Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency

Holly O Witteman, Michael Hendricks, Sharon Straus, Cara Tannenbaum

Summary

Background **Across countries and disciplines, studies show male researchers receive more research funding than their female peers. Because most studies have been observational, it is unclear whether imbalances stem from evaluations of female research investigators or of their proposed research. In 2014, the Canadian Institutes of Health Research created a natural experiment by dividing investigator-initiated funding applications into two new grant programmes: one with and one without an explicit review focus on the calibre of the principal investigator.**

Methods **We analysed application success among 23 918 grant applications from 7093 principal investigators in all investigator-initiated Canadian Institutes of Health Research grant programmes between 2011 and 2016. We used generalised estimating equations to account for multiple applications by the same applicant and compared differences in application success between male and female principal investigators under different review criteria.**

Findings **Overall application success across competitions was 15·8%. After adjusting for age and research domain, the predicted probability of success in traditional programmes was 0·9 percentage points lower for female applicants than male applicants (95% CI 2·0 lower–0·2 higher; odds ratio 0·934, 95% CI 0·854–1·022). In the new programme, in which review focused on the proposed science, the gap remained 0·9 percentage points (3·2 lower–1·4 higher; 0·998, 0·794–1·229). In the new programme with an explicit review focus on the calibre of the principal investigator, the gap was 4·0 percentage points (6·7 lower–1·3 lower; 0·705, 0·519–0·960).**

Interpretation **Gender gaps in grant funding are attributable to less favourable assessments of women as principal investigators, not of the quality of their proposed research. We discuss reasons less favourable assessments might occur and strategies to foster fair and rigorous peer review.**

Funding **None.**

Lancet 2019; 393: 531–40

See **Comment** page 494

Department of Family and Emergency Medicine, Faculty of Medicine, Université Laval, Québec City, QC, Canada (H O Witteman PhD); **Department of Biology, Faculty of Science, McGill University, Montreal, QC, Canada** (M Hendricks PhD); **Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, ON, Canada** (Prof S Straus MD); **and Institute for Gender and Health, Canadian Institutes of Health Research, Ottawa, ON, Canada** (Prof C Tannenbaum MD)

Correspondence to: Dr Holly O Witteman, Department of Family and Emergency Medicine, Faculty of Medicine, Université Laval, Québec City, QC G1V 0A6, Canada **holly.witteman@fmed.ulaval.ca**

Introduction

For more than two decades, studies have shown that women in academia must perform to a higher standard than men to receive equivalent recognition,[1–3] especially Indigenous and racialised women.[4–7] Compared with men, women are more often characterised as lacking the brilliance necessary for discovery,[6] and are less likely to be viewed as scientific leaders.[8–10] Women in academia contribute more labour for less credit on publications,[11,12] receive less compelling letters of recommendation,[13] receive systematically lower teaching evaluations despite no differences in teaching effectiveness,[14] are more likely to experience harassment,[15,16] do more service work,[17] and are expected to do special favours for students.[18] After taking parental leave, men in academia have more successful careers whereas women's careers suffer.[19] Women receive less start-up funding as biomedical scientists[20] and are under-represented in invitations to referee papers.[21] Compared with publications led by men, those led by women take longer to publish[22] and are cited less often,[23] even when published in high-impact journals.[24] Articles[25] and conference abstracts[26] led by women are accepted more frequently when reviewers are unaware of authors' identities. Women are under-represented as invited speakers at conferences,[3] universities,[27] and grand rounds.[28] When women are invited to give these prestigious talks, they are less likely to be introduced with their formal title of Doctor.[29] Female faculty are less likely to reach higher ranks in medical schools than male faculty, even after accounting for age, experience, specialty, and research productivity.[30] When fictitious or real people are presented as women in randomised experiments, they receive lower ratings of competence from scientists,[31] worse teaching evaluations from students,[32] and fewer email responses from professors after presenting as students seeking a PhD adviser[33] or as scientists seeking copies of an article or data for a meta-analysis.[34]

Conversely, other research has shown advantages experienced by women in academia; for example, achieving tenure with fewer publications than men.[35] In assessments of potential secondary and post-secondary teachers and professors, women are favoured in male-dominated fields, as are men in female-dominated fields.[36] When fictitious people are presented as women in randomised experiments, they receive higher rankings as potential science faculty than men.[37,38] This aligns with evidence from other contexts showing that high-potential women are favoured over high-potential men[39] and

Research in context

Evidence before this study

We searched MEDLINE and Google Scholar with no date restrictions, using search terms that included combinations of MeSH terms, such as "Peer Review, Research" and "Women", and unstructured terms, such as "funding" and "sex or gender". We also manually searched references of key articles. We found that funding agencies around the world show gender gaps in grant success, with women typically receiving less funding than men. An international meta-analysis of 21 studies published in 2007, showed that, overall, men had 7% higher odds of funding success than women. Because previous research has been observational, and some studies did not account for confounders, such as field of research, it has been difficult to draw robust conclusions about the reasons for such discrepancies in funding.

Added value of this study

Our study offers the first quasiexperimental evidence regarding whether gender gaps in research funding stem from evaluations of the science or of the scientist. Analysing data from a large, natural experiment in which we adjusted for key confounders of age and research domain, we found that sex-specific funding success was not significantly different in a new grant programme in which peer review focused on the proposed science, whereas another new grant programme focused on the scientist showed a significantly larger sex-related difference that disadvantaged women.

