

Stereotype Threat and Gender Bias in Internal Medicine Residency: It is Still Hard to be in Charge



Annabel K. Frank, MD¹, Jackie J. Lin, MD², Sophia Bellin Warren, MD³, Justin L. Bullock, MD⁴, Patricia O'Sullivan, EdD², Lauren E. Malishchak, MD⁵, Rebecca A. Berman, MD¹, Maria A. Yialamas, MD⁵, and Karen E. Hauer, MD, PhD¹

¹Department of Medicine, University of California, San Francisco, San Francisco, CA, USA; ²School of Medicine, University of California, San Francisco, San Francisco, CA, USA; ³Department of Medicine, MetroHealth Medical Center, Cleveland, OH, USA; ⁴Department of Medicine, University of Washington, Seattle, WA, USA; ⁵Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

ABSTRACT

BACKGROUND: Despite similar numbers of women and men in internal medicine (IM) residency, women face unique challenges. Stereotype threat is hypothesized to contribute to underrepresentation of women in academic leadership, and exploring how it manifests in residency may provide insight into forces that perpetuate gender disparities.

OBJECTIVE: To quantify the prevalence of stereotype threat in IM residency and explore experiences contributing to that stereotype threat.

DESIGN: We used a mixed methods study design. First, we surveyed IM residents using the Stereotype Vulnerability Scale (SVS) to screen for stereotype threat. Second, we conducted focus groups with women who scored high on the SVS to understand experiences that led to stereotype threat.

PARTICIPANTS: The survey was sent to all IM residents at University of California, San Francisco (UCSF), in September–November 2019. Focus groups were conducted at UCSF in Spring 2020.

APPROACH: The survey included an adapted version of the SVS. For focus groups, we developed a focus group guide informed by literature on stereotype threat. We used a thematic approach to data analysis. The mixed methods design enabled us to draw meta-inferences by integrating the two data sources.

KEY RESULTS: Survey response rate was 61% (110/181). Women were significantly more likely than men to have a score indicating stereotype threat vulnerability (77% vs 0%, $p < 0.001$). Four themes from focus groups characterized women's experiences of gender bias and stereotype threat: gender norm tension, micro-aggressions and sexual harassment, authority questioned, and support and allyship.

CONCLUSIONS: Gender-based stereotype threat is highly prevalent among women IM residents. This phenomenon poses a threat to confidence and ability to execute patient care responsibilities, detracting from well-being and professional development. These findings indicate that, despite robust representation of women

in IM training, further attention is needed to address gendered experiences and contributors to women's vulnerability to stereotype threat.

KEY WORDS: stereotype threat; gender bias; internal medicine residency; women

J Gen Intern Med 39(4):636–42

DOI: 10.1007/s11606-023-08498-5

© The Author(s) 2023

INTRODUCTION

Women have achieved tremendous success in medical training and careers, and now comprise an increasing percentage of the physician workforce.¹ Studies across specialties have demonstrated improvements in various patient outcomes such as mortality, readmission, and procedural measures in patients treated by women compared to men physicians.^{2–4} Despite these successes, women continue to face unique challenges and inequities.

Women physicians experience lower sense of belonging, unequal work-life responsibilities, and higher rates of burnout than their male counterparts.^{1, 5–9} Even women in academic leadership positions experience bullying which impacts their performance, sense of inclusion, and desire for continued advancement.¹⁰ While women faculty hold more leadership positions than in the past, men are more likely to hold clinical affairs and research affairs deanships whereas women are more likely to hold admissions, diversity, and student affairs deanships, which command lower compensation.¹¹ Disparities in compensation exist at all ranks and across specialties, even when controlling for clinical volume, publications, and external funding.¹² Furthermore, troubling descriptions of sexual harassment and exclusionary culture persist, more commonly described in surgical specialties.^{7, 13, 14} In specialties with greater gender parity such as internal medicine (IM), where in academic year 2021–2022 women accounted for 47% of residents,¹⁵ it may be assumed the culture is favorable for women. However, descriptions of inequity and bias persist. Gender bias in IM residency, encountered from both patients and other providers, has been shown to pose a threat to fair assessment, clinical learning, and leadership development.^{16–18}

Prior presentations An abstract describing our preliminary findings was presented at the April 2021 AAMC Virtual Group on Educational Affairs (GEA) Joint Regional Conference.

