

The Relationship Between Applicant Gender and Internal Medicine Residency Interview Scores

Robert M. Stern, MD, EdM
 Mary W. Montgomery, MD
 Nora Y. Osman, MD
 Joel T. Katz, MD
 Maria A. Yialamas, MD

ABSTRACT

Background Gender inequity is widespread in academic medicine, including in the promotion, academic recognition, and compensation of female faculty.

Objective To assess whether these inequities extend to the GME intern selection process, this study examines differences in the interview scores assigned to male and female applicants at one large internal medicine residency program.

Methods Subjects include 1399 applicants who completed 3099 interviews for internship positions for the Brigham and Women's Hospital internal medicine residency in Electronic Residency Application Service (ERAS) cycles 2015–2016, 2017–2018, 2018–2019, and 2019–2020. Unadjusted and multivariable linear regressions were used to assess the simultaneous effect of applicant gender, interviewer gender, and applicant academic characteristics on pre-interview, post-interview, and change in interview scores.

Results Our analysis included 3027 interviews (97.7%) of 1359 applicants (97.1%). There were no statistically significant differences in the interview scores assigned to female versus male applicants. This was true across pre-interview scores (difference = 0.03, $P = .61$), post-interview scores (difference = 0.00, $P = .98$), and change in interview scores (difference = 0.01, $P = .24$) as well as when adjusting for the baseline academic characteristics of both male and female applicants. This was also true when analyzing individual application years, individual residency tracks, and accounting for the gender of the faculty interviewers.

Conclusions The findings do not support the presence of gender inequity in the interview scores assigned to male and female applicants included in this study.

Introduction

In January 2020, the Association of American Medical Colleges (AAMC) unveiled a new initiative calling on medical schools and teaching hospitals to “identify and address gender inequities in academic medicine.”¹ As part of the initiative, the AAMC listed numerous examples of gender inequity within academic medicine, including underrepresentation of women in the physician and scientific workforce, exclusion of women from leadership roles in academic medicine, and decreased financial compensation and academic recognition of female faculty as compared to their male colleagues. Motivated by the AAMC Statement on Gender Equity, we sought to determine whether the gender inequities pervading academic medicine extend into one critical entry point in academic medicine—the evaluation processes used by graduate medical education (GME) programs for intern selection.

Numerous studies have demonstrated gender inequities in hiring and selection processes outside of medicine.^{2–5} For example, in a randomized double-blind study assessing the hiring practices of scientific faculty, female applicants were deemed less competent and were less likely to be hired than male applicants with identical skill sets.⁴ Similarly, an analysis of symphony orchestras found that orchestras that adopted blind auditions hired more women when compared to orchestras with open auditions.⁵ In addition, gender inequities have been documented in the materials that make up the Electronic Residency Application Service (ERAS) applications for GME positions, including in letters of recommendation.^{6–8}

Nevertheless, there are only a few studies examining gender bias in the GME applicant review or interview processes, and these studies did not find evidence of gender inequity.^{9–11} While very well done, these studies are now older and did not incorporate into their analyses many important academic characteristics of applicants. This suggested the need for a more updated and detailed analysis given the subjective nature of this interview process and the high risk of bias.

DOI: <http://dx.doi.org/10.4300/JGME-D-21-00270.1>

Editor's Note: The online version of this article contains pre-interview, post-interview, and change in score regressions that are unadjusted and adjusted for baseline academic characteristics.