Implications of all the available evidence

Our study provides empirical evidence that gender gaps in grant funding stem from women being evaluated less favourably as principal investigators, not from differences in the quality of proposals led by men and women. Women might be evaluated less favourably because of conscious or unconscious bias on the part of reviewers, systemic bias in the form of review criteria reflecting cumulative disadvantage, or differences in women's effort or descriptions of accomplishments. Future research should advance knowledge by evaluating possible reasons for discrepancies in evaluations of principal investigators, controlling for possible confounders, and including other potential sources of bias beyond gender. To ensure the best research is funded, funders should ensure the design and execution of their grant programmes do not reproduce or exacerbate biases.

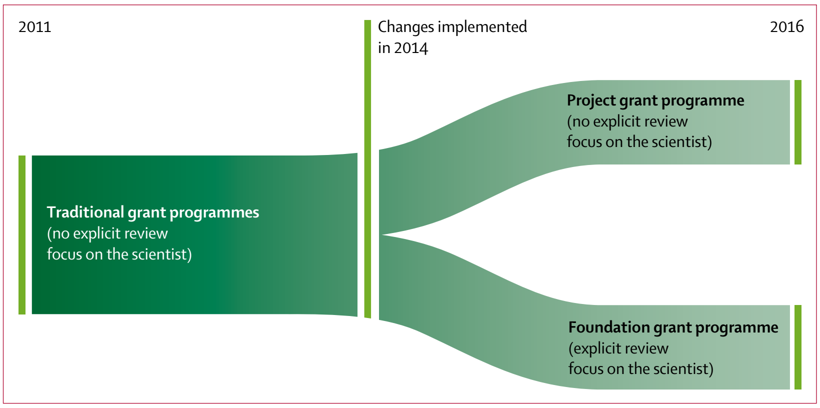

Figure 1: **Changes in Canadian Institutes of Health Research grant programmes**

that, although women face discrimination at earlier stages, once women have proven themselves in a male-dominated context, they are often favoured over men.[40]

Considering this evidence and the crucial importance of obtaining grant funding for academic career success, we consider the question: what causes gender gaps in research grant funding? Previous research has shown such gaps occur, but not why. A 2007 international meta-analysis of 21 studies found an overall gender gap in favour of men, with 7% higher odds of fellowship or grant funding for male applicants than female applicants.[41] Research has since shown that, compared with their male colleagues, female principal investigators have less grant success,[42] less success in some but not all programmes,[43,44] equivalent success after adjusting for academic rank[45,46] but fewer funds requested and received,[45–48] less representation among first grant awardees but equivalent longevity,[49] or equivalent funding rates.[50] These previous studies were observational, making it difficult to draw robust conclusions about whether discrepancies in funding success rates are attributable to evaluations of female researchers or of their proposed research. Furthermore, some previous studies did not account for potential confounding variables like domain of research.[41,51]

Our study objective was therefore to establish whether gender gaps in grant funding are attributable to differences in evaluations of male and female researchers or of their proposed research, 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 federal health research funding agency, enabling the comparison of grant success among male and female applicants between three grant programmes: traditional grant programmes, which had shown younger male principal investigators to have the highest percentages of success in 2001–11,[42] but for which no subsequent analysis had yet been done; and two new competitions, one with and one without an explicit review focus on the calibre of the principal investigator.

Our hypotheses regarding comparisons of gaps in grant success between female and male principal investigators after accounting for age and research domain were as follows: H_0, gaps will be similar under traditional review criteria and both sets of new review criteria; H_1, the gap will be larger in favour of male principal investigators in the new competition with more focus on the proposed science; and H_2, the gap will be larger in favour of male principal investigators in the new competition with more focus on the scientist. Support for H_0 would suggest that any gaps might reflect different career choices made by women and men,[52,53] differences between the types of research proposed by female and male principal investigators, equally poorer evaluations of female researchers and their proposed research, or might be spurious. Support for H_1 would suggest that gaps are partly or wholly driven by female principal investigators writing less compelling grants or otherwise proposing science assessed as lower quality than that of their male colleagues. Support for H_2 would suggest that gaps are partly or wholly driven by women being assessed less favourably as principal investigators compared with their male colleagues. Other potential alternative hypotheses such as gaps in favour of female applicants were not considered because publicly available programme statistics showed such results to be impossible.

Methods

Population, setting, and data source

Beginning in 2014, the Canadian Institutes of Health Research (CIHR) phased out traditional open grant programmes, dividing investigator-initiated funding into two new programmes: Project (originally planned to use 55% of available funding for investigator-initiated grants) and Foundation (45% of funding; figure 1). Both programmes used a staged review process in which proposals ranked below a threshold were rejected from continuing to further stages. In the Project programme, the same application was assessed at the first and second stages and, like traditional programmes, reviewers were instructed to assess primarily the research proposed, with 25% of the score representing reviewers' assessments of the applicants' expertise, experience, and resources. In contrast, Foundation was about people, not projects. Reviewers of applications at the first stage were instructed to assess primarily the principal investigator, with 75% of the score representing reviewers' assessments of the applicant's leadership, productivity, and the significance of their contributions. Applicants who were successful at this stage (about one-third) were then invited to submit a detailed research proposal (table 1).

During the years included in this study (2011–16), new investigators within the first 5 years of their independent career and those who had never held CIHR funding could apply to programmes of their choice. Principal investigators who already held CIHR funding were eligible for the Foundation programme if they held an active CIHR grant scheduled to end within a specific date range. Otherwise, they could only apply to the Project programme. This meant that a portion of allocation was dependent on an arbitrary variable. Principal investigators could apply to multiple programmes, with some restrictions. In the first cycle of the Foundation programme, applicants 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 programmes. In the second cycle of the Foundation programme, applicants could apply to both Foundation and Project, providing they did not submit the same research proposal to both programmes. If both grants were awarded, the applicant would need to choose which to accept.