Received October 11, 2023

Accepted October 16, 2023

Published online November 20, 2023

To understand impediments to women physicians' career success, we can look towards theoretical frameworks about how women's identities shape their experiences. Feminist scholars advocate that, to inform social change, we need to examine the power structures and intersectional forces that underly social injustices.^{19, 20} A feminist theoretical approach demands examination of the cultural context and lived experiences of women. One theoretical framework critical to understanding women's experiences is stereotype threat. Stereotype threat is a phenomenon in which performance is impaired due to fear of confirming negative stereotypes, and women in historically male-dominated fields, such as math, are particularly susceptible.²¹ Stereotype threat illuminates how women's confidence and performance are influenced by gender stereotypes, and scholars have hypothesized it may contribute to underrepresentation of women in academic leadership positions.²²

Survey assessment has shown that women in academic medicine are vulnerable to stereotype threat;⁶ however, no studies have explored the experiences that promote this stereotype threat in IM training. Through this mixed methods study, we sought to understand the experiences of gender bias in IM residency and how these experiences perpetuate stereotype threat vulnerability. The objective of this mixed methods study was to identify women IM residents experiencing stereotype threat based on quantitative data and explore their experiences through qualitative data.

METHODS

Design

Using an interpretivist research paradigm, we conducted mixed methods research employing an explanatory sequential design with two phases.^{23, 24} This methodology enables understanding the prevalence of stereotype threat and exploring the concept in greater depth through qualitative data collection specifically with those who experienced stereotype threat. In phase 1, we administered a quantitative survey to assess for stereotype threat and inform our sampling for phase 2. In phase 2, we recruited women who endorsed stereotype threat in the survey. We used focus groups to stimulate discussion about cultural phenomena and enable participants to compare experiences. We chose a general thematic approach to address the research question around women experiencing stereotype threat and gender bias during residency training.²⁵

Team Composition

The study team included eight investigators who identify as women and one man. Three team members are IM clinical faculty with experience in program leadership and education scholarship and one is a faculty leader dedicated to education scholarship. Throughout the project, we had team members

at several stages of training including three IM residents who went on to fellowship, one medical student, and one chief resident.

Setting and Participants

The study was conducted at University of California, San Francisco (UCSF). We surveyed all IM residents at UCSF between September and November 2019. We invited survey respondents who endorsed stereotype threat, defined as a score > 18, to participate in focus groups.^{21, 26} Focus group participants were compensated \$25. The UCSF institutional review board approved the study (IRB #19-28,224).

Phase 1: Quantitative Survey

All UCSF IM residents received email invitations to a survey (Appendix 1) with questions about stereotype threat adapted from the Stereotype Vulnerability Scale (SVS), a survey tool initially developed to measure stereotype threat among women in math classes.²⁷ In a recent study, the SVS was adapted for medical students to assess students' perceptions of stereotype against their race as predictor of vulnerability to stereotype threat.²⁶ We similarly adapted the SVS by specifying residency as the context and substituting medical ability for math ability. Additional survey questions included demographic characteristics: assigned gender at birth, gender identity, sexual orientation, race/ethnicity, and age. We chose to survey all residents to compare responses among genders.

We calculated descriptive statistics for all demographics. The SVS score was calculated by summing points from each of the six items with 1 (strongly disagree) to 5 (strongly agree). Two items were reverse-coded so that a higher score indicated increased vulnerability to stereotype threat. SVS score could range from 6 to 30. We characterized SVS scores > 18 as a positive screen for vulnerability to stereotype threat based on prior implementation of the SVS.²⁶ A two-tailed Fisher exact test was used to compare the proportion of women versus men who scored positive for stereotype threat.