Our study examines whether there are systematic differences in the pre-interview, post-interview, and change in interview rating (ie, “score”) assigned to male and female applicants who interviewed for positions within the Brigham and Women’s Hospital (BWH) internal medicine (IM) residency program between 2015 and 2020. We chose to focus specifically on interview scores as GME program directors consistently rate the interview as one of the most important factors when determining the final rank list.¹²

Methods

The BWH IM residency is a large academic program with 174 residents spread across 5 unique residency tracks (categorical, primary care, preliminary, medicine-dermatology, and a separately accredited medicine-pediatrics). The BWH IM residency program participates in the National Resident Matching Program (NRMP) through which it receives more than 4000 applications annually. Program leadership reviews all applications and selects approximately 350 applicants (220 categorical, 50 primary care, 80 preliminary) to interview for residency based on a predetermined combination of an applicant’s academic performance and extracurricular activities, with an emphasis on a holistic review that prevents any single factor from determining the application outcome. No screening filters (eg, United States Medical Licensing Examination [USMLE] scores) are used. Each year, applicants invited to interview are enrolled in more than 80 allopathic medical schools across the United States and a smaller number of international medical schools.

On the interview day, each categorical and preliminary applicant completes two 25-minute interviews with faculty interviewers; primary care applicants complete 4 such interviews. Prior to the interview, each faculty interviewer is required to review an applicant’s full ERAS file and assign the applicant a pre-interview score based on the materials presented in the ERAS application. Following completion of the interview, each faculty member then assigns the applicant a post-interview score based on a combination of the applicant’s ERAS application and interview performance. Pre-interview scores and post-interview scores are assigned on a 1 to 5 scale in 0.5-point increments, with 1 being defined as an “absolutely top candidate” and 5 being defined as “not suitable” for the residency. Each year, approximately 75 different faculty serve as interviewers for the BWH IM residency. Faculty are required to attend a yearly interview workshop during which program leadership review the interview form completed by

Objectives

To assess whether the gender inequities that are widespread in academic medicine extend to the graduate medical education intern selection process.

Findings

There were no statistically significant differences in the interview scores assigned to female versus male applicants in the pre-interview scores, post-interview scores, or change in interview scores.

Limitations

The study is focused on the application process of a single internal medicine residency program and does not compare residency programs across different specialties or different institutions.

Bottom Line

The findings do not support the presence of gender inequity in the interview scores assigned to male and female applicants in the intern selection process, but whether this is generalizable across different specialties or different institutions is yet to be determined.

faculty, expectations of faculty during the interview process, and suggested questions for interviewees. Starting in 2016–2017, the program also incorporated an implicit bias training into the interview workshop.

Our subjects include all applicants who interviewed for the BWH IM residency in the ERAS application cycles 2015–2016, 2017–2018, 2018–2019, and 2019–2020. Applicants who interviewed in 2016–2017 were excluded from the analysis as the residency program no longer had the complete set of ERAS applications for that year. In addition, applicants who interviewed for a position in the medicine-dermatology and medicine-pediatrics residency tracks were not included in the study due to differences in the interview processes for these tracks.

To collect the data for this study, 2 investigators (M.W.M., R.M.S.) reviewed and recorded all relevant ERAS application and interview scoresheet data. This included pre-interview score, post-interview score, change in interview score, application year, self-reported gender, self-reported race, and the residency track applied to. It also included academic characteristics: medical school ranking (based on the 2020 *US News & World Report* Medical School Research rankings), advanced degrees in addition to MD, most recent USMLE Step 1 score, most recent USMLE Step 2 CK score, clerkship grades, sub-internship grade in medicine, number of accepted or printed peer-reviewed publications (first, middle, and last authored), number of accepted or completed poster presentation and oral presentations (first and last authored), Alpha Omega Alpha (AOA) status, Gold Humanism Honor Society status, Medical Student Performance Evaluation (MSPE) ranking, and departmental ranking. The MSPE ranking and departmental

rankings were only included for applicants from medical schools who rank students or put them into percentiles and include these rankings/percentiles in their MSPE letter or departmental letter. Self-reported gender and self-reported race were pulled directly from the ERAS application, with options for gender on the ERAS application including only male or female.

In addition to the data collected for each interviewee, investigators collected data on each faculty interviewer, including gender. The data on faculty interviewers was drawn from a combination of online hospital biographies and internal records kept by the BWH IM residency and the BWH Department of Medicine. Following completion of data collection and prior to data analysis, all data was deidentified and assigned a study number.