Data were held internally and analysed by staff at the CIHR within their mandate as a national funding agency, respecting national guidelines on research ethics.[54] Our study had the objective of evaluating CIHR's investigator-initiated programmes, and was not therefore within the scope of research ethics board review in Canada. Nevertheless, applicants were informed 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.

Statistical analysis

We analysed data from all applications submitted to CIHR grant programmes across all investigator-initiated competitions in 2011 through 2016. We excluded

	Foundation (stage 1)*	Project (all stages)*
Application form		
Summary	3500 characters, or about one page†	3500 characters, or about one page
Main text	Leadership (1750 characters, or about half a page) Significance of contributions (1750 characters, or about half a page) Productivity (1750 characters, or about half a page) Vision and programme direction (3500 characters, or about one page)	Quality of the idea (1750 characters, or about half a page) Importance of the idea (3500 characters, or about one page) Approach (15 750 characters, or about four and a half pages) Expertise, experience, and resources (3500 characters, or about one page)
References	7000 characters, or about two pages	7000 characters, or about two pages
Budget	Expected to reflect annual Canadian Institutes of Health Research open funding held over the past 5–7 years, but requests for increases permitted	No official limit, but the top 2% of grants by total budget were evaluated separately and funded out of a more competitive large grant envelope
Appendices	None (no figures, no tables, no papers, no letters of support)	Up to two pages of figures (no tables, no papers, no letters of support)
Statement of most significant contributions	3500 characters, or about one page	3500 characters, or about one page
Career contributions (a one-page electronic form accepting numbers only)	Publications Individuals trained or mentored Intellectual property	None

(Table 1 continues on next page)

	Foundation (stage 1)*	Project (all stages)*
(Continued from previous page)		
CV elements available to reviewers in Canadian Common CV (CCV)‡		
Identification	Name, degrees, credentials, employment, leaves of absence	Name, degrees, credentials, employment, leaves of absence
Funding history	All funding active within the past 7 years, as principal investigator or other role	All funding active within the past 5 years, as principal investigator or other role
Recognitions (eg, awards)	Up to five (cycle 1) or ten (cycle 2)§	Up to five
Publications	Up to 25 (cycle 1) or all publications from the past 7 years¶ (cycle 2)	Up to ten
Intellectual property	Up to ten	Up to five
Presentations	Up to ten (cycle 1) or 25 (cycle 2)	Up to five
Supervisory activities (eg, trainees)	Up to 25 (cycle 1) or all supervisory activities within the past 7 years (cycle 2)	Up to ten
Knowledge and technology translation activities	Up to ten	Up to five
Information about other research team members available to reviewers		
Coprincipal investigators	Permitted but only shared previous work considered as evidence of leadership, contributions, and productivity (no individual work from either principal investigator considered)	Permitted with shared and individual previous work considered as evidence of expertise and experience
Team members	Not listed at stage 1	Listed with brief CCVs (five publications, no grants) and statement of most significant contributions (3500 characters, about one page) from each co-applicant
Adjudication criteria		
Criterion 1 (25% of score)	"Leadership: Is the applicant recognized in their field, demonstrating a history of holding influential roles, inspiring others, mobilizing communities and advancing the direction of a field? Has the applicant demonstrated the ability to successfully establish, resource, and direct programs of research, which should include: securing the required resources, ensuring effective collaboration, and/or incorporating knowledge translation strategies?"	"Quality of the Idea: Are the overall goal and objectives of the project well-defined and clear; with distinct outputs that support advances in health-related knowledge, health research, health care, health systems, and/or health outcomes? Is the rationale of the project idea sound, logical and valid?"
Criterion 2 (25% of score)	"Significance of Contributions: Has the applicant significantly advanced knowledge and/or its translation into improved health care, health systems, and/or health outcomes? Has the applicant engaged, trained, and/or launched the career paths of promising individuals in research and/or other health-related non-academic fields?"	"Importance of the Idea: Is the proposed contribution(s) of the project well-defined, clear, and significant with respect to advancing health-related knowledge, health research, health care, health systems, and/or health outcomes? To what extent does the project respond to the objective(s) of the Funding Opportunity?"
Criterion 3 (25% of score)	"Productivity: Has the applicant demonstrated an outstanding level (quantity) of research outputs in their field based on prior work? Has the applicant's previous work generated high quality research outputs?"	"Approach: Are the approaches and methods appropriate to deliver the proposed output(s) and achieve the proposed contribution(s)? Are the timelines and related deliverables of the project realistic? Does the proposal identify potential challenges and appropriate mitigation strategies?"
Criterion 4 (25% of score)	"Vision and Program Direction: Are the vision, goal, overall objective(s), and potential contribution(s) of the proposed research program well-defined and well-articulated in the context of a logical career progression for the Program Leader(s)? Is the vision forward-looking, creative, and appropriately ambitious? Does the vision aim to significantly advance knowledge and/or its translation to improved health care, health systems, and/or health outcomes?"	"Expertise, Experience, and Resources: Does the applicant(s) bring the appropriate expertise and experience to lead and deliver the proposed output(s), and to achieve the proposed contribution(s)? Is there an appropriate level of engagement and/or commitment from the applicant(s)? Is the environment (academic institution and/or other organization) appropriate to enable the conduct and success of the project?"

