally – we report the correlations since very few previous studies have been able to correlate reviewer observables (e.g. scientific competence and exposure to female PhD students) beyond gender with biases in competence assessments.[39]

Finally, we estimated the change in inter-reviewer agreement of the competence assessments scores for male and female applicants just before and after the disclosure of bias.[40] Using Gwet's AC chance corrected inter-rater agreement measure (Gwet, 2014; and Klein, 2018), we find an insignificant decline in the reviewers' agreements for males (0.84 to 0.82, n.s.), but a significant decline in agreement for female applicants (0.90 to 0.82, $p<0.01$). In other words, overall, the review committees agreed more about the (lower) competence of

[39] For example, that the share of male reviewers is (marginally) positively correlated with male bias should not be e.g. interpreted as indicating that other reviewers are necessarily influencing the individual reviewers scores, since it could reflect other unobserved factors that also differ across fields. The reviewers first grade the applicants independently. We cannot, of course, rule out that the reviewers were in contact via e.g. telephone to discuss application before grading and meeting to discuss the applicants.

[40] The simplest measure is the percentage agreement between reviewers, i.e., the number of times any pair of reviewers agree on the same grade divided by the number of possible pairs.

female applicants before WW than after, when the reviewer-agreement rates converged to the male applicant level.[41]

What caused these changes in reviewer behavior? Internal and external documents reveal that the MFR adopted a three-pronged approach in response to the revelation of gender bias: (i) an immediate internal discussion about the competence assessment criteria, (ii) internal-follow up and quality control of the evaluation process, and (iii) an increase in the share of female reviewers (cf. MFR, 1997; and Appendix B). Our finding of an immediate demise of bias and no significant differences in bias by reviewer gender, could suggest that the (relatively slow) increase in the share of female reviewers is unlikely to be the main cause of the change in reviewer behavior, but the rapid phase-out of male-only committees could. Besides the actions taken by the MFR, there could be direct effects of the disclosure of the bias on the individual decision-making process (Pope et al., 2018; Alesina et al., 2024).[42]

Summing up: After WW, the high male competence assessment premium vanished. At the same time reviewers agreed less with one another about female applicants, while reviewer agreement about male applicants did not change. High-competence reviewers exhibit less bias, and so do reviewers with greater exposure to female PhD students over her career. Moreover, it is interesting to note that the only time we observe a reversal of bias (i.e. favoring women) is in the immediate year after WW, and then only among female reviewers. In the next section we outline a model of reviewer behavior, that explains these findings and offer a test for the impact on allocative efficiency.

Part 2: The Impact on Allocative Efficiency

The key stated objective of the Medical Research Council was to support basic research of high international quality.[43] *A priori*, it is not clear how research output will be affected by the disclosure of bias – as it depends on the underlying source of discrimination. If female researchers were previously displaced in the ranking by less talented male researchers, a sharp reduction in bias should, all else equal, raise the average research output of grantees. By contrast, if reviewers held accurate priors about gender differences in expected output conditional on funding and used this information efficiently, eliminating male competence premia could reduce expected output, creating a potential trade-off between gender equity and

[41] The reported numbers reflect the agreements using the Gwet's AC2 chance corrected ACs with ordinal weights (see Gwet, 2014; and Klein, 2018). Overall agreement before WW: .86 (range .76-.93 across committees), and after WW: .82 (.64 -.91 across committees). We refer the reader to Gwet (2014) for further details.

[42] Moreover, the structure of the grading system, which had been adapted from the NIH 30 years earlier came under scrutiny and was updated. The grading scales were extended allowing for a more nuanced assessment, and the subscores (competence, relevance, and method) which jointly determined the final within review board ranking, were previously viewed as complements in the calculations of the final score changed to being viewed as substitutes. However, the grading system changes were not implemented until years after WW (in 1998), and are therefore not a cause of the immediate elimination of the male premium.

[43] See MFR project catalogues 1988-1994 (e.g. MFR, 1994)

research output.[44] To organize our thinking, we develop a simple conceptual framework (with full details in the Online Appendix) and then test for allocative efficiency using within-MFR variation in the expected impact of the WW revelation and realized research output at the end of the grant period.

Conceptual framework: Discriminating Reviewers and Allocative Efficiency

In the model, MFR chooses which *K* applicants to fund, but delegates the ranking of applicants to reviewers. MFR's objective is to maximize the expected sum of ex post research output among the *K* funded projects. Reviewers hold (possibly group-specific) beliefs about the latent quality q_i of applicants from groups $g \in \{W, M\}$. They observe a noisy signal s_i (e.g. pre-application citations, publications, impact measures), update their priors using Bayes' rule, and rank applicants by their posterior expectation of output conditional on funding. MFR then funds the top *K* applicants according to this ranking.

Because reviewers use group-specific priors, two applicants with the same signal s_i can be treated differently if they belong to different groups. In particular, if reviewers believe that men have higher expected quality than women ($\tilde{\mu}_M > \tilde{\mu}_W$), they implement a lower effective cutoff for men ($C_M > C_W$). Appendix Figure A5 provides correlations that could motivate such differential beliefs in the present context using data on pre-WW applicants' post-application careers[45] Figure A5 shows substantial gender gaps in research impact, citations, promotions and earnings over years post-application. We assume that these correlations are well-known to reviewers, who are all senior and well-established researchers at Swedish universities.

Whether the use of such group-specific cutoffs are efficient depends on the nature of the priors. If reviewer priors reflect true predictive differences in expected output (i.e. $\tilde{\mu}_g = \mu_g$), then allowing for group-specific cutoffs can be *ex ante* efficient: it uses all available information and maximizes expected aggregate research output. On the other hand, if priors are misinformed, or reviewers have taste-based preferences in favor of men, then selection is pushed away from true-quality. In that case, a gender-blind rule that equalizes cutoffs across groups raises expected output among grantees, Thus, the net impact of the WW revelation and the associated shift toward gender-neutral assessment on research output is ultimately an

[44] Constraints outside the control of the reviewers could play an important role - such as unequal family commitments or higher subsequent career barriers (Sarsons, 2017, Card et al., 2020).

[45] Figure A5 show that, conditional on pre-application research output, review committee and year of application fixed effects, the over-all gender gap in research output increases substantially until around 8 years after application where it levels off at a cumulative citation gender gap of around 60 percent and a cumulative impact factor gap of 45 percent. It also shows that the substantial difference in research output trajectories is followed by a large gender gap in promotions to full professorship which peaks at 10 percentage points, 14 years after the application. The bottom right panel of Figure A5, shows a persistent earnings gap between men and women, starting at approximately 40 percent and narrowing to 20 percent eighteen years after the grant application (when the average applicant is around age 50).

empirical question.

Our empirical findings so far line up closely with the mechanisms in the model. First, pre-WW reviewers were biased against women but exhibited more coherence in their assessment of women than of men. After the reform, evaluations rely relatively more on noisy individual signals, male premia vanish, and measured agreement within review committees declines for women. Second, we find that pre-WW male premia were smaller among reviewers with higher screening ability (greater expertise and greater exposure to female researchers).

This pattern—together with the asymmetric changes in agreement—is difficult to reconcile with purely accurate statistical discrimination, and instead points toward misperceived priors or taste-based discrimination. Third, the only clear reversal of bias (favoring women) occurs in the immediate round after WW, and only among female reviewers. Such a temporary reversal is consistent with belief-based partiality under misspecified inference, as in (Bohren, Imas, and Rosenberg, 2019) rather than with a stable, fully rational updating process.[46]

Most importantly, the simple framework captures the core trade-off in MFR's allocation problem. A gender-blind rule raises the share of women among funded applicants. If pre-reform priors were biased or reflected taste-based discrimination, the model predicts that this shift should also increase the long-run research output gap between grantees and non-grantees. By contrast, if reviewers' priors were accurate and informative, discarding group information could reduce the quality gap between funded and unfunded applicants.

Viewed through the lens of the model, our results—on the disappearance of male premia, the role of expertise and exposure, suggest that the reform is may have increased allocative efficiency. The next section directly tests the impact of the WW revelation on research output.

5.2 Empirical Strategy

To assess whether the elimination of the male bias caused higher or lower research output, we exploit within-MFR variation in the expected intensity of the impact of the WW revelation. Our within-MFR intensity measure is the review-committee specific bias ($\hat{\beta}_j^{Pre}$) estimated for the full pre-WW period (1990-1994). Our choice of intensity measure is motivated by the large differences in male premiums across review-committees (see Appendix Figure A7), that the

[46] Why do female but not male reviewers' overshoot in the immediate aftermath of WW? One explanation could be that when observing the public revelation of bias, in the subsequent round, reviewers becomes aware of the higher standard for women. In the next round female reviewers on average overshoots to compensate for this bias if they believe that male reviewers' male premiums were preference based (hence persistent) – while in fact the increased risk of sanctions eliminated it. Overshooting in this case indicates mis-specified beliefs about *how others* will respond to the revelation of bias. If female reviewers update their beliefs about male reviewers' response to the revelation, the female reviewer overshoot should decay over time. We see indications of such decay in Figure 3, panel C.

applicants are assigned to review committees, and that neither the original op-ed, the subsequent academic publication, nor the official internal evaluations by the MFR provide evidence on *which* review committees were most biased. We use the pre-reform committee bias in the following triple-difference model estimated by OLS:

$$y_{ijt_{+5}} = \alpha + \gamma\left(\hat{\beta}_j^{Pre} \times POST_t \times GRANTED_i\right) + \mathbf{X}_{ijt} + \theta_j + \lambda_t + \varepsilon_{ijt} \quad (1)$$

Where y_{ijt} is the long-run cumulative research output at the end of the grant period plus one year to allow for lags in the publication process (i.e. five years after application). $POST_t$ is an indicator equal to 1 if the applicant (*i*), applied after the 1994 application round, $GRANTED_i$ is equal to 1 if the individual were awarded the grant, $\boldsymbol{X}_{ijt}$ is a vector of pre-application individual controls (as well as the two-way interactions between $\hat{\beta}_j^{Pre}, POST_t, GRANTED_i$).γ is the parameter of interest and it captures the change in the long-run research output gap across grantees and non-grantees, before and after the revelation of bias, across applicants assigned to review committees with different levels of male premium in the pre-revelation period. ε_{ijt} is the error term.[47]

The specification above allows us to account for θ_j review-committee fixed effects, and thereby account for any research output differences across committees that are constant across time, due to e.g. field specific career trajectories.[48] Even more importantly the within-MFR variation allows us to also take year of application fixed effects (λ_t) into account, as well as any general pre- vs post differences across funded and non-funded applicants into account. General application-cohort differences could be of key importance in this setting. The changes in the academic landscape over the later 1990s and early 2000s, especially for female researchers (e.g. general improvements in female research opportunities unrelated to the changes at MFR) could otherwise lead to biased estimates of the impact of the disclosure of bias at the MFR.

The key identifying assumption is that the differences in research output of funded and non-funded applicants assessed by to committees with different pre-reform bias would have

[47] The focus of our analysis is the general call. However, for some years targeted calls (specific topics) are issued as well, and we don't observe whether an applicant (if applying to both specialized and general call and if granted) is receiving the grant from the main or specialized call. To deal with this, we use "any grant in year t" to measure application success, and include an indicator variable for those applying both to main and specialized call (16% of applicants).

[48] A complication for our long-run analysis is that from 1998 round the number of committees increase from 11 to 14 and in some cases change names. To be able to assign pre-reform bias to these later cohorts and thereby retain a balanced panel of committees, we created 10 time-consistent review committee groups using information on the overlap of review members across the years. To account for these changes, we include (time-consistent) review committee group fixed effects in the full pre-reform bias estimation. We average the pre-reform bias for the Medicine and Surgery committees to assign a pre-reform bias estimate to the three Clinical science committees replacing the Medicine and Surgery committees in 1998.

evolved similarly in the absence of the WW-revelation. The parallel trend assumption would be violated if (1) the applicants assessed by committees with different levels of pre-disclosure bias have differential publication trajectories already before WW, or (2) if the committee specific bias is correlated with changes in applicant or reviewer quality – i.e. the sorting of applicants across review committees' changes, based on the level of pre-reform bias, or if the quality of the reviewers changes systematically with the level of bias.

The parallel trends assumption cannot be tested *directly*. However, in support of it, in addition to the absence of information on which review committee were most biased, recall that applicants are assigned to review committees based on the topic of the research proposal. In other words, sorting into specific review committees based on private information on pre-reform biases is likely difficult. We also test this notion below, by assessing if female applicants are more or less likely to be assigned to more or less biased committees after the revelation than before. Also recall that applicant quality did not change notably (Figure 2D) and that retained and exiting reviewers had similar biases in the pre-period (Table 3, col 2 and 3), indicating that changes in the review committees were not based on the pre-WW committee specific bias.[49]

Moreover, we can assess the plausibility of the parallel trends assumption *indirectly* by testing for systematic changes in research output across committees *before* WW. Specifically, we test whether the research output gap of grantees vs. non-grantees changed in the two years before WW is also related to the pre-WW bias. In these specifications we amend specification (1) with a $\text{PlaceboWW}_{t=93.94}$ indicator equal to 1 for those applying just before WW, and interact that with the intensity measure and Granted indicators, and include the corresponding two-way interaction terms as in the baseline model and estimate:

$$y_{ijt_{+5}} = \alpha + \gamma\left(\hat{\beta}_j^{\text{Pre}} \times \text{GRANTED}_i \times \text{POSTWW}_t\right) + \\ + \rho\left(\hat{\beta}_j^{\text{Pre}} \times \text{GRANTED}_i \times \text{PlaceboWW}_t\right) + \mathbf{X}_{ijt} + \theta_j + \lambda_t + \varepsilon_{ijt} \quad (2)$$

If the estimate $\hat{\rho} \neq 0$, this would indicate that any changes detected among the true post-WW applicants could be driven by unobserved trends in research output across review committees with different pre-reform bias. A difference between the eq. (2) specification and the standard event-type model is our need to estimate the pre-reform bias intensity measure. To retain precision in the intensity measure estimate, but to still avoid the endogeneity of competence assessments in these "placebo" specifications we reestimate the pre-WW $\hat{\beta}_j^{Pre}$ after excluding the placebo cohorts (1993-94) from the review-committee specific bias estimation (i.e. we use

[49] Figure 2C showed that there was no overall change in reviewer quality around the reform.

$\hat{\beta}^{Pre}_{j,t\in 90-92}$). In other words, the eq. (2) estimates effectively compares the changes in research output gap across funded and non-funded applicants in the placebo and the true-post period in comparison with the corresponding research output gap for applicants in 1990-1992 application rounds across committees with differing male bias in the start of the sample period.

