Female grant applicants are equally successful when peer reviewers assess the science, but not when they assess the scientist

Holly O. Witteman, PhD[1]; Michael Hendricks, PhD[2];

Sharon Straus, MD FRCPC MSc HBSc[3]; Cara Tannenbaum, MD MSc[4]

[1]Associate Professor, Department of Family and Emergency Medicine, Faculty of Medicine, 1050 avenue de la Médecine, Université Laval, Quebec City, Canada, G1V 0A6

[2]Assistant Professor, Department of Biology, Faculty of Science, 1205 av du Docteur-Penfield, McGill University, Montreal, Canada, H3A 1B1

[3]Professor, Li Ka Shing Knowledge Institute of St. Michael's Hospital, 30 Bond Street, Toronto, Canada, M3B 2T9

[4]Scientific Director, Institute for Gender and Health, Canadian Institutes of Health Research, Ottawa, Canada, H3A 1W4

Corresponding author: Holly O. Witteman, 1050 avenue de la Médecine, Pavillon Ferdinand-Vandry 2881, Université Laval, Québec City, Québec, Canada, G1V 0A6, telephone: 418.656.2131 x3981, fax: 418.656.2465, email: holly.witteman@fmed.ulaval.ca

Keywords (MeSH terms): Peer Review, Research/methods; Sex Factors

Keywords (non-MeSH terms): Gender Bias

Running title: Gender Bias in Grant Review

Word count: 3249

ABSTRACT

Background: Funding agencies around the world show gender gaps in grant success, with women often receiving less funding than men. However, these studies have been observational and some have not accounted for potential confounding variables, making it difficult to draw robust conclusions about whether gaps were due to bias or to other factors. In 2014, the Canadian Institutes of Health Research (CIHR) phased out traditional investigator-initiated programs and created a natural experiment by dividing all investigator-initiated funding into two new grant programs: one with and one without an explicit review focus on the caliber of the principal investigator. In this study, we aimed to determine whether these differently-structured grant programs had different success rates among male and female applicants.

Methods: We analyzed results of 23,918 grant applications from 7,093 unique applicants in a 5-year natural experiment across all open, investigator-initiated CIHR grant programs in 2011-2016. Our primary outcome was grant application success. We used Generalized Estimating Equations to account for multiple applications by the same applicant and an interaction term between each principal investigator's self-reported sex and grant program group to compare any gaps in success rates among male and female applicants in the two new programs to the baseline gap in traditional programs. Because younger cohorts of investigators and fields such as health services research and population health have higher proportions of women, our analysis controlled for principal investigators' ages and applications' research domains.

Results: The overall grant success rate across all competitions was 15.8%. After adjusting for age and research domain, the predicted probability of funding success among male principal investigators' applications in traditional programs was 0.9 percentage points higher than it was among female principal investigators' applications (OR 0.934, 95% CI 0.854-1.022). In the new program in which review focused on the quality of the proposed science, the gap was 0.9% in favour of male principal investigators and not significantly different from traditional programs (OR 0.998, 95% CI 0.794-1.229). In the new program with an explicit review focus on the caliber of the principal investigator, the gap was 4.0% in favour of male principal investigators, significantly larger than in traditional programs (OR 0.705, 95% CI 0.519-0.960).

Conclusions: Avoiding bias in grant review is necessary to ensure the best research is funded, regardless of who proposes it. In this study, gender gaps in grant success rates were significantly larger when there was an explicit review focus on the principal investigator. Because of the quasi-experimental study design, these findings offer more conclusive evidence

than was previously available about the causes of gender gaps in grant funding. Specifically, this study suggests that such gaps are attributable to differences in how women are assessed as principal investigators, not differences in the quality of science led by women.

INTRODUCTION

For decades, studies have shown that women in academia and science must perform to a higher standard than men to receive equivalent recognition,[1–4] especially Indigenous and racialized women.[5–11] Compared to men, women are more often characterized as lacking the brilliance, drive, and talent required to carry a novel line of inquiry through to discovery,[8,12] with children as young as six years old endorsing such stereotypes.[13] Women are less likely than men to be viewed as scientific leaders[14–17] or depicted as scientists.[18] Women in academia contribute more labor for less credit on publications,[19,20] receive less compelling letters of recommendation,[21–24] receive systematically lower teaching evaluations despite no differences in teaching effectiveness,[25] are more likely to experience harassment,[11,26–28] and are expected to do more service work[29,30] and more special favors for students.[31] While men in academia have more successful careers after taking parental leave, women's careers suffer after the same.[32] Women receive less start-up funding as biomedical scientists[33] and are underrepresented in invitations to referee papers.[34] Compared to publications led by men, those led by women take longer to publish[35] and are cited less often,[36,37] even when published in higher-impact journals.[38] Papers[39] and conference abstracts[40] led by women are accepted more frequently when reviewers are blinded to the identities of the authors. Women are underrepresented as invited speakers at major conferences,[4] prestigious universities,[41] and grand rounds.[42,43] When women are invited to give these prestigious talks, they are less likely to be introduced with their formal title of Doctor.[44] Female surgeons have been shown to have better patient outcomes overall,[45] yet, when a patient dies in surgery under the care of a female surgeon, general practitioners reduce referrals to her and to other female surgeons in her specialty, whereas they show no such reduced referrals to male surgeons following a patient's death.[46] Women are less likely to reach higher ranks in medical schools even after accounting for age, experience, specialty, and measures of research productivity.[47] When fictitious or real people are presented as women in randomized experiments, they receive lower ratings of competence from scientists,[48] worse teaching evaluations from students,[49] and fewer email responses from professors after presenting as students seeking a PhD advisor[9] or as scientists seeking copies of a paper or data for a meta-analysis.[50]

Conversely, other research has demonstrated advantages experienced by women in academia; for example, achieving tenure with fewer publications than men.[51] In assessments of potential secondary and postsecondary teachers and professors, women are favored in male-dominated fields, as are men in female-dominated fields.[52] When fictitious people are presented as women

in randomized experiments, they receive higher rankings as potential science faculty.[53,54] This aligns with evidence from other contexts showing that high-potential women are favored over high-potential men[55] and that, while women face discrimination at earlier stages, once women have proven themselves in a male-dominated context, they are favored over men.[56]

In sum, there is considerable evidence that women face or have faced persistent, pervasive barriers in many aspects of academia and science. There is also recent evidence that women in academia and science may fare better in merit evaluations than equally-qualified men after they have progressed beyond postsecondary education. In light of this evidence as a whole, we consider the question: does bias in favor of men or women influence current research funding?