*Although Project used the same application and review criteria at all stages, Foundation stage 2 involved a second application and new criteria. Because only a minority of applicants were invited to Foundation stage 2 and the observed effects reported in this paper reflect differences at stage 1 (figure 3), this table focuses on stage 1 application and review criteria. Full details of Foundation stage 2 are available in the appendix. †Applications were submitted via structured web forms. Each section had limits on total character count, including spaces. ‡Full details of CCV elements are available in the appendix. The CCV is a structured form generated through an online database. §The Foundation application form and CCV changed slightly from cycle 1 (2014–15 competition) to cycle 2 (2015–16 competition). ¶No additional publications, grants, or other activities were permitted to be listed if the applicant had taken a leave (eg, a typical Canadian 1-year maternity and parental leave) during that time. Applicants were permitted to list each instance of leave and explain its career impact in a 900-character (approximately 180 words) section in the CCV.

Table 1: **Foundation grant stage 1 and project grant application and adjudication**

withdrawn applications, because these did not receive full peer review. We also excluded applications if the principal investigator had not reported their sex, birth date, domain of research of their application, or if their self-reported birth year was before 1920 (the default when people do not enter a birth date) or after 2000 (the result when people erroneously enter the current year as their birth year).

We coded application funding success as true (1) if the application was approved after the peer review process, and false (0) if not. Because our aim was to analyse the effects of peer review, we coded applications

	All programmes in dataset		Traditional programmes		Project		Foundation	
	Male applicants	Female applicants	Male applicants	Female applicants	Male applicants	Female applicants	Male applicants	Female applicants
Total	2592/15 775 (16·4%)	1189/8143 (14·6%)	2059/11 879 (17·3%)	991/6326 (15·7%)	334/2469 (13·5%)	134/1119 (12·0%)	199/1427 (13·9%)	64/698 (9·2%)
Research domain of funded and submitted applications								
Biomedical	1882/10 891 (17·3%)	505/3268 (15·5%)	1464/8066 (18·2%)	416/2481 (16·8%)	274/1876 (14·6%)	60/529 (11·3%)	144/949 (15·2%)	29/258 (11·2%)
Clinical	388/2639 (14·7%)	254/1858 (13·7%)	331/2030 (16·3%)	208/1412 (14·7%)	37/356 (10·4%)	34/281 (12·1%)	20/253 (7·9%)	12/165 (7·3%)
Health systems and services	158/1059 (14·9%)	232/1550 (15·0%)	126/858 (14·7%)	202/1268 (15·9%)	16/111 (14·4%)	17/157 (10·8%)	16/90 (17·8%)	13/125 (10·4%)
Population and public health	164/1186 (13·8%)	198/1467 (13·5%)	138/925 (14·9%)	165/1165 (14·1%)	7/126 (5·6%)	23/152 (15·1%)	19/135 (14·1%)	10/150 (6·7%)
Age of applicants, years								
Applications submitted	52 (9)	50 (9)	52 (9)	51 (8)	50 (9)	48 (8)	49 (9)	47 (9)
Applications funded	52 (9)	50 (8)	52 (9)	51 (8)	51 (9)	48 (8)	53 (9)	50 (9)

Data are n/N (%) or mean (SD).

Table 2: **Application characteristics**

that were deemed fundable but not initially approved as unsuccessful, even if they were later awarded funds through bridge grants or priority announcements for specific funding areas.

We used generalised estimating equations to fit a logistic model that accounted for applicants submitting multiple applications during these 5 years.[55] We modelled grant success as a function of grant programme, applicants' self-reported binary sex, self-reported age, self-declared domain of research, and an interaction term between applicant sex and grant programme. All predictors were categorical variables, except for applicant age, which was continuous, and mean-centred before analysis to facilitate interpretation. Model coefficients were used to compute contrasts and associated odds ratios (ORs) for success by sex within each programme.

The interaction term allowed us to address the objective of this study by establishing whether there was any effect of different review criteria on relative percentages of success between male and female applicants after controlling for age and domain of research. We controlled for these criteria because younger cohorts of investigators include larger proportions of women than men, as do domains of health research other than biomedical. Adjusting for age also helped to account for the Foundation and Project programmes having predefined minimal allocations to new investigators. 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. At the time of this study, CIHR did not routinely collect data on other applicant characteristics associated with disparities in funding and career progression; for example, career stage, race, ethnicity, indigeneity, and disability.

We did analyses in R statistical computing software, version 3.4.0, using the geepack package to fit models.[56] We then used the fitted model to test the pairwise effect of sex within each programme, using the lsmeans package.[57] This allowed us to compute marginal effects for specific contrasts of interest. We then estimated 95% CIs around differences in predicted probabilities as Wald intervals.

Additional methods and results, and R code used are given in the appendix.

See Online for appendix

Results

Out of 25 706 applications during the 5 years of this study, we excluded 1788 consisting primarily of principal investigators with birth years before 1920 (n=1631). The final dataset contained 23 918 applications from 7093 unique applicants (63% male applicants, 37% female applicants). Summary statistics by programme, sex, age, and research domain are shown in table 2.

