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

Holly O. Witteman, $\mathrm{PhD}^{1}$; Michael Hendricks, $\mathrm{PhD}^{2}$; 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<br>${ }^{2}$ Assistant Professor, Department of Biology, Faculty of Science, 1205 av du Docteur-Penfield, McGill University, Montreal, Canada, H3A 1B1<br>${ }^{3}$ Professor, Li Ka Shing Knowledge Institute of St. Michael's Hospital, 30 Bond Street, Toronto, Canada, M3B 2T9<br>${ }^{4}$ Scientific Director, Institute for Gender and Health, Canadian Institutes of Health Research, Ottawa, Canada, H3A 1W4

Corresponding author: 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: 2691


#### 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 different grant programs had differing success rates among male and female applicants.


Methods: We analyzed results of 23,918 grant applications from 7,093 unique applicants in a 5year natural experiment across all open, investigator-initiated CIHR grant programs in 20112016. Our primary outcome was whether or not each grant application was successful. 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 the grant program group to determine the extent to which gender gaps in health research grant funding differed by grant program. 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 (OR $0.934,95 \% \mathrm{Cl} 0.854-1.022$ ). In the new program in which review focused on the quality of the proposed science, the gap was $0.9 \%$ and not different from traditional programs (OR $0.998,95 \% \mathrm{Cl} 0.794-1.229$ ). In the new program with an explicit review focus on the caliber of the principal investigator, the gap was $4.0 \%$ (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, supporting the hypothesis that gender gaps in grant funding are partly or wholly attributable to women being assessed less favourably as principal investigators.

## INTRODUCTION

For decades, studies have shown that women in academia, science, and medicine 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 labour for less credit on publications, ${ }^{19,20}$ receive less compelling letters of recommendation, ${ }^{21-24}$ are expected to do more service work, ${ }^{25,26}$ receive systematically lower teaching evaluations despite no differences in teaching effectiveness, ${ }^{27}$ and are more likely to experience harassment. ${ }^{11,28-30}$ While men in academia have more successful careers after taking parental leave, women's careers suffer after the same. ${ }^{31}$ Women receive less startup funding as biomedical scientists ${ }^{32}$ and are underrepresented in invitations to referee papers. ${ }^{33}$ Compared to publications led by men, those led by women take longer to publish ${ }^{34}$ and are cited less often, ${ }^{35,36}$ even when published in higher-impact journals. ${ }^{37}$ Women's papers ${ }^{38}$ and conference abstracts ${ }^{39}$ are accepted more frequently when reviewers are blinded to the identities of the authors. Women are underrepresented as invited speakers at major conferences ${ }^{4}$ or presenters at grand rounds, ${ }^{40,41}$ and when women are invited to give these prestigious talks, they are less likely to be introduced with their formal title of Doctor. ${ }^{42}$ Female surgeons have been shown to have better patient outcomes overall, ${ }^{43}$ 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. ${ }^{44}$ Women are less likely to reach higher ranks in medical schools even after accounting for age, experience, specialty, and measures of research productivity. ${ }^{45}$ When fictitious or real people are presented as women in randomized experiments, they receive lower ratings of competence from scientists, ${ }^{46}$ worse teaching evaluations from students, ${ }^{47}$ 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 metaanalysis. ${ }^{48}$ In sum, there is considerable evidence that women face persistent, pervasive barriers in academia, science, and medicine.

In light of this evidence, we consider the question: does gender bias influence research funding? A 2007 meta-analysis of 21 studies from a range of countries found an overall gender gap, with $7 \%$ higher odds of fellowship or grant funding for male applicants. ${ }^{49}$ Research since
then has documented that, compared to their male colleagues, female principal investigators have lower grant success rates, ${ }^{50}$ lower grant success rates in some but not all programs, ${ }^{51,52}$ equivalent grant success rates after adjusting for academic rank ${ }^{53,54}$ but fewer funds requested and received, ${ }^{53-56}$ or equivalent funding rates. ${ }^{57}$ 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 of gender gaps in grant funding have been observational, making it difficult to draw robust conclusions about the causes of gaps when they are observed. Furthermore, some previous studies have not accounted for potential confounding variables; for example, domain of research. ${ }^{49,58,59}$

In this study, we aimed to determine whether or not gender gaps in grant funding are attributable to women being evaluated less favourably than men as principal investigators by analyzing data from a natural experiment at a national health research funding agency.