Previous research on this question has been suggestive but not conclusive. A 2007 meta-analysis of 21 studies from a range of countries found an overall gender gap in favor of men, with 7% higher odds of fellowship or grant funding for male applicants.[57] Research since then has documented that, compared to their male colleagues, female principal investigators have lower grant success rates,[58] lower grant success rates in some but not all programs,[59,60] equivalent grant success rates after adjusting for academic rank[61,62] but fewer funds requested and received,[61–64] or equivalent funding rates.[65] To the best of our knowledge, no such study has yet found women to experience higher grant success rates nor to receive more grant funding than men. These previous studies have been observational, making it difficult to draw robust conclusions about the causes of gender gaps when they are observed. Furthermore, some previous studies have not accounted for potential confounding variables; for example, domain of research.[57,66,67]

Our objective in this study was therefore to determine whether gender gaps in grant funding are attributable to potential differences in how male and female principal investigators are evaluated, using real-world data and a study design that would allow for stronger conclusions than those from observational studies. Our study made use of a natural experiment at a national health research funding agency which allowed for the comparison of grant success rates among male and female principal investigators between three grant programs: traditional grant programs, which had demonstrated higher success rates among younger male principal investigators for applications submitted in 2001-2011[58] but for which no such analysis had been conducted after 2011, and two new competitions, one with and without an explicit review focus on the quality of the principal investigator.

We identified three potential results of our comparisons of gaps in success rates between female and male principal investigators. First, if gaps were to be similar under traditional review criteria and both sets of new review criteria, this would suggest that such gaps, when present, may reflect different career paths and choices made by women and men,[68,69] differences between the types of research proposed by female and male principal investigators, or may be spurious. Second, if there were to be a larger gap in favour of male principal investigators in the competition with more focus on the science, this would suggest that gender gaps were due to female principal investigators proposing science that peer reviewers assessed as being of lower quality. Third, if there were to be a larger gap in favour of men in the competition with more focus on the scientist, this would suggest that gender gaps in research funding are partly or wholly driven by women being assessed less favorably as principal investigators compared to their male colleagues. Other potential results such as gaps in favor of female principal investigators were not considered a priori because publicly-available summary statistics of the programs showed these results to be impossible.

METHODS

Beginning in 2014, the Canadian Institutes of Health Research (CIHR) phased out traditional open grant programs and divided all investigator-initiated funding into two new programs: the Project grant program and Foundation grant program. Both new programs used a staged review process in which lower-ranked applications were rejected from progressing to further stages. As in traditional programs, reviewers in the new Project grant program were instructed to primarily assess the research proposed. Seventy-five percent of the application score was based on reviewers' assessments of ideas and methods while 25% was based on reviewers' assessments of principal investigators' and teams' expertise, experience, and resources. In contrast, the Foundation grant program was about 'people, not projects' and was designed to provide grants to fund programs of research. At the first stage of the Foundation review process, reviewers were instructed to primarily assess the principal investigator, with 75% of the score allocated to reviewers' assessments of principal investigators' leadership, productivity, and the significance of their contributions, and 25% to a one-page summary of their proposed 5- or 7-year research program. Only principal investigators who passed this stage were invited to submit a detailed proposal describing their research. Thus, these new programs enabled a direct, quasi-experimental comparison of success rates of male and female applicants in grant programs with and without an explicit focus on the caliber of the principal investigator.

New investigators and those who had never held CIHR funding could apply to programs of their choice. Established principal investigators who already held CIHR funding were eligible for the Foundation program if one or more of their active CIHR grants was scheduled to end within a specific date range. Principal investigators could apply to multiple programs, with some restrictions. In the first cycle of the Foundation program, principal investigators who passed the first stage and were accepted to submit a full description of their research could not simultaneously apply to the last cycle of traditional programs. In the second cycle of the Foundation program, principal investigators could apply to Foundation and Project programs, providing they did not submit the same research proposal to both programs.

We analyzed data from all applications submitted to CIHR grant programs across all investigator-initiated competitions in 2011 through 2016. We excluded applications that were withdrawn, as these did not receive full peer review. We also excluded applications if the principal investigator, referred to as the nominated principal applicant in the CIHR system, had not reported their sex, their age, the domain of research of their application, or if their self-reported age was unrealistic. We defined unrealistic ages as occurring when a principal investigator's self-reported birth year was prior to 1920 or after 2000. Ensuring correct date entry in web-based forms is a known challenge in human-computer interaction.[70] In the online system used to collect the data in this study, the default birth date is prior to 1920, suggesting that such self-reported birth years were most likely to occur when people did not enter a birth date. Birth years in the 2000s were deemed to be errors and may have occurred when people accidentally entered the current year rather than their birth year.

We used Generalized Estimating Equations to fit a logistic model that accounted for the same principal investigator submitting multiple applications,[71,72] including principal investigators who applied to both Project and Foundation programs. We conducted analyses in R statistical computing software, version 3.4.0,[73] using the geepack package to fit models.[74] We then used the fitted model to test the pairwise effect of sex within each program, using the lsmeans package.[75] This allowed us to compute marginal effects for specific contrasts of interest.

We modeled grant success rates as a function of the grant program, principal investigators' self-reported binary sex, self-reported age, self-declared domain of research, and an interaction term between each principal investigator's sex and the grant program to which they were

applying. The interaction term allowed us to address the objective of this study by determining whether there was any effect of different review criteria on relative success rates between male and female applicants after controlling for age and domain of research. We controlled for these because younger cohorts of investigators included larger proportions of female principal investigators, as did domains of health research other than biomedical. The adjustment for age also helped account for the fact that both the Foundation program and Project program had a predefined minimal allocation to new investigators, meaning those within their first five years as independent investigators. The CIHR collected data about binary sex, not gender; therefore, our study assumes that people who self-reported as female or male identified as women or men, respectively. The CIHR did not collect complete data on other applicant characteristics that have been shown to be associated with disparities in funding and career progression; for example, career stage, race, ethnicity, Indigeneity and disability.[6,76] Further analytical details are available in the online appendix.

RESULTS

There were a total of 25,706 applications during the five years of this study. We excluded 1,788 applications consisting primarily of principal investigators with unrealistic years of birth; i.e., birth years prior to 1920 (n=1,631) or after 2000 (n=12). The final dataset analyzed contained 23,918 applications from 7,093 unique principal investigators. There were 15,775 applications from 4,472 male principal investigators and 8,143 applications from 2,621 female principal investigators. Twenty-eight percent of principal investigators submitted a single application during the five-year study period, 20% submitted two applications, 25% submitted three or four applications, and the remaining 27% of principal investigators submitted five or more applications. The maximum number of applications from a principal investigator during the 5-year period was 40.