The overall grant success across the dataset was 15·8%. Female applicants experienced significantly lower percentages of success than male applicants in the Foundation programme, but not in Project nor in traditional programmes. After adjusting for age and research domain, the predicted probability of funding was 0·9 percentage points lower for female applicants than for male applicants in traditional grant programmes (95% CI 2·0 lower–0·2 higher; OR 0·934, 95% CI 0·854–1·022), 0·9 percentage points lower in the Project programme (3·2 lower–1·4 higher; 0·998, 0·794–1·229), and 4·0 percentage points lower in the Foundation programme (6·7 lower–1·3 lower; 0·705, 0·519–0·960; figure 2). As shown in figure 3, the gap in Foundation was driven by discrepancies at the first review stage, in which review focused primarily on the

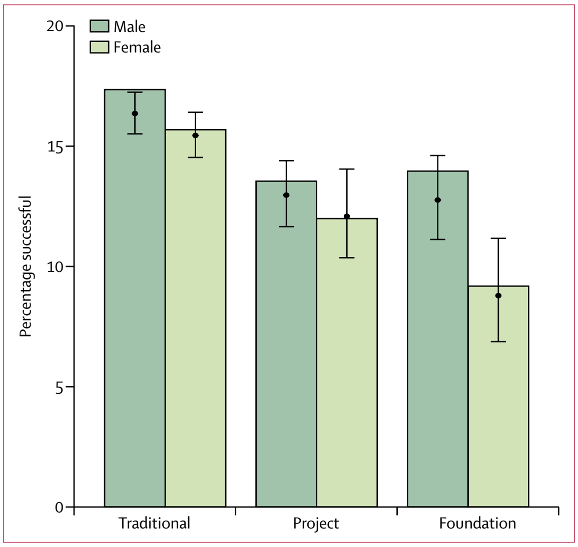

Figure 2: **Funding success by grant programme**
Points indicate model-predicted means and error bars indicate 95% CIs.

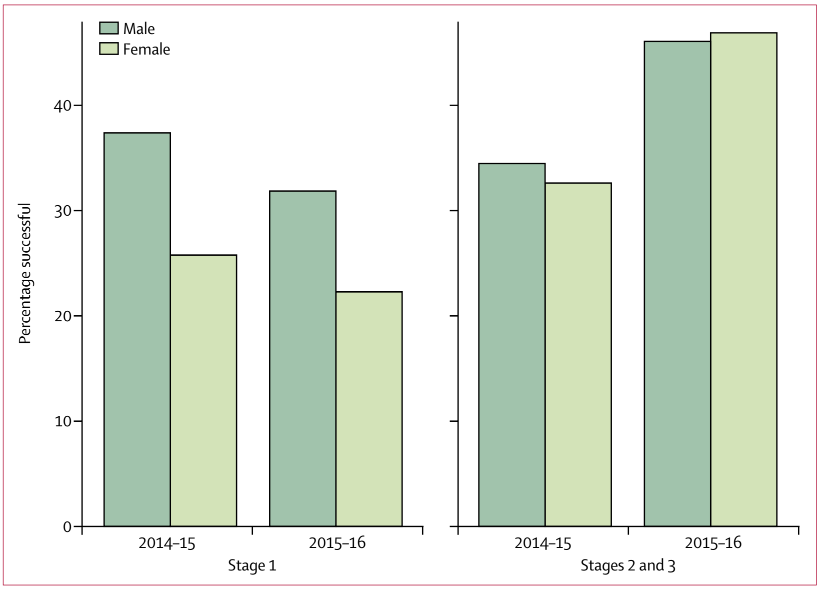

Figure 3: **Foundation programme results by review stage**
The review focus in stage 1 was the calibre of the researcher, and in stages 2 and 3 was quality of research.

principal investigator. Odds of funding success were lower in the Project and Foundation programmes than in traditional programmes, in non-biomedical domains than biomedical domains, and for younger applicants than older applicants (table 3).

Discussion

Our study provides stronger evidence than was previously available regarding why gender gaps occur in grant funding. In this natural experiment, when reviewers primarily assessed an applicant's proposed science, no statistically significant differences existed between percentages of success 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 second of our alternative hypotheses; namely, that gender gaps in funding stem from female principal investigators being evaluated less favourably than male principal investigators, not from differences in evaluations of the quality of their proposals. This finding aligns with previous studies in other countries. In the USA, female grant applicants to the National Institutes of Health's R01 programme were less likely than male applicants to be described as leaders,[58] and review criteria including subjective assessments of leadership favoured men in a prestigious award.[59,60] In the Netherlands, grant reviewers gave equal scores to men's and women's proposed research, but assigned lower scores to women as researchers.[61]

We identified three main explanations for why female principal investigators might be evaluated less favourably than male principal investigators: individual bias, systemic bias, and lower performance. Individual bias refers to reviewers' subjective evaluations of principal investigators reflecting conscious or unconscious gender bias. Previous randomised experiments have used hypothetical scenarios in which participants were asked to predict how they would respond in a given situation. These showed that scientists favoured men as laboratory managers[31] and men as junior faculty in an older experiment,[62] but women as junior faculty in recent experiments.[37,38] Other randomised experiments used actual behaviours as study outcomes. These experiments all favoured men,[32,34] particularly white men.[33] In real-world observational studies, reviewers assessed equal productivity less positively for female versus male fellowship applicants[1] and assigned lower grant scores to female versus male applicants even after accounting for more than 20 potential confounders, including applicants' h-indices (indicating publication success) and history of funding success.[63] Thus, evidence of scientists favouring women comes exclusively from hypothetical scenarios, whereas evidence of scientists favouring men comes from hypothetical scenarios and real behaviour. This might reflect academics' growing awareness of the social desirability of achieving gender balance, while real academic behaviour might not yet put such ideals into action.

Systemic bias refers to grant programme design, such as review criteria that unfairly favours male principal investigators because of cumulative advantage. Research shows that women's manuscripts[25] and conference abstracts[26] were more often accepted when reviewers were unaware of authors' identities, suggesting that productivity criteria might unfairly disadvantage women. Compared with female academics, male academics have received more start-up and grant funding,[20,41] had articles accepted more often and more quickly,[22] had lighter service loads,[17] been less likely to be charged the productivity tax[64] of dealing with sexual harassment, and been more frequently invited to speak at conferences,[3] universities,[27] and grand rounds,[28] selected for awards,[65] and perceived as leaders.[8–10] These discrepancies have made it easier for the average male researcher to build his CV compared with his average female peer. Specific to our study, male principal investigators might have had stronger CVs because of historically higher CIHR grant funding yielding more data for them to publish.[42] Designing a grant programme to include leadership as a criterion might especially introduce systemic bias. Women receive negative responses for being ambitious[66] and must often choose between being seen as a competent leader or being liked.[67,68] This complicates career advancement within academia because advancement to leadership often depends on collegial relationships and others' endorsements.