Finally, we also expand the basic event-study structure in Eq. (2) in a simple way by splitting the true post-WW period into to two parts, i.e. we allow the post-WW effect to vary with distance to the disclosure of bias in the post-period. This final specification allows us to assess whether any changes in the research output gap across granted and non-granted are temporary and fade-out, or whether they can be considered more permanent.[50]

In our main specification, we use a summary index of the cumulative research output as our key dependent variable. Specifically, we create the index using the first principal component of the (log of) cumulative number of citations, publications, and impact factor at the end of the grant period (t+5) and observation period (t+11). In Appendix Table A2, we also present estimates for the individual subcomponents of the long-run research output index.

5.3. The Impact on Allocative Efficiency

Table 4 provides the OLS results from four different versions of equation (1). As a prelude to our main findings, columns (1) show that females are not more or less likely to be assessed by more or less biased committees before than after WW. This finding confirms our prior that applicants are not aware which committees are most biased and/or are unable to self-select across committees based on private information about committee bias. The column (2) estimates show that the MFR did not allocate *additional or fewer* resources specifically to the most biased committees, but instead that funding were reallocated to female applicants (col. 3) in the post-period as predicted. This conclusion does not change even after including review committee specific trends (col. 4). The reallocation effect is sizable: the average gender bias in the pre-WW period was around .5 of a standard deviation, suggesting that the average probability of females receiving the grant increased by 7 percentage points (.5*.14) relative to the change in an on average unbiased committee.

Columns (5) to (7) present our main results on research output. First, in Column (5) we establish that there is no impact on research activity at the extensive margin: the pre-reform bias is not differentially linked to changes in the probability of having *any* new publications recorded five years post-application (i.e. at the end of the grant period), nor at the end of the observation window (t+11) (not reported). In other words, funding primarily allows for more research time rather than a complete exit from academic publication process.

[50] The other differences compared to a standard event-study model, is that we always pool at least two application round cohorts rather than providing an estimate for each application round cohort to increase power.

The column (5) finding is also important since they allow us to directly interpret the main intensive margin effects (columns 6-7) without sample selection concerns. Columns (6) shows that at the end of the grant period the cumulative research output increases by approximately 23 percent of a standard deviation.

Columns (7) also assess the validity of the parallel trends assumption using the eq.(2) specification. While the point estimates for the true treated cohort are similar to the baseline findings, the placebo reform estimates are small and insignificant. These results suggest that our baseline findings are not simply driven by differential pre-reform trends across applicants assigned to committees with differing pre-WW biases.

Figure 4 provides event-study estimates (solid black diamonds) using the same specification as in Col. (7), but now also allow the impact of the disclosure of bias to vary over time post WW. The estimates show that: (i) there is no change in the research output gaps between funded and non-funded across more or less biased committees applying before WW. (ii) Directly after WW, when the male premiums collapsed, the aggregate research output gap between funded and non-funded applicants increase significantly more in committees with higher pre-WW male premiums, and (iii) that the higher gap fades somewhat towards the end of the sample period (cohorts 1997-99). While point-wise confidence bands are wide, jointly the post-WW period research output gap is significantly higher than before WW ($p<.05$). In other words – the allocative efficiency, if anything, increased after the revelation of bias.

Figure 4 also provides the same set of estimates from a series sensitivity checks: Excluding review-committee specific trends, including applicants who leave academia (no post application publication by t+5), allowing the effect of pre-application publication history to vary by application round, or balancing or changing the post-WW cohorts to short or long-run effects changes our main conclusion, etc. The revelation of biased significantly increased the research output gap in committees with the highest pre-WW male premiums in the post period.

We provide additional results and robustness checks in the Appendix. In summary, Appendix Table A2 shows that our main conclusions hold for the subcomponents of the research output index individually (i.e. the impact factor, number of publications, and citations, see cols. 5-10). Although precision is poor – we do not find any indication that research output decreased for any of these subcomponents of the index. Our analysis focus on the research output accumulated by the end of the granted period (t+5), which from the view-point of the research council should be most important. Yet, even 11 years out from application the point estimates remain above zero, although at that time they are no longer statistically significant.[51]

[51] To account for the uncertainty of the pre WW committee-specific bias estimates, we also report standard errors (in brackets) that take the sampling variability into account by bootstrapping the two step procedure jointly (1000 repetitions). While the adjusted errors are larger, the key conclusions remain.

Finally, despite that we use within MFR variation, and thus are able to control for general time trends, one may be concerned that the disclosure of bias at MFR could lead to reduction of bias in other related settings e.g. later tenure decisions, etc. However, importantly, recall that the findings in Table 4 and Figure 4 on top of using within-MFR variation builds on a comparison between individuals who experienced very similar (potential) changes in the research environment later in their (post-WW) careers. The main difference is that they applied just before and after WW. In other words, to the extent that there were major improvements in the research environment (international journals publication process, tenure decisions, etc. due to WW) specifically in the fields with highest bias, our findings of the changes in research output induced by the disclosure of bias at MFR is likely, if anything, *understated*, unless these general changes *only* benefitted those applying just after WW and then *only* the granted applicants.

Overall, the substantial improvement in the research output documented in Table 4 allows us to draw two important conclusions. First, disclosure of bias improved selection on true-research potential and thereby provided a more efficient allocation of the public funds channeled through the MFR. Second, interpreted together with the findings on the changes in reviewer assessments and through the lens of our model, the increase in research output gap is inconsistent with accurate beliefs-based motives. If so, we would expect the elimination of bias to have decreased the research output gap.

Part 3: The Impact on the Medical Research Community and the Direction of Research

The Impact on PhD Enrollment

One month *before* WW disclosed the bias – the government's budget bill (1994/95:100) proposed extra resources to increase the number of PhD students in academic faculties that had experienced low growth over the past ten years, citing labor market conditions (Proposition 1994/95:100). One month *after* WW disclosed the bias – the government stated that "after further consideration, a gender equality perspective should be applied in the allocation of these [new] resources".[52] Before WW, preparatory work (Ds 1994:30) focused on the lack of women in STEM fields, with no mention of medicine. In the final bill, published one month after WW's disclosure, the gender imbalance among senior medical faculty was specifically highlighted (Proposition 1994/95:164).[53] Finally, directly after WW on top of the primary objective, an

[52] Out of the extra 37,5 million, 30 million should be earmarked to increase the doctoral positions held by women, and that the same amounts could be expected in 1996 and 1997. The government, stated that the doctoral positions should primarily be allocated to research fields where the process from undergraduate degree to doctoral degree results in a significant underrepresentation of women.

[53] Moreover, in November, a government report (SOU 1995:110) called for less focus on STEM and more focus on women in higher positions in academia. The report repeatedly highlights the op-ed article by WW and call WW's findings *"[...] of immense importance in getting what was previously more or less regarded as women's whining*

additional goal was to specified: "to increase the number of *female* researchers" (emphasis in original) (MFR, 1995).

The timing of these changes suggests WW's revelation was a catalyst the resource allocation toward female biomedical PhD students. To examine the impact on female PhD enrollment, we link data from Statistics Sweden's graduation register to population registers. The final dataset includes information from 1971 to 2017 and contains the gender, the year and semester of PhD program enrollment, graduation year, level, program, and university for all PhD graduates. We compute the share of female PhD graduates based on the year and semester of program enrollment. Our analysis is restricted to the eleven largest universities in Sweden, as measured by the number of PhD graduates (in total 23,000 graduates over the 1990-2000 enrollment period).[54]

Figure 5A illustrates the evolution of the share of female PhD graduates in biomedicine, social sciences, and other disciplines throughout the 1990s. Across all fields, the proportion of female graduates increased steadily—rising from approximately 35 percent to nearly 60 percent in medicine, from 35 percent to 50 percent in the social sciences, and from 25 percent to 35 percent in other fields.[55] Notably, around the time of WW's publication, the rate of female enrollment in medicine begins to outpace that of the other fields. This divergence is especially pronounced when comparing biomedicine to the social sciences, which not only share a similar pre-trend but also started from a comparable level of female representation prior to the WW findings. Appendix Figure A5 illustrates the divergence, and the statistical difference, by relating the average changes in medicine relative other fields using the last pre-WW enrollment cohort as the reference point (DiD-estimate 7.5 percentage points, $p<.01$).

Figure 5B presents the change in the standard deviation of female graduate shares across universities in medicine relative to other fields, using the final pre-WW enrollment cohort as the reference point. Following the WW publication, the dispersion in female enrollment across universities begins to decline in medicine compared to other disciplines. This reduction in the dispersion suggests a more coordinated and system-wide response within the biomedical field aimed at increasing female participation in PhD programs.

---

accepted as a legitimate problem." (p.37, our translation) and that "*the major focus on MFR [after WWs findings] put a focus on a problem previously not recognized by the research community […]*." (p.53). The report resulted in two important initiatives the 1995 so-called Tham reform and the 1997 introduction of quantitative recruitment targets for female professors (Silander, 2023). In medicine the so-called designated Tham-professorships increased the number of professors at medical universities by on average one professor or 1.08% (range 0.7%-2%, Göteborg - Umeå) as of year 2000.

[54] Karolinska institute (KI) is the largest university among these with 3220 graduates, and the smallest is Lund technical university (LTH) with around 500 PhD graduates.

[55] During this time period, there were a number of initiatives taken to increase the female representation in academia in general. For example, between 1985 and 1995 around 950 "gender inequality projects" were carried out at the universities and colleges to document and increase female researchers. These were mostly in the form courses, seminars, lectures and meetings in local associations, networks and mentor programs and in policy statements in equality plans (jämställdhetsplaner) and "other documents". A special emphasis on improving the gender imbalance in technical and natural sciences is again noted – with limited impact (Ds 1994:130).

Together, Figures 5A and 5B illustrate broader changes in the biomedical sciences compared to other academic fields in the wake of the WW findings. In contrast, Figure 5C delves into developments within biomedicine, focusing on enrollment patterns across subfields. It plots the change in the share of female PhD enrollment between the pre- and post-WW periods against the estimated male competence assessment premium observed in each subfield prior to the disclosure.[56] The figure reveals that female enrollment increased most in subfields where the pre-WW bias was greatest. Given that these field-specific biases were not publicly known to prospective PhD students, this pattern suggests that the post-reform increases in female enrollment were likely driven by university initiatives aimed at increasing female enrollment.

The Impact on Direction of Research

The significant increase in the share of female PhD students in the wake of WW could result in a higher share of research focused on women's health. To investigate this question in our setting, we link the PhD graduation data to publication records in Author-ity/Medline up until 2018.[57] Using the PMIDs, we scraped PubMed for the MeSH terms for each publication and created indicators for whether the article contained the MeSH term "female" or "male". The MeSH term "Female" is defined as "female organs, diseases, physiologic processes, genetics, etc." (Koning et al. 2021).

As a starting point, Figure 6 shows that female PhD students (during the observation period) are approximately 5 percentage points more likely to publish any articles focused on female health compared to peers from the same enrollment cohort during their first 7 years after PhD enrollment.[58] Comparing students enrolled in the same year, university and field (by including year × university × field fixed effects) does not change this conclusion. There is an even clearer pattern when focusing on first-author publications, indicating that the gender gap in research focus is especially large when the focal PhD student contributed most to the publication. These findings are consistent with previous findings that biomedical researchers

[56] The sub-field specific male premium is estimated using the entire pre-reform period (1990-1994) using the col. 3 Table 2 specification. Physiology and Chemistry have two review committees each, and to map the estimates to the enrollment rates we use the inverse of the standard errors of the bias estimate to calculate a weighted average bias.

[57] We link PhD students to publications using Author-ity 2018 by matching full names (and name changes), year PhD of enrollment, field, and university of graduation. We managed to merge 63% of the graduates to their publication records. 87 percent of the matches are unique matches using surname and initial, among Author-ity subjects with any affiliation registered to Swedish universities. The rest of the matches we keep are individuals which we get two matches by surname and initial in Authority. To find the correct match (or combination due to e.g. name changes) we use two-way combinations of research interests/university affiliation/timing of PhD (relative to time of first publication) criterions. Finally, we manually searched Author-ity for a few obvious matches not captured by these two-way criteria's (e.g. based on name-changes/middle names). We have excluded all graduates with three or more matches in Authority, and those with two-matches that do not comply with any of the two-way combinations of criterias for inclusion.

[58] The average time from enrollment to graduation was 6 years during the observation period.

are more likely to focus on topics more relevant for their own demographic group (Koning et al., 2021; Dossi, 2024).[59]

Whether increasing the share of female PhD students influenced the direction of research is, however, a different question.[60] It is not clear that a higher share of female graduate students will necessarily translate into increased attention to women's health. For instance, female students may displace male colleagues already working on such topics, potentially resulting in little net change. Alternatively, a growing female presence in doctoral programs might generate incentives to diversify research portfolios, leading some students to pursue other research areas.[61] Truffa and Wong (2025) tackles this causal question by investigates how undergraduate gender diversity at US universities that transitioned from all-male to coeducation in the 1960-1980s influenced research direction of the universities over the next 6 years.

We complement and extends Truffa and Wong (2025) in several ways: first by extending the follow up period from 6 to 17 years and by investigating the effects among PhD students rather than at the university level as a whole. Second, and more importantly, our findings complement their findings by focusing on female (PhD) enrollment changes at the intensive margin rather than the extensive margin. This is an important difference for policy purposes since most universities today are already co-educated. In our setting gender balance was already achieved at the undergraduate level, and close to being achieved at the graduate level before WW.

To investigate the impact of the increasing share of female PhD students on the direction of research, we use an instrumental variables approach based on the sharp increase in female enrollment following WW. Specifically, we use the share of female PhD enrollees by university and field in the five years preceding the WW publication as an instrument for the change in female enrollment after WW. Prior to WW, there was substantial variation in female enrollment rates across university-field combinations. The logic behind the instrument is straightforward. Departments with already high shares of female PhD students before WW needed to make smaller changes to achieve gender balance in PhD programs. In contrast, departments with initially low female representation had to make larger changes.[62] Figure 5D plots the relationship between the pre-WW female enrollment rates (x-axis) and the increase in female enrollment after WW. As expected, there is a convergence in the share of female graduate students across medical research departments post-WW.

[59] See Koning et al. (2021) for male/female teams, and Dossi (2024) for black/white teams.
[60] Shifts in research direction could occur directly, due to gender based differences in research interests, or indirectly, via peer effects within research environments.
[61] Similar to Goldin's pollution theory of discrimination (Goldin, 2014), it is also possible that an increased share of women working on women's health has negative impact on the prestige for men who work on the same topic.
[62] See for example Ahern and Dittmar (2012) for a similar approach.

Table 5 provides our main findings on the resulting changes in research direction. The basis for the analysis is a panel dataset at the university-field-period of enrollment level.[63] Columns (1) and (2) provide first-difference estimates for the change in female PhD enrollment and the change in the share of female health-focused research among PhD students. For the years leading up to WW (1985-89 vs 1990-94) (column 1), there is an insignificant and small positive correlation between changes in female enrollment and female health focused research.