The overall grant success rate across the data set was 15.8%. As shown in Figure 1, after adjusting for age and research domain, the predicted probability of funding was 0.9 percentage points higher for male principal investigators than female principal investigators in traditional programs (OR 0.934, 95% CI 0.854-1.022). This gap was 0.9% in the Project program (OR 0.998, 95% CI 0.794-1.229) and 4.0% in the Foundation program (OR 0.705, 95% CI 0.519-0.960). Figure 2 shows how the gap in the Foundation program was driven by discrepancies at the first review stage, where review focused on the principal investigator. Across all grant programs, odds of receiving funding were also lower in the three non-biomedical research

domains and for younger principal investigators. Tabular results are available in the online appendix.

Figure 1 about here

Figure 2 about here

DISCUSSION

Our study provides stronger evidence than was previously available regarding the likely causes of gender gaps in grant funding. When reviewers primarily assessed the science, there were no statistically significant differences between success rates for male and female principal investigators. When reviewers explicitly assessed the principal investigator as a scientist, the gap was significantly larger. These data support the third of three possible findings we described in our introduction; namely, that gender gaps in funding stem from female principal investigators being evaluated less favourably than male principal investigators, not from differences in the quality of their science.

Our findings align with previous studies that have similarly observed that reviewers may assess the characteristics of female funding applicants less favourably. Data from the United States showed that female grant applicants to the National Institutes of Health's flagship R01 program were less likely than male applicants to be described as leaders.[14] In the Netherlands, grant reviewers gave equal scores to men's and women's proposed research but assigned lower scores to women as researchers.[77] In Sweden, similar biases have been shown among evaluators' assessments of applicants for governmental venture capital.[78] Our findings may also be placed in the context of evidence from other domains in which observed gender gaps at the highest levels of achievement are explained by attitudes, not ability. For example, when gender equality improves in a country, the gender gap in top mathematics performers disappears.[79] Similarly, women became more successful in orchestra auditions when auditioning musicians' identities were concealed behind a screen.[80]

The hypothesis that gender gaps in peer review outcomes are rooted in less favourable evaluations of female applicants was further supported by the observed effects of subsequent actions taken by the CIHR. Following the grant cycles analyzed in our study and as part of a broader Equity Framework,[81] the CIHR implemented new policies in an attempt to eliminate the

observed gender gap in Foundation grants. The policies included instructions that reviewers should complete an evidence-based[82] reviewer training module about multiple forms of unconscious bias.[83] Training has been previously shown to help mitigate the effects of bias.[15,84] Additionally, the Foundation grant program regulations were revised such that, should reviewer training not have the desired effects at Stage 1, a proportional number of female applicants would nonetheless proceed to the next stage at which their proposed research would be evaluated. In other contexts, such quotas based on the available pool of candidates have been shown to increase women's representation, the overall quality of candidates, or both.[85–87] In the following Foundation grant cycle, which was underway during this project, success rates were equivalent for male and female principal investigators. (See summary data in the online appendix.)

Our study has four main strengths. First, it was quasi-experimental. To the best of our knowledge, this is the first evidence from a study design that was not fully observational, enabling stronger conclusions than were previously possible. Second, while quasi-experimental studies have potential for selection bias,[88] in this study, selection bias was somewhat limited by eligibility rules. Specifically, for principal investigators who had already established their careers and received funding from the CIHR, only those whose grants from traditional programs were expiring within a specific time period were permitted to apply to Foundation. This meant that a portion of allocation was dependent on an external, arbitrary variable. Third, we controlled for age and domain of research, two key confounders in studies of gender bias in grant funding. Academic rank, which correlates with age, has been shown to account for gender gaps in grant funding in other studies, as has domain of research.[66,67] Having accounted for these key confounders strengthens our findings. Fourth and finally, our study analyzed all available data over a period of five years from a major national funding agency, thus offering evidence from a large data set of real-world grant review.

Our study also had two main limitations. First, principal investigators were not randomized to one grant program or the other. Although a number of aspects of our study minimized the potential to observe the results we found, the non-randomized design leaves open the possibility that unobserved confounders or selection bias may have contributed to the observed differences. For example, due to the unavailability of these data, we were unable to account for principal investigators' publication records. Publication record is a potential confounder because men tend to publish more than women overall.[89] Inclusion of such a variable could therefore