## 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 continuing on 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 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 being 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 selfreported 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. ${ }^{60}$ 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, ${ }^{61,62}$ including principal investigators who applied to both Project and Foundation programs. We conducted analyses in R statistical computing software, version 3.4.0, ${ }^{63}$ using the geepack package to fit models. ${ }^{64}$ We then used the fitted model to test the pairwise effect of sex within each program, using the Ismeans package. ${ }^{65}$ 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' selfreported 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. ${ }^{6,66}$ 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 5year 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 $0.998,95 \% \mathrm{Cl} 0.794-1.229$ ) and $4.0 \%$ in the Foundation program (OR $0.705,95 \% \mathrm{CI} 0.519-$ 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 hypothesis 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 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. ${ }^{67}$ In Sweden, similar biases have been shown among evaluators' assessments of applicants for governmental venture capital. ${ }^{68}$ 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. ${ }^{69}$ Similarly, women became more successful in orchestra auditions when auditioning musicians' identities were concealed behind a screen. ${ }^{70}$

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, ${ }^{71}$ 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 ${ }^{72}$ reviewer training module about multiple forms of unconscious bias. ${ }^{73}$ Training has been previously shown to help mitigate the effects of bias. ${ }^{15,74}$ 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. ${ }^{75-77}$ 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, ${ }^{78}$ 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. ${ }^{58,59}$ 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. ${ }^{79}$ 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 previous research has shown that female funding applicants received systematically lower scores compared to male applicants with equivalent publication records. ${ }^{2}$ 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. ${ }^{80}$ 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, and consider reviewer training and other policies to mitigate the effects of all forms of bias. Future research should investigate other potential sources of bias and evaluate methods of reducing bias 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/tcps2eptc2/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 fundinganalytics@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 VicePresident, for her comments on drafts 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 MW. The Matthew Matilda Effect in Science. Soc Stud Sci 1993; 23: 325-41.
2 Wenneras C, Wold A. Nepotism and sexism in peer-review. Nature 1997; 387: 341-3.
3 Reuben E, Sapienza P, Zingales L. How stereotypes impair women's careers in science. Proc Natl Acad Sci U S A 2014; 111: 4403-8.

4 Klein RS, Voskuhl R, Segal BM, et al. Speaking out about gender imbalance in invited speakers improves diversity. Nat Immunol 2017; 18: 475-8.

5 Settles IH, Cortina LM, Malley J, Stewart AJ. The Climate for Women in Academia: the Good, the Bad, and the Changeable. Psychol Women Q 2006; 30: 47-58.

6 Ginther DK, Schaffer WT, Schnell J, et al. Race, ethnicity, and NIH research awards. Science 2011; 333: 1015-9.

7 Williams J, Phillips KW, Hall EV. 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 2015; 347: 262-5.

9 Milkman KL, 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 2015; 100: 1678-712.

10 Henry F, Dua E, Kobayashi A, et al. Race, racialization and Indigeneity in Canadian universities. Race Ethnicity and Education 2017; 20: 300-14.

11 Clancy KBH, Lee KMN, Rodgers EM, Richey C. Double jeopardy in astronomy and planetary science: Women of color face greater risks of gendered and racial harassment. $J$ Geophys Res Planets 2017; 122: 2017JE005256.

12 Elmore KC, Luna-Lucero M. Light Bulbs or Seeds? How Metaphors for Ideas Influence Judgments About Genius. Soc Psychol Personal Sci 2016; published online Oct 7. 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 2017; 355: 389-91.

14 Magua W, Zhu X, Bhattacharya A, 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 2017; 26: 560-70.

15 Carnes M, Devine PG, Baier Manwell L, 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 2015; 90: 221-30.

16 Smyth FL, Nosek BA. On the gender-science stereotypes held by scientists: explicit accord with gender-ratios, implicit accord with scientific identity. Front Psychol 2015; 6: 415.

17 Carli LL, Alawa L, Lee Y, Zhao B, Kim E. Stereotypes about gender and science: Women $\neq$ scientists. Psychol Women Q 2016; 40: 244-60.

