

Research Conducted in Women Was Deemed More Impactful but Less Publishable than the Same Research Conducted in Men

Sohad Murrar, PhD,¹ Paula A. Johnson, MD, MPH,² You-Geon Lee, PhD,³ and Molly Carnes, MD, MS⁴⁻⁶

Abstract

Background: Female scientists, who are more likely than their male counterparts to study women and report findings by sex/gender, fare worse in the article peer review process. It is unknown whether the gender of research participants influences the recommendation to publish an article describing the study.

Materials and Methods: Reviewers were randomly assigned to evaluate one of three versions of an article abstract describing a clinical study conducted in men, women, or individuals. Reviewers assessed the study's scientific rigor, its level of contribution to medical science, and whether they would recommend publishing the full article. Responses were analyzed with logistic regression controlling for reviewer background variables, including sex and experience level.

Results: There was no significant difference in perceived research rigor by abstract condition; contribution to medical science was perceived to be greater for research conducted in women than men (odds ratio=1.7; $p=0.030$). Nevertheless, reviewers were almost twice as likely to recommend publication for research conducted in men than the same research conducted in women (predicted probability 0.606 vs. 0.322; $p=0.000$).

Conclusions: These results are consistent with abundant data from multiple sources showing a lower societal value placed on women than men. Because female investigators are more likely than male investigators to study women, our findings suggest a previously unrecognized bias that could contribute to gender asymmetries in the publication outcomes of peer review. This pro-male publication bias could be an additional barrier to leadership attainment for women in academic medicine and the advancement of women's health.

Keywords: gender, equity, academic medicine, peer review

Introduction

GENDER IS A pervasive status cue that explicitly or implicitly places a lower value on women and female-gendered roles than on men and male-gendered roles.¹⁻⁶ Evidence of women's lower status throughout medicine includes among other things salary inequity⁷⁻¹⁴ with a strong negative correlation between physician salary and the percentage of women in the discipline¹⁵⁻¹⁷; underrepresentation in leadership (even in female-dominated fields)^{18,19}; less al-

location of research funding for ovarian, cervical, and uterine cancer than prostate cancer despite their greater lethality when standardized for incidence, mortality, and person-years of life lost²⁰; higher Medicare reimbursement rates for 42 out of 50 procedures performed in men than pair matched procedures performed in women²¹; generalization of research findings in men to all adults²²⁻²⁵; and disregard for clinical outcomes in women if they differ from those in male research participants.²⁶ Women, particularly women of color, are underrepresented among participants in clinical research,²⁷

¹Division of Psychology and Counseling, College of Education, Governors State University, University Park, Illinois, USA.

²Wellesley College, Wellesley, Massachusetts, USA.

³Wisconsin Center for Education Research, University of Wisconsin-Madison, Madison, Wisconsin, USA.

⁴Center for Women's Health Research, University of Wisconsin-Madison, Madison, Wisconsin, USA.

⁵Women in Science and Engineering Leadership Institute (WISELI), University of Wisconsin-Madison, Madison, Wisconsin, USA.

⁶Departments of Medicine, Psychiatry, and Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA.

and studies that do report sex/gender analyses are published in lower impact journals than those that do not.²⁸

Female scientists are more likely than their male counterparts to study women and to report findings by sex/gender.^{28,29} Women may fare worse than men in the article peer review process, the outcomes of which are a critical contributor to successful career advancement in academic medicine.^{28,30–33} Given the devaluation of women relative to men, we asked whether articles describing research conducted on women might be viewed as less valuable and less publishable than those describing research conducted on men. If this were the case, it would simultaneously repress dissemination of research findings relevant to the health of women, negatively affect female scientists' research productivity, and adversely affect attainment of future grant support.

This chain of events would impede the career advancement of women in academic medicine and contribute to their lower representation in senior ranks and leadership (Fig. 1).^{34,35} To answer this question, we investigated whether article reviewers' assessment of a study's scientific rigor, contribution to medical science, and recommendation to publish are influenced by the specified gender of the participants in the study.

Materials and Methods

Research design

We conducted an online randomized controlled experiment in which participants received a cover story *via* email about a new developing journal, *Interdisciplinary Journal of Health*. Participants came from a previously-developed database of all R01 grant awardees from 2010 to 2014 in which we manually retrieved information from NIH RePORTER public database using the method of Jaggi et al.³⁶ and Kaatz et al.³⁷ to assign sex and race/ethnicity. This involved searching the internet for pictures, biographical information, including country of origin, and text with pronouns.

The invitation said that the journal was piloting a new approach that was mindful of reviewers' time limitations and would thus evaluate whether reviewers could make assessments about an article based solely on its abstract. The review was said to be a double-blinded review process in which the reviewers and authors were unknown to each other. The invitation was sent to four groups of awardees separately with 2 days between them. Each group was randomly chosen from the overall database of 17,870 R01 awardees. There were three groups of 5,000 awardees and one group with 2,870