Phase 2: Focus Groups

Three investigators (A.K.F., J.L.B., K.E.H.) developed a semi-structured focus group guide (Appendix 2) informed by best practices for focus group development and literature on stereotype threat and gender disparities in medicine.^{5, 6, 17, 28, 29} A.K.F. conducted one pilot focus group in-person with seven recently graduated IM residents including several chief residents. The pilot focus group yielded insight into potential themes which informed probing questions for subsequent focus groups. Pilot participants were eligible for and agreed to participate in the study, and investigators included the pilot in the analysis. A.K.F. conducted two

additional focus groups via videoconference platform. We audio-recorded and transcribed all focus groups' verbatim and deidentified transcripts before analysis. These three focus groups achieved sufficiency using conceptual depth criteria based on the range of clinical experiences and interactions described by participants that informed development of themes: investigators' identification of a range of exemplars of stereotype threat and gendered experiences that they compared, contrasted, and discussed to refine themes; and resonance of participants' stories and our themes with literature on gendered experiences.³⁰

We used thematic analysis.³¹ A.K.F. reviewed all three transcripts to develop a preliminary codebook. Three investigators (A.K.F., S.B.W., K.E.H.) used the draft codebook to code a transcript individually. After discussion, A.K.F. and K.E.H. refined the codebook for clarity and completeness. Using the finalized codebook, two investigators coded each transcript; A.K.F. coded all transcripts and one other investigator coded each transcript (J.J.L., S.B.W., K.E.H.). We reconciled differences in coding through discussion. Interviews were coded using Microsoft Word and uploaded into Dedoose for further analysis.³²

Five investigators (A.K.F., J.J.L., J.L.B., P.O'S., K.E.H.) reviewed excerpts grouped by code and summarized themes. Investigators iteratively reviewed excerpts grouped by code and discussed the findings to ensure we sufficiently captured the essential themes and variations in perspective.³⁰

Reflexivity

We considered reflexivity throughout the analysis by discussing and documenting our reactions and emotions to the data informed by our personal experiences and perspectives.³³ Team members at different stages of their careers shared how the data resonated with them and reflected on how gender bias manifests at each stage.

Data Integration and Metainferences

Two essential characteristics of mixed methods research are integration of the two data sources and drawing metainferences from the integration.^{23, 34} We used survey results to inform questions and probes in the focus groups. We drew insights through analysis of survey and focus group data, and these metainferences are summarized in the results and elaborated in the discussion.

RESULTS

Phase 1: Quantitative Survey Assessing Stereotype Threat

The survey response rate was 61% (110/181). Of the respondents, 64 identified as women, 1 as non-binary, and 45 as men. We limited statistical analyses to participants who

identified as women or men due to sample size limitations; of note, the non-binary individual did screen positive for stereotype threat (SVS score > 18). Women were significantly more likely than men to have a score indicating experiencing gender-based stereotype threat (77% vs 0%, $p < 0.001$).

Phase 2: Focus Groups

Of the 110 survey respondents, 39 offered to participate in focus groups. Thirteen were ineligible because they were men or non-binary, and 6 were ineligible because they screened negative for vulnerability to stereotype threat. We recruited from the 20 willing and eligible survey respondents and conducted 2 focus groups with 7 and 3 participants. The pilot focus group included 7 recently graduated residents. In total, 17 women participated in 3 focus groups. We identified 4 themes describing women's experiences of gender bias and impacting vulnerability to stereotype threat: gender norm tension, microaggressions and sexual harassment, authority questioned, and support and allyship.