The first-difference result (column 2) for the pre- vs post-WW period is strikingly different. Changes in female enrollment shares now have a large and significant correlation with changes in research direction, with a 10-percentage point increase in female enrollment share associated with a 2.2 percentage point increase in the share of research published during the first seven years after enrollment (a 17 percent increase). The differences in estimates across the first two columns highlight the importance of using the policy induced changes in female enrollment rates around WW to avoid confounding by unobserved trends or shocks correlated with both female enrollment and research direction. In other words, comparing col. 1 and 2, suggests that the findings in col. 2 is not caused by differential trends across university programs with differing initial shares of female students, which is important to keep in mind when interpreting our results below.

Column (3) provides the first stage of our 2SLS estimation. It shows – similar to the pattern in Figure 5D – that departments with lower pre-WW shares experienced greater increases in the share of female PhD students, conditional on department and enrollment fixed effects and university-specific trends. Column (4) shows the reduced form – indicating a negative relationship between the instrument and the share of research with female health focus.

Columns (5) and (6) provide our main result. In Column (5) we show that a 10 percentage point increase in the female enrollment share increases the share of female health-focused publications by 3.5 percentage points (0.345*0.1) within the first seven years after PhD enrollment. Relative to the pre-WW mean (13 percent) this corresponds to a 26 percent increase. In Column (6) we show that the effect remains positive when we extend the follow-up period to 17 years, suggesting that the increase in female PhD enrollment impacts the long-run direction of research.

Finally, Columns (7) and (8) report the corresponding estimates for the share of research with a male health-focus: finding no effects. Taken together, these results suggest that the increasing share of female PhD students resulted in an increasing attention to women's health, without reducing attention to men's health.

[63] Odontology moved from Lund to Malmö during the observation period and we recode all Odontology students to Malmö university for consistency.

Despite these key differences in method, setting, time-period, and female enrollment margin, ours and Truffa and Wongs (2025) estimates are of comparable magnitudes. Truffa and Wong finds that transitioning from male-only to coeducation increases the undergraduate female graduation share by 28 percentage points, which led to a 44 percent increase in research focused on female subjects and gender differences – i.e. a 15 percent change for a 10 percentage point increase in female graduation share. An estimate just shy of the range of our estimates (17-26 percent).[64]

6 Conclusions

Our paper shows that the public revelation of gender bias in competence assessments of biomedical researchers in Sweden led to meaningful changes across several important dimensions. The publication of evidence revealing substantial bias in the evaluation of applicant quality triggered a series of reforms within the Medical Research Council: the share of female reviewers increased significantly, the competence assessments became, on average, independent of applicant gender, and more merit-based. Prior to the revelation, both the proportion of female reviewers and the scientific competence of the committees were negatively correlated with the degree of male bias. After the disclosure, individual reviewers adapted their decision-making processes, and on average, gender bias in assessments disappeared.

These institutional changes translated into a more efficient allocation of public research funds—funds that were explicitly intended to "support basic research of high international quality." Notably, the post-application publication trajectories of the grantees improved, especially for applicants evaluated by the most biased committees prior to the disclosure. This improvement in long-run research output is arguably our most important finding. Our results, when considered alongside evidence from both internal and publicly available documents, suggest that more rigorous internal discussions and clearer guidance regarding evaluation criteria played a key role in reducing male bias in the assessment process.

We also find evidence that the WW revelation prompted broader changes within the medical research community. Immediately after WW, political attention shifted more explicitly toward gender equality in PhD enrollment, and the Medical Research Council (MFR) introduced a new core objective: "to increase the number of female researchers". The timing of this policy shift closely coincides with a marked increase in the enrollment of female PhD students, diverging from the pre-WW trajectory. Compared to other fields, universities in the

[64] We also use a slightly different way of measuring female health focus in research. We use the MeSH terms directly, while Truffa and Wong designates "female focused" if "Female" is one of the top Medical Subject Headings (MeSH) of the publication. Note that the share of female related research using their measure is almost identical to the share of female focused research using our measure.

biomedical sciences began enrolling more female students after the revelation of gender bias. Moreover, the variation in female enrollment across universities within the biomedical field decreased compared to other disciplines, suggesting a uniform and coordinated response. These patterns point to deliberate institutional efforts to recruit more female PhDs.

Finally, the rising share of female graduate students was associated with a significant increase in research focused on women's health topics—implying that the effects of the WW revelation extended beyond academia, yielding broader societal benefits for women.

A few concluding specificities of our setting are worth highlighting. First, while biomedicine is a highly competitive field, the level of bias in competence assessments prior to the WW revelation was substantial. Second, gender parity in PhD enrollment had largely been achieved—even exceeding that in many fields today, such as Economics in the United States. Third, Sweden had a long-standing commitment to gender equality as a distinct policy objective, dating back to the 1970s. The revelation of bias occurred in the mid-1990s - a period during which gender equality was high on the policymakers' agenda. This fertile ground for change likely played a role in facilitating the sharp and rapid institutional responses that followed WW. On the other-hand the repeated referral to WWs findings in the Government report which resulted in the most radical measures to date (the one-time preferential hiring initiative, the so-called Tham-professorships, implemented in 1997) could indicate a potentially important cross-fertilization channel – with WW's explosive findings paving the way for radical reforms by putting *"a focus on a problem previously not recognized by the research community"(*SOU 1995:110, p.53*).*

In either case, our results demonstrate that the public revelation of bias can serve as a powerful catalyst for institutional change. In this case, the WW findings helped move the Swedish Medical Research Council toward a new equilibrium—marked by a dramatic reduction in male bias, a higher share of female grantees, and increased research output. In the field of biomedicine, the revelation enabled a broad push for change, reshaping the demographic composition of PhD cohorts and steering research agendas toward topics related to women's health.

Finally, the impact of WW's revelation extended beyond Sweden. As Abbott (1997) observed "[t]he response [to WW] across the world could be measured on the Richter scale […]" prompting research councils in other countries — such as the United Kingdom – to reexamine their own evaluation procedures.[65]

[65] In 2012, a major European Commission expert group report on the need to for structural change in research institutions promoting gender equality and excellence in research referred to the WW article as a landmark study (European Commission, 2012). WW's findings also extended beyond Europe, being cited and discussed in several reports by the U.S. National Academies, shaping US policy conversations on bias and fairness in scientific evaluation (c.f . NAS, 2007).

6 References

Abbott, Alison (1997), “Equality not taken for Granted”, *Nature*, 390, 204

Ahern, Kenneth. and Amy Dittmar (2012), “The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation”, *Quarterly Journal of Economics*, 127(1): 137-97.

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti. 2024. "Revealing Stereotypes: Evidence from Immigrants in Schools." *American Economic Review* 114 (7): 1916–48.

Altonji, Joseph G., and Rebecca M. Blank (1999). "Race and Gender in the Labor Market." In *Handbook of Labor Economics*, edited by O. Ashenfelter and D. Card, 3143-3259. Vol. 3C. Amsterdam: Elsevier Science B.V., 1999.

Azmat, Ghazala and Rosa Ferrer (2017) “Gender Gaps in Performance: Evidence from Young Lawyers”, *Journal of Political Economy,* 125(5): 1306-55.

Bagues, M., Sylos-Labini, M., & Zinovyeva, N. (2017). “Does the gender composition of scientific committees matter?”. *American Economic Review*, 107(4), 1207-1238.

Belloni, A., V. Chernozhukov, and C. B. Hansen (2014), “Inference on treatment effects after selection among high-dimensional controls”, *Review of Economic Studies* 81: 608–650.

Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American economic review*, 109(10), 3395-3436.

Boring A., and A. Phillipe (2021), “Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching”: *Journal of Public Economics*, Volume 193,

Card David, Stefano DellaVigna, Patricia Funk, Nagore Iriberri (2020), “Are Referees and Editors in Economics Gender Neutral?”, *The Quarterly Journal of Economics*, Volume 135, Issue 1, February 2020, Pages 269–327,

Carson, Richard T., Joshua Graff Zivin, and Jeffrey G. Schrader. (2024) “Choose Your Moments: NIH Peer Review and Scientific Risk Taking”, NBER working paper No. 31409.

Chen, Jiafeng, and Jonathan Roth (2024). "Logs with zeros? Some problems and solutions." *The Quarterly Journal of Economics* 139.2 (2024): 891-936.

Coate, Stephen, and Glenn C. Loury (1993) "Will Affirmative-Action Policies Eliminate Negative Stereotypes?", *American Economic Review*, 83, no. 5: 1220-1240.

Cornell, Bradford, and Ivo Welch (1996). "Culture, Information, and Screening Discrimination, *Journal of Political Economy*, 104, no. 3 (1996): 542-571.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach”. *Econometrica*, 64(5), 1001–1044.

Dossi, Gaia (2024), “Race and Science”, *mimeo*, London School of Economics.

Dupas, Pascaline, Alicia Sasser Modestino, Muriel Niederle, and Justin Wolfers (2021). "Gender and the dynamics of economics seminars. No. w28494. *National Bureau of Economic Research*, 2021

Ds 1994:30 "Kartläggning och utvärdering av jämställdhetsprojekt vid universitet och högskolor".

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). "Unconditional quantile regressions". *Econometrica*, 77(3), 953-973.

Freeman R, Weinstein E, Marincola E, Rosenbaum J, Solomon F. Careers (2001) "Competition and careers in biosciences". *Science*. Dec 14;294(5550)

Ginther, D. K., & Kahn, S. (2009). "Does science promote women? Evidence from academia 1973-2001". In *Science and engineering careers in the United States: An analysis of markets and employment* (pp. 163-194). University of Chicago Press

Goldin, Claudia (2014) "A pollution theory of discrimination: male and female differences in occupations and earnings." *Human capital in history: The American record*. University of Chicago Press. 313-348.

Gwet, K. L. 2014, *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. 4th ed. Gaithersburg, MD: Advanced Analytics.

Hargens, Lowell L.; McCann, James C. and Reskin, Barbara F. (1978). "Productivity and reproductivity: fertility and professional achievement among research scientists." *Social Forces*. 57:154-165

Klein D, (2018), "Implementing a general framework for assessing interrater agreement in Stata", *The Stata Journal* 18, Number 4, pp. 871–901.

Koffi, Marlène (2021), "Gendered Citations at Top Economic Journals". 2021*, AEA Papers and Proceedings*, 111: 60–64.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson, (2021). "Who do we invent for? Patents by women focus more on women's health, but few women get to invent." *Science*, 372(6548): 1345–1348.

Lazear, Edwards, Sherwin Rosen (1981), "Rank and Order Tournament: An Optimal Labor Contract" in *Journal of Political Economy* vol. 89. 1981, p 841-864

Lazer, Edwards and Sherwin Rosen (1991), "Male-Female Wage Differentials in Job Ladders", *Journal of Labor Economics*, 8 (1).

Lee, D. S. (2009), "Training, wages, and sample selection: Estimating sharp bounds on treatment effects", *Review of Economic Studies* 76: 1071–1102.

Lundberg, Shelly. J., and Startz, Richard. (1983). "Private discrimination and social intervention in competitive labor market". *The American economic review*, 73(3), 340-347.

MFR (1994), "Project Catalogue 1993/94", ISBN-91-85546-90-9.

MFR (1995), “Project Catalogue, 1995/96”, ISBN-91-85547-03-4

Mishra S, Torvik VI. (2016) “Quantifying Conceptual Novelty in the Biomedical Literature”. *D-Lib magazine : the magazine of the Digital Library Forum*. 22(9-10).

NAS (2007), *Beyond bias and barriers fulfilling the potential of women in academic science and engineering*, Committee on Maximizing the Potential of Women in Academic Science and Engineering, National Academy of Sciences. ISBN-13: 978-0-309-10042-7

Price, Joseph, and Justin Wolfers (2007) "Racial Discrimination Among NBA Referees.", *Quarterly Journal of Economics*, 122, no. 4: 1859-1887.

Pope, Devin G., Joseph Price, Justin Wolfers (2018), “Awareness Reduces Racial Bias”, *Management Science*, 64(11), 4988–4995.

Prop. 1993/94: 147, “Jämställdhetspolitiken: Delad makt – delat ansvar”, Government Proposition, February 17, 1994.

Prop. 1994/95:100 ”Förslag till statsbudget för budgetåret 1995/96”, bilaga 9., Government Proposition, December 22, 1994.

Prop 1994/95:164 “Jämställdhet mellan kvinnor och män inom utbildningsområdet”, Government Proposition, February 16, 1995.

Quach, Simon and Zhengyi Yu (2025) “How Do Women Respond to Gender Biases in Recognition for Teamwork?”, mimeo.

Sandström, Ulf, and Martin Hällsten (2008) "Persistent nepotism in peer-review." *Scientometrics* 74: 175-189.

Sarsons, Heather (2017), "Recognition for Group Work: Gender Differences in Academia." *American Economic Review* 107 (5): 141–45.

Torvik, V.I., Weeber, M., Swanson, D.R. and Smalheiser, N.R., (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), pp.140-158.

Torvik, V.I. and Smalheiser, N.R., (2009). ”Author name disambiguation in MEDLINE”. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(3), pp.1-29.

Truffa, Francesca and Wong, Ashley (2024), Undergraduate Gender Diversity and the Direction of Scientific Research. Forthcoming in AER. (CESifo Working Paper No. 11294, Available at SSRN: https://ssrn.com/abstract=4991792 or http://dx.doi.org/10.2139/ssrn.4991792

Wennerås, Christine, and Agnes Wold. "Nepotism and Sexism in Peer Review.", *Nature*, 387 (1997): 341-343.

Wold, Agnes, and Cecilia Chrapkowska (2004), ”Förbjuden frukt på kunskapens träd”, Atlantis. [in Swedish]

Wu, Alice (2018), "Gendered Language on the Economics Job Market Rumors Forum." *AEA Papers and Proceedings* 108: 175–79.

Vetenskapsrådet (2015), “Reformer inom forskning och forskarutbildning 1990-2014”, Vetenskapsrådets rapporter, ISBN 978-91-7307-279-3.

News articles

DN (1995a), ”Därför forskar inte kvinnor”, *Dagens Nyheter*, 1995-01-22

DN (1995b) ”Svag grund för kritiken mot forskningsrådet”, DN, 1995-01-30

DN (1995c), ”Råd lovar bättring” – *Dagens Nyheter*, 1995-03-29

NYT (1997), ”Swedish Study finds sex bias in Getting Science Jobs”, The New York Times – May 22, 1997.

The Economist, ”Shameful”, May 22, 1997.