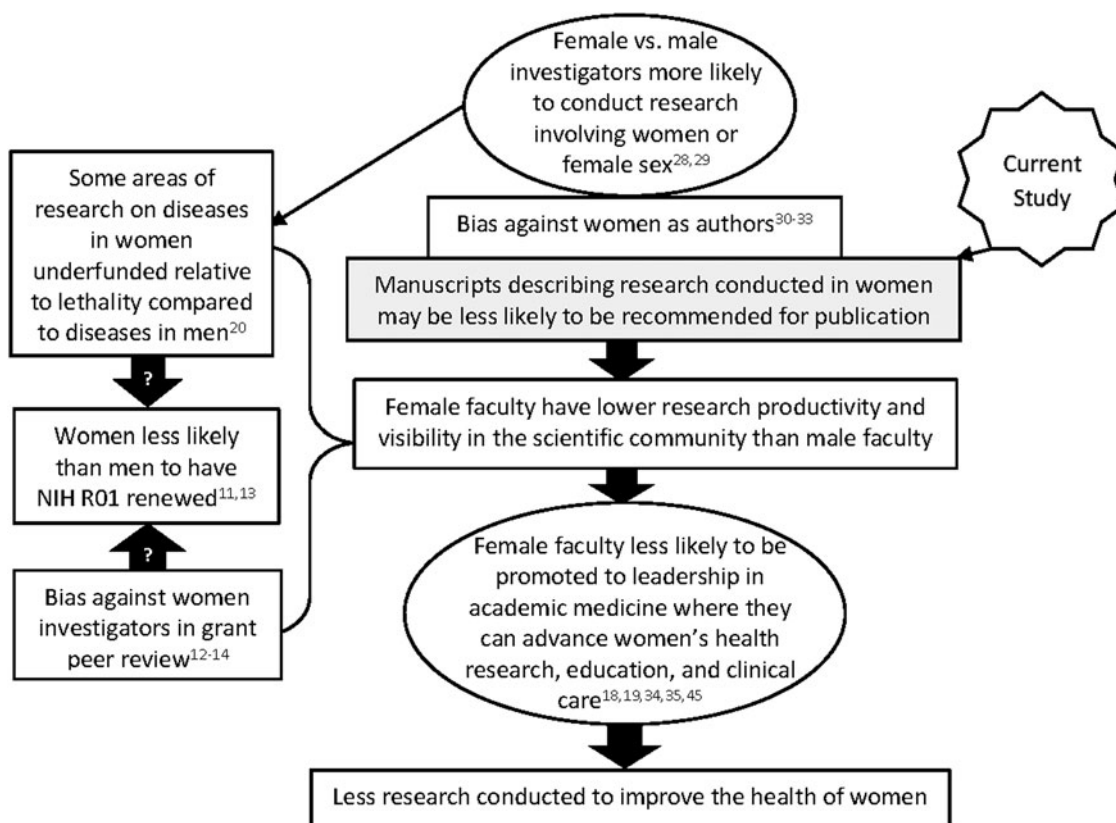


FIG. 1. Conceptual model of contributors to women's lower productivity and academic career advancement compared to their male counterparts that conspire to reduce research on women's health. Although each of these contributors may be small, women investigators are disadvantaged by the multiple ways their lower status synergistically affects their ability to engage in research. Women may be disadvantaged as investigators trying to secure research funding because of their own gender and/or because they may be studying a relatively underfunded women's health issue. Women may be disadvantaged in article peer review due to their own gender and because the research described in their articles was conducted on women. These biases result in fewer publications that may further impede success in garnering increasingly competitive research funding. The combination of fewer publications and less research funding reduce the likelihood that women will be promoted to senior faculty or leadership positions where they could have multiple positive effects on women's health, including advocating for more research funding for women's health issues.

awardees. The invitation included a link to the study for people to click and start the supposed review process. For all awardees with a valid email address who did not click on the link to participate in the review, a reminder email about the invitation was sent 1 week later.

Once participants clicked on the link to complete their review (*i.e.*, clicked into the study), they were randomly assigned to read one of three modifications of the same abstract taken from PubMed pertaining to hypothalamic-pituitary-adrenal (HPA) axis reactivity to intimacy in people with a history of sexual trauma.³⁸ The participants were randomized and equally distributed across the three conditions. The manipulated variable was the gender of the subjects in the study described in the abstract; the abstracts only differed in who the subjects of the study ostensibly were: individuals (control), men (male abstract), or women (female abstract). The participants then read their randomly-assigned abstract and completed a series of questions that encompassed three predetermined outcome measures: scientific rigor, contribution to medical science, and recommendation to publish. Upon completing the outcome measures, participants were debriefed and told that the study sought to evaluate gender bias in medical research. After the debriefing, reviewers were prompted to consent to participation in the study by submitting their data or to decline participation and retract their data by closing out of the window in which they had completed the review. Data were only recorded for participants who consented to participate in the study after the debriefing had taken place. The Institutional Review Board at the University of Wisconsin-Madison approved this study.

Sample

Of 17,870 emails sent to all R01 grant awardees from 2010 to 2014, 17,296 had valid email addresses that received the initial

invitation, although there is no way to verify how many of the emails were opened or read. A total of 358 recipients participated in the study ($358/17,296=0.021\%$). While the overall response rate was low, we were satisfied with the number of respondents. One explanation for the low response rate is that our email could have been mistaken for one of the many predatory journals that target our research population daily. Among them, we excluded 42 where participants did not complete the consent form after debriefing and 1 with no completed outcome measures. Of the final analytic sample of 315 participants, 34.3% were women ($N=108/315$), 66.7% were White ($N=210/315$), 27.9% were Asian ($N=88/315$), 77.5% held PhDs ($N=244/315$), and 64.4% were “experienced” investigators ($N=203/315$), meaning they had previously obtained an R01 or equivalent award at the time of application.

Table 1 summarizes participants’ demographics by the three abstract conditions. Our analytic sample included 36.8% of participants in the control condition, 33.7% in the male abstract condition, and 29.5% in the female abstract condition. Since participants were randomly assigned to one of the three abstract conditions, baseline characteristics were relatively well balanced across abstract conditions for participants’ sex, experience level, and citizenship (no statistically significant difference among conditions). However, significant imbalances were observed for participants’ race (white: 75.9% in control abstract vs. 58.1% in female abstract, $p<0.05$) and training background (PhDs: 70.7% in control abstract vs. 82.8% in female abstract, $p<0.05$). We accounted for this imbalance in the statistical analyses by including each baseline characteristic as a covariate.