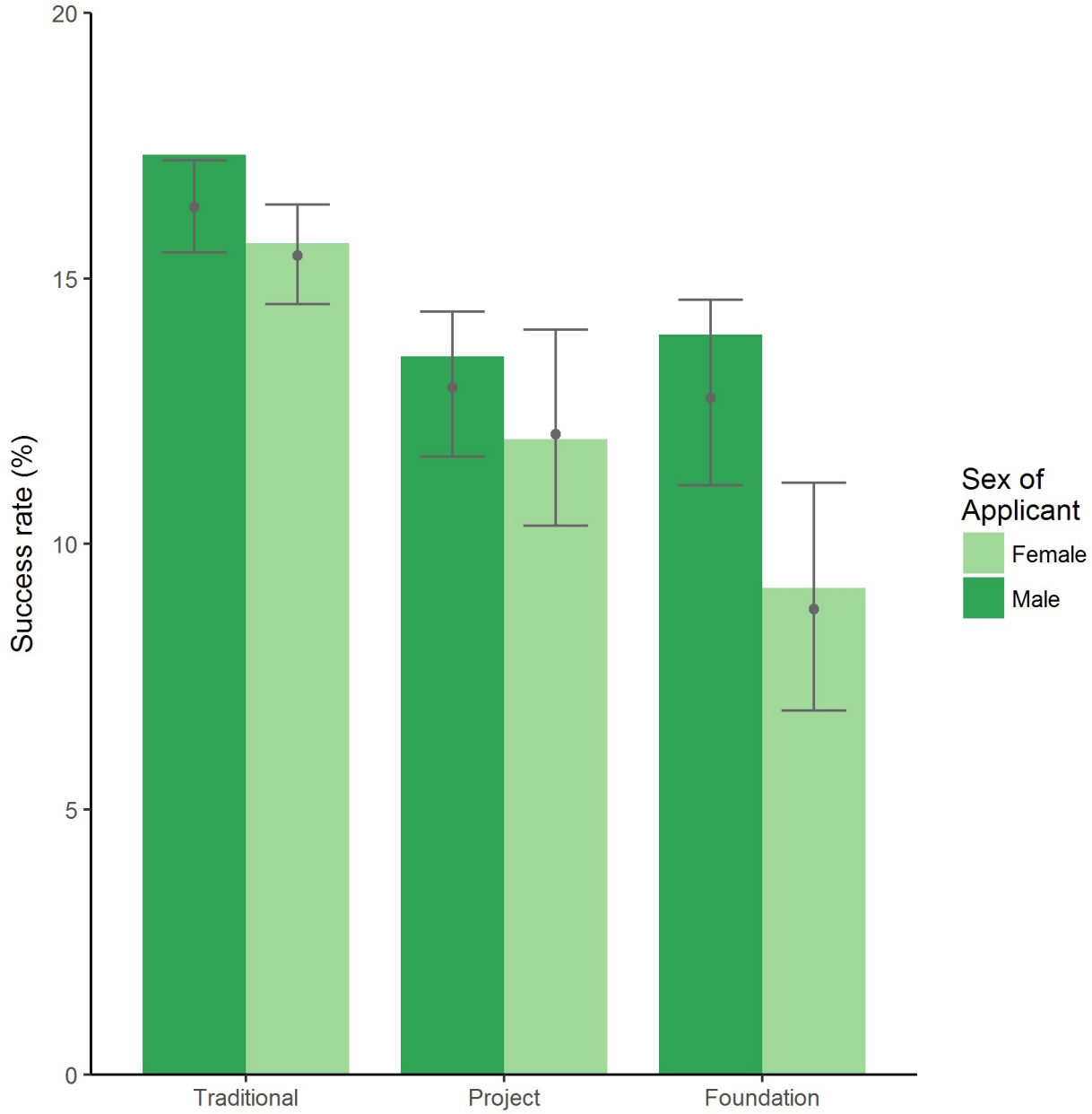
account for all or part of the observed differences. Alternatively, it could increase the observed differences, given that female funding applicants have been shown to receive systematically lower scores compared to male applicants with equivalent publication records.[2] More recent research is examining whether male and female principal investigators with equivalent productivity records were evaluated equivalently by Foundation reviewers (R. Tamblyn, personal communications, July 3, 2017 and December 11, 2017). Second, our assumption that people who self-report as female or male also identify as women or men, respectively, may not be true in all cases. Data are lacking regarding how many people identify as transgender or non-binary in Canada. However, context may be offered by a recent analysis suggesting that transgender people were 0.4% of the US population.[90] If this proportion were reflected in our study and if all transgender applicants reported their sex as their assigned sex at birth, this would represent 28 people in total within this study, a number unlikely to substantially change the results of our analyses.

Conclusions

Bias in grant review prevents the best research from being funded. When this occurs, lines of research go unstudied, careers are damaged, and funding agencies are unable to deliver the best value for money, not only within a given funding cycle, but also long term as small differences compound into cumulative disadvantage. To encourage rigorous, fair peer review that results in funding the best research, we recommend that funders minimize opportunities for bias by focusing assessment on the science rather than the scientist, measure and report funding by applicant characteristics, including potential confounding variables, and consider reviewer training and other policies to mitigate the effects of all forms of bias. Future research should investigate the potential for other types of bias and evaluate methods of reducing bias and increasing fairness in peer review.

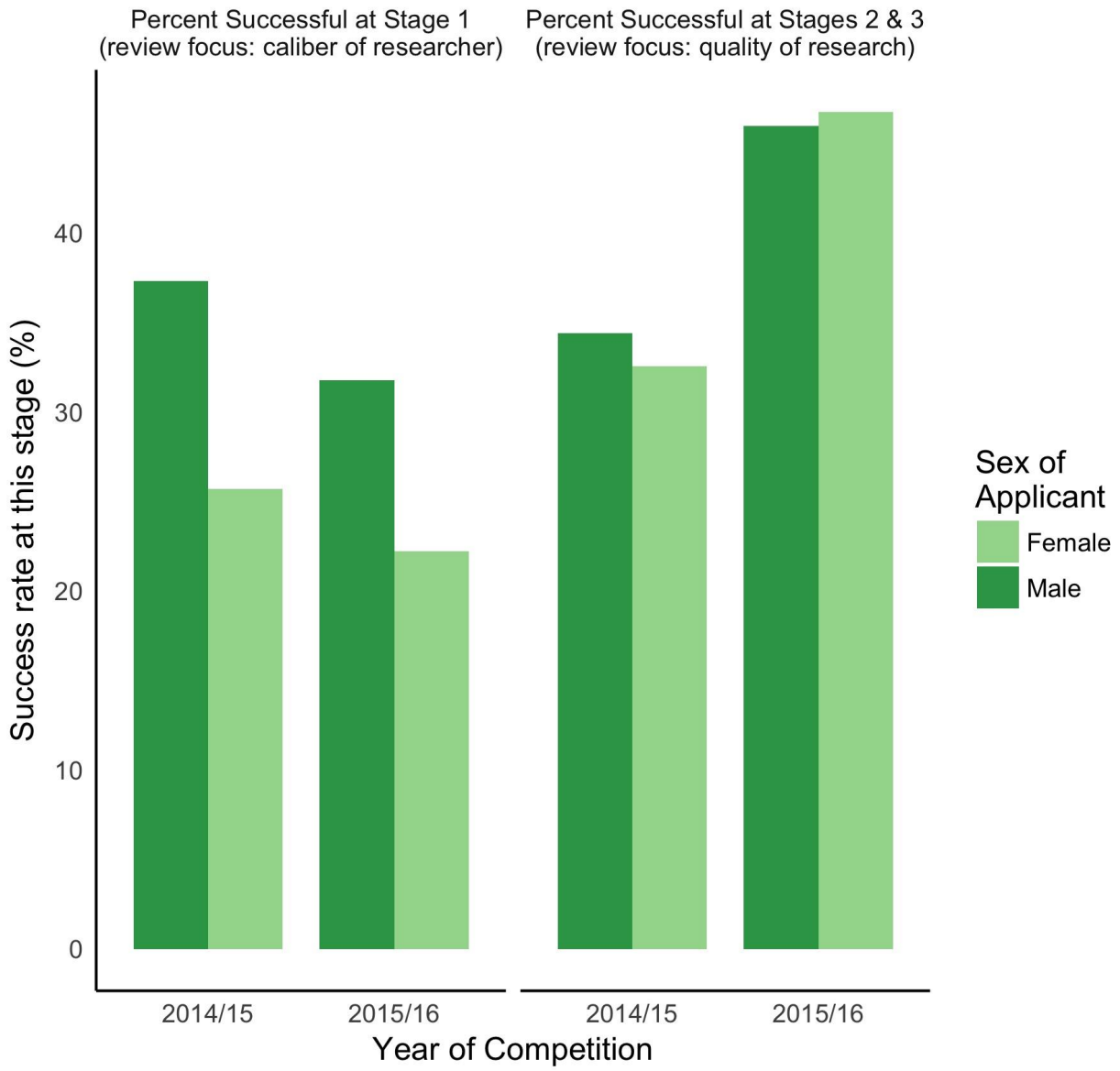
FIGURES

Figure 1. Funding success rate by grant program



Columns indicate observed success rates. Points and error bars indicate model-predicted means and 95% confidence intervals, respectively.