Lower performance refers to female principal investigators submitting weaker applications than male peers because of lower ability or effort in the criteria assessed or less compelling descriptions of criteria-specific performance in the application. Research does not support the hypothesis that women have inherently lower ability in male-dominated fields;[69] however, it is possible that although women have equivalent ability to answer research questions, they might face challenges or dissuasion in securing institutional resources, putting themselves forward for awards, or pursuing leadership roles without all listed qualifications. Given that trainees publish more when supervised by female principal investigators,[70] female principal investigators might also devote more effort to mentoring trainees at cost to their own CVs. Female Foundation applicants in our study might also have unconsciously adopted modest, less compelling language often used when describing women's accomplishments.[13] Women have been shown to be less likely to boast about their intellectual performance.[2]

In summary, our study provides empirical evidence from a natural experiment that gender gaps in grant funding stem from women being evaluated less favourably as principal investigators, not from differences in the quality of proposals led by men and women. The natural experiment does not allow for estimation of the contributions of three possible sources—individual bias, systemic bias, or lower performance—to these lower evaluations. As detailed earlier, previous research suggests all three explanations are plausible, with more literature supporting the first two. If the first explanation is true, effective solutions might include funders instituting reviewer anti-bias training, blind review, or automatic adjustments to correct for known biases. If the second explanation is true, effective solutions require broader action. Funders could use more equitable review criteria or automatic adjustments, journals and societies could ensure equitable publication and awards processes, and institutions could provide all faculty with equitable resources and support; for example, funding for pilot studies, protected time for research, mentorship, sponsorship, and leadership training. If the third explanation is true, female principal investigators might need to focus more on research outputs to ensure that funders' objectives are maximally attained, secure additional resources, seek leadership opportunities, and describe their accomplishments in compelling language.

	Odds ratio (95% CI)	Predicted probability (95% CI)
Applicant sex by programme		
Male applicants in traditional programmes (ref)	1·000	16·3% (15·5–17·2)
Female applicants in traditional programmes	0·934 (0·854–1·022)	15·4% (14·5–16·4)
Male applicants in Project	0·762 (0·675–0·860)	12·9% (11·6–14·4)
Female applicants in Project	0·703 (0·587–0·841)	12·1% (10·3–14·0)
Male applicants in Foundation	0·748 (0·641–0·873)	12·7% (11·1–14·6)
Female applicants in Foundation	0·493 (0·375–0·647)	8·8% (6·9–11·2)
Research domain		
Biomedical (ref)	1·000	NA
Clinical	0·815 (0·738–0·900)	NA
Health systems and services	0·846 (0·747–0·959)	NA
Population and public health	0·772 (0·681–0·877)	NA
Age		
Age (mean-centred)	1·005 (1·001–1·010)	NA

Predicted probabilities and odds ratios (95% CIs) calculated with lsmeans package. Predicted probabilities are adjusted for age and research domain.

Table 3: **Associations between predictors and funding success**

Our study had three main limitations. First, principal investigators were not randomly assigned to one grant programme or the other, meaning unobserved confounders or selection bias might have contributed to observed differences. Because we analysed the interaction between applicant sex and grant programme, not a main effect, for a variable to be a confounder, it must vary between male and female applicants and its influence must vary between differing evaluation criteria. The most probable potential unobserved confounders are therefore previous funding and publication records; however, even after controlling for these, research shows female applicants still received lower grant scores.[63] Other potential confounders include leadership roles and protected time for research (versus clinical, teaching, administrative, or other responsibilities). We would expect these to be highly correlated with funding and publication records and also subject to gender bias in their distribution. Other variables might also influence grant success rates; for example, race, ethnicity, institutional affiliation, coapplicants, whether the principal investigator is a clinician scientist, clinical specialty, and more fine-grained detail about domains of research.

Without data to examine whether these variables vary significantly between male and female health researchers in Canada, we cannot rule out the possibility that these might be confounders. Second, because of smaller numbers of senior female health researchers, our planned analyses did not include analyses of three-way interactions between grant programme group, self-reported sex, and age. It is possible that the two-way interaction observed might be driven by differences in evaluations of men and women who are more advanced in their careers. Third, we assumed that people who self-report as female or male also identify as women or men, respectively. If Canadian statistics are similar to US statistics[71] and if all transgender or non-binary applicants reported their assigned sex at birth, this would represent 28 applicants, a number unlikely to meaningfully change results of our analyses.

Our study also has four main strengths. First, it was quasiexperimental and thus provides evidence from a study design that was not fully observational. This design enables stronger conclusions than were previously possible. Second, although quasiexperimental studies have potential for selection bias, in this study, selection bias was limited by grant programme eligibility rules. Third, we controlled for age and domain of research, both of which are key confounders in studies of gender bias in grant funding. Fourth, our study analysed all available data over 5 years from a major national funding agency, thus offering robust evidence from a large dataset of real-world grant review.