Measures

The outcome measures of interest were article reviewers’ assessment of a study’s scientific rigor, contribution to

TABLE 1. DESCRIPTIVE STATISTICS OF PARTICIPANTS’ BACKGROUND BY ABSTRACT CONDITIONS

Variable	All		Abstract condition					
	N	%	Control		Male abstract		Female abstract	
			N	%	N	%	N	%
Total	315	100	116	36.8	106	33.7	93	29.5
Sex								
Male	207	65.7	75	64.7	69	65.1	63	67.7
Female	108	34.3	41	35.3	37	34.9	30	32.3
Race								
White	210	66.7	88	75.9	68	64.2	54	58.1
Asian	88	27.9	26	22.4	32	30.2	30	32.3
Other minorities	17	5.4	2	1.7	6	5.7	9	9.7
Experience level								
Experienced	203	64.4	80	69.0	68	64.2	55	59.1
New	112	35.6	36	31.0	38	35.8	38	40.9
Training background								
PhD	244	77.5	82	70.7	85	80.2	77	82.8
MD	33	10.5	14	12.1	11	10.4	8	8.6
MD/PhD	23	7.3	13	11.2	7	6.6	3	3.2
Others	15	4.8	7	6.0	3	2.8	5	5.4
Citizenship								
Non Citizen	132	41.9	45	38.8	45	42.5	42	45.2
U.S. Citizen	183	58.1	71	61.2	61	57.5	51	54.8

TABLE 2. DESCRIPTIVE STATISTICS FOR OUTCOME VARIABLES AND MODERN SEXISM SCALE BY ABSTRACT CONDITIONS

Variable	All		Abstract condition					
	Mean	SD	Control		Male abstract		Female abstract	
			Mean	SD	Mean	SD	Mean	SD
Scientific rigor	3.40	1.28	3.34	1.22	3.31	1.33	3.58	1.30
Perceived contribution	3.52	1.28	3.61	1.39	3.32	1.24	3.65	1.16
Publication recommendation	0.44	0.50	0.39	0.49	0.60	0.49	0.31	0.47
Modern sexism ^a	5.37	0.99	5.38	0.96	5.34	0.84	5.37	1.17

^aThe average score of 8 modern sexism items ranges from 1 to 7. SD, standard deviation.

medical science, and recommendation to publish (Table 2). Scientific rigor was measured on a seven-point Likert-type scale (*How scientifically rigorous would you consider the research presented in the abstract?*; 1 = not at all rigorous to 7 = very rigorous) as was contribution to medical science (*How much of a contribution to medical science do you consider the research presented in the abstract?*; 1 = not at all a contribution to 7 = a major contribution). Recommendation to publish the article was measured with yes or no choices (*Based on the abstract, would you recommend that this research be published in the Interdisciplinary Journal of Health?*). Analyses controlled for (1) sex, (2) race/ethnicity, (3) research experience level, (4) training background, and (5) citizenship (Table 1).

Sex was measured as a dichotomous variable (0 = Male, 1 = Female). Because of the low proportion of some ethnic/racial groups, we merged seven racial categories into three: White, Asian, and Other minorities. Participants' experience level was measured as a dichotomous variable (0 = new, 1 = experienced investigator). While participants' training background was measured as PhD, MD, MD/PhD, DDS, DVM, Other, and Unknown, it was included as a dichotomous training background variable (MD or MD/PhD = 1) in the analyses. US citizenship was measured as a dichotomous variable (1 = US citizen).

Although not one of our predetermined outcomes, we wanted to explore whether participants' explicit views on sex/gender might influence their responses. Therefore, in addition to participants' demographics and background, we asked them to complete a "social survey" after their review which couched the eight items of the Modern Sexism Scale (MSS) within a total of 21 questions on social, environmental, economic, and political issues.³⁹

The MSS intends to measure participants' denial of continued discrimination, antagonism toward women's demands, and lack of support for policies designed to help women. The MSS includes items such as, "It is easy to understand the anger of women's groups in America" and "Discrimination against women is no longer a problem in the United States" (reverse coded). Item responses (1 = strongly disagree to 7 = strongly agree) were recorded so that a higher value indicates fewer sexist responses, while a lower score indicates more sexist responses. The average score for eight items was used as an independent variable in the analysis (Table 2).

Statistical analyses

We tested whether article reviewers' assessment of a study's scientific rigor, contribution to medical science, and

recommendation to publish differed across the three experimental conditions: individuals (control abstract), men (male abstract), and women (female abstract). To adjust baseline differences, we included participants' demographic and background variables (sex, race, experience level, training background, citizenship) as control variables in our regression models. In addition, we included an average of the MSS (eight items) as a covariate in ordinal and binary logistic models to control for participants' explicit beliefs about women. We controlled for these variables because research on gender bias (Fig. 1) in peer review^{40,41} suggests that they could confound the outcomes of our experimental intervention.

Since article reviewers' assessment of a study's scientific rigor and contribution to medical science was measured on seven-point Likert-type scales, that is, ordinal scales, we used ordinal logistic regression models to test the differences of these outcomes among three abstract conditions and present results as proportional odds ratios (ORs), as well as logit coefficients.^{42,43} For the binary outcome of recommendation to publish, we used binary logistic regression models to test the outcome differences between abstract conditions.[†]

For the exploratory analyses of interactions of recommendation to publish with reviewers' explicit views of sex/gender, we included an average of the MSS (eight items) as a covariate in ordinal and binary logistic models to control for participants' explicit beliefs about women. We next constructed interaction models by adding interaction terms between a composite measure of the MSS and the three abstract conditions and examined predicted probabilities of outcome measures from logistic regressions for these interactions.