For person-level comparison of proportions by gender we used the Pearson's chi-square test. For comparison of means of continuous variables, we used the Student's *t* test for the equality of means. To account for non-independence of scores within each applicant in interview-level data we clustered the standard errors at the level of individual to allow for within-individual correlation in scores. We performed unadjusted and multivariable linear regressions to assess the simultaneous effect of gender and all recorded academic factors on pre-interview, post-interview, and within-person change in interview scores. All hypothesis tests were 2-sided with the significance level set at $P < .05$. All analyses were performed using STATA/MP 16 (StataCorp LLC, College Station, TX).

The Brigham and Women's Hospital Institutional Review Board approved the study.

Results

We reviewed 1399 applicants who completed a total of 3099 interviews for internship positions within the BWH IM residency program. Forty applicants and 72 interviews were excluded from our analysis either because of missing interview score data or missing ERAS application data. Our analytic sample included 3027 interviews (97.7%) of 1359 applicants (97.1%) that were completed by 170 unique interviewers.

The demographic information and baseline academic characteristics for male vs female applicants are presented in TABLE 1. Of the 1359 applicants, 669 (49.2%) self-reported as female and 690 (50.8%) self-reported as male; 12.7% ($N = 85$) of the female applicants and 15.7% ($N = 108$) of the male applicants were from groups defined by the AAMC as underrepresented in medicine (ie, applicants from racial and ethnic populations underrepresented in the

medical profession relative to their numbers in the general population).

There were no differences between male and female applicants across many of the baseline academic characteristics assessed, including clerkship grades, sub-internship grade in medicine, AOA status, Gold Humanism Honor Society status, MSPE designation, or departmental designation. As compared to female applicants, a higher proportion of male applicants were found to have a PhD (17.8% vs 10.5%, $P < .001$), to have a higher Step 1 score (mean 251.5 vs 247.4, $P < .001$), to have 1 or more first/last authored publications (66.3% vs 60.3%, $P = .023$), and to have 1 or more middle-authored publications (80.1% vs 75.2%, $P = .030$).

When assessing the unadjusted pre-interview and post-interview scores and the change in the interview scores, there were no significant differences in scores between female and male applicants. Female applicants had an average pre-interview score of 2.22 as compared to 2.19 for male applicants ($P = .61$). Female applicants had an average post-interview score of 2.08 as compared to 2.08 for male applicants ($P = .98$). Female applicants had an average improvement of their interview score of 0.13 as compared to 0.12 for male applicants ($P = .24$). There were also no differences between the unadjusted pre-interview and post-interview scores and change in interview scores of female and male applicants when considering each year individually (TABLE 2) or when assessing applicants within an individual residency application track (eg, primary care, categorical, or preliminary; TABLE 3).

When adjusting for the baseline academic characteristics of female and male applicants, including PhD status, Step 1 score, and both first/last and middle authored publications, there continued to be no statistically significant differences between the pre-interview, post-interview, and the change in interview scores between female and male applicants (provided as online supplementary data).

Finally, of the 170 faculty interviewers over 4 years, 44.1% ($n = 75$) were female. We found that there were no statistically significant differences in the pre-interview or post-interview scores or the change in interview scores that female or male faculty assigned to male vs female applicants (TABLE 4).

Discussion

We found no difference in the interview scores assigned to female and male applicants attributable to gender of the applicant. This was true across pre-interview scores, post-interview scores, and change in interview scores as well as when we adjusted for the

TABLE 1

Demographic Information and Baseline Academic Characteristics of Male and Female Applicants Across All Years