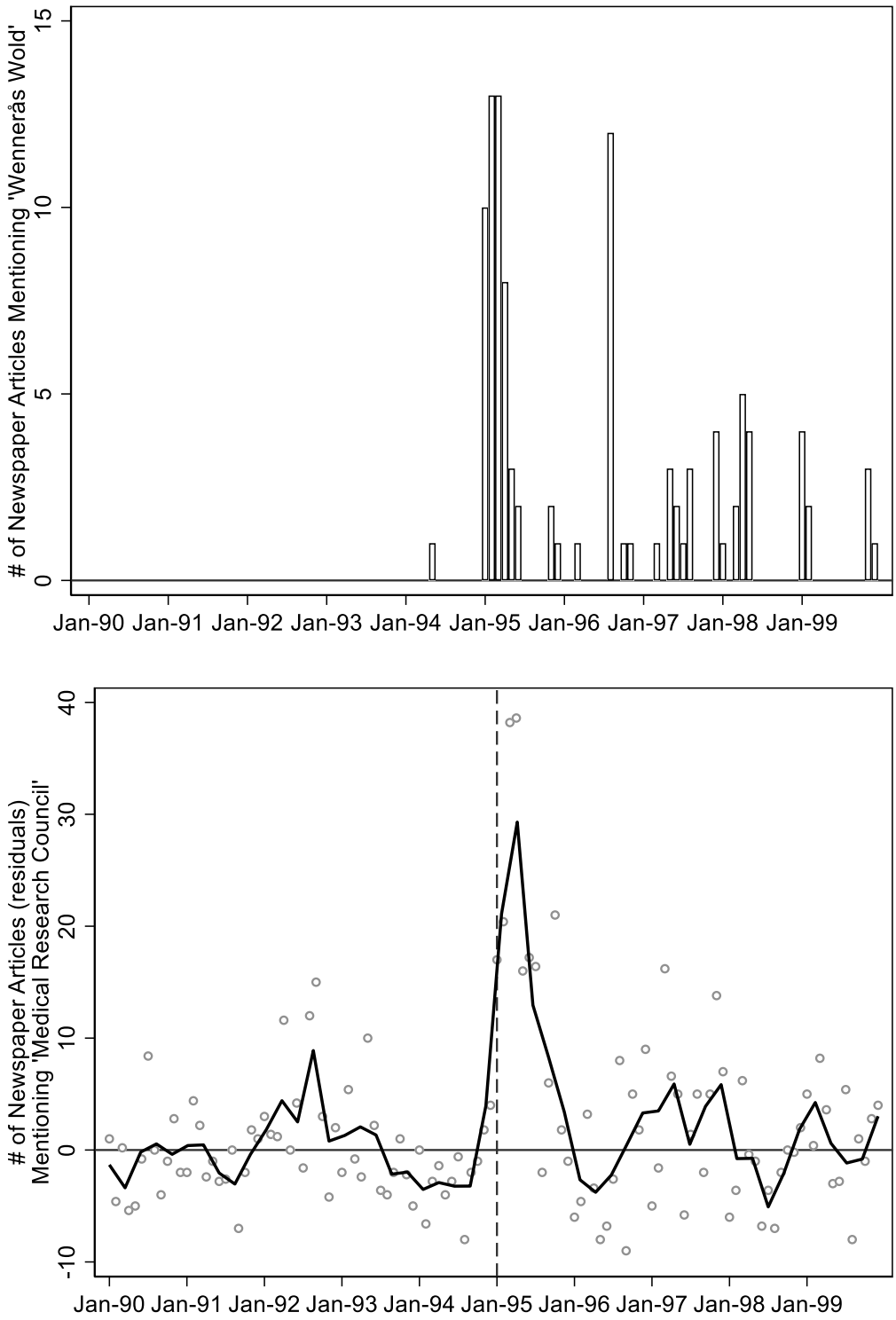


Figure 1: Media attention following Revelation of Bias at the Medical Research Council.
Note: the top figure plots the raw counts of monthly articles published in major Swedish newspapers mentioning "Wennerås Wold". The bottom figure plots the residual counts for article mentioning "Medical Research Council" after controlling for calendar month fixed effects using the pre 1995 data, and the local mean (solid black line) (bandwidth 1 month). (source Swedish Royal Library article search for "Wennerås Wold" and "Medicinska forskningsrådet", tidningar.kb.se accessed 2025-04-08)

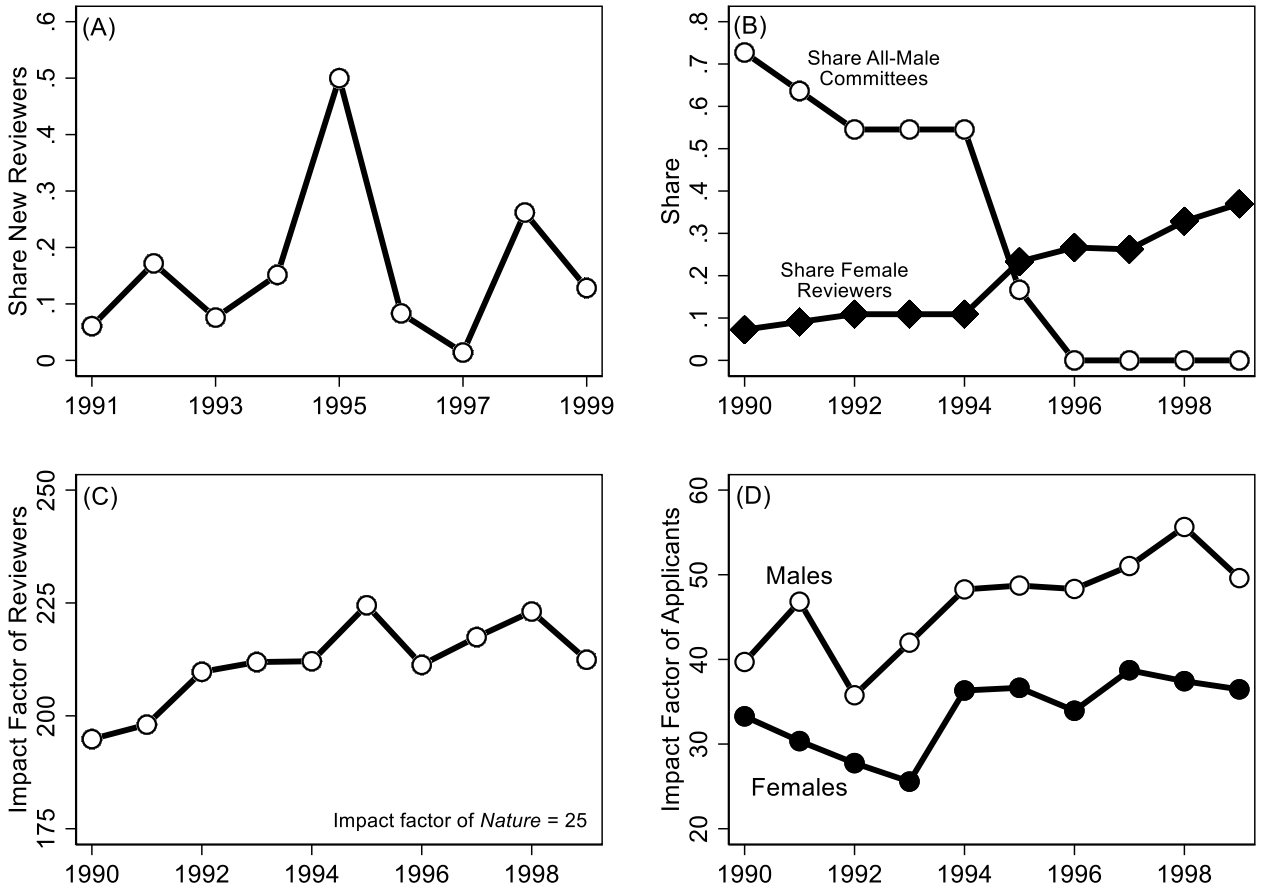


Figure 2: Trends in Reviewer and Applicant Characteristics by Application Round

Note: Figure A plots the share of new reviewers. Figure B Plots the gender composition of review the review committees. Figure C Plots the average cumulative impact factor of the reviewers during the past 10 years. Figure D plots the average cumulative impact factor over the past 10 years of the male and female applicants in the year of application.

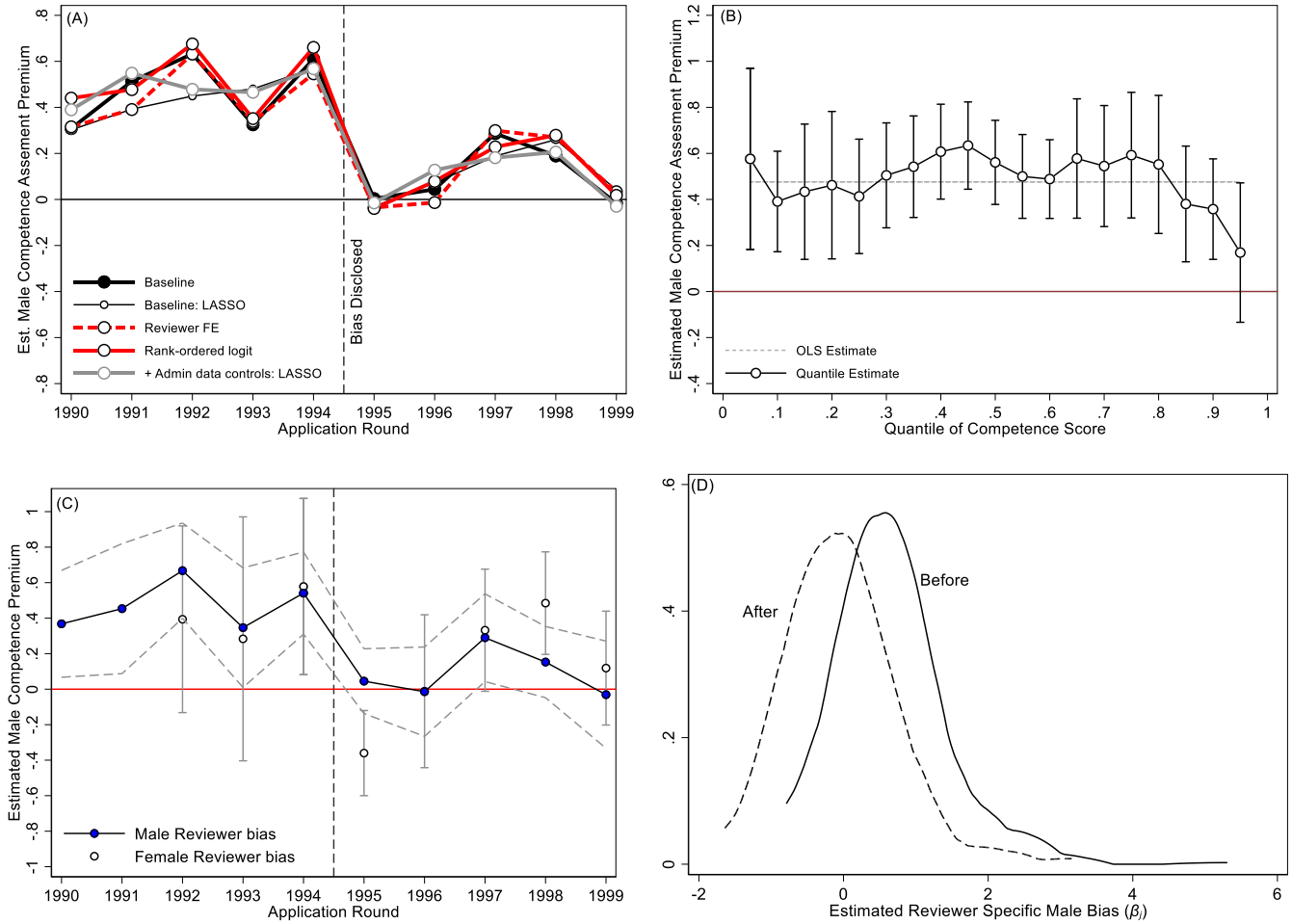


Figure 3 Male Competence Assessment Premiums around the Disclosure of Bias.

Note: Figure A plots the estimated male bias before and after the disclosure using the different specifications indicated by the figure. Figure B plots the estimated bias across the competence distribution, obtained using unconditional quantile regressions. Figure C shows the bias for male (black) and female (white) reviewers. Figure D shows the kernel densities of the reviewer bias estimates before and after the disclosure of bias.

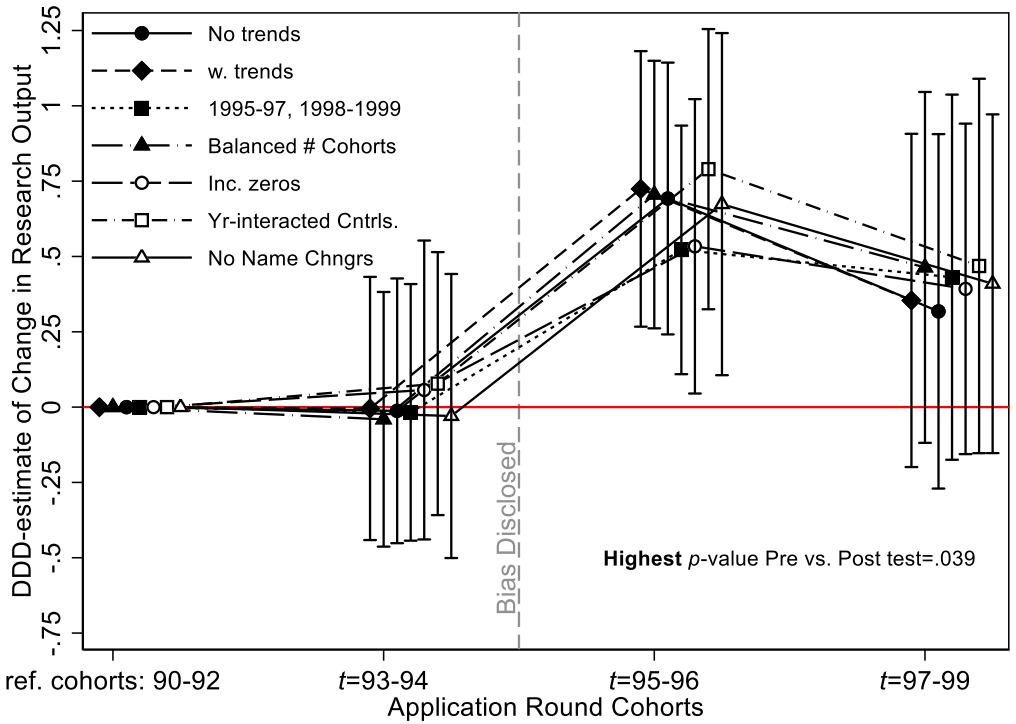


Figure 4: The Impact on Allocative Efficiency Five Years after Application: Event-study using Within-MRC variation in Exposure

Notes: Event study effect of the revelation of bias using within-research council variation in exposure. Regressions compares the changes in research output of grantees vs non-grantees among applicants before vs after the revelation of bias depending on the pre-reform review committee specific bias (estimated for the period 1990-1992 mean: .43 sd(.52). The baseline estimated model groups the pre- and post-revelation application cohorts (1990-92 (t-1) vs.: 1993-94 (t=0), 1995-96 (t+1), 1997-99 (t+2)). Regressions include baseline controls and application year and review group fixed effects (black circles), review group specific trends (diamonds), changes the t+1 cohorts to include 1997 (black squares), balances the post cohorts by dropping 1999 (triangles), includes individuals with zero post application publications after 5 years (hollow circles), a specification where all controls are interacted with year of application fixed effects (hollow squares), and when excluding name changers (hollow triangles). Error-bars provide 90% point-wise confidence intervals. Standard errors cluster at the review-committee group level. The Figure also reports the p-value of the joint pre-vs-post test of the specification which provided the *highest* p-value (hollow circles). The point estimates should be interpreted as the changes in the long-run research output difference across granted and non-granted per 1 std. dev. increase in pre-reform review committee male bias.