*In our ordinal logistic regression models,

$$\text{logit}[P(Y \leq j|x)] = \log \frac{P(Y \leq j|x)}{1 - P(Y \leq j|x)} = \alpha_j + B'x, \quad j = 1, \dots, 7$$

where Y is the scientific rigor or contribution to medical science ($j = 1, \dots, 7$; 7-point Likert-type scale), x is a vector of covariate variables (abstract conditions [ref. control abstract], sex, race/ethnicity, experience level, training background, citizenship, and modern sexism scale), B' is a vector of corresponding coefficients.

†In our binary logistic regression models,

$$\text{logit}[P(Y = 1|x)] = \log \frac{P(Y = 1|x)}{1 - P(Y = 1|x)} = \alpha_j + B'x$$

where $Y = 1$ is to recommend to publication ($Y = 0$, no recommendation), x is a vector of covariate variables (abstract conditions [ref. control abstract], sex, race/ethnicity, experience level, training background, citizenship, and modern sexism scale), B' is a vector of corresponding coefficients.

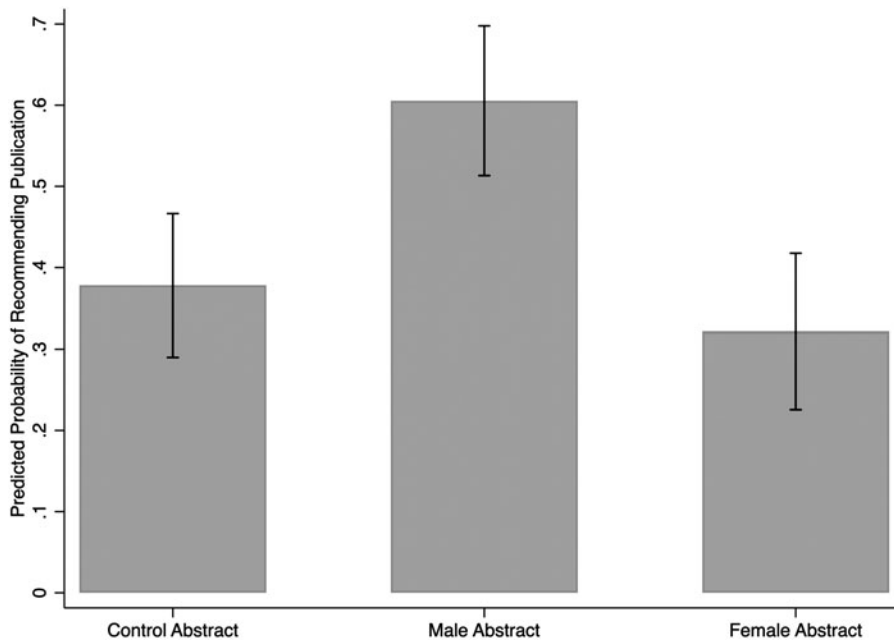


FIG. 2. Predicted probability of recommending publication by abstract conditions. Reviewers were nearly twice as likely to recommend publication of the male abstract than control or female abstract ($p < 0.001$). Predicted probabilities to recommend publication are 0.377 for control abstract, 0.606 for men abstract, and 0.322 for women abstract. Logistic regression models control for reviewer sex, race/ethnicity, experience level, training background (MD, PhD, MD-PhD), and U.S. citizenship.

We performed statistical analyses using STATA software release 16 with “logit,” “ologit,” “margins,” and “marginsplot” commands.⁴⁴ We assigned statistical significance when the p -values were < 0.05 and did not make adjustments for multiple comparisons.

Results

Reviewers were almost twice as likely to recommend publication when the abstract described research conducted in men ($N = 64/106$) than in women ($N = 29/93$) (predicted probabilities; 0.606 vs. 0.322; $p = 0.000$) (Fig. 2). The predicted probability that reviewers recommended publishing research conducted in individuals (control abstract) was not significantly different from research conducted in women ($p = 0.404$) but was significantly lower than research conducted in men ($p = 0.001$).

Ordinal logistic regression found that reviewers were significantly more likely to consider research conducted on women as a greater contribution to medical science than the same research conducted on men ($p = 0.030$; with no significant difference between male or female abstract and control abstract, $p = 0.086$ and $p = 0.618$, respectively). Specifically, the odds of being in a higher category of the perceived contribution to medical science scale ($1 = \text{not at all a contribution to } 7 = \text{a major contribution}$) were 1.73 times greater for research conducted on women than research conducted on men ($b = 0.55$, $OR = 1.73$, $p = 0.030$).

There was no significant difference in perceived rigor of the research by abstract condition (Table 3). There was a significant positive correlation between recommendation to publish and contribution to medical science (Pearson $r = 0.53$, $p = 0.000$) with no significant difference in this correlation between abstract type (Fig. 3).

In addition to our three primary outcomes, we conducted an exploratory analysis to probe whether the positive bias for recommending publication for research conducted on men versus women was heterogeneous across reviewers' explicit

beliefs about women. To test interactions between publication recommendation and MSS scores, we added interaction terms of abstract conditions with the average MSS scores in our final models across each outcome.

There were no significant interactions for scientific rigor and contribution to medical science with MSS scores, but the probabilities of the recommendation to publish in each abstract condition were significantly heterogeneous across MSS scores [Log likelihood ratio test of interaction effect: $\chi^2(2) = 10.8$, $p = 0.005$].[‡] That is, reviewers with more sexist responses were less likely to recommend publishing the female than other abstract conditions, while those with fewer sexist responses showed relatively less difference in recommending publication among three abstract conditions. These exploratory findings suggest that reviewers' explicit views of sex/gender should be further investigated for their potential influence on the outcomes of peer review.