Characteristic	Females (N=669)	Males (N=690)	P Value
Underrepresented in medicine, ^a No. (%)	85 (12.7)	108 (15.6)	.12
Residency track, No. (%)			<.001
Categorical	372 (55.6)	469 (68.0)	
Primary care	133 (19.9)	66 (9.6)	
Preliminary	164 (24.5)	155 (22.5)	
Medical school ranking, ^b No. (%)			.20
1–25	437 (65.3)	457 (66.2)	
26–50	99 (14.8)	81 (11.7)	
50–unranked	133 (19.9)	152 (22.0)	
PhD or equivalent, No. (%)	70 (10.5)	123 (17.8)	<.001
Step 1 score	247.41 (12.50)	251.45 (12.57)	<.001
Step 2 CK score	259.85 (10.71)	260.92 (11.28)	.12
Clerkship grades, No. (%)			.94
100% honors/highest available grade	255 (38.1)	260 (37.7)	
75–99% honors/highest available grade	166 (24.8)	176 (25.5)	
50–74% honors/highest available grade	125 (18.7)	120 (17.4)	
25–49% honors/highest available grade	31 (4.6)	40 (5.8)	
<25% honors/highest available grade	18 (2.7)	19 (2.8)	
Pass/fail only	74 (11.1)	75 (10.9)	
Sub-internship grade in medicine			.12
Honors/highest available grade	433 (64.7)	411 (59.6)	
Less than honors	17 (2.5)	25 (3.6)	
Pass/fail only	27 (4.0)	41 (5.9)	
Not completed at time ERAS application submitted	192 (28.7)	213 (30.9)	
First/last author publications, No. (%)			.023
0	266 (39.7)	233 (33.7)	
≥1	403 (60.3)	457 (66.3)	
Middle author publications, No. (%)			.030
0	503 (75.2)	553 (80.1)	
≥1	166 (24.8)	137 (19.9)	
First author oral presentations/posters, No. (%)			.37
0	84 (12.6)	98 (14.2)	
≥1	585 (87.4)	592 (85.8)	
Alpha Omega Alpha status, No. (%)			.76
Yes	358 (53.5)	358 (51.9)	
No	108 (16.1)	110 (15.9)	
Medical school does not participate	203 (30.3)	222 (32.2)	
Gold Humanism Honor Society Status, No. (%)			.45
Yes	117 (17.5)	111 (16.0)	
No	267 (39.9)	290 (42.0)	
Medical school does not participate	285 (42.6)	289 (41.9)	
MSPE designation, No. (%)			.13
Highest designation	326 (48.7)	311 (45.1)	
Second highest designation	40 (6.0)	43 (6.2)	
Third highest designation	3 (0.4)	10 (1.4)	
Fourth highest designation	2 (0.3)	0 (0.0)	
No designation present in MSPE letter	298 (44.5)	326 (47.2)	

TABLE 1

Demographic Information and Baseline Academic Characteristics of Male and Female Applicants Across All Years (continued)

Characteristic	Females (N=669)	Males (N=690)	P Value
Medicine departmental designation, No. (%)			.18
Highest designation	450 (67.4)	439 (63.6)	
Second highest designation	68 (10.2)	87 (12.6)	
Third highest designation	16 (2.4)	10 (1.4)	
No designation present in medicine department letter	134 (20.1)	154 (22.3)	

^a Defined by the AAMC as applicants from racial and ethnic populations underrepresented in the medical profession relative to their numbers in the general population.

^b Based on the 2020 *US News & World Report* Medical School Research rankings.

TABLE 2

Unadjusted Pre-Interview Scores, Post-Interview Scores, and Change in Interview Scores Across All Years and by Individual Year