18 Kerkhoven AH, Russo P, Land-Zandstra AM, Saxena A, Rodenburg FJ. Gender Stereotypes in Science Education Resources: A Visual Content Analysis. PLoS One 2016; 11: e0165037.

19 Macaluso B, Larivière V, Sugimoto T, Sugimoto CR. Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. Acad Med 2016; 91: 1136-42.

20 Feldon DF, Peugh J, Maher MA, Roksa J, Tofel-Grehl C. Time-to-Credit Gender Inequities of First-Year PhD Students in the Biological Sciences. CBE Life Sci Educ 2017; 16. DOI:10.1187/cbe.16-08-0237.

21 Trix F, Psenka C. Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty. Discourse \& Society 2003; 14: 191-220.

22 Schmader T, Whitehead J, Wysocki VH. A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants. Sex Roles 2007; 57: 509-14.

23 Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. J Appl Psychol 2009; 94: 1591-9.

24 Dutt K, Pfaff DL, Bernstein AF, Dillard JS, Block CJ. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. Nat Geosci 2016; 9: 805-8.

25 Babcock L, Recalde MP, Vesterlund L, Weingart L. Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability. Am Econ Rev 2017; 107: 714-47.

26 Guarino CM, Borden VMH. Faculty Service Loads and Gender: Are Women Taking Care of the Academic Family? Res High Educ 2017; 58: 672-94.

27 Stark P, Ottoboni K, Boring A, Cetinkaya-Rundel M. Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. 2016; published online Jan 7. DOI:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.

28 Berdahl JL. The sexual harassment of uppity women. J Appl Psychol 2007; 92: 425-37.
29 Jagsi R, Griffith KA, Jones R, Perumalswami CR, Ubel P, Stewart A. Sexual Harassment and Discrimination Experiences of Academic Medical Faculty. JAMA 2016; 315: 2120-1.

30 Nelson RG, Rutherford JN, Hinde K, Clancy KBH. Signaling Safety: Characterizing Fieldwork Experiences and Their Implications for Career Trajectories. Am Anthropol 2017; 119: 710-22.

31 Antecol H, Bedard K, Stearns J. Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies? IZA Discussion Papers, 2016 http://ftp.iza.org/dp9904.pdf.

32 Sege R, Nykiel-Bub L, Selk S. Sex Differences in Institutional Support for Junior Biomedical Researchers. JAMA 2015; 314: 1175-7.

33 Lerback J, Hanson B. Journals invite too few women to referee. Nature 2017; 541: 455-7.
34 Hengel E. Publishing while Female: Gender Differences in Peer Review Scrutiny. 2016; published online Oct 1. http://www.erinhengel.com/research/publishing_female.pdf.

35 Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: global gender disparities in science. Nature 2013; 504: 211-3.

36 Caplar N, Tacchella S, Birrer S. Quantitative evaluation of gender bias in astronomical publications from citation counts. Nature Astronomy 2017; 1: 0141.

37 Ghiasi G, Larivière V, Sugimoto CR. On the Compliance of Women Engineers with a Gendered Scientific System. PLoS One 2015; 10: e0145931.

38 Budden AE, Tregenza T, Aarssen LW, Koricheva J, Leimu R, Lortie CJ. Double-blind review favours increased representation of female authors. Trends Ecol Evol 2008; 23: 4-6.

39 Roberts SG, Verhoef T. Double-blind reviewing at EvoLang 11 reveals gender bias. Journal of Language Evolution 2016; 1: 163-7.

40 Boiko JR, Anderson AJM, Gordon RA. Representation of Women Among Academic Grand Rounds Speakers. JAMA Intern Med 2017; 177: 722-4.

41 Buell D, Hemmelgarn B, Straus SE. Proportion of women presenters at medical grand rounds at major academic centres in Canada, a retrospective observational study (in press). BMJ Open.

42 Files JA, Mayer AP, Ko MG, et al. Speaker Introductions at Internal Medicine Grand Rounds: Forms of Address Reveal Gender Bias. J Womens Health 2017; 26: 413-9.

43 Wallis CJ, Ravi B, Coburn N, Nam RK, Detsky AS, Satkunasivam R. Comparison of postoperative outcomes among patients treated by male and female surgeons: a population based matched cohort study. BMJ 2017; 359: j4366.