Discussion

In an experimental study manipulating the gender of participants in a report of clinical research, we found that article peer reviewers were significantly more likely to recommend publishing research conducted in men than the same research conducted in either women or individuals. This difference occurred despite comparable ratings of scientific rigor and higher ratings of scientific impact for the research conducted in women than men.

[‡]In the interaction model of MSS with abstract conditions (reference group = control abstract), main effects for male abstract, female abstract, and MSS were $b = -2.59$ (standard error [SE] = 1.78, $p = 0.147$), -5.40 (SE = 1.72, $p = 0.002$), and -0.70 (SE = 0.23, $p = 0.002$), respectively. Interaction effects for male and female abstracts were $b = 0.66$ (SE = 0.33, $p = 0.045$) and 0.96 (SE = 0.31, $p = 0.002$), respectively. Interaction of MSS with a contrast between male and female abstract was $b = 0.30$ (SE = 0.32, $p = 0.359$).

TABLE 3. THE SUMMARY RESULTS FROM ORDINAL/BINARY LOGISTIC REGRESSION OF PERCEIVED SCIENTIFIC RIGOR, CONTRIBUTION TO MEDICAL SCIENCE, AND PUBLICATION RECOMMENDATION ON ABSTRACT CONDITIONS

Abstract condition	Scientific rigor			Contribution to medical science			Publication recommendation		
	b	OR	p	b	OR	p	b	OR	p
(Ref. male abstract)									
Control abstract	0.07 (0.24)	1.07 (0.26)	0.783	0.42 ⁺ (0.25)	1.53 ⁺ (0.37)	0.086	-0.94** (0.28)	0.39** (0.11)	0.001
Women abstract	0.46 ⁺ (0.26)	1.58 ⁺ (0.42)	0.079	0.55* (0.25)	1.73* (0.44)	0.030	-1.20** (0.30)	0.30** (0.09)	0.000

⁺*p*<0.1; **p*<0.05; ***p*<0.001. *b* refers to logistic regression coefficient; SE in parentheses. Ordinal logistic regression was used on scientific rigor and contribution to medical science, respectively. Binary logistic regression was used on publication recommendation. Each model includes sex, race/ethnicity, experience level, training background, citizenship, and modern sexism scale as covariates (not reported). OR, odds ratio; SE, standard error.

Figure 1 provides a conceptual model of how bias against research conducted in women—if found in further investigations to generalize across other areas of health and disease—may contribute to the lower productivity of women in academic science and medicine. This relatively lower pro-

ductivity in turn impedes their promotion to leadership positions³⁴ where they would be in positions to benefit women’s health in multiple ways.^{24,45}

Abundant evidence substantiates the lower status, prestige, and value of women compared with men in the prevailing