Gender Norm Tension

Residents described gender norm tensions while leading teams, making clinical decisions, and giving orders to nurses and other staff. They felt that societal expectations of their behavior as women were at odds with requirements for successfully performing their job. Participants universally described challenges to fulfill stereotypically feminine approaches of being collaborative and friendly, while still being "*decisive and authoritative*" (FG3-7). One participant described "*I get feedback, I'm not assertive enough....so then I'm more assertive. And then I get feedback that I'm being confrontational...so then I'm way less assertive. And then I get feedback again that I smile too much and that I am not assertive enough. And so it's impossible to strike this perfect balance*" (FG3-1). This type of negative feedback reinforced to participants "*how thin of a tight rope you are traversing*" (FG2-3) and they felt "penalized" (FG3-4) when attempting to be authoritative.

Such gender norm tension also manifested in participants' descriptions of the differing expectations of them compared to men peers: "*the expectation is that you always have to be emotionally present as a woman, and when you're not, it's like you're failing. And then for our male colleagues, the expectation is not that they always need to be emotionally present. So when they aren't, that's fine...when they are, it's extra amazing*" (FG2-3). Another participant described how men colleagues received more credit, especially with regard to stereotypically feminine qualities: "*everyone's like, 'Wow, he's a really good listener...People are impressed with him for listening instead of just expecting it like they would of you*" (FG3-2). These multifaceted expectations consumed mental energy: "*exhausting...to be in that space where*

you're like, am I nice enough? Am I supportive enough? Am assertive enough? Am I all these things enough?" (FG2-3). These feelings led women to conclude that they were held to higher standards than men, with wide-ranging and conflicting expectations for their behavior.

Nonetheless, many participants took pride in the perception that they brought valuable interpersonal skills and high emotional intelligence. One participant shared her ability to *"empower different members of the team and bring in everyone's skills"* (FG2-2). Still, they felt when their communication and collaborative skills shined too much, their clinical strengths would go underrecognized. One participant recalled getting feedback from an attending, *"What a great job you do bringing snacks for the team'...I was so offended...write something about my clinical acumen"* (FG2-2). Despite taking pride in their emotional strengths, altogether participants felt burdened by the need to be both warm and authoritative simultaneously, in an environment in which other people seemed to prioritize their warmth and judge them differently than men's.

Microaggressions and Sexual Harassment

Participants experienced frequent gender-based microaggressions from other care providers and patients, and overt sexual harassment from patients. These experiences caused participants to doubt their own potential. A common form of microaggression was being misidentified as non-physicians and *"incessantly being called nurse"* (FG1-4). Microaggressions threatened participants' sense of worth and capability as physicians: *"it gets reinforced, these notions that, 'Oh, you're not supposed to do this.' Because, 'Oh, are you a nurse?' Or, 'Are you sure you're a doctor? Have you graduated high school yet?'"* (FG3-2). These constant assumptions that participants were non-physicians seeded doubt in their abilities, promoting vulnerability to stereotype threat.

Participants recalled distracting, derogatory comments regarding their speech and appearance from other physicians and patients. One participant recalled *"I was doing an echo on a patient. The whole time, he's like, '... You're so pretty. I'm so glad I get to watch you while you do this echo'...completely degrades my position and my professional place...I'm just doing the echo and getting sexually harassed the whole time"* (FG3-1). While the constant microaggressions eroded participants' confidence, overtly sexist incidents stood out in their intensity: *"those big mega experiences, that shook my confidence for months afterwards and still makes me afraid for when I'm going to be in that situation again"* (FG3-5). Another participant worried about the impact on learning: *"If you're being harassed by a patient, what is the chance that you're actually going to remember any of the clinical knowledge that happened in that encounter?... Not a lot of memory forming when the cortisol is coursing through you"* (FG3-7). Altogether, microaggressions and sexual assault made participants feel they did not belong in the physician role, detracted from their learning, and caused self-doubt.