Figure 2. Foundation results by review stage



Columns indicate observed success rates.

APPENDICES

Appendix 1. Methodological and Analytical Details

DECLARATIONS

Abbreviations

CIHR: Canadian Institutes of Health Research

Ethics Approval, Consent to Participate and Consent for Publication

The views expressed in this paper are those of the authors and do not necessarily reflect those of the CIHR or the Government of Canada. Data were held internally and analyzed by staff at the CIHR within their mandate as a national funding agency. Research and analytical studies at the CIHR fall under the Canadian Tri-Council Policy Statement 2: Ethical Conduct for Research Involving Humans (available: pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/, accessed 2017 July 13.) This study had the objective of evaluating CIHR's Investigator-Initiated programs, and thus fell under Article 2.5 of TCPS-2 and not within the scope of Research Ethics Board review in Canada. Nevertheless, applicants were informed through ResearchNet, in advance of peer review, that CIHR would be evaluating its own processes. All applicants provided their electronic consent; no applicant refused to provide consent.

Availability of Data and Materials

Data are confidential due to Canadian privacy legislation. Researchers interested in addressing other research questions related to grant funding may contact the CIHR at funding-analytics@cihr.ca.

Competing Interests

This work was unfunded. HW holds grant funding from the CIHR as principal investigator of a Foundation grant. HW and MH are two of the three founding national co-chairs of the Association of Canadian Early Career Health Researchers, an organization that has published statements critical of aspects of the CIHR grant program changes, including the Foundation grant program. CT is a Scientific Institute director at the CIHR and is therefore partially employed by the CIHR. SS holds grant funding from the CIHR, including a Foundation grant,

and also received contract funding from the CIHR to lead the scoping review described herein and to analyze applicant and reviewer survey responses. HW receives salary support from a Research Scholar Junior 1 Career Development Award from the Fonds de Recherche du Québec-Santé. SS receives salary support from a Tier 1 Canada Research Chair in Knowledge Translation and Quality of Care.

Contributions

Conceptualization: HW MH DG (see Acknowledgements); Methodology: HW JW RD AC MH DG; Formal Analysis: JW; Investigation: HW JW RD AC; Data Curation: JW RD AC; Writing – Original Draft: HW JW; Writing – Review & Editing: HW JW RD AC MH CT SS DG; Visualization: HW JW RD; Project Administration: HW RD. Contributions are listed according to the CReDIT taxonomy (docs.casrai.org/CRediT).

Acknowledgements

The authors gratefully acknowledge early work by Ms. Anne-Sophie Julien, statistician, who worked with HW to outline potential analytical approaches, drafted preliminary R code for the proposed research question, and reviewed a draft of the manuscript. We also thank Ms. Anne Strong, CIHR staff, for extracting the dataset, and Dr. Anne Martin-Matthews, CIHR Vice-President, for her comments on drafts of this manuscript, and many colleagues who commented online in response to previous preprint versions of this manuscript. The authors are extremely grateful to Dr. Jonathan A. Whiteley (JW), Dr. Rachelle E. Desrochers (RD), Dr. Alysha Croker (AC), and Dr. Danika Goosney (DG), all of whom were employed at the CIHR during this project and assisted with the scoping of the analysis and interpretation of the results. JW conducted the analysis at the CIHR as the data used consisted of confidential applicant information that could not be shared externally. He also drafted details of the methodology in collaboration with HW. The CIHR is committed to collaborating with the research community to advance knowledge on best practices in peer review without unduly influencing the conclusions drawn by its collaborators. For this reason, CIHR employees are not currently permitted to be co-authors on papers describing analyses of the agency's programs. Without the active participation of these CIHR employees and their contributions, this paper would not have been possible.

REFERENCES

1. Rossiter, M. W. The Matthew Matilda Effect in Science. *Soc. Stud. Sci.* **23,** 325–341 (1993).
2. Wenneras, C. & Wold, A. Nepotism and sexism in peer-review. *Nature* **387,** 341–343 (1997).
3. Reuben, E., Sapienza, P. & Zingales, L. How stereotypes impair women's careers in science. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4403–4408 (2014).
4. Klein, R. S. *et al.* Speaking out about gender imbalance in invited speakers improves diversity. *Nat. Immunol.* **18,** 475–478 (2017).
5. Settles, I. H., Cortina, L. M., Malley, J. & Stewart, A. J. The Climate for Women in Academia: the Good, the Bad, and the Changeable. *Psychol. Women Q.* **30,** 47–58 (2006).
6. Ginther, D. K. *et al.* Race, ethnicity, and NIH research awards. *Science* **333,** 1015–1019 (2011).
7. Williams, J., Phillips, K. W. & Hall, E. V. *Double jeopardy?: Gender bias against women of color in science*. (Hastings College of the Law, Center for WorkLife Law, 2014).
8. Leslie, S.-J., Cimpian, A., Meyer, M. & Freeland, E. Expectations of brilliance underlie gender distributions across academic disciplines. *Science* **347,** 262–265 (2015).
9. Milkman, K. L., Akinola, M. & Chugh, D. What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *J. Appl. Psychol.* **100,** 1678–1712 (2015).
10. Henry, F. *et al.* Race, racialization and Indigeneity in Canadian universities. *Race Ethnicity and Education* **20,** 300–314 (2017).
11. Clancy, K. B. H., Lee, K. M. N., Rodgers, E. M. & Richey, C. Double jeopardy in astronomy and planetary science: Women of color face greater risks of gendered and racial harassment. *J. Geophys. Res. Planets* **122,** 2017JE005256 (2017).