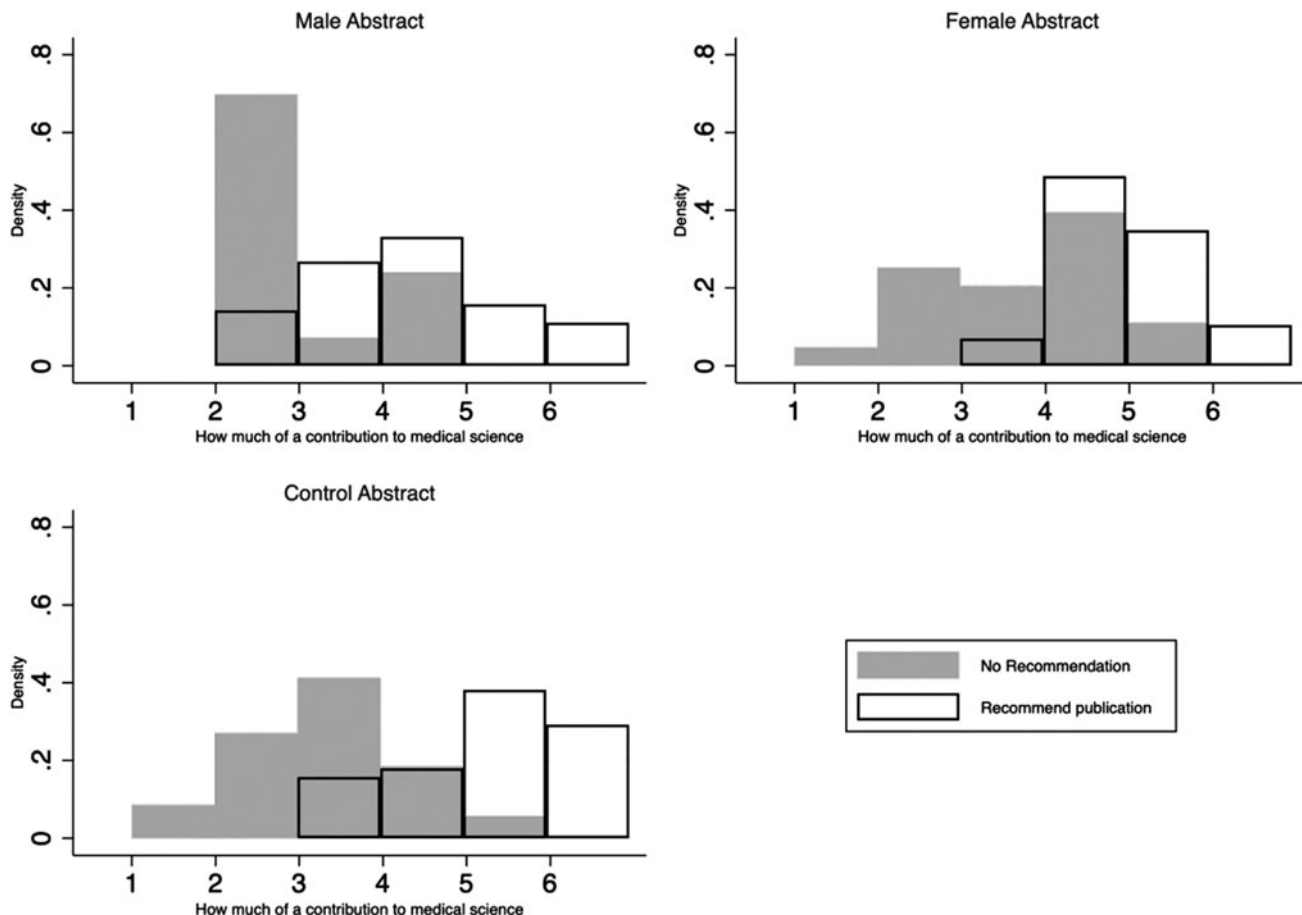


FIG. 3. Histograms of “contribution to medical science” (1=none to 7=major) by publication recommendation and abstract conditions. Gray histogram for publication not recommended and white histogram for publication recommended by the three abstract conditions (male, female, and control). The y-axis refers to the proportion of recommendations to publish for each assessment of contribution to medical science so that they sum to 1.0. This makes it possible to use the same vertical scale. So, for example, a y-scale of 0.5 means that 50% of overall responses occurred at a certain range of x-axis (e.g., 1 to ~1.99, 2 to ~2.99, 3 to ~3.99, and so on). Each histogram shows the positive relationship between “contribution to medical science” and “recommendation for publication,” which is relatively consistent across the three abstract conditions.

social hierarchy.^{1,17} Even the Bible states that a man is worth 50 shekels of silver, while a woman is worth only 30.³ Our assumption, therefore, was that if research conducted in women was less likely to be recommended for publication than the same research conducted in men, it would be because women are generally devalued relative to men—even as research subjects.

However, the greater likelihood that the male abstract would be recommended for publication than either the female or control abstract reveals the possibility that the difference in recommendations was related to implicitly higher value placed on research in men rather than lower value placed on research in women. Greater credence to the results of research findings in men compared with those found in women has been found in studies of hypertension, peripheral vascular disease, and depression.^{22,25,26} The regression models found no impact of reviewer gender, consistent with the vast majority of research since both men and women absorb the same societal values and have similar awareness of gender stereotypes.^{1,5,18}

We are unaware of any other study that assessed peer reviewers' assessment of an abstract or article with experimental manipulation of the gender of participants. The closest research we could find were two studies that found bias among article reviewers when research topics dealt with gender bias. Cisak et al.⁴⁵ found that research articles on gender bias were published significantly less often and in lower impact journals than research describing race bias, and Handley et al. demonstrated that male scientists deemed articles on gender bias as less meritorious than did female scientists, who found such articles more valuable.⁴⁴

Limitations of our study include nonresponse bias, in which study participants may not represent typical article reviewers. This concern is mitigated by random assignment of participants to the abstract versions, having a pool of participants that only included NIH-funded scientists who are undoubtedly familiar with the article peer review process, and the fact that scientists frequently decline article review invitations. It is likely that our email invitation was deleted without response by many who may have perceived it as one of the multiple invitations they receive to participate in reviewing activities for predatory journals.

Analyses of interactions of recommendations to publish with responses to MSS were not among our three primary outcomes and thus were exploratory and should be viewed in that context.[§] We did not adjust the level of statistical sig-

[§]Since the MSS is related to gender bias, it is interesting that the MSS found no differences among the three abstract conditions (Table 2), while there were significant differences in the outcomes (e.g., publication recommendation, contribution to medical science) across abstract conditions. Our exploratory interaction model suggested greater heterogeneity of publication recommendation across abstract conditions for those with more sexist responses, implying a conditional relationship between the MSS and our outcome differences across abstract conditions. However, there are several different possibilities for this: (1) it is possible that the MSS is not sensitive enough for measuring gender bias as it might emerge in this study; (2) the clinical topic used in this study could possibly be related to the MSS and our outcomes in different ways; and (3) the MSS and our outcomes could measure different aspects of gender bias. While further scrutiny regarding the MSS is beyond the scope of our study, future research may need to consider these issues in their design and analysis. We appreciate the thoughtful observation and interpretation by the statistical reviewer.

nificance for multiple comparisons because any correction to limit the likelihood of a type 1 error (finding something to be significant at the traditionally accepted $p < 0.05$ level when it is actually due to chance) will increase the likelihood of type 2 errors (missing something that is an important finding).^{48–50}

Martell et al.⁵¹ integrate research from a number of fields to conclude that gender segregation in organizations arises from collective behavior of individuals who express only a small pro-male bias. They conclude that small effects that favor men (as low as 1% in their computer simulations) can have real world consequences for women. In this context, if only a small bias exists against publication of research conducted in women, since women investigators are more likely to conduct such research, it contributes to the other biases noted in Figure 1. Because this is the first study, to our knowledge, that looks for (and finds evidence to support) bias against publishing research conducted in women, we felt that the chance of missing a significant finding that might close off future research in this area would be worse than the likelihood of a type 1 error.

Another potential limitation is the research topic of the abstract, which was HPA axis reactivity in survivors of sexual trauma with and without posttraumatic stress disorder (PTSD). While it is true that women are overrepresented among victims of sexual assault, subsequent PTSD occurs at similar rates in both men and women.⁵² It is unlikely that the topic accounts for the difference in recommendation to publish since rigor was assessed as comparable in the abstract conditions. Furthermore, if recommendations to publish are based on the practical value the research offers, given that sexual trauma is more common in women than men, we might expect this particular abstract to be recommended to publish at a higher rate when the research was conducted in women—which was not the case.