	Pre-Interview Scores			Post-Interview Scores			Change in Interview Scores		
	Female	Male	P Value	Female	Male	P Value	Female	Male	P Value
All Years	(N=1548)	(N=1471)		(N=1548)	(N=1471)		(N=1548)	(N=1471)	
	2.22 (0.66)	2.19 (0.72)	.61	2.08 (0.71)	2.08 (0.77)	.98	-0.13 (0.36)	-0.12 (0.40)	.24
2015–2016	(N=387)	(N=390)		(N=387)	(N=390)		(N=387)	(N=390)	
	2.26 (0.60)	2.25 (0.74)	.84	2.13 (0.70)	2.13 (0.79)	.95	-0.14 (0.35)	-0.12 (0.42)	.58
2017–2018	(N=419)	(N=370)		(N=419)	(N=370)		(N=419)	(N=370)	
	2.25 (0.68)	2.27 (0.72)	.79	2.12 (0.72)	2.16 (0.78)	.54	-0.13 (0.35)	-0.11 (0.35)	.41
2018–2019	(N=367)	(N=364)		(N=367)	(N=364)		(N=367)	(N=364)	
	2.21 (0.66)	2.18 (0.71)	.63	2.07 (0.71)	2.06 (0.74)	.91	-0.14 (0.37)	-0.12 (0.37)	.40
2019–2020	(N=375)	(N=347)		(N=375)	(N=347)		(N=375)	(N=347)	
	2.13 (0.66)	2.09 (0.72)	.59	1.98 (0.70)	1.96 (0.75)	.77	-0.14 (0.36)	-0.13 (0.45)	.78

TABLE 3

Unadjusted Pre-Interview Scores, Post-Interview Scores, and Change in Interview Scores by Residency Track Across All Years

	Pre-Interview Scores			Post-Interview Scores			Change in Interview Scores		
	Female	Male	P Value	Female	Male	P Value	Female	Male	P Value
Categorical Track	(N=711)	(N=905)		(N=711)	(N=905)		(N=711)	(N=905)	
	2.28 (0.66)	2.21 (0.72)	.08	2.14 (0.66)	2.07 (0.76)	.14	-0.14 (0.35)	-0.13 (0.37)	.63
Primary Care Track	(N=516)	(N=259)		(N=516)	(N=259)		(N=516)	(N=259)	
	2.11 (0.68)	2.09 (0.73)	.77	1.97 (0.71)	1.94 (0.73)	.78	-0.14 (0.36)	-0.14 (0.43)	.99
Preliminary Track	(N=321)	(N=307)		(N=321)	(N=307)		(N=321)	(N=307)	
	2.23 (0.64)	2.26 (0.69)	.71	2.11 (0.74)	2.19 (0.79)	.31	-0.12 (0.24)	-0.07 (0.35)	.14

TABLE 4

Unadjusted Pre-Interview Scores, Post-Interview Scores, and Change in Interview Scores by Gender Across All Years

	Pre-Interview Scores			Post-Interview Scores			Change in Interview Scores		
	Female	Male	P Value	Female	Male	P Value	Female	Male	P Value
Female Interviewer	2.19	2.14	.36	2.04	2.00	.47	-0.15	-0.14	.48
Male Interviewer	2.24	2.22	.58	2.11	2.11	.86	-0.13	-0.11	.66

baseline academic characteristics of male and female applicants. This was also true when we analyzed individual application years and individual residency tracks and accounted for the gender of the faculty interviewers.

This lack of difference in interview scores for female vs male applicants was true despite statistically significant differences in a few of the baseline academic characteristics for female vs male applicants, including PhD status, Step 1 score, and number of publications. There are multiple possible explanations for why differences in baseline academic characteristics did not result in differences in interview scores. First, while the differences in Step 1 score and number of publications were statistically significant, the absolute differences were quite small and therefore unlikely to influence the score assigned by an interviewer. Second, there were several applicant characteristics not accounted for in our analysis, including letters of recommendations, extracurricular activities, or personal statements. These unaccounted-for characteristics may have balanced out the observed differences in baseline academic characteristics.