44 Sarsons H. Interpreting Signals in the Labor Market: Evidence from Medical Referrals. 2017. https://scholar.harvard.edu/files/sarsons/files/sarsons_jmp.pdf.

45 Jena AB, Khullar D, Ho O, Olenski AR, Blumenthal DM. Sex Differences in Academic Rank in US Medical Schools in 2014. JAMA 2015; 314: 1149-58.

46 Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. Science faculty's subtle gender biases favor male students. Proc Natl Acad Sci U S A 2012; 109: 16474-9.

47 MacNell L, Driscoll A, Hunt AN. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. Innov High Educ 2015; 40: 291-303.

48 Massen JJM, Bauer L, Spurny B, Bugnyar T, Kret ME. Sharing of science is most likely among male scientists. Sci Rep 2017; 7: 12927.

49 Bornmann L, Mutz R, Daniel H-D. Gender differences in grant peer review: A metaanalysis. J Informetr 2007; 1: 226-38.

50 Tamblyn R, McMahon M, Girard N, Drake E, Nadigel J, Gaudreau K. Health services and policy research in the first decade at the Canadian Institutes of Health Research. CMAJ Open 2016; 4: E213-21.

51 Ley TJ, Hamilton BH. Sociology. The gender gap in NIH grant applications. Science 2008; 322: 1472-4.

52 Pohlhaus JR, Jiang H, Wagner RM, Schaffer WT, Pinn VW. Sex differences in application, success, and funding rates for NIH extramural programs. Acad Med 2011; 86: 759-67.

53 Waisbren SE, Bowles H, Hasan T, et al. Gender differences in research grant applications and funding outcomes for medical school faculty. J Womens Health 2008; 17: 207-14.

54 Bedi G, Van Dam NT, Munafo M. Gender inequality in awarded research grants. Lancet 2012; 380: 474.

55 Head MG, Fitchett JR, Cooke MK, Wurie FB, 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 2013; 3: e003362.

56 Svider PF, D'Aguillo CM, White PE, et al. Gender differences in successful National Institutes of Health funding in ophthalmology. J Surg Educ 2014; 71: 680-8.

57 Mutz R, Bornmann L, Daniel H-D. Does Gender Matter in Grant Peer Review? Zeitschrift für Psychologie 2012; 220: 121-9.

58 Albers CJ. Dutch research funding, gender bias, and Simpson's paradox. Proceedings of the National Academy of Sciences 2015; 112: E6828-9.

59 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 2015; 112: E7036-7.

60 Bargas-Avila JA, Brenzikofer O, Tuch AN, Roth SP, Opwis K. Working towards Usable Forms on the World Wide Web: Optimizing Date Entry Input Fields. Advances in HumanComputer Interaction 2011; 2011. DOI:10.1155/2011/202701.

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

62 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.

63 R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2017 https://www.R-project.org/.

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

65 Lenth RV, Others. Least-squares means: the R package Ismeans. J Stat Softw 2016; 69: 1-33.

66 Hewitt T. Open Letter to University Presidents from the Canada Research Chairs Program. 2016; published online April 28. http://www.chairs-chaires.gc.ca/whats_new-quoi_de_neuf/2016/letter-lettre-eng.aspx (accessed July 7, 2017).

67 van der Lee R, Ellemers N. Gender contributes to personal research funding success in The Netherlands. Proc Natl Acad Sci U S A 2015; 112: 12349-53.

68 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; published online Feb 1. DOI:10.1111/etap. 12275.

69 Hyde JS, Mertz JE. Gender, culture, and mathematics performance. Proc Natl Acad Sci U S A 2009; 106: 8801-7.

70 Goldin C, Rouse C. Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians. Am Econ Rev 2000; 90: 715-41.

71 CIHR's Gender Equity Framework. 2017; published online Feb 10. http://www.cihrirsc.gc.ca/e/50238.html (accessed Aug 2, 2017).

72 Tricco AC, Thomas SM, Antony J, et al. Strategies to Prevent or Reduce Gender Bias in Peer Review of Research Grants: A Rapid Scoping Review. PLoS One 2017; 12: e0169718.

73 Unconscious Bias in Peer Review. http://www.cihr-irsc.gc.ca/lms/e/bias/ (accessed Aug 2, 2017).

74 Moss-Racusin CA, van der Toorn J, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. A 'Scientific Diversity' Intervention to Reduce Gender Bias in a Sample of Life Scientists. CBE Life Sci Educ 2016; 15. DOI:10.1187/cbe.15-09-0187.