We hope that the results of our study stimulate others to investigate whether similar gender bias occurs in reviewing and recommending publication of articles describing research in other conditions that occur in both men and women, including diseases of the heart, lung, bone, kidney, skin, brain, and gastrointestinal tract.

Because female investigators are more likely than male investigators to study women,^{28,29} our findings suggest a previously unrecognized bias that could contribute to gender asymmetries in the publication outcomes of peer-review⁵³ and gender disparities in clinical research.²⁸ Our findings further affirm how gendered assumptions and expectations that arise from cultural stereotypes can subtly and insidiously influence academic careers.

Achieving gender equity in academic medicine will require a cultural change with interventions at multiple levels in all systems that evaluate, reward, and promote scientists and the area of science they study.⁵⁴ Interventions that motivate faculty in academic medicine, science, and engineering to practice cognitive strategies to “break the bias habit” are among the few tested experimentally and found to make a long-term individual and institutional impact.^{55,56} Although this strategy has not been studied in the context of article peer review, the career impact of publishing in peer reviewed journals suggests that such an intervention might be important to study in this context.

Authors' Contributions

S.M., M.C., and Y.-G.L. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: S.M., P.A.J., and M.C. Acquisition, analysis, or interpretation of data: S.M., M.C., and Y.-G.L. Drafting of the article: S.M., M.C., and Y.-G.L. Critical revision of the article for important intellectual content: S.M., M.C., P.A.J., and Y.-G.L. Statistical analysis: Y.-G.L. Obtained funding: P.A.J. and M.C. Administrative, technical, or material support: S.M., M.C., and Y.-G.L.

Disclaimer

Any opinions expressed herein do not necessarily reflect the opinions of the National Institutes of Health.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

This work was funded by discretionary research funds (Dr. Johnson) and grant R35 GM122557 (Dr. Carnes). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the article; or decision to submit the article for publication.

References

- Ridgeway CL, Bourg C, Eagly AH, Beall AE, Sternberg RJ. Gender as status: An expectation states theory approach. In: Eagly AH, Beall AE, Sternberg RJ, eds. *The Psychology of Gender*. 2nd ed. New York, NY: Guilford Press; 2004:217–241.
- Alksnis C, Desmarais S, Curtis J. Workforce segregation and the gender wage gap: Is 'women's' work valued as highly as 'men's'? *J Appl Soc Psychol* 2008;38:1416–1441.
- The Holy Bible, New Revised Standard Version Bible. Leviticus 27. 3.4. Nashville, TN: Thomas Nelson, Inc., 1989, p. 79.
- Goodman J, Loftus EF, Miller M, Greene E. Money, sex, and death: Gender bias in wrongful death damage awards. *Law Soc Rev* 1991;25:263–285.
- West TV, Heilman ME, Gullett L, Moss-Racusin CA, Magee JC. Building blocks of bias: Gender composition predicts male and female group members' evaluations of each other and the group. *J Exp Soc Psychol* 2012;48:1209–1212.
- Block K, Croft A, Schmader T. Worth less?: Why men (and women) devalue care-oriented careers. *Front Psychol* 2018; 29:1353.
- Lo Sasso AT, Richards MR, Chou CF, Gerber SE. The \$16,819 pay gap for newly trained physicians: The unexplained trend of men earning more than women. *Health Affairs (Project Hope)* 2011;30:193–201.
- Jena AB, Olenski AR, Blumenthal DM. Sex differences in physician salary in US public medical schools. *JAMA Intern Med* 2016;176:1294–1304.
- Jagsi R, Griffith KA, Stewart A, Sambuco D, DeCastro R, Ubel PA. Gender differences in the salaries of physician researchers. *JAMA* 2012;307:2410–2417.
- Mensah M, Beeler W, Rotenstein L, et al. Sex differences in salaries of department chairs at public medical schools. *JAMA Intern Med* 2020;180:789–792.
- Kolehmainen C, Carnes M. Who resembles a scientific leader-Jack or Jill? How implicit bias could influence research grant funding. *Circulation* 2018;137:769–770.
- 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.
- 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;26:1–10.
- Witteman HO, Hendricks M, Straus S, Tannenbaum C. Gender bias in CIHR Foundation grant awarding. *Lancet* 2019;394:e41–e42.
- Kane L. Medscape Female Physician Compensation Report 2019, 2019:3. Available at: <https://www.medscape.com/slideshow/2019-compensation-female-physician-6011698> Accessed March 8, 2021.
- Medscape. Female Physician Compensation Report 2019:3.
- Pelley E, Carnes M. When a specialty becomes "women's work": Trends in and implications of specialty gender segregation in medicine. *Acad Med* 2020;95:1499–1506.
- Hofler LG, Hacker MR, Dodge LE, Schutzberg R, Ricciotti HA. Comparison of women in department leadership in obstetrics and gynecology with those in other specialties. *Obstet Gynecol* 2016;127:442–447.
- Spector ND, Asante PA, Marcelin JR, et al. Women in pediatrics: Progress, barriers, and opportunities for equity, diversity, and inclusion. *Pediatrics* 2019;144:e20192149.
- Spencer RJ, Rice LW, Ye C, Woo K, Uppal S. Disparities in the allocation of research funding to gynecologic cancers by Funding to Lethality scores. *Gynecol Oncol* 2019;152:106–111.
- Benoit MF, Ma JF, Upperman BA. Comparison of 2015 Medicare relative value units for gender-specific procedures: Gynecologic and gynecologic-oncologic versus urologic CPT coding. Has time healed gender-worth? *Gynecol Oncol* 2017;144:336–342.
- McFalls EO, Ward HB, Moritz TE, et al. Coronary-artery revascularization before elective major vascular surgery. *N Engl J Med* 2004;351:2795–2804.
- Carnes M, Sarto GE, Springer K. Managing depression in outpatients. *N Engl J Med* 2001;344:1252–1253.
- Carnes M, Morrissey C, Geller SE. Women's health and women's leadership in academic medicine: Hitting the same glass ceiling? *J Womens Health* 2008;17:1453–1462.
- Whooley MA, Simon GE. Managing depression in medical outpatients. *N Engl J Med* 2000;343:1942–1950.
- Wing LM, Reid CM, Ryan P, et al. A comparison of outcomes with angiotensin-converting—Enzyme inhibitors and diuretics for hypertension in the elderly. *N Engl J Med* 2003;348:583–592.
- Geller SE, Koch AR, Roesch P, Filut A, Hallgren E, Carnes M. The more things change, the more they stay the same: A study to evaluate compliance with inclusion and assessment of women and minorities in randomized controlled trials. *Acad Med* 2018;93:630–635.
- Sugimoto CR, Ahn YY, Smith E, Macaluso B, Lariviere V. Factors affecting sex-related reporting in medical research: A cross-disciplinary bibliometric analysis. *Lancet* 2019; 393:550–559.