Authority Questioned

Participants reported that patients, other physicians, and other care providers regularly questioned their authority as resident physicians. Lacking authority made it difficult to carry out their responsibilities, particularly in high-acuity clinical situations. Multiple participants described leading code blues as particularly challenging: *"I was doing it in the way that you're supposed to, which is speak loudly and clearly, and make clarifications...the feedback I was getting was, 'Don't you dare tell me what to push. We've been running codes for 20 years. We know how to do this. Don't you dare tell us how to do compressions'...later, I found out that the nurses who were on that night had given feedback to the ICU supervisor 'The code leader had been very bossy'"* (FG3-5). One participant summarized, *"when you know you're supposed to be in charge and there's no ambiguity, it's still hard to be in charge"* (FG3-4). Another participant described being interrupted and ignored while trying to give instructions to a team of nurses during a rapid response: *"[I] felt... what am I doing wrong here as a leader that I can't get this done?...I still struggle with that a little bit....every time I get that feedback or those types of interactions where my role is questioned, it shakes my confidence a little bit. Same with patients, when patients question it too, I'm like, do you really want me as your doctor?"* (FG2-1). Doubt from patients about their competence also negatively affected participants, including one who reported hearing: *"You seem very nice, but you really seem like you don't know what you're doing"* (FG2-2). Participants' experiences with lack of authority across multiple contexts diminished their confidence and led them to question their ability to perform their job, though all felt retrospectively that they were capable.

Because participants felt it was difficult to gain respect, some avoided situations that made them vulnerable, even at the expense of missing learning opportunities. Although a common physician practice at all levels is to run questions by colleagues, participants reported that engaging in this practice could prompt others to assume they were insufficiently competent. One resident described asking a colleague to check a patient with her: *"I'd been managing this patient the whole night and then had done a lot of good initial triaging and management, afterwards I left the room feeling like I had completely failed in a way...the way that this backup person came in...quite condescending, being like well, did you think about this? Should we do this too?"* (FG2-3). Similarly, participants hesitated to ask questions because it undermined their authority: *"asking questions, it's an invitation for someone to mansplain to me essentially"* (FG2-2). Participants deliberately avoided asking questions and considered this to have a direct negative impact on their learning. Even as senior team members, they had difficulty earning respect from interns, colleagues, and attendings: *"One of my male co-residents was like, 'No, that's absolutely wrong. This person is in a different type of shock'.... then the attending was like,*

'Oh, yeah. That's not right.' ...then we went into the room and did the exam all together. It was clear that my assessment was correct...[I] had been totally shut down in front of the entire 20 people there, and then actually I was right and no one bothered to say a damn thing" (FG3-3). These dynamics led participants to feel they needed to prove themselves through double checking and gathering evidence to be sure they were right before speaking up, or not speak at all.

Participants described ways of trying to exude authority. Many opted to wear a white coat and display their badges prominently: *"wear that badge that says, 'DOCTOR', in all capital letters like a shield"* (FG2-2). Participants ruminated about how to modify their appearance to gain respect and resented that they had to consider these matters: *"if I look more put together, will people respect me more or less?"* (FG3-3). Nonetheless, any authority they did have felt tenuous and rumination about how to establish authority consumed mental energy.

Support and Allyship

Participants related acts of support and allyship that helped overcome the consequences of gender bias and stereotype threat. One ICU attending pre-emptively acknowledged that men tend to crowd out women when viewing x-rays, encouraging men to step back and women to step up. Attendings also exhibited allyship by affirming clinical decisions or standing behind the participant so the team would address her during rounds. Participants recalled that both men and women displayed acts of allyship, though women tended to recognize and acknowledge microaggressions more frequently. Many participants described interactions in which no one responded in the moment but later acknowledged that gender bias had occurred; participants appreciated that they noticed but also highlighted this missed opportunity to educate the team in the moment.