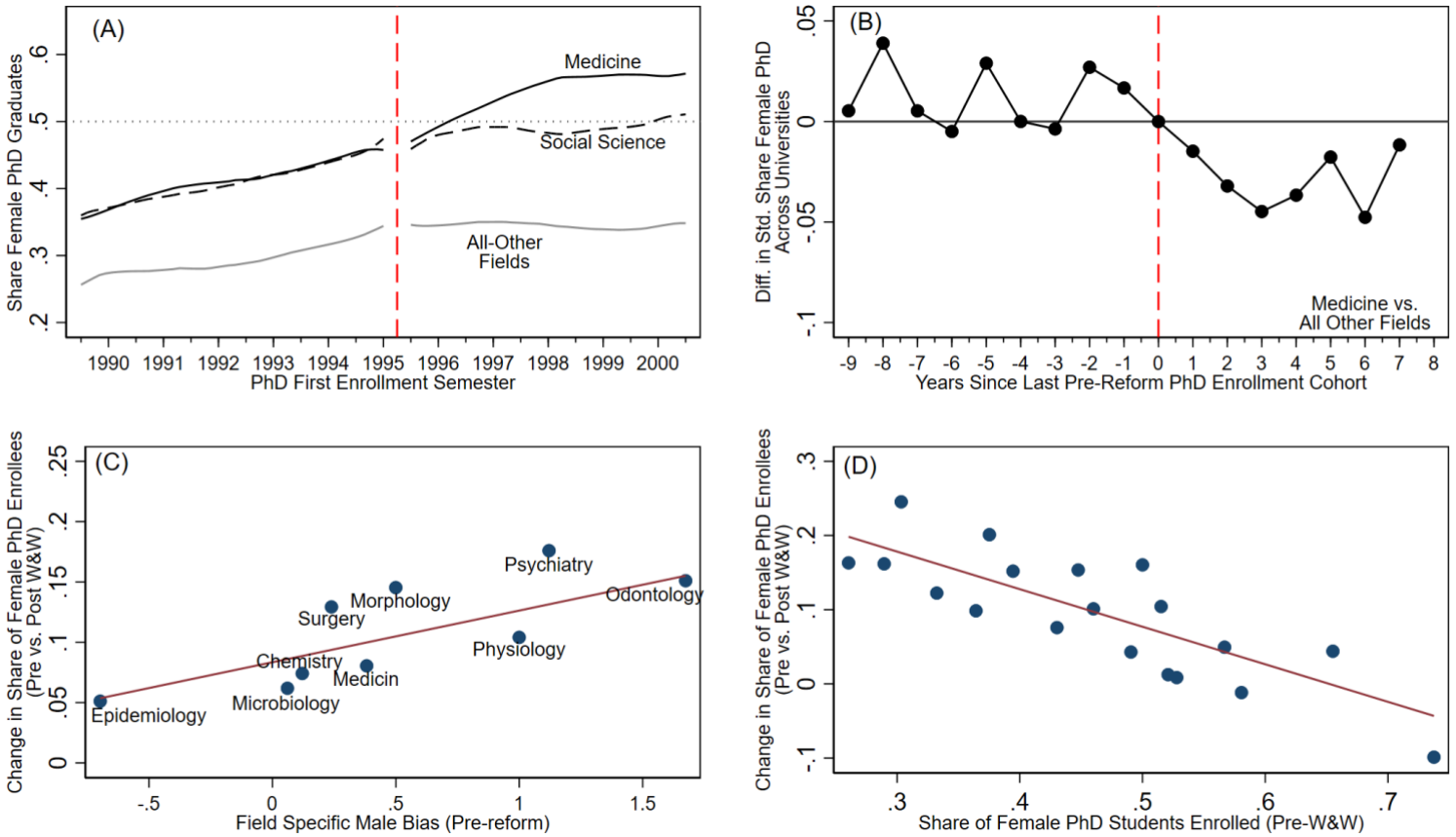


Figure 5: The Impact on the Enrollment of Female PhD-students

Note: Figure A plots the share of female PhD graduates by year of enrollment in PhD program in medicine, social science, and in other fields. The solid thick lines are local averages. Figure B plots the Standard deviation in the share of female PhD graduates by year of enrollment *across* universities in medicine vs all-other fields, relative to the last enrollment cohort just before WW. Figure C plots the Changes in Share of Female PhD Graduates (pre-post WW) across fields vs the estimated field specific male pre-reform bias. Figure D shows a bin scatter plot of the Changes in Share of Female PhD Graduates (pre-post WW) and the pre-reform share of Female PhD Graduates by university-field. The analysis sample is restricted to the eleven largest universities in terms of the number of PhD graduates (SU, LU, LTU, KI, GBG, UU, CTH, KTH, GU, UMU, LIU), in total 20,777 graduates with year of enrollment in the 1990-2000 period. All estimates are weighted by field-by-university PhD enrollment.

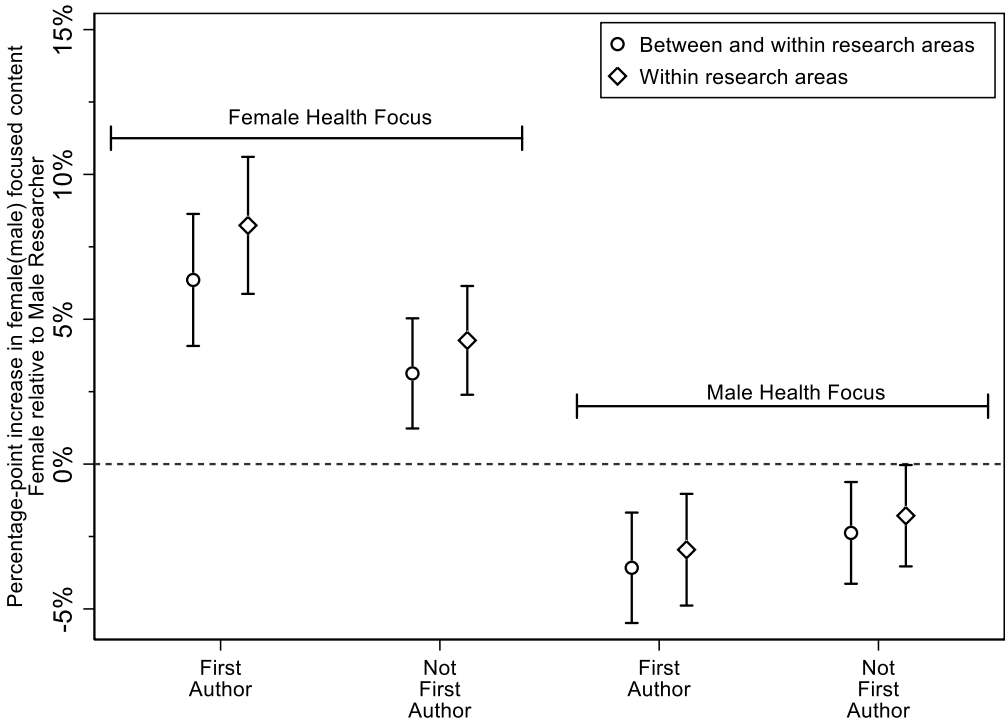


Figure 6: Gender Difference in the Direction of Research

Note The figure plots point estimates and 95% confidence intervals from a specification regressing the share of published articles during the first 7 years after enrolling in a PhD program (average duration of PhD completion is around 6 years) that are focused on female health (left part) and male health (right part) on a female author indicator variable. The graph shows points estimates and 95% confidence intervals and provides estimates from separate regressions when using the share of publications when the author is listed as the first author vs any other order in the author list. The first author in medicine is an important indicator for the relative contribution of each author. The hollow circle estimates are from a model only including a year of enrollment fixed effects and thus compares both within and across research fields/universities. The hollow diamond estimates controls for Year-of-enrollment x University x Field fixed Effects, effectively comparing the difference between male and female researchers research interest who started the PhD program in the same year in the same university and in the same field. The sample consists of researchers enrolling in the academic year 1990 to 2001.

Table 1 Descriptive Statistics Applicants Characteristics and Publication History			
	All	Males	Females
Share Males	.57		
Share Granted (Whole Period)	.15	.18	.12
Share Granted Year Before WW	.18	.26	.08
Share Granted Year After WW	.18	.23	.14
Academic Age (yrs since first publication)	5.99	6.09	5.86
	(3.2)	(3.3)	(3.0)
Age at Application*	36.1	36.0	36.2
	(5.3)	(4.9)	(5.7)
Age at PhD*	34.0	33.8	34.2
	(5.2)	(4.8)	(5.7)
Share with kids at application*	.60	.62	.58
Age at birth of first kid*	31.6	32.0	30.8
	(5.2)	(5.4)	(4.6)
Share born in Sweden*	.69	.67	.73
Share connected w. review committee	.10	.11	.08
Share reapply. t+1, if not granted in t	.41	.39	.44
Ever granted (for first time applicants in the 1991-1995 rounds)	.27	.32	.19
Share last name changers	.08	.015	.13
Publication History at application [t-10 to t]:			
Number of Publications	12.9	14.7	10.4
	(10.1)	(11.3)	(7.7)
Impact Factor	41.8	47.8	33.8
	(36.2)	(40.7)	(27.2)
Number of First-author Publications	5.96	6.68	4.99
	(4.37)	(4.97)	(3.15)
First-Author Impact Factor	18.4	21.0	14.9
	(15.3)	(17.1)	(11.7)
Network Size*	11.6	13.2	9.5
	(9.1)	(10.12)	(6.9)
Share articles w. most freq. Last-author	.54	.53	.56
Share w. any US NIH-supported publ.	.30	.32	.27
Share w. any Pub. in a Top† Journal	.13	.15	.11
Total Citations at Application	93	108	75
	(159)	(181)	(121)
Total *Post*-Application Citations of	485	560	381
Pre-Application articles	(618)	(717)	(429)
Number of Applications	1,282	733	549

Note: * indicates that the statistics are calculated after successful matching to Statistic Sweden administrative data (>95% match rate). NIH funded is an indicator for if the article is indicated as having received support from the (US) National Institute of Health. † Top journal is here defined as having any article published in *Cell*, *Lancet*, *New England Journal of Medicine*, *Nature*, *Science*, *Nature Genetics*, *Nature Medicine* in the pre-application period. * The number of unique Co-authors.

Table 2. Male Competence Assessment Premiums Just Before vs Just After Disclosure of Bias

Dependent variable:	*Competence Score*	*Competence Score*	*Competence Score*	*Competence Score*
Panel A: Main Sample	Before Disclosure	Before Disclosure (Replicating W&W)	Before Disclosure	After Disclosure
	(1)	(2)	(3)	(4)
Male Premium	.809***	.650***	.610***	.003
	(.157)	(.138)	(.137)	(.143)
Total Impact Factor		.009***	.005*	.006*
		(.002)	(.003)	(.004)
First Author Impact			.017***	.022**
			(.005)	(.009)
Number of publications			.0010	.012
			(.011)	(.008)
Committee Connection		.441**	.465**	.293
		(.184)	(.182)	(.202)
R-squared	.40	.54	.59	.47
Observations	123	123	123	151
Review committee FE	Yes	Yes	Yes	Yes
Panel B: Excl. name changers				
Male Premium	.856****	.703***	.665***	-.009
	(.160)	(.140)	(.139)	(.148)
Total Impact Factor		.009***	.005*	.006
		(.002)	(.003)	(.004)
First Author Impact			.017***	.026***
			(.005)	(.009)
Numb. of publications			.002	.010
			(.011)	(.009)
Committee Connection			.412**	.399*
			(.187)	(.212)
R-squared	.41	.57	.60	.50
Observations	117	117	117	143

Note: The table shows cross-sectional estimate of the return to being male researcher on competence assessment just before the disclosure of bias (1994), and just after (1995), conditional on the cumulative impact factor of the applicants' publications, first-author impact factor, and the number of publications published between 10 years before to the year of application. Connected is an indicator for if at least one of the review board members ruled themselves out from evaluating the applicant due to conflict of interest (e.g. the applicant is a previous co-author). All specifications also includes fixed effects for the review committee the application was assigned to, and an indicator for whether the applicant submitted more than one grant application in the same year. The outcome variable is the standardized competence score (mean 0, sd. 1).

Table 3. Male Competence Premiums by Reviewer Group

Dependent variable:	*Competence Score*	*Competence Score*	*Competence Score*
	DiD Before vs After WW	Year Before WW	Year After WW
	(1)	(2)	(3)
Male Premium	.546***		
	(.131)		
Male Premium x Post	-.581***		
	(.156)		
Male Premium (Exiting Reviewers)		.573***	
		(.157)	
Male Premium (Retained Reviewers)		.521***	-.007
		(.167)	(.143)
Male Premium (Entering Reviewers)			-.059
			(.130)
R-squared	.35	.35	.37
Observations	1,427	669	758
Reviewer FE	Yes	Yes	Yes

Note: The table presents findings using reviewer-applicant level data for 1994-95 application rounds. The outcome variable is the standardized competence score. Observations are weighted by inverse of the number of reviews per applicant to mimic baseline specification. Column 1 use data from before and after WW and report the pre-WW male bias and the change in bias after WW, controlling for reviewer fixed effects, and a post dummy and the other controls (see col.3, Table 2). Column 2 reports the marginal male assessment premiums effects in 1994 for reviewers who stopped serving after WW and for the retained reviewers (stayed on at least until the next application round) separately. Column 3 reports the same effects for the retained reviewers (i.e. active before and after WW), and for the new reviewers. The standard errors in parenthesis are clustered at the review level and */**/*** corresponds to 10/5/1% significance levels.