29. Nielsen MW, Andersen JP, Schiebinger L, Schneider JW. One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis. *Nat Hum Behav* 2017;1:791–796.
30. Hengel E. Publishing while female. In: Lundberg S, ed. *Women in economics*. London, United Kingdom: Centre for Economic Policy Research Press. A VoxEU.org Book, 2020:80–90.
31. Fox CW, Paine CET. Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. *Ecol Evol* 2019;9:3599–3619.
32. Knobloch-Westerwick S, Glynn CJ, Hume M. The Matilda Effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Sci Commun* 2013;35:603–625.
33. Silbiger NJ, Stubler AD. Unprofessional peer reviews disproportionately harm underrepresented groups in STEM. *PeerJ* 2019;7:e8247.
34. 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;93:1694–1699.
35. Blumenthal DM, Olenski AR, Yeh RW, et al. Sex differences in faculty rank among academic cardiologists in the United States. *Circulation* 2017;135:506–517.
36. Jagsi R, Motomura AR, Griffith KA, Rangarajan S, Ubel PA. Sex differences in attainment of independent funding by career development awardees. *Ann Intern Med* 2009;151:804–811.
37. Kaatz A, Magua W, Zimmerman DR, Carnes M. A quantitative linguistic analysis of National Institutes of Health R01 application critiques from investigators at one institution. *Acad Med* 2015;90:69–75.
38. Martinson A, Craner J, Sigmon S. Differences in HPA axis reactivity to intimacy in women with and without histories of sexual trauma. *Psychoneuroendocrinology* 2016;65:118–126.
39. Swim JK, Aikin KJ, Hall WS, Hunter BA. Sexism and racism: Old-fashioned and modern prejudices. *J Personal Soc Psychol* 1995;68:199–241.
40. Guthrie S, Ghiga I, Wooding S. What do we know about grant peer review in the health sciences? *F1000Res* 2017;6:1335.
41. Brezis ES, Birukou A. Arbitrariness in the peer review process. *Scientometrics* 2020;123:393–411.
42. Agresti A. *Categorical data analysis*. Hoboken, NJ: Wiley & Sons, 2002.
43. Long JS, Freese J. *Regression models for categorical dependent variables using stata*, 3rd ed. College Station, TX: Stata Press, 2014.
44. StataCorp. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC, 2019.
45. Carnes M, Johnson P, Klein W, Jenkins M, Bairey Merz CN. Advancing women's health and women's leadership with endowed chairs in women's health. *Acad Med* 2017;92:167–174.
46. Cislak A, Formanowicz M, Saguy T. Bias against research on gender bias. *Scientometrics* 2018;115:189–200.
47. Handley IM, Brown ER, Moss-Racusin CA, Smith JL. Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proc Natl Acad Sci U S A* 2015;112:13201–13206.
48. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. *J Res Educ Effectiveness* 2012;5:189–211.
49. Streiner DL, Norman GR. Correction for multiple testing: Is there a resolution? *Chest* 2011;140:16–18.
50. Feise RJ. Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol* 2002;2:8.
51. Martell RF, Emrich CG, Robison-Cox J. From bias to exclusion: A multilevel emergent theory of gender segregation in organizations. *Res Organ Behav* 2012;32:137–162.
52. Kang H, Dalager N, Mahan C, Ishii E. The role of sexual assault on the risk of PTSD among Gulf War veterans. *Ann Epidemiol* 2005;15:191–195.
53. Helmer M, Schottdorf M, Neef A, Battaglia D. Gender bias in scholarly peer review. *Elife* 2017;6:e21718.
54. Carnes M. The American College of Physicians is working hard to achieve gender equity, and everyone will benefit. *Ann Intern Med* 2018;168:741–743.
55. Carnes M, Devine PG, Manwell LB, et al. Effect of an intervention to break the gender bias habit: A cluster randomized, controlled trial. *Acad Med* 2015;90:221–230.
56. Devine PG, Forscher PS, Cox WTL, Kaatz A, Sheridan J, Carnes M. A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *J Exp Soc Psychol* 2017;73:211–215.

Address correspondence to:

Molly Carnes, MD, MS

Departments of Medicine, Psychiatry, and Industrial

and Systems Engineering, University of Wisconsin

Center for Women's Health Research

700 Regent Street, Suite 301

Madison, WI 53726

USA

E-mail: mlcarnes@wisc.edu