Though support from colleagues was appreciated, some participants lamented that they did not have a stronger internal sense of worth. When seeking jobs and networking opportunities, they hesitated more than they observed their male colleagues doing: *"it wasn't until my husband, who's a man, was like, 'Your worth is here, and you're asking for things here' I feel like it's unfortunate that it takes a man to remind me that I'm worth more than I think I am"* (FG1-1). Participants appreciated when mentors, often women, shared their own struggles with confidence and offered encouragement. Although support in the form of allyship and mentorship helped participants cope with gender bias, they remained vulnerable to stereotype threat in an environment where the threats to confidence were frequent.

DISCUSSION

Our survey demonstrated that women IM residents are highly vulnerable to stereotype threat and focus group data elucidated the mechanisms and impacts of this stereotype

threat vulnerability. Stereotype threat affected self-esteem and performance through gender norm tension, microaggressions and sexual harassment, and questioning women's authority. While numeric representation of women and men in IM may suggest progress towards gender equity, our findings reveal that significant inequities persist.

We identified two meta-inferences encapsulating findings from our mixed methods study. First, gender expectations contradicted and therefore diminished women residents' capacity to be authoritative. Second, because others questioned their abilities and showed disrespect, women residents questioned their own capability. They ruminated over how to change perceptions of them, which detracted from honing their medical competency. Other qualitative studies have similarly shown that women residents had precarious relationships giving orders to nurses, particularly in high-acuity situations, and had trouble with how they were perceived when being assertive, which inhibited their ability to influence patient care.^{18, 35, 36} Data show women emergency medicine residents are rated lower across milestones, with particular gender differences found in procedures, airway management, and emergency stabilization.^{29, 37} A simulation study showed that when female and male actors led the same scripted code blue scenario, the male code leader was rated more highly in communication and leadership skills.³⁸ These findings alongside our data support the notion that simply being a woman in appearance, voice, and demeanor is enough to make that physician appear less of a leader in high-acuity medical scenarios.

Our data demonstrate that gender dynamics in residency had immense impact on confidence and cognitive load, which is a concern because these are mechanisms by which stereotype threat can impair performance and growth.³⁹ Though the participants in our study were succeeding, as evidenced by their status as women physicians and multiple having been selected as chief residents at a competitive residency program, their experiences of stereotype threat are concerning. They face substantial external and internalized obstacles every day. Participants described frequent microaggressions and sexual harassment; we grouped these together in our results as offenses existing along a continuum of frequent minor insults to infrequent, high-severity incidents.⁴⁰ These events take a toll on the affected residents, who often do not see a mechanism to report these events or potential benefit of doing so.⁴¹ Though it may seem that these women have "made it," our findings indicate there remains painful inequity.

Creating solutions to gender bias and the stereotype threat it perpetuates is challenging because of unconscious bias and external as well as internalized stereotypes. Women in our focus groups identified factors that helped them persevere: (1) role models and mentors who genuinely believed in them and pushed them, and (2) gaining experience, knowledge, and confidence in their abilities despite external skepticism.

Prior literature supports the notion that role models have an important role in mitigating stereotype threat and improving women's performance, for example, in mathematics.⁴²⁻⁴⁴ In addition to interactions with salient role models, several other interventions show promise for reducing stereotype threat, including education on bias and stereotype threat, self-affirmations, and promoting growth mindsets.^{22, 45-50} In a randomized study, a "gender bias habit-changing" workshop improved faculty awareness of gender bias and changed their behaviors to improve the climate for women in intervention departments and not in the control departments.⁴⁵ Importantly, stereotype threat can be reduced and academic performance improved when learners are encouraged to view intelligence as malleable rather than fixed.^{47, 48} Accordingly, women physicians should be encouraged to view leadership skills as teachable and attainable. Chronic stereotype threat can lead to domain disidentification which can manifest as abandonment of ambition.⁵¹ The goal of any intervention should be to reduce stereotype threat and prevent women residents from doubting their potential as leaders, a concerning possibility raised by our study and similar research,^{18, 35, 37} and a viewpoint that will make it hard for women to close leadership gaps in academic medicine.