Bias in grant review, whether individual or systemic, prevents the best research from being funded. When this occurs, lines of research go unstudied, careers are damaged, individual rights and potential go unrealised, and funding agencies are unable to deliver the best value for money both within funding cycles and longer term as small differences compound into cumulative disadvantage. Programmes that fund people, not projects, might reproduce common societal biases. Funders, institutions, journals, societies, and individual researchers might all have roles to play to ensure that peer review is fair and rigorous. We recommend funders focus assessment on the science rather than the scientist, measure and report funding by relevant applicant characteristics, and develop policies to mitigate effects of all forms of bias. To further advance knowledge, future research should investigate why female principal investigators might be evaluated less favourably, forms of bias beyond gender, and the effectiveness of strategies for facilitating peer review that consistently supports the best research.

Contributors

HOW and MH conceptualised the study and contributed to study methods. HOW investigated the data, wrote the original draft, visualised data, and did project administration. HOW, MH, CT, and SS reviewed and edited the Article, and were responsible for the decision to submit the manuscript. Contributions are listed according to the CRediT taxonomy (docs.casrai.org/CRediT).

Declaration of interests

This work was unfunded. HOW holds grant funding from the CIHR as principal investigator of a Foundation grant and receives salary support from a Research Scholar Junior 2 Career Development Award from the Fonds de Recherche du Québec-Santé. HOW and MH are two of the three founding national cochairs of the Association of Canadian Early Career Health Researchers, an organisation that has published statements critical of aspects of the CIHR grant programme changes, including the Foundation grant programme. SS holds grant funding from the CIHR, including a Foundation grant, receives salary support from a Tier 1 Canada Research Chair in Knowledge Translation and Quality of Care, and received contract funding from the CIHR to lead the scoping review described in the appendix and to analyse applicant and reviewer survey responses. CT is a scientific institute director at the CIHR and is therefore partially employed by the CIHR. The views expressed in this Article are those of the authors and do not necessarily reflect those of the CIHR or the Government of Canada.

Data sharing

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

Acknowledgments

The authors gratefully acknowledge early work by Anne-Sophie Julien, statistician, who worked with HOW to outline potential analytical approaches, drafted preliminary R code for the proposed research question, and reviewed a draft of the Article. We thank Anne Strong, CIHR staff, for extracting the dataset, and many colleagues who commented online in response to previous preprint versions of this manuscript. We thank Jonathan A Whiteley, Rachelle E Desrochers, Alysha Croker, and Danika Goosney, all of whom were employed at the CIHR during this study. Danika Goosney also conceptualised the study. Jonathan A Whiteley, Rachelle E Desrochers, Alysha Croker, and Danika Goosney also contributed to the study methods. Jonathan A Whiteley did the formal data analysis at the CIHR, because the data used consisted of confidential applicant information that could not be shared externally. Jonathan A Whiteley and Rachelle E Desrochers also investigated the data. Jonathan A Whiteley, Rachelle E Desrochers, and Alysha Croker also curated data. Jonathan A Whiteley also wrote the original draft. Jonathan A Whiteley, Rachelle E Desrochers, Alysha Croker, and Danika Goosney also reviewed and edited the Article. Jonathan A Whiteley and Rachelle E Desrochers also visualised the data. Rachelle E Desrochers also did project administration. Contributions are listed according to the CRediT taxonomy (docs.casrai.org/CRediT). 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 coauthors on articles describing analyses of the agency's programmes. Without the active participation of these CIHR employees and their contributions, this Article would not have been possible.

References

1 Wenneras C, Wold A. Nepotism and sexism in peer-review. *Nature* 1997; **387:** 341–43.
2 Reuben E, Sapienza P, Zingales L. How stereotypes impair women's careers in science. *Proc Natl Acad Sci USA* 2014; **111:** 4403–08.
3 Klein RS, Voskuhl R, Segal BM, et al. Speaking out about gender imbalance in invited speakers improves diversity. *Nat Immunol* 2017; **18:** 475–78.
4 Ginther DK, Schaffer WT, Schnell J, et al. Race, ethnicity, and NIH research awards. *Science* 2011; **333:** 1015–19.
5 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.
6 Leslie S-J, Cimpian A, Meyer M, Freeland E. Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 2015; **347:** 262–65.
7 Henry F, Dua E, Kobayashi A, et al. Race, racialization and indigeneity in Canadian universities. *Race Ethn Ed* 2017; **20:** 300–14.

8 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.

9 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.

10 Carli LL, Alawa L, Lee Y, Zhao B, Kim E. Stereotypes about gender and science: women ≠ scientists. *Psychol Women Q* 2016; **40:** 244–60.

11 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.

12 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:** ar4.

13 Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. *J Appl Psychol* 2009; **94:** 1591–99.

14 Boring A, Ottoboni K, Stark P. Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research* 2016; published online Jan 7, 2017. DOI:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.

15 Jagsi R, Griffith KA, Jones R, Perumalswami CR, Ubel P, Stewart A. Sexual harassment and discrimination experiences of academic medical faculty. *JAMA* 2016; **315:** 2120–21.

16 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.

17 Guarino CM, Borden VMH. Faculty service loads and gender: are women taking care of the academic family? *Res High Educ* 2017; **58:** 672–94.

18 El-Alayli A, Hansen-Brown AA, Ceynar M. Dancing backwards in high heels: female professors experience more work demands and special favor requests, particularly from academically entitled students. *Sex Roles* 2018; **79:** 136–50.

19 Antecol H, Bedard K, Stearns J. Equal but inequitable: who benefits from gender-neutral tenure clock stopping policies? *Am Econ Rev* 2018; **108:** 2420–41.

20 Sege R, Nykiel-Bub L, Selk S. Sex differences in institutional support for junior biomedical researchers. *JAMA* 2015; **314:** 1175–77.