Table 4: The Impact on Allocative Efficiency as Measured by Post-application Research Output

Dependent Variable:	p(Male)	p(Granted)			Any New Research Output At end of Grant Period (t+5) (Ext. Margin)	End of Grant Period Output Index (t+5) (Intensive Margin)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reform Intensity Measure:							
$\hat{\beta}_j^{PRE}$ *: Pre-WW Committee Specific Bias*							
$\hat{\beta}_j^{PRE} \times \text{PostWW}_{t\geq 95}$	.029	-.019	.138**	.138			
	(.052)	(.045)	(.056)	(.105)			
$\hat{\beta}_j^{PRE} \times \text{PostWW}_{t\geq 95} \times \text{Male}_i$			-.237***	-.234***			
			(.068)	(.070)			
$\hat{\beta}_j^{PRE} \times \text{PostWW}_{t\geq 95} \times \text{Granted}_{ijt}$					-.044	.447*	.524**
					(.027)	(.232)	(.225)
$\hat{\beta}_j^{PRE} \times \text{PlaceboWW}_{t=93,94} \times \text{Granted}_{ijt}$							-.004
							(.260)
Mean of Dependent var.	.62	.18	.18	.18	0.95	0	0
R-squared	.09	.13	.14	.14	.04		.31
Application Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Review Committee Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Review Committee Trends				Yes	Yes	Yes	Yes
Observations	1,366	1,366	1,366	1,366	1,366	1,309	1,309

Note: The Table provides DD (and DDD) estimates using pre-reform male bias as a measure of the expected intensity of the disclosure of bias across review-committees. $\hat{\beta}_j^{PRE}$ is the pre-reform field specific competence assessment premium estimated using the full set of baseline controls (Column 3, Table 2), PostWW is an indicator variable for the application rounds 1995-1999. Granted is an indicator for if the applicant was awarded any grant from MFR (main or specialized call). Controls same set of controls as col. 3 Table 2, and we also include an indicator for missing information on committee connection (1998/99), an indicator for multiple applications (i.e. to main and specialized call – since we do not observe (if granted) which committee the grant stems from) in the same application round - interacted with the male dummy. In addition to field, and application year fixed effects in column 3-4 we also control for the two-way interactions of intensity/post indicator with the male dummy. Columns (5-9) controls for the corresponding two-way interactions between POSTWW x Granted, POSTWW x $\hat{\beta}_j^{PRE}$, and Granted x $\hat{\beta}_j^{PRE}$. In Column 5-8 the $\hat{\beta}_j^{PRE}$ reflects the committee specific bias in the full pre-period (1990-94) (mean: .43, std.dev: 42), and in column (9) it is estimated for the reference cohorts' period (1990-92) (mean: .41, std.dev: .51). To account for changes in the number/names of the review committees (post 1998) we create time-consistent review committee groups and include fixed effects for those review groups. For consistency over time the pre-reform field specific bias is assigned applicants based on their review committee groups. Outcomes The research output index is standardized first principal component of the (log) cumulative citations, impact factor, and the number of publications published within 5 years (grant period+1 year). See Table Appendix A2 for the results estimated for each research output component individually and for further robustness checks of the baseline results. Inference Standard errors in parenthesis clustered at the review committee group in parenthesis (*/**/*** corresponds to 10/5/1% significance levels).

Table 5: Female Enrollment and Changes in PhD Students Research Direction during the first seven years after PhD enrollment

Outcome:	Changes in Share Research w. Female Health Focus		Share Female PhD Students		Share Research w. Female Health Focus		Share Research w. Male Health Focus	
Sample Restrictions:	**Enrolled 1985-94**	**Enrolled 1990-2001**				17-year follow up		17-year follow up
	Before WW	Around WW						
Specification:	OLS	OLS	First stage	Reduced Form	2SLS	2SLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Changes $\overline{\text{Female PhD Students}}_{ijt}$	.046	.221***						
	(.063)	(.079)						
Post-WW × $\overline{\text{PreFemPhDStud}}_{ijt}$			-.586***	-.203*				
			(.155)	(.106)				
$\overline{\text{Female PhD Students}}_{ijt}$					.345***	.220**	-.031	-.023
					(.123)	(.103)	(.175)	(.125)
K.-P. First-stage *F*-stat					14.27	14.27	14.27	14.27
Observations	52	52	155	155	155	155	155	155
Mean of dependent variable (PreWW)			.46	.13	.13	.13	.15	.15

Note: The unit of observation is at the university-field-period level, all regressions are all weighted by the number of enrolled students in each field and university. The full sample consists of PhD students enrolled in the academic years between 1985 and 2001. Column 1 provides the first difference estimates in the years before WW (1985-89 vs 1990-94). The WW sample is split into three distinct enrollment periods pre-WW (1990-1994), Phase-in period (1995-1996), and 1997-2001 Post-WW). Column 2 provides the first difference estimates for pre vs. post WW periods. Columns 3-8 provides level regressions, all including period and department (university x field) FE, (where field is defined by the two-digit field of study codes) and university specific trends. The main outcome variable is the share of research that is focused on female (male) health published during the first seven (seventeen) years after PhD enrollment. PostWW is an indicator equal to 1 if enrolled after WW, $\overline{\text{FemalePhD Students}}_{ijt}$ is the female share of graduate students enrolled in the university/field/period. $\overline{\text{PreFemPhDStud}}_{ijt}$ is the pre-WW Share of Female PhD Students. Columns (3-8) also include a phase-in period indicator (enrollment cohorts 1995-96, see figure) and its interaction with the pre-WW share female students. First-stage *F* stat is Kleibergen-Paap Wald statistic, standard errors in cols.3-8 clustered at the university-field level presented in parenthesis (heteroscedasticity robust in first-differences columns), 10/5/1 percent level significance reported */**/*** respectively.

APPENDICES

Appendix A: Additional Results
Appendix B: Additional Background Information

Table A1. Which reviewer/review board characteristics are **correlated** with the male competence premiums?

Dependent variable:	*Competence Score*	*Competence Score*	*Competence Score*	*Competence Score*	*Competence Score*
Sample period:	Pre-reform all years	Pre-reform all years	Pre-reform all years	Pre-reform all years	Pre-reform all years
	(1)	(2)	(3)	(4)	(5)
Male Applicant	.436***	.442***	.481***	.492***	.529***
	(.086)	(.091)	(.094)	(.094)	(.092)
Reviewer Characteristics					
Male x Reviewer Female			-.285	-.299	-.292
			(.190)	(.194)	(.180)
Male x Reviewer Scientific Competence			-.116**	-.121**	-.112**
			(.047)	(.047)	(.045)
Male x Reviewer Exposure to Female PhD Students				-.150**	-.126**
				(.061)	(.062)
Review Board Characteristics					
Male x Share Male Reviewers					.124*
					(.064)
Male x Committee Scientific Competence					-.217***
					(.069)
Adj R-squared	.12	.27	.28	.28	.29
Observations	2,875	2,875	2,875	2,875	2,875
Field FE x Year FE	Yes				
Reviewer x Year FE		Yes	Yes	Yes	Yes

Note: Standard errors two-way cluster at the reviewer level by year level and application level. All continuous variables are standardized (mean 0, Std.1) for comparability. Exposure to female PhDs is measured as the share of female PhD graduating between 1975 and 1990 in the field-by-university of employment of the reviewer. The review board characteristics are computed after excluding the focal reviewer. The specifications also included the baseline controls (see Table 3).

Table A2: The Impact on Long-run Research Output using within Council variation in Reform Intensity - Additional Findings and Robustness

Dependent Variable:	**Long run Research Output Index (Standardized)**		**Subcomponents of the Research Output Index:** Log Cumulative Publications		Log Cumulative Impact Factor		Log Cumulative Citations	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Timing of Outcome Measurement:	End of Grant Period (t+5) (baseline)	End of observation Period (t+11)	At (t+5)	At (t+11)	At (t+5)	At (t+11)	At (t+5)	At (t+11)
$\hat{\beta}_j^{Pre} \times \text{Post}_{t\geq 95} \times \text{Granted}_{ijt}$	.447*	.292	.331*	.356	.284	.254	.670*	.400
	(.232)	(.209)	(.165)	(.201)	(.178)	(.256)	(.364)	(.326)
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Review Committee Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Review Committee Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations								

Note: For convenience Column 1 reiterates the baseline findings in the main text. Column (2) use the output index at t+11 (end of observation period) as the outcomes. Columns 3 to 8 provides the results for the subcomponents of the Research output index. See notes to Table 4 for further details.

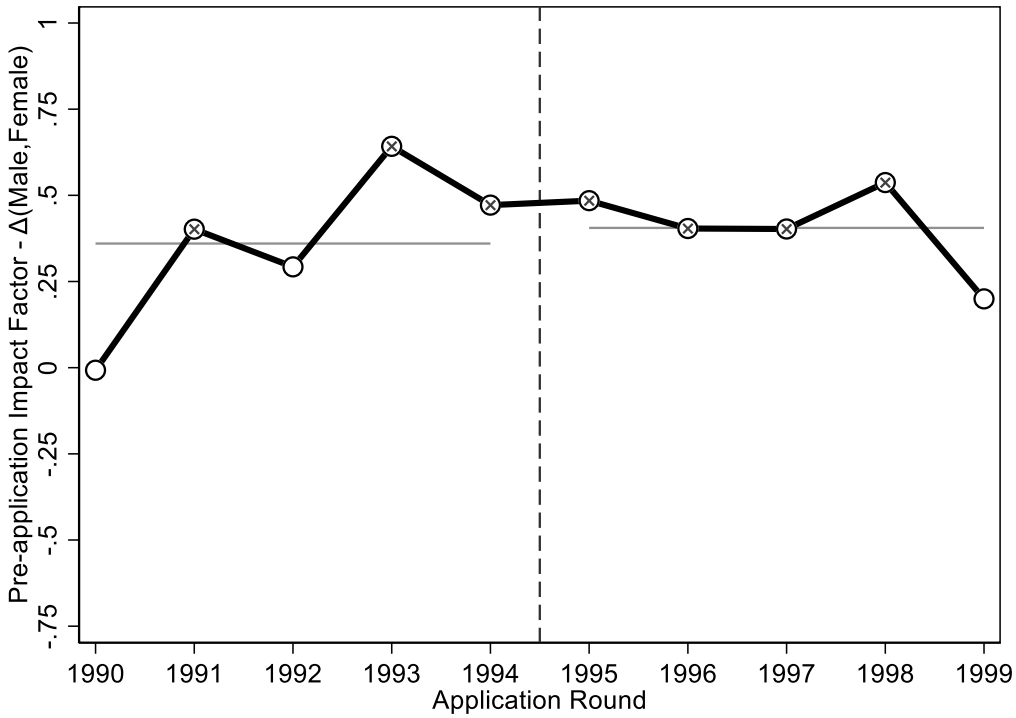


Figure A1: Absence of changes in gender difference in Applicant Competence around the WW revelation (See Figure 2D for the trends in levels for males and females)

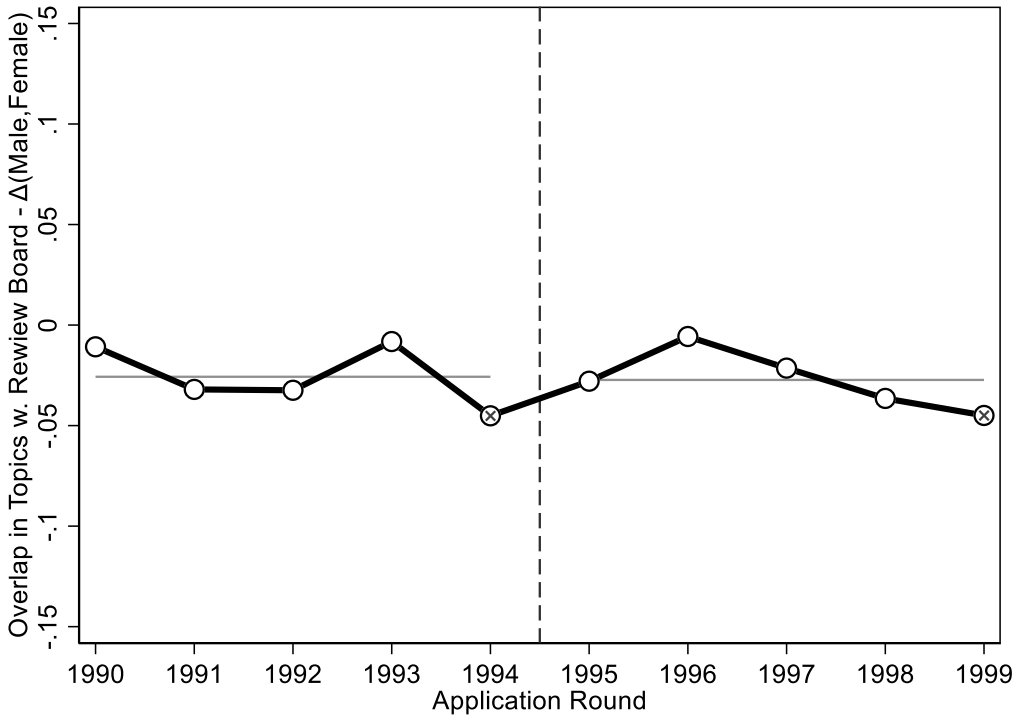


Figure A2: Gender difference in Overlap in Research Interests between applicants and committee reviewers around the reform

Note: Research interest overlap is based on the Medical Subject Headings of the publications of the applicants and reviewers during the past 10 years before the application year. It is computed by calculating the share of MeSH terms that overlaps between the applicant and each committee member, and then we take the average overlaps over the entire review board as the final outcome.