12. Elmore, K. C. & Luna-Lucero, M. Light Bulbs or Seeds? How Metaphors for Ideas Influence Judgments About Genius. *Soc. Psychol. Personal. Sci.* (2016). doi:10.1177/1948550616667611
13. Bian, L., Leslie, S.-J. & Cimpian, A. Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* **355,** 389–391 (2017).
14. Magua, W. *et al.* Are Female Applicants Disadvantaged in National Institutes of Health Peer Review? Combining Algorithmic Text Mining and Qualitative Methods to Detect Evaluative Differences in R01 Reviewers' Critiques. *J. Womens. Health* **26,** 560–570 (2017).
15. Carnes, M. *et al.* The effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Acad. Med.* **90,** 221–230 (2015).
16. Smyth, F. L. & Nosek, B. A. On the gender-science stereotypes held by scientists: explicit accord with gender-ratios, implicit accord with scientific identity. *Front. Psychol.* **6,** 415 (2015).
17. Carli, L. L., Alawa, L., Lee, Y., Zhao, B. & Kim, E. Stereotypes about gender and science: Women ≠ scientists. *Psychol. Women Q.* **40,** 244–260 (2016).
18. Kerkhoven, A. H., Russo, P., Land-Zandstra, A. M., Saxena, A. & Rodenburg, F. J. Gender Stereotypes in Science Education Resources: A Visual Content Analysis. *PLoS One* **11,** e0165037 (2016).
19. Macaluso, B., Larivière, V., Sugimoto, T. & Sugimoto, C. R. Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. *Acad. Med.* **91,** 1136–1142 (2016).
20. Feldon, D. F., Peugh, J., Maher, M. A., Roksa, J. & Tofel-Grehl, C. Time-to-Credit Gender Inequities of First-Year PhD Students in the Biological Sciences. *CBE Life Sci. Educ.* **16,** (2017).
21. Trix, F. & Psenka, C. Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty. *Discourse & Society* **14,** 191–220 (2003).

22. Schmader, T., Whitehead, J. & Wysocki, V. H. A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants. *Sex Roles* **57,** 509–514 (2007).

23. Madera, J. M., Hebl, M. R. & Martin, R. C. Gender and letters of recommendation for academia: agentic and communal differences. *J. Appl. Psychol.* **94,** 1591–1599 (2009).

24. Dutt, K., Pfaff, D. L., Bernstein, A. F., Dillard, J. S. & Block, C. J. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nat. Geosci.* **9,** 805–808 (2016).

25. Stark, P., Ottoboni, K., Boring, A. & Cetinkaya-Rundel, M. Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. (2016). doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

26. Berdahl, J. L. The sexual harassment of uppity women. *J. Appl. Psychol.* **92,** 425–437 (2007).

27. Jagsi, R. *et al.* Sexual Harassment and Discrimination Experiences of Academic Medical Faculty. *JAMA* **315,** 2120–2121 (2016).

28. Nelson, R. G., Rutherford, J. N., Hinde, K. & Clancy, K. B. H. Signaling Safety: Characterizing Fieldwork Experiences and Their Implications for Career Trajectories. *Am. Anthropol.* **119,** 710–722 (2017).

29. Babcock, L., Recalde, M. P., Vesterlund, L. & Weingart, L. Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability. *Am. Econ. Rev.* **107,** 714–747 (2017).

30. Guarino, C. M. & Borden, V. M. H. Faculty Service Loads and Gender: Are Women Taking Care of the Academic Family? *Res. High. Educ.* **58,** 672–694 (2017).

31. El-Alayli, A., Hansen-Brown, A. A. & Ceynar, M. Dancing Backwards in High Heels: Female Professors Experience More Work Demands and Special Favor Requests, Particularly from Academically Entitled Students. *Sex Roles* 1–15 (2018).

32. Antecol, H., Bedard, K. & Stearns, J. *Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?* (IZA Discussion Papers, 2016).
33. Sege, R., Nykiel-Bub, L. & Selk, S. Sex Differences in Institutional Support for Junior Biomedical Researchers. *JAMA* **314,** 1175–1177 (2015).
34. Lerback, J. & Hanson, B. Journals invite too few women to referee. *Nature* **541,** 455–457 (2017).
35. Hengel, E. Publishing while Female: Gender Differences in Peer Review Scrutiny. (2016).
36. Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C. R. Bibliometrics: global gender disparities in science. *Nature* **504,** 211–213 (2013).
37. Caplar, N., Tacchella, S. & Birrer, S. Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy* **1,** 0141 (2017).
38. Ghiasi, G., Larivière, V. & Sugimoto, C. R. On the Compliance of Women Engineers with a Gendered Scientific System. *PLoS One* **10,** e0145931 (2015).
39. Budden, A. E. *et al.* Double-blind review favours increased representation of female authors. *Trends Ecol. Evol.* **23,** 4–6 (2008).
40. Roberts, S. G. & Verhoef, T. Double-blind reviewing at EvoLang 11 reveals gender bias. *Journal of Language Evolution* **1,** 163–167 (2016).
41. Nittrouer, C. L. *et al.* Gender disparities in colloquium speakers at top universities. *Proceedings of the National Academy of Sciences* (2017). doi:10.1073/pnas.1708414115
42. Boiko, J. R., Anderson, A. J. M. & Gordon, R. A. Representation of Women Among Academic Grand Rounds Speakers. *JAMA Intern. Med.* **177,** 722–724 (2017).
43. Buell, D., Hemmelgarn, B. & Straus, S. E. Proportion of women presenters at medical grand rounds at major academic centres in Canada, a retrospective observational study (in press). *BMJ Open*
44. Files, J. A. *et al.* Speaker Introductions at Internal Medicine Grand Rounds: Forms of Address Reveal Gender Bias. *J. Womens. Health* **26,** 413–419 (2017).

45. Wallis, C. J. *et al.* Comparison of postoperative outcomes among patients treated by male and female surgeons: a population based matched cohort study. *BMJ* **359,** j4366 (2017).

46. Sarsons, H. Interpreting Signals in the Labor Market: Evidence from Medical Referrals. (2017).

47. Jena, A. B., Khullar, D., Ho, O., Olenski, A. R. & Blumenthal, D. M. Sex Differences in Academic Rank in US Medical Schools in 2014. *JAMA* **314,** 1149–1158 (2015).

48. Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 16474–16479 (2012).

49. MacNell, L., Driscoll, A. & Hunt, A. N. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innov High Educ* **40,** 291–303 (2015).

50. Massen, J. J. M., Bauer, L., Spurny, B., Bugnyar, T. & Kret, M. E. Sharing of science is most likely among male scientists. *Sci. Rep.* **7,** 12927 (2017).

51. Lutter, M. & Schröder, M. Who becomes a tenured professor, and why? Panel data evidence from German sociology, 1980–2013. *Res. Policy* **45,** 999–1013 (2016).

52. Breda, T. & Hillion, M. Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science* **353,** 474–478 (2016).

53. Williams, W. M. & Ceci, S. J. National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 5360–5365 (2015).

54. Ceci, S. J. & Williams, W. M. Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men. *Front. Psychol.* **6,** 1532 (2015).