21 Lerback J, Hanson B. Journals invite too few women to referee. *Nature* 2017; **541:** 455–57.

22 Hengel E. Publishing while female: are women held to higher standards? Evidence from peer review. Cambridge Working Paper Economics 1753, Faculty of Economics, University of Cambridge 2017; published online Dec 4. https://www.repository.cam.ac.uk/handle/1810/270621.

23 Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: global gender disparities in science. *Nature* 2013; **504:** 211–13.

24 Ghiasi G, Larivière V, Sugimoto CR. On the compliance of women engineers with a gendered scientific system. *PLoS One* 2015; **10:** e0145931.

25 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.

26 Roberts SG, Verhoef T. Double-blind reviewing at EvoLang 11 reveals gender bias. *J Lang Evol* 2016; **1:** 163–67.

27 Nittrouer CL, Hebl MR, Ashburn-Nardo L, Trump-Steele RCE, Lane DM, Valian V. Gender disparities in colloquium speakers at top universities. *Proc Natl Acad Sci USA* 2018; **115:** 104–08.

28 Boiko JR, Anderson AJM, Gordon RA. Representation of women among academic grand rounds speakers. *JAMA Intern Med* 2017; **177:** 722–24.

29 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:** 413–19.

30 Carr PL, Raj A, Kaplan SE, Terrin N, Breeze JL, Freund KM. Gender differences in academic medicine: retention, rank, and leadership comparisons from the national faculty survey. *Acad Med* 2018; **93:** 1694–99.

31 Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci USA* 2012; **109:** 16474–79.

32 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.

33 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.

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

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

36 Breda T, Hillion M. Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science* 2016; **353:** 474–78.

37 Williams WM, Ceci SJ. National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proc Natl Acad Sci USA* 2015; **112:** 5360–65.

38 Ceci SJ, Williams WM. Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men. *Front Psychol* 2015; **6:** 1532.

39 Leslie LM, Manchester CF, Dahm PC. Why and when does the gender gap reverse? diversity goals and the pay premium for high potential women. *Acad Manage J* 2017; **60:** 402–32.

40 Bohren JA, Imas A, Rosenberg M. The dynamics of discrimination: theory and evidence. 2017; published online Nov 18. DOI:10.2139/ssrn.3081873 (accessed Jan 10, 2018).

41 Bornmann L, Mutz R, Daniel H-D. Gender differences in grant peer review: a meta-analysis. *J Informetr* 2007; **1:** 226–38.

42 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.

43 Ley TJ, Hamilton BH. Sociology. The gender gap in NIH grant applications. *Science* 2008; **322:** 1472–74.

44 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.

45 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.

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

47 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.

48 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:** 680–88.

49 Hechtman LA, Moore NP, Schulkey CE, et al. NIH funding longevity by gender. *Proc Natl Acad Sci USA* 2018; **115:** 7943–48.

50 Mutz R, Bornmann L, Daniel H-D. Does gender matter in grant peer review? *Zeitschrift für Psychologie* 2012; **220:** 121–29.

51 Albers CJ. Dutch research funding, gender bias, and Simpson's paradox. *Proc Natl Acad Sci USA* 2015; **112:** E6828–29.

52 Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. *Proc Natl Acad Sci USA* 2011; **108:** 3157–62.

53 Gino F, Wilmuth CA, Brooks AW. Compared to men, women view professional advancement as equally attainable, but less desirable. *Proc Natl Acad Sci USA* 2015; **112:** 12354–59.

54 Canadian Tri-Council Policy Statement 2: Ethical Conduct for Research Involving Humans. pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/default/ (accessed July 13, 2017).

55 Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM. Mixed effects models and extensions in ecology with R. Springer, 2009.

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

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

58 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.

59 Carnes M, Geller S, Fine E, Sheridan J, Handelsman J. NIH director's pioneer awards: could the selection process be biased against women? *J Womens Health* 2005; **14:** 684–91.

60 Barres BA. Does gender matter? *Nature* 2006; **442:** 133–36.

61 van der Lee R, Ellemers N. Gender contributes to personal research funding success in The Netherlands. *Proc Natl Acad Sci USA* 2015; **112:** 12349–53.

62 Steinpreis RE, Anders KA, Ritzke D. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study. *Sex Roles* 1999; **41:** 509–28.

63 Tamblyn R, Girard N, Qian CJ, Hanley J. Assessment of potential bias in research grant peer review in Canada. *CMAJ* 2018; **190:** E489–99.

64 Zepeda L. The harassment tax. *Science* 2018; **359:** 126.

65 Lincoln AE, Pincus S, Koster JB, Leboy PS. The matilda effect in science: awards and prizes in the US, 1990s and 2000s. *Soc Stud Sci* 2012; **42:** 307–20.

66 Okimoto TG, Brescoll VL. The price of power: power seeking and backlash against female politicians. *Pers Soc Psychol Bull* 2010; **36:** 923–36.

67 Eagly AH, Karau SJ. Role congruity theory of prejudice toward female leaders. *Psychol Rev* 2002; **109:** 573–98.

68 Rudman LA, Moss-Racusin CA, Phelan JE, Nauts S. Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *J Exp Soc Psychol* 2012; **48:** 165–79.

69 Hyde JS, Mertz JE. Gender, culture, and mathematics performance. *Proc Natl Acad Sci USA* 2009; **106:** 8801–07.

70 Pezzoni M, Mairesse J, Stephan P, Lane J. Gender and the publication output of graduate students: a case study. *PLoS One* 2016; **11:** e0145146.

71 Meerwijk EL, Sevelius JM. Transgender population size in the United States: a meta-regression of population-based probability samples. *Am J Public Health* 2017; **107:** e1–8.