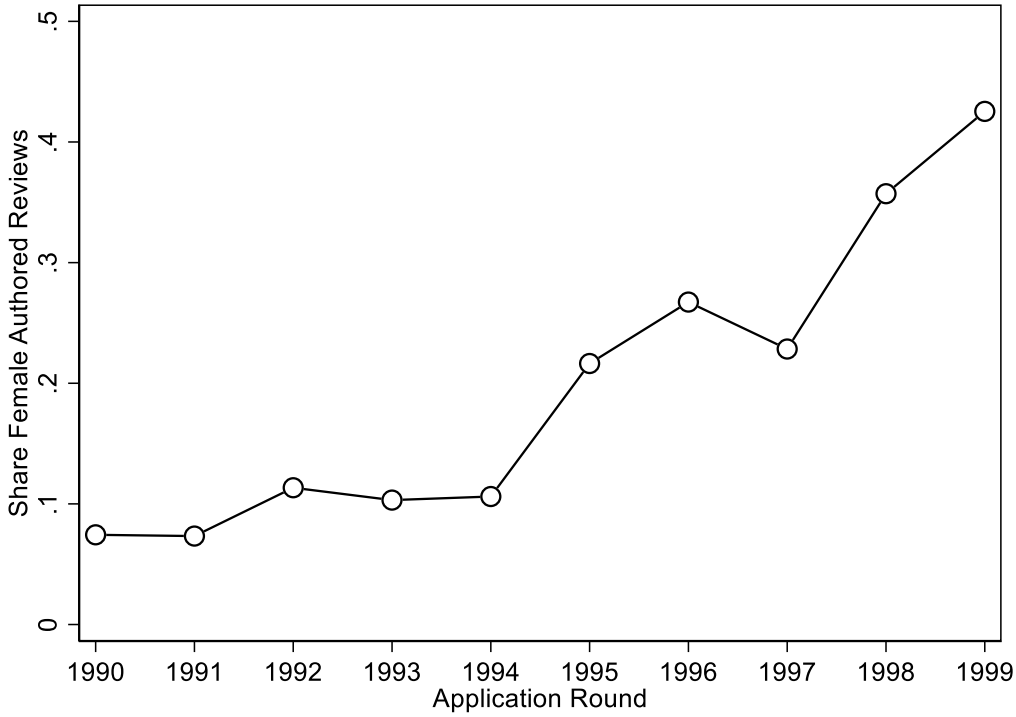


Figure A3: Share Female Authored Reviews

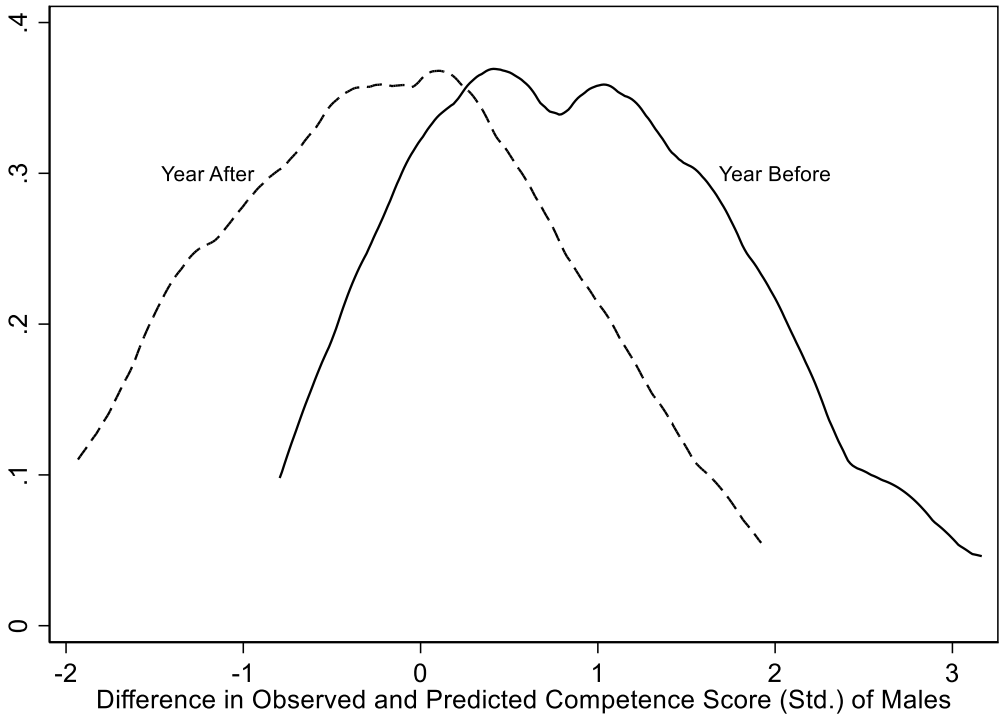


Figure A4: Difference in the Observed and Predicted Competence Score of Males
Note: the figure use data from the 1994 (just before) and 1995 (just after) application rounds and estimates the baseline model (Col. 3, table 2) for the female sample, forming predictions based on the female sample returns to observables (separate for each year) and then taking the differences between the observed and predicted competence scores for the males.

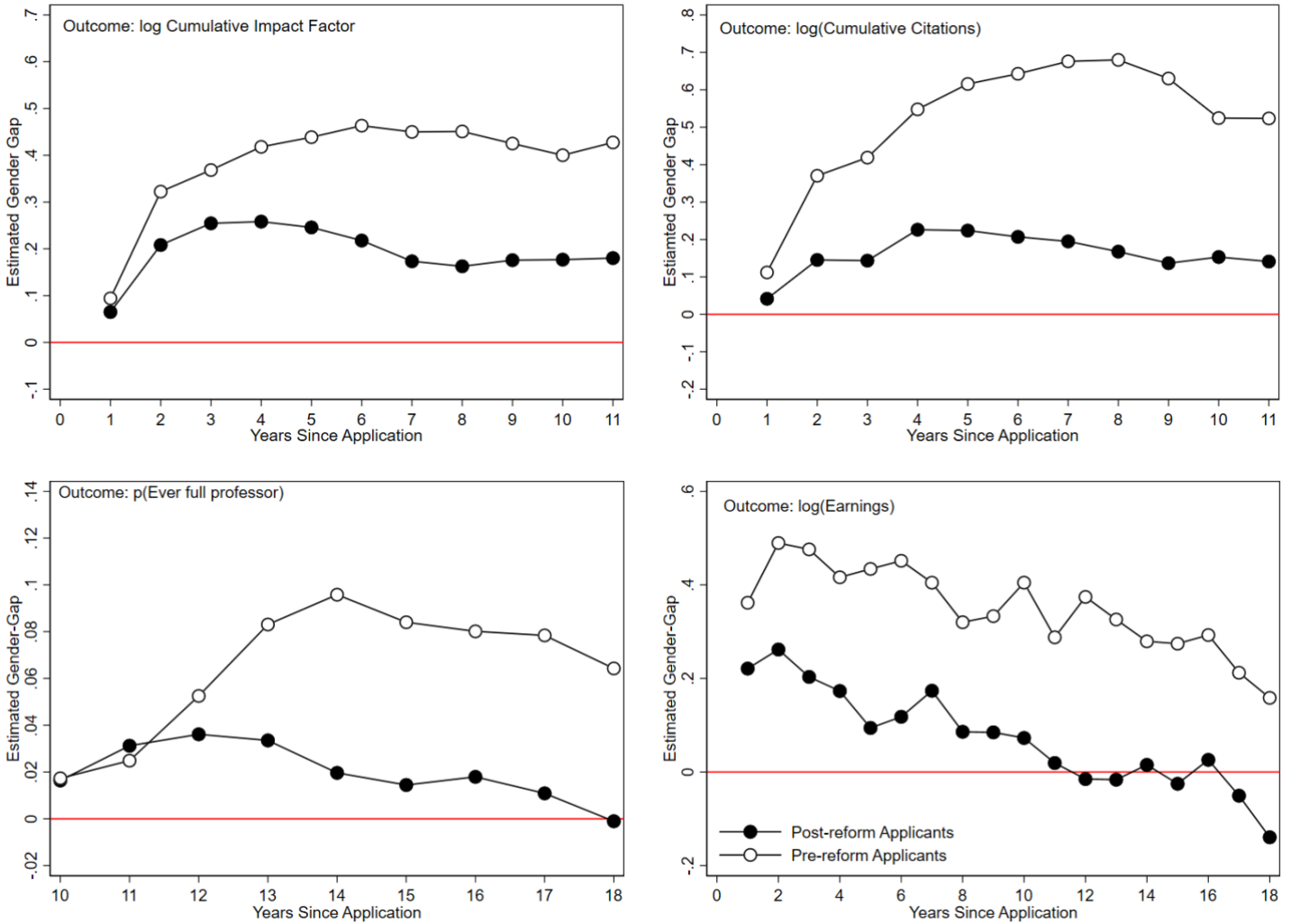
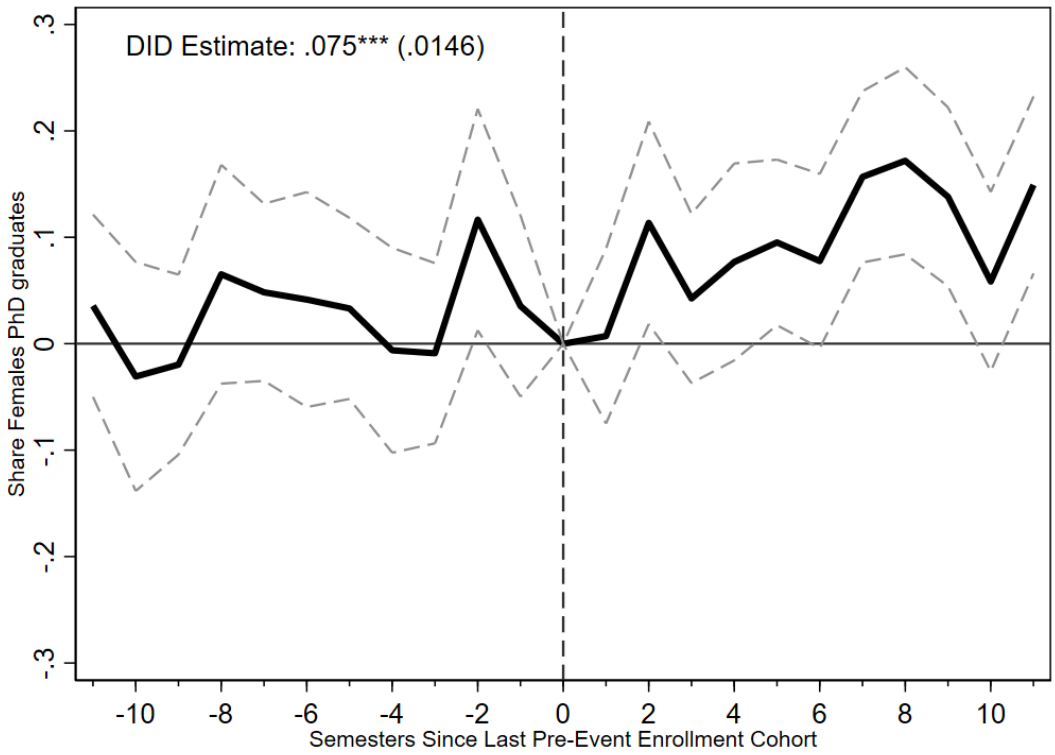
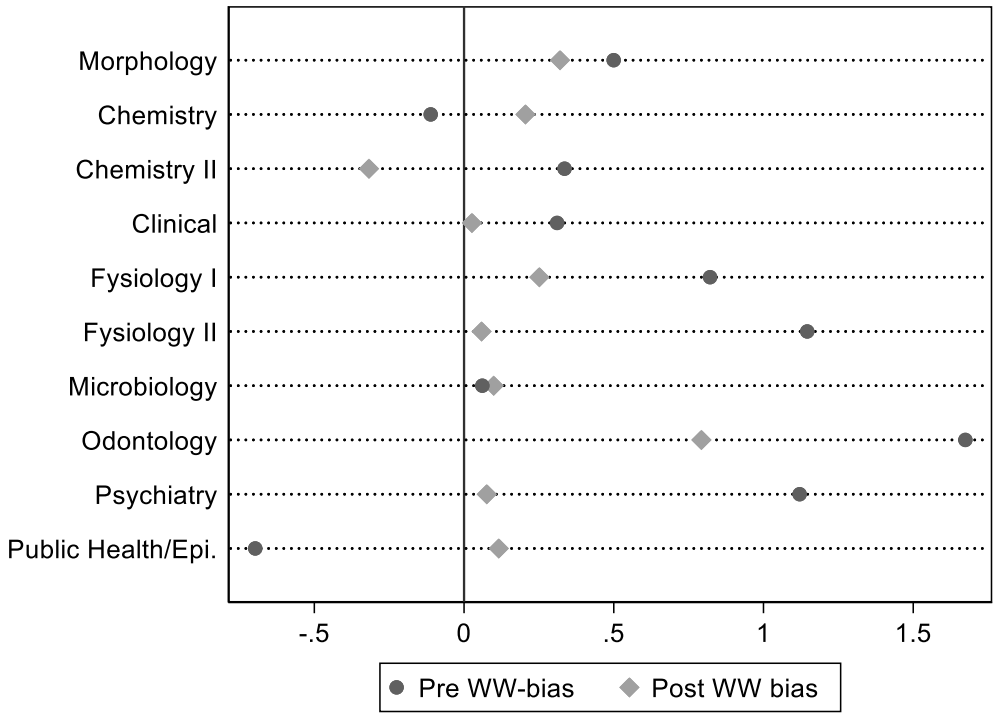


Figure A5 The Estimated Pre- and Post-Reform **Gender Gap** in Research Output and Careers (conditional on pre-application characteristics)



Appendix Figure A6:



Appendix Figure A7: Changes in bias By Review-committee Groups cells.

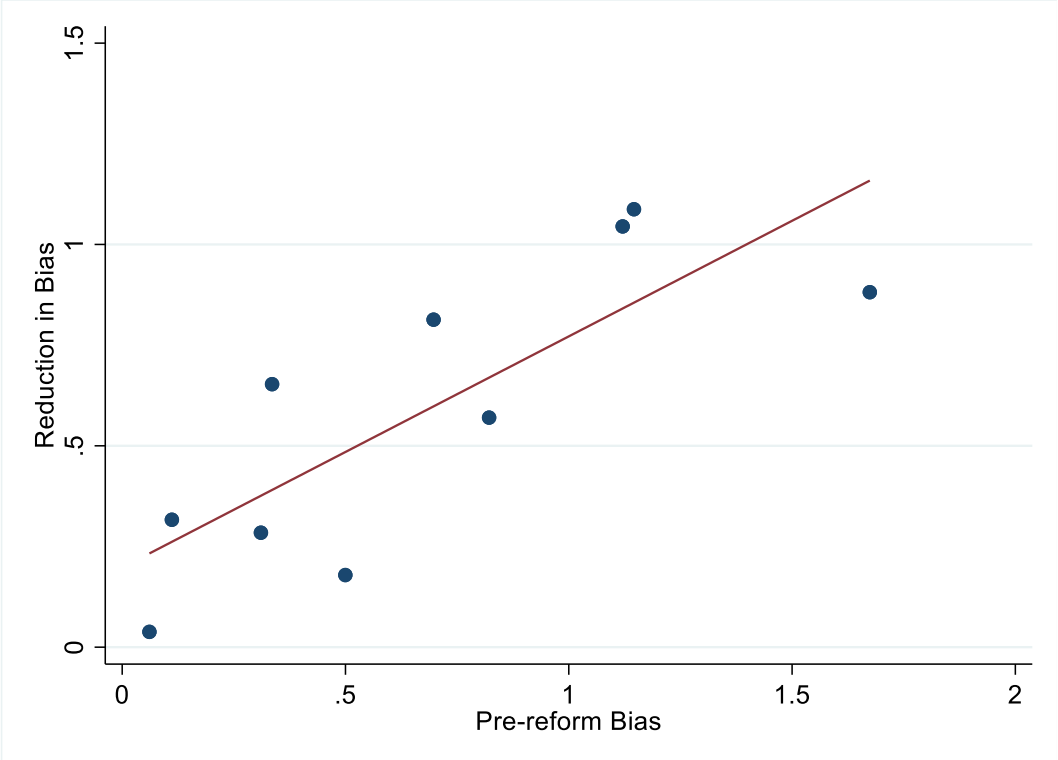


Table A3: Key results from Wennerås and Wold op-ed.

	Share granted						Number of Applicants			
		Women	**Men**			**Women**	**Men**			**Women**	**Men**
…by citation frequency in international journals (in 1994)	**0-4**	0%	17%	***… when applicants research program is assessed to be: (if stated on their review summary)***	**Less interesting**	0%	0%	***Applicants with lack of independence (if stated on their review summary)***	Total	16	17
	5-14	11%	21%		**Interesting**	0%	36%				
	15-	13%	39%		**Very Interesting**	30%	52%		Granted	0	6

Source: DN (1995a). This table reproduces the main table reported in the original article by W&W in Dagens Nyheter published in January 1995. Note that far from all review summaries includes the information on interest, or independence. See Table 2 for the replication of the 1997 *Nature* article main results.