75 Balafoutas L, Sutter M. Affirmative action policies promote women and do not harm efficiency in the laboratory. Science 2012; 335: 579-82.

76 Besley T, Folke O, Persson T, Rickne J. Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden*. http://perseus.iies.su.se/~tpers/papers/Draft170122AER.pdf.

77 Hyll W. Gender Quotas and Human Capital Formation: A Relative Deprivation Approach. Ger Econ Rev 2017; 18: 302-26.

78 Salzberg AJ. Removable Selection Bias in Quasi-Experiments. Am Stat 1999; 53: 103-7.
79 West JD, Jacquet J, King MM, Correll SJ, Bergstrom CT. The role of gender in scholarly authorship. PLoS One 2013; 8: e66212.

80 Meerwijk EL, Sevelius JM. Transgender Population Size in the United States: a MetaRegression of Population-Based Probability Samples. Am J Public Health 2017; 107: e1-8.

## 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 Ismeans 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 <br> Ratio <br> (OR) | OR lower bound <br> of 95\% <br> confidence <br> interval | OR upper <br> bound of 95\% <br> confidence <br> interval |
| :--- | :--- | ---: | ---: |
| (Intercept): Male applicants in Program: <br> traditional programs, Research Domain: <br> 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 <br> Research Domain: Health Systems and <br> Services 0.846 0.747 0.959 <br> Research Domain: Population and Public <br> Health 0.772 0.681 0.877 <br> Sex * Program: Female applicants in Project 0.988 0.794 1.229 <br> Sex * Program: Female applicants in <br> Foundation 0.705 0.518 0.960 |  |  |  |

Table S2. Associations between Predictors and Funding Success

|  | Number of applications in data set | $\begin{array}{r} \text { Percent } \\ \text { successful } \\ \text { (unadjusted) } \end{array}$ | Predicted probabilities* (after adjusting for age and research domain) | Odds <br> Ratio <br> (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 <br> and Public <br> Health 2653 $13.6 \%$ $\mathrm{n} / \mathrm{a}$ 0.772 0.681 0.877 <br> Age       <br> Age (mean- <br> centered) 23918 $15.8 \%$ $\mathrm{n} / \mathrm{a}$ 1.005 1.001 1.010*Predicted probabilities and odds ratios (including 95\% confidence intervals) calculated with |
| :--- |

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 | $\begin{array}{r} 317 \\ (317 / 850=37 \%) \end{array}$ | $\begin{array}{r} 126 \\ (126 / 490=26 \%) \end{array}$ |
| Review focus: quality of research | Funded | $\begin{array}{r} 109 \\ (109 / 317=34 \%, \\ 109 / 850=13 \%) \end{array}$ | $\begin{array}{r} 41 \\ (41 / 126=33 \%, \\ 41 / 490=8 \%) \end{array}$ |
| 2015/16 | Applied at Stage 1 | 623 | 279 |
| Review focus: caliber of applicant | Invited to Stage 2 and applied | $\begin{array}{r} 198 \\ (198 / 623=32 \%) \end{array}$ | $\begin{array}{r} 62 \\ (62 / 279=22 \%) \end{array}$ |
| Review focus: quality of research | Funded | $\begin{array}{r} 91 \\ (91 / 198=46 \%, \\ 91 / 623=15 \%) \end{array}$ | $\begin{array}{r} 29 \\ (29 / 62=47 \%, \\ 29 / 279=10 \%) \end{array}$ |

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: <br> $\mathbf{n}(\%$ of previous <br> stages $)$ | Female applicants: <br> $\mathbf{n}(\%$ of previous <br> stages $)$ |
| :--- | :--- | ---: | ---: |
| $\mathbf{2 0 1 6 / 1 7}$ | Applied at Stage 1 | 428 | 172 |
| Review focus: caliber <br> of applicant | Invited to Stage 2 <br> and applied | 165 | 69 |
| Review focus: quality <br> of research | Funded | $(165 / 428=39 \%)$ | $(69 / 172=40 \%)$ |

## 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, leno 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 Ismeans. J Stat Softw 2016; 69: 1-33.