When taken in their entirety, our results argue against significant gender inequities in the interview scores assigned to female and male applicants within the BWH IM residency. This finding is discordant with gender inequities found elsewhere in academic medicine and within hiring and evaluation processes outside of medicine, but consistent with the small body of research on gender inequity within the evaluation of applicants for GME training positions. For example, in a study of how faculty characteristics affected interview scores within the University of Chicago Internal Medicine Residency Program, Oyler et al found that neither the gender of the faculty interviewer nor the gender of the applicant affected assigned interview scores.¹⁰ Similarly, in an analysis of the application process within the diagnostic radiology program at the Medical University of South Carolina, Hewett et al found that the proportion of female applicants invited to interview and eventually ranked in the top 25% of the rank list exceeded the overall proportion of female applicants.⁹

Our findings, along with others,^{9,10} raise important questions. Why have gender inequities identified in both academic medicine and in hiring processes outside of medicine not been found in the GME applicant evaluation process? Is this a problem with the methodologies used to answer the question or is there something unique about the residency application process? While additional studies are needed to appropriately answer these questions, there are potential hints drawn from both the medical and

sociology literature.^{13,14} Gender biases against women in hiring and admissions processes have been found to be mitigated when evaluators are provided with “individuating proof of competence and past performance excellence that are relevant to the employment opportunity” and when women represent 25% or more of the applicant pool.¹⁴ The ERAS application and largely uniform interview days provide a standardized format for female and male applicants to demonstrate both their prior accomplishments and their level of competence. In addition, while not all GME application pools have enough female applicants, many do achieve this 25% threshold. In fact, during our study period, 41.9% to 43.2% of the resident physicians in Accreditation Council for Graduate Medical Education-accredited internal medicine programs identified as female.^{15–18} It is therefore likely that women represented at least 40% of the overall internal medicine applicants.

Our study has multiple limitations. It is focused on the application process of a single internal medicine residency program. It does not compare residency programs across different specialties or different institutions. Therefore, it is not clear whether our findings can be extrapolated to other internal medicine or non-internal medicine training programs across the country. This is especially true given particular aspects of our residency program. We have women in visible leadership positions (half of our associate program directors are women), and there is a roughly equal gender balance of male and female trainees. In addition, beginning in 2016–2017, the residency leadership began requiring implicit bias training for all residency interviewers in order to minimize bias. These aspects of the residency program may serve to reduce bias, compared to residency programs with less gender diversity or programs without implicit bias training.

In addition to the above, 1 year of data during our study period was missing as the residency program did not have the complete set of ERAS applications for that year. Next, while we adjusted for many of the objective baseline academic characteristics included in an ERAS application, we did not adjust for the more subjective components of the ERAS application, including letters of recommendations, personal statements, extracurricular activities, or leadership accomplishments which could have altered our findings. Finally, we evaluated only one component of the applicant evaluation process. We did not assess for gender inequities, for example, in the initial screening process of applicants or in the formation of the final rank list and eventual Match results.

Despite these limitations, we believe that our study contributes to much-needed research on gender

inequity in academic medicine and more specifically on the GME applicant evaluation process. It also invites additional studies of the GME application process, including a multisite study of gender inequity that includes residency programs across hospitals and specialties and individual site and multisite studies assessing inequities that may arise from other demographic characteristics, including race, sexual orientation, or disability status.

Within our own residency, our next step is to evaluate the association of these other demographic characteristics. We also intend to set up a system of annual review of our interview data and to implement changes in the applicant evaluation process based on any biases that are uncovered. We call on other GME programs across the country to do the same.

Conclusions

While gender inequity is widespread in academic medicine, within one large internal medicine residency, we found no statistically significant differences in the interview scores assigned to female vs male applicants. This was true even when adjusting for the academic characteristics of male and female applicants and when analyzing individual residency tracks and accounting for the gender of faculty interviewers.