55. Leslie, L. M., Manchester, C. F. & Dahm, P. C. Why and When Does the Gender Gap Reverse? Diversity Goals and the Pay Premium for High Potential Women. *Acad. Manage. J.* **60,** 402–432 (2017).

56. Bohren, J. A., Imas, A. & Rosenberg, M. The Dynamics of Discrimination: Theory and Evidence. (2017). doi:10.2139/ssrn.3081873

57. Bornmann, L., Mutz, R. & Daniel, H.-D. Gender differences in grant peer review: A meta-analysis. *J. Informetr.* **1,** 226–238 (2007).

58. Tamblyn, R. *et al.* Health services and policy research in the first decade at the Canadian Institutes of Health Research. *CMAJ Open* **4,** E213–21 (2016).

59. Ley, T. J. & Hamilton, B. H. Sociology. The gender gap in NIH grant applications. *Science* **322,** 1472–1474 (2008).

60. Pohlhaus, J. R., Jiang, H., Wagner, R. M., Schaffer, W. T. & Pinn, V. W. Sex differences in application, success, and funding rates for NIH extramural programs. *Acad. Med.* **86,** 759–767 (2011).

61. Waisbren, S. E. *et al.* Gender differences in research grant applications and funding outcomes for medical school faculty. *J. Womens. Health* **17,** 207–214 (2008).

62. Bedi, G., Van Dam, N. T. & Munafo, M. Gender inequality in awarded research grants. *Lancet* **380,** 474 (2012).

63. Head, M. G., Fitchett, J. R., Cooke, M. K., Wurie, F. B. & Atun, R. Differences in research funding for women scientists: a systematic comparison of UK investments in global infectious disease research during 1997-2010. *BMJ Open* **3,** e003362 (2013).

64. Svider, P. F. *et al.* Gender differences in successful National Institutes of Health funding in ophthalmology. *J. Surg. Educ.* **71,** 680–688 (2014).

65. Mutz, R., Bornmann, L. & Daniel, H.-D. Does Gender Matter in Grant Peer Review? *Zeitschrift für Psychologie* **220,** 121–129 (2012).

66. Albers, C. J. Dutch research funding, gender bias, and Simpson's paradox. *Proceedings of the National Academy of Sciences* **112,** E6828–E6829 (2015).

67. Volker, B. & Steenbeek, W. No evidence that gender contributes to personal research funding success in The Netherlands: A reaction to van der Lee and Ellemers. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E7036–7 (2015).

68. Ceci, S. J. & Williams, W. M. Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences* **108,** 3157–3162 (2011).

69. Gino, F., Wilmuth, C. A. & Brooks, A. W. Compared to men, women view professional advancement as equally attainable, but less desirable. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 12354–12359 (2015).

70. Bargas-Avila, J. A., Brenzikofer, O., Tuch, A. N., Roth, S. P. & Opwis, K. Working towards Usable Forms on the World Wide Web: Optimizing Date Entry Input Fields. *Advances in Human-Computer Interaction* **2011,** (2011).

71. Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R*. (Springer, 2009).

72. Hubbard, A. E. *et al.* To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* **21,** 467–474 (2010).

73. R Development Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2017).

74. Halekoh, U., Højsgaard, S., Yan, J. & Others. The R package geepack for generalized estimating equations. *J. Stat. Softw.* **15,** 1–11 (2006).

75. Lenth, R. V. & Others. Least-squares means: the R package lsmeans. *J. Stat. Softw.* **69,** 1–33 (2016).

76. Hewitt, T. Open Letter to University Presidents from the Canada Research Chairs Program. (2016). Available at: http://www.chairs-chaires.gc.ca/whats_new-quoi_de_neuf/2016/letter-lettre-eng.aspx. (Accessed: 7th July 2017)

77. van der Lee, R. & Ellemers, N. Gender contributes to personal research funding success in The Netherlands. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 12349–12353 (2015).

78. Malmström, M., Johansson, J. & Wincent, J. Gender Stereotypes and Venture Support Decisions: How Governmental Venture Capitalists Socially Construct Entrepreneurs' Potential. *Entrepreneurship Theory and Practice* (2017). doi:10.1111/etap.12275
79. Hyde, J. S. & Mertz, J. E. Gender, culture, and mathematics performance. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8801–8807 (2009).
80. Goldin, C. & Rouse, C. Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians. *Am. Econ. Rev.* **90,** 715–741 (2000).
81. CIHR's Gender Equity Framework. (2017). Available at: http://www.cihr-irsc.gc.ca/e/50238.html. (Accessed: 2nd August 2017)
82. Tricco, A. C. *et al.* Strategies to Prevent or Reduce Gender Bias in Peer Review of Research Grants: A Rapid Scoping Review. *PLoS One* **12,** e0169718 (2017).
83. Unconscious Bias in Peer Review. Available at: http://www.cihr-irsc.gc.ca/lms/e/bias/. (Accessed: 2nd August 2017)
84. Moss-Racusin, C. A. *et al.* A 'Scientific Diversity' Intervention to Reduce Gender Bias in a Sample of Life Scientists. *CBE Life Sci. Educ.* **15,** (2016).
85. Balafoutas, L. & Sutter, M. Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science* **335,** 579–582 (2012).
86. Besley, T., Folke, O., Persson, T. & Rickne, J. Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden∗ .
87. Hyll, W. Gender Quotas and Human Capital Formation: A Relative Deprivation Approach. *Ger. Econ. Rev.* **18,** 302–326 (2017).
88. Salzberg, A. J. Removable Selection Bias in Quasi-Experiments. *Am. Stat.* **53,** 103–107 (1999).
89. West, J. D., Jacquet, J., King, M. M., Correll, S. J. & Bergstrom, C. T. The role of gender in scholarly authorship. *PLoS One* **8,** e66212 (2013).

90. Meerwijk, E. L. & Sevelius, J. M. Transgender Population Size in the United States: a Meta-Regression of Population-Based Probability Samples. *Am. J. Public Health* **107,** e1–e8 (2017).

APPENDIX 1. Methodological and Analytical Details

Methods

We analysed data from applications submitted to the Canadian Institutes of Health Research (CIHR) for competitions in investigator-initiated grant programs in fiscal years 2011/12 through 2015/16. We fit a logistic regression model to the data, with application success as the binary outcome of interest, as a function of the following predictors: self-reported binary sex (male or female) and age of the principal investigator (nominated principal applicant or "applicant"), the self-declared primary domain of research of the application within one of four categories (biomedical; clinical; health systems and services; or social, cultural, environmental and population health),[1] and the grant program group to which the application was submitted. Grant programs were grouped into three categories: traditional investigator-initiated grant programs programs (includes regular open operating grant programs from fiscal years 2011/12 to 2014/15 which account for 88% of grants in the Traditional programs group plus 6 smaller programs, some of which continued into fiscal year 2015/16); Project (spring 2015 competition, fiscal year 2015/16); and Foundation (two competitions: 2014/15 and 2015/16). All predictors were categorical variables, except for applicant age, which was continuous and was mean-centered prior to analysis.