Table X:

Outcome:	Change in Review Board Specific Bias		Change in Research Output During Grant Period			
Specification:	Unweighted First-Stage	Weighted First-Stage	**Not Granted** Unweighted 2SLS	**Granted** Unweighted 2SLS	**Not Granted** Weighted 2SLS	**Granted** Weighted 2SLS
Pre-WW Field Specific Male Premiums	.847***	.939***				
	(.136)	(.135)				
Change in Review Board Specific Bias			.195***	.624**	.098	.340*
			(.072)	(.252)	(.135)	(.199)
K.-P. First-stage *F*-stat	38.58	74.63				

Note: The unit of observation is at the university-field-period level, all regressions are all weighted by the number of enrolled students in each field and university. The full sample consists of PhD students enrolled in the academic years between 1985 and 2001. Column 1 provides the first difference estimates in the years before WW (1985-89 vs 1990-94). The WW sample is split into three distinct enrollment periods pre-WW (1990-1994), Phase-in period (1995-1996), and 1997-2001 Post-WW). Column 2 provides the first difference estimates for pre vs. post WW periods. Columns 3-8 provides level regressions, all including period and department (university x field) FE, (where field is defined by the two-digit field of study codes) and university specific trends. The main outcome variable is the share of research that is focused on female (male) health published during the first seven (seventeen) years after PhD enrollment. PostWW is an indicator equal to 1 if enrolled after WW, $\overline{\text{FemalePhD Students}}_{ijt}$ is the female share of graduate students enrolled in the university/field/period. $\overline{\text{PreFemPhDStud}}_{ijt}$ is the pre-WW Share of Female PhD Students. Columns (3-8) also include a phase-in period indicator (enrollment cohorts 1995-96, see figure) and its interaction with the pre-WW share female students. First-stage *F* stat is Kleibergen-Paap Wald statistic, standard errors in cols.3-8 clustered at the university-field level presented in parenthesis (heteroscedasticity robust in first-differences columns), 10/5/1 percent level significance reported */**/*** respectively.

Table A4. Assessing Male Competence Assessment Premiums – **Excluding name changers**

Dependent variable:	*Competence Score (std.)*	*Competence Score (std.)*	*Competence Score (std.)*	*Competence Score (std.)*
	Before Revelation	Before Revelation (Replicating W&W)	Before Revelation	After Revelation
	(1)	(2)	(3)	(4)
Male Premium	.856****	.703***	.665***	-.009
	(.160)	(.140)	(.139)	(.148)
Total Impact Factor		.009***	.005*	.006
		(.002)	(.003)	(.004)
First Author Impact			.017***	.026***
			(.005)	(.009)
Numb. of publications			.002	.010
			(.011)	(.009)
Committee Connection			.412**	.399*
			(.187)	(.212)
R-squared	.41	.57	.60	.50
Observations	117	117	117	143
Review committee FE	Yes	Yes	Yes	Yes

Note: See table 2 for details – this table replicates Table 2 after excluding name changers as recorded in statistics Sweden name change database.

Table A5. Assessing Male Method and Relevance Score Male Assessment Premiums

Dependent variable:	***Method*** *score (std.)*	***Method*** *score (std.)*	***Relevance*** *Score (std.)*	***Relevance*** *Score (std.)*	***Total*** *Score (std)*	***Total*** *Score (std)*
	Before Disclosure	After Disclosure	Before Disclosure	After Disclosure	Before Disclosure	After Disclosure
	(1)	(2)	(3)	(4)	(5)	(6)
Male Premium	.359**	.180	.087	-.060	.392**	.054
	(.166)	(.159)	(.158)	(.162)	(.148)	(.147)
Total Impact Factor	.009***	.010***	.012	.010**	.010***	.010
	(.003)	(.003)	(.004)	(.004)	(.003)	(.004)
First Author Impact	.020***	.013	.014	.023***	.018***	.025
	(.005)	(.008)	(.006)	(.007)	(.005)	(.009)
Numb. of publications	-.026**	-.017*	-.035	-.016	-.025**	-.009
	(.012)	(.010)	(.014)	(.011)	(.012)	(.010)
Committee Connection	-.004	.468**	.343*	.286	.243	.392*
	(.226)	(.223)	(.198)	(.221)	(.194)	(.211)
R-squared	0.44	0.31	0.46	0.28	0.54	0.41
Observations	123	151	123	151	123	151
Review committee FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: See table 2 for details – this table replicates Table 2 after excluding name changers as recorded in statistics Sweden name change database.

Appendix B: **Brita Wahren's Letter to the Council Chairman on the Content of the Scores**

Brita Wahren's Letter to the Council Chairman on the Content of the Scores
In November 1995, Britta Wahren, the Chairwoman of Microbiology II committee sends a letter to the Chairman of the council Olle Stendahl, stating:

Our translation/summary (Original letter on the following pages)

> *"Dear Olle,*
> *As a direct result of today's discussion about how the prioritization should take place I attach my list of the variables that I use to try to identify for the subscores. […] I believe most reviewers use similar variables to prioritize the candidates" […]*
>
> ***The Applicant's Competence:***
>
> *Novelty*
> ***The applicant's competence for the specific task***
> *can be judged by the publication list,*
> *impact factors in the publication list, the applicants order in the author list*
> *how the project is described*
> ***The Applicant's general competence***
> *can be judged by the activity in regards to*
> *publications (increasing or decreasing)*
> *education*
> *administrative skill*
> *teaching skill*
> *Other grants – over years, size, source*
> *International activity*
>
> *The group's competence"*
>
> (Our translation, her emphasis in bold letters in the original letter), 1995-11-21, Dnr (1179/95),

Stored at the National Archive in:

Medicinska forskningsrådet (Statens Medicinska forskningsråd), Rådets föredragningslistor SE/RA/420710/A/A 1/A 1 B/21

Ink. den 1995-11-23
Dnr 1179/95

SWEDISH INSTITUTE FOR INFECTIOUS DISEASE CONTROL

KAROLINSKA INSTITUTE
DEPARTMENT OF CLINICAL VIROLOGY

Stockholm 951121

Bäste Olle,

som ett direkt resultat av dagens diskussion om hur prioriteringsarbetet bör gå till, skickar jag min lista på de variabler jag brukar försöka identifiera inom de större sammanfattande rubrikerna. De kan vara till fördel om någon nisch skall identifieras eller särskild fördel spåras eller sökas inom ett ämnesområde. Annars tror jag att de flesta prioriterar med liknande variabler.

Det vore intressant för mig om det finns ytterligare egenskaper som borde beaktas, sådana diskussioner har säkert förts tidigare på MFR.

Med bästa hälsning Britta Wahren

951121 B Wahren

Variabler i bedömning av anslagsansökningar

Frågeställning/projektet

Originell, klar
Intressant frågeställning
Klar plan
Genomförbar plan
Kan frågeställningen lösas med föreslagen strategi
Originell, nyskapande, innovativ
Förväntad betydelse
Fokuserad
Ifrågasättande
Aktuell (prioriterat område)
Betydelsefullt område
Område i stark utveckling
Repetitiv, dvs upprepar kända fakta
Har resultatens reproducerbarhet eller sanning verifierats
Funktionell frågeställning eller deskriptiv
Klinisk relevans
Utvecklingspotential
Framtidsutsikter för forskaren, för fältet

Metoder/material

Adekvat teknik
Originell teknik
Originell, nyskapande
Finns material och metoder,
eller klara planer på hur det skall samlas, behandlas
Materialet originellt
Materialet svåråtkomligt, unikt
God utvärderingsteknik
Egen metod
Hur skicklig är sökande i de beskrivna metoderna
Är samarbete planerat i metoddelen

Sökandes kompetens

Nyskapande

Sökandes kompetens för den direkta uppgiften

kan bedömas av publikationslista,

impactfaktorer i publikationslista, sökandes placering på arbetena

hur projektet beskrivs

Sökandes allmänna kompetens

kan bedömas av aktuell aktivitet i

publikationshänseende (ökande eller minskande)

utbildning

administrativ skicklighet

lärarskicklighet

Andra anslag - över åren, storlek, källa

Internationell aktivitet

Gruppens kompetens

(huvudmannen för gästforskare,

eller mottagande grupp internationellt)

Bedömning /uppföljning

Vid projektansökan är projektets höjd huvudsaken, tätt följd av kompetensen

Vid tjänst är den personliga kompetensen huvudsaken, tätt följt av projektet

Beröm särskilt goda, intressanta sidor

Kritik av felaktig frågeställning, metodik

Konstruktiv kritik eller feedback

Site visits

Table A3: Comparing the Wahren list for what goes into the Competence Score (in order of appearance) with the measures we have at our disposal

The Wahren List:	**Our measures**	**Data Source**
novelty	Average Novelty scores of publications past 10 years	Torvik et al.
impact factors	Cumulative Impact Factor past 10 years	Journal Impact Factor
the applicants order in the author list	Overall First-author Impact Factor Number of First-author publications	PubMed
how the project is described	*No measure (applications destroyed)*	-
activity in regards to publications (increasing or decreasing)	Share of impact from last 3 years over total Impact 10 years	PubMed
education	PhD granting institution FE 2-digit Field of PhD FE	Statistics Sweden
administrative skill	*No measure (applications destroyed)*	-
teaching skill	*No measure (applications destroyed)*	-
other grants – over years, size, source	Presence of publication collaboration with NIH grant	Azoulay et al.
international activity	Presence of publication collaboration with NIH grant	Azoulay et al.
the group's competence	*No measure – not applicable to the post-doc grants*	-

Examples of formulations in the review summary statements of 1994 applicants:

R1564 "[the applicant] lists 8 papers after PhD graduation, however most with more senior colleagues in Lund". "[The applicant] defended his thesis in 1991, he has trouble demonstrating his independence"

R1589 "After the thesis 4 papers has been published or are in press and on one of these he is first-author"

R1616 " [the applicant] lists 15 publications after thesis, however only 2 without his former thesis advisor"

R1569 "The project is of interest. The scientific production is within the research proposals area and is published in good journals."

R1583 "… several pieces published in good journals. The total number of publications amounts to 29 articles. In summary the applicant has without a doubt the competence needed to complete the project…"

R1650 "Her scientific production is quantitatively small and can considering the competition only be given a medium priority"

R1666 "despite … defending as recently as 1994 he has a relatively high production. List 15 publications with several in prominent journals".

R1562 "The applicants own scientific production is hence relatively low. Considering this the application is given low priority".

R1560 "The Project is very strong and well worth support. [the applicants] role in the project in relation to [supervisor] should be discussed. High priority"

R1609 "[applicant] writes very competently and adequately. Collaborations has been established with researchers in England and USA", "All previous work with senior collaborators in highly ranked international journals. Also this project in close collaboration with [supervisor]. Independence therefore difficult to judge"

R1625 "He shows a very high scientific productivity beyond the thesis work. Certain weakness in the application is the lack of focus in the research program. In summary the applicant is well suited for a research position."

R1567 "The applicant defended in May 1994 at Karolinska Institute". "In the current tough competition, where many applicants have reached considerable independence competence and created own profiles after dissertation, this application cannot be placed in the absolute top but is given a good priority".

R1570 "Research program of great interest in the forefront of the subject". "Publications in very good journals but the thesis quality is diminished by that

one thesis manuscript was never submitted and was apparently premature." "Are now at a very good laboratory, gives an independent and energetic impression, and should after returning home be able to get established. The application is given high priority".

R1587 "[The applicant] has during the last year submitted a couple of papers to eminent journals (good luck!), indicating that the research has now begun." "High priority"

R1591 "The research question is well defined, but not original. Not prioritized in the very tough competition"

R1628 "The applicants own competence, documented through her scientific publication record, is at present not particularly voluminous"

R1567 "The project is well described and focused in the application. Despite newly graduated he seems to have excellent competence and has very good merits (The two *Science* papers)."

R1614 "The applicant's competence for the task needs to be assessed as good or very good and is well-documented in independently produced original work".

R1651 "The applicant's competence for the task is good based on the rather voluminous publication record displayed"

Assigning reviewer numbers to reviewer gender and competence.

The evaluation protocol of each application contains a number assigned to each reviewer (1 to 6) and the scores that each reviewer assigns to each candidate for competence of the applicant, method, and relevance of the research proposal. To find out which reviewer corresponds to which individual, we applied two approaches. First, we used the name of the author and the information about which reviewer (anonymous number) abstained from grading the author due to a conflict of interest.

The order of the reviewers is as follows: reviewer 1 is the chair of the committee, the remaining regular reviewers are ordered based on their surnames, and the suppleant (alternate) is registered as number 6. The names of the members of the review committees are published by the Medical Research Council in their yearly project catalogues (e.g., MFR, 1997, p. 557 ff). With this information, we linked each review board member to their assessment of each applicant.

To verify that this assumption of the ordering was correct, we searched the Author-ity PubMed database for connections between the applicants and reviewers. In many cases, though not all, we found that the author and the reviewer had co-authored work before the application date. Hence, the reviewer had indicated a conflict of interest on the evaluation sheet and abstained from scoring the applicant. The imperfect match is not surprising, since a reviewer is considered to have a conflict of interest not only if they have co-authored work. For example, in some cases, evaluations are missing because, as stated in the evaluation sheets, the reviewer worked at the same institution or was a second-degree co-author (e.g., N1566, N1655, and N1633).

As additional verification, we looked at the protocols from the evaluation committees in the 1993–1998 application rounds (MFR, 1993), which are stored at the National Archives. Between 1993–1998, the protocols from the evaluation committee meetings indicate which reviewer (named) registered a conflict of interest for each applicant (with applicant registration number and name). Using this information, we checked in the reviewer-level data whether the reviewer number and assigned reviewer name corresponded to the missing information on the reviewer evaluation. It corresponded perfectly in the cases where there was an indication in the original white sheets that there had been a declared conflict of interest.

By linking the reviewer name to the reviewer number, we could then search the Author-ity database for the publications of the reviewers during the 10 years prior to the application year, to measure their scientific competence in the same way as for the applicants (i.e., cumulative impact factor) and determine their gender based on their name (with no errors). We were also able to follow the reviewers across multiple application years.

References

MFR 1997, Project Catalogue 1995/1996, The Swedish Medical Research Council, Stockholm , ISBN 91-85547-03-4.

MFR 1993, ”Protokoll från sammanträde med prioriteringskommittén för […] 1993-11-01”, 1993-11-25, Dnr 698/93, National Archives reference code: SE/RA/420710/A/A 1/A 1 A/29

MFR 1995,”Beslutslista Sammanträde Medicinska forskningsrådet”, August 23, 1995, dnr 691/95, National Archives reference code SE/RA/420710/A/A 1/A 1 A/