References

1. Association of American Medical Colleges. AAMC Statement on Gender Equity. <https://www.aamc.org/what-we-do/diversity-inclusion/aamc-statement-gender-equity>. Accessed August 19, 2021.
2. Petit P. The effects of age and family constraints on gender hiring discrimination: a field experiment in the French financial sector. *Labour Econ*. 2007;14(3):371–391. doi:10.1016/j.labeco.2006.01.006
3. González MJ, Cortina C, Rodríguez J. The role of gender stereotypes in hiring: a field experiment. *Eur Sociol Rev*. 2019;35(2):187–204. doi:10.1093/esr/jcy055
4. Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci U S A*. 2012;109(41):16474–16479. doi:10.1073/pnas.1211286109
5. Goldin C, Rouse C. Orchestrating impartiality: the impact of “blind” auditions on female musicians. *Am Econ Rev*. 2000;90(4):715–741. doi:10.1257/aer.90.4.715
6. Filippou P, Mahajan S, Deal A, et al. The presence of gender bias in letters of recommendations written for urology residency applicants. *Urology*. 2019;134:56–61. doi:10.1016/j.urology.2019.05.065
7. Grimm LJ, Redmond RA, Campbell JC, Rosette AS. Gender and racial bias in radiology residency letters of recommendation. *J Am Coll Radiol*. 2020;17(1):64–71. doi:10.1016/j.jacr.2019.08.008
8. Turrentine FE, Dreisbach CN, St Ivany AR, Hanks JB, Schroen AT. Influence of gender on surgical residency applicants' recommendation letters. *J Am Coll Surg*. 2019;228(4):356–365.e3. doi:10.1016/j.jamcollsurg.2018.12.020
9. Hewett L, Lewis M, Collins H, Gordon L. Gender bias in diagnostic radiology resident selection, does it exist? *Acad Radiol*. 2016;23(1):101–107. doi:10.1016/j.acra.2015.10.018
10. Oyler J, Thompson K, Arora VM, Krishnan JA, Woodruff J. Faculty characteristics affect interview scores during residency recruitment. *Am J Med*. 2015;128(5):545–550. doi:10.1016/j.amjmed.2015.01.025
11. Smith CJ, Rodenhauer P, Markert RJ. Gender bias of Ohio physicians in the evaluation of the personal statements of residency applicants. *Acad Med*. 1991;66(8):479–481. doi:10.1097/00001888-199108000-00014
12. National Residency Matching Program. Results of the 2018 NRMP Program Director Survey. <https://www.nrmp.org/main-residency-match-data/>. Accessed August 19, 2021.
13. Bohnet I, Van Geen A, Bazerman M. When performance trumps gender bias: joint vs. separate evaluation. *Manage Sci*. 2016;62(5):1225–1234. doi:10.1287/mnsc.2015.2186
14. Isaac C, Lee B, Carnes M. Interventions that affect gender bias in hiring: a systematic review. *Acad Med*. 2009;84(10):1440–1446. doi:10.1097/ACM.0b013e3181b6ba00
15. Brotherton SE, Etzel SI. Graduate medical education, 2015–2016. *JAMA*. 2016;316(21):2291–2310. doi:10.1001/jama.2016.13513
16. Brotherton SE, Etzel SI. Graduate Medical Education, 2017–2018. *JAMA*. 2018;320(10):1051–1070. doi:10.1001/jama.2018.10650
17. Brotherton SE, Etzel SI. Graduate Medical Education, 2018–2019. *JAMA*. 2019;322(10):996–1016. doi:10.1001/jama.2019.10155
18. Brotherton SE, Etzel SI. Graduate Medical Education, 2019–2020. *JAMA*. 2020;324(12):1230–1250. doi:10.1001/jama.2020.14635



All authors are with the Department of Medicine, Brigham and Women's Hospital and Harvard Medical School. **Robert M. Stern, MD, EdM***, is Instructor in Medicine; **Mary W. Montgomery, MD***, is Instructor in Medicine; **Nora Y. Osman, MD**, is Assistant Professor in Medicine; **Joel T. Katz, MD**, is Associate Professor in

Medicine; and **Maria A. Yialamas, MD**, is Assistant Professor in Medicine.

*Drs Stern and Montgomery served as co-first authors and contributed equally to the work.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

Corresponding author: Robert M. Stern, MD, EdM, Brigham and Women's Hospital, rstern@bwh.harvard.edu

Received March 5, 2021; revision received July 26, 2021; accepted July 28, 2021.