We did not expect the success of applications from the same applicant to be independent of each other. We therefore used Generalized Estimating Equations (GEE) to fit the logistic model, with an exchangeable working correlation structure within applicants to account for the lack of independence of these applications[2,3] using the geepack package to fit models.[4] We then used the fitted model to test the pairwise effect of sex within each program, using the lsmeans package in R.[5] This allowed us to compute marginal effects for specific contrasts of interest, averaged over other terms in the model.

The binary response in the model was application success, which is true (1) if the application was approved after the peer review process, and false (0) if not. Because our aim was to analyze the effects of peer review, we coded applications that were deemed fundable but not approved in the competition to which they were applied as unsuccessful, even if they were later awarded money through other administrative processes such as bridge grants or priority announcements for specific funding areas.

Data

Our dataset was a full export of CIHR competition data from its Electronic Information System (EIS). This dataset does not include withdrawn applications. It only includes applications submitted that were fully assessed by peer review and either approved or not.

Results

Within our sample, male applicants applying to traditional open programs in the biomedical domain experienced the highest success rates. There was a small increase in the odds of success with age. Applications in non-biomedical domains had lower odds of success. Female applicants experienced significantly lower success rates than male applicants in Foundation, but

not in Project nor in traditional programs. This was confirmed by using the model coefficients to compute contrasts and associated odds ratios for sexes within each program.

Tables S1 and Table S2 provide different ways of viewing these results. Table S1 shows the raw GEE results while Table S2 shows the observed success rates, predicted probabilities, and calculated odds ratios for the interaction between applicant sex and program at uniform values of age, averaged across domains.[5] Table S3 shows results by review stage for the cycles included in the quasi-experiment and thus in the analyses in the paper. Table S4 shows results by review stage for the following cycle after changes were made to the program, specifically, after a reviewer learning module on the topic of unconscious bias was implemented in the program.

TABLES

Table S1. Odds of Grant Success: GEE Results

	Odds Ratio (OR)	OR lower bound of 95% confidence interval	OR upper bound of 95% confidence interval
(Intercept): Male applicants in Program: traditional programs, Research Domain: Biomedical, Mean age	0.229	0.216	0.242
Female applicants	0.934	0.854	1.022
Program: Project	0.762	0.675	0.860
Program: Foundation	0.748	0.641	0.873
Age (mean-centered)	1.005	1.001	1.010
Research Domain: Clinical	0.815	0.738	0.900
Research Domain: Health Systems and Services	0.846	0.747	0.959
Research Domain: Population and Public Health	0.772	0.681	0.877
Sex * Program: Female applicants in Project	0.988	0.794	1.229
Sex * Program: Female applicants in Foundation	0.705	0.518	0.960

Table S2. Associations between Predictors and Funding Success

	Number of applications in data set	Percent successful (unadjusted)	Predicted probabilities* (after adjusting for age and research domain)	Odds Ratio (OR)*	OR lower bound of 95% confidence interval*	OR upper bound of 95% confidence interval*
Applicant Sex x Program Interaction						
Male applicants in traditional programs (reference)	11879	17.3%	16.3%	1.000		
Female applicants in traditional programs	6326	15.7%	15.4%	0.934	0.854	1.022
Male applicants in Project	2469	13.5%	12.9%	0.762	0.675	0.860
Female applicants in Project	1119	12.0%	12.1%	0.703	0.587	0.841
Male applicants in Foundation	1427	13.9%	12.7%	0.748	0.641	0.873
Female applicants in Foundation	698	9.2%	8.8%	0.493	0.375	0.647
Research Domain						
Biomedical (reference)	14159	16.9%	n/a	1.000		
Clinical	4497	14.3%	n/a	0.815	0.738	0.900
Health Systems and Services	2609	14.9%	n/a	0.846	0.747	0.959

Population and Public Health	2653	13.6%	n/a	0.772	0.681	0.877
Age						
Age (mean-centered)	23918	15.8%	n/a	1.005	1.001	1.010

*Predicted probabilities and odds ratios (including 95% confidence intervals) calculated with lsmeans package.

Table S3. Foundation results by review stage during study

Cycle	Stage	Male applicants: n (% of previous stages)	Female applicants: n (% of previous stages)
2014/15	Applied at Stage 1	850	490
Review focus: caliber of applicant	Invited to Stage 2 and applied	317 (317/850 = 37**%**)	126 (126/490 = 26**%**)
Review focus: quality of research	Funded	109 (109/317 = 34%, 109/850 = 13**%**)	41 (41/126 = 33%, 41/490 = 8**%**)
2015/16	Applied at Stage 1	623	279
Review focus: caliber of applicant	Invited to Stage 2 and applied	198 (198/623 = 32%)	62 (62/279 = 22%)
Review focus: quality of research	Funded	91 (91/198 = 46%, 91/623 = 15%)	29 (29/62 = 47%, 29/279 = 10%)

A subset of applications also underwent additional Stage 3 review after Stage 2, again with a focus on the quality of the research, before being funded or rejected. There were no differences by applicant's self-reported sex.

Table S4. Foundation results by review stage after implementation of unconscious bias reviewer training module

Cycle	Stage	Male applicants: n (% of previous stages)	Female applicants: n (% of previous stages)
2016/17	Applied at Stage 1	428	172
Review focus: caliber of applicant	Invited to Stage 2 and applied	165 (165/428 = 39%)	69 (69/172 = 40%)
Review focus: quality of research	Funded	54 (54/165 = 33%, 54/428 = 13%)	22 (22/69 = 32%, 22/172 = 13%)

REFERENCES

1 The Four Themes of CIHR Funded Health Research - CIHR. 2014; published online Nov 24. http://www.cihr-irsc.gc.ca/e/48801.html (accessed July 8, 2017).

2 Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM. Mixed Effects Models and Extensions in Ecology with R. Springer, 2009.

3 Hubbard AE, Ahern J, Fleischer NL, *et al.* To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* 2010; **21**: 467–74.

4 Halekoh U, Højsgaard S, Yan J, Others. The R package geepack for generalized estimating equations. *J Stat Softw* 2006; **15**: 1–11.

5 Lenth RV, Others. Least-squares means: the R package lsmeans. *J Stat Softw* 2016; **69**: 1–33.