Limitations to our study include that women who are most impacted by gender bias and vulnerable to stereotype threat may have been more likely to complete the survey and join focus groups. The number of focus group participants was modest. We were not able to explore intersectionality. The study was not designed to explore experiences of gender diverse trainees, including but not limited to transmen, transwomen, non-binary, or genderfluid individuals. Given our respondent pool, with a single non-binary participant, our team balanced the intended focus of the study, the risk of epistemic violence via exclusion of gender diverse people, and the risk of overgeneralizing the experience of a single individual.

CONCLUSIONS

This study demonstrates that women in IM residency are highly vulnerable to stereotype threat. The mixed methods design enabled us to elucidate the mechanisms by which stereotype threat arises and persists. Women are treated with a lack of authority and made to feel unfit for the physician role, impacting their confidence and efficacy. In addition to cultural efforts to reduce gender stereotypes and workplace bias, targeted interventions should be undertaken to alleviate resident vulnerability to stereotype threat.

Supplementary Information: The online version contains supplementary material available at <https://doi.org/10.1007/s11606-023-08498-5>.

Corresponding Author: Annabel K. Frank, MD; Department of Medicine, University of California, San Francisco, San Francisco, CA, USA (e-mail: annabel.frank@ucsf.edu).

Funding This research was supported by a grant from the AAMC Western Group on Educational Affairs (WGEA).

Conflict of Interest None of the authors have relationships with organizations that have a financial interest in the subject or content of the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

1. **Butkus R, Serchen J, Moyer DV, et al.** Achieving gender equity in physician compensation and career advancement: a position paper of the American College of Physicians. *Ann Intern Med.* 2018;168(10):721-723.
2. **Tsugawa Y, Jena AB, Figueroa JF, Orav EJ, Blumenthal DM, Jha AK.** Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians. *JAMA Intern Med.* 2017;177(2):206-213.
3. **Mehrotra A, Morris M, Gourevitch RA, et al.** Physician characteristics associated with higher adenoma detection rate. *Gastrointest Endosc.* 2018;87(3):778-786.e5.
4. **Wallis CJ, Ravi B, Coburn N, Nam RK, Detsky AS, Satkunasivam R.** Comparison of postoperative outcomes among patients treated by male and female surgeons: a population based matched cohort study. *BMJ.* 2017;359:j4366.
5. **Pololi LH, Civian JT, Brennan RT, Dottolo AL, Krupat E.** Experiencing the culture of academic medicine: gender matters, a national study. *J Gen Intern Med.* 2013;28(2):201-207.
6. **Fassiotto M, Hamel EO, Ku M, et al.** Women in academic medicine: measuring stereotype threat among junior faculty. *J Womens Health.* 2016;25(3):292-298.
7. **Hu YY, Ellis RJ, Hewitt DB, et al.** Discrimination, abuse, harassment, and burnout in surgical residency training. *N Engl J Med.* 2019;381(18):1741-1752.
8. **Winkel AF, Telzak B, Shaw J, et al.** The role of gender in careers in medicine: a systematic review and thematic synthesis of qualitative literature. *J Gen Intern Med.* 2021;36(8):2392-2399.
9. **Stead W, Manne-Goehler J, Blackshear L, et al.** Wondering if I'd get there quicker if I was a man: factors contributing to delayed academic advancement of women in infectious diseases. *Open Forum Infect Dis.* 2023;10(1):ofac660.
10. **Iyer MS, Way DP, MacDowell DJ, Overholser BM, Spector ND, Jaggi R.** Bullying in academic medicine: experiences of women physician leaders. *Acad Med.* 2023;98(2):255.
11. **Dandar VM, Lautenberger DM.** Exploring Salary Equity Among Medical School Leadership. Washington, DC: Association of American Medical Colleges; 2022. <https://store.aamc.org/leadership-compensation-equity-report-exploring-salary-equity-among-medical-school-leadership.html>. Accessed October 26, 2023.
12. **Jena AB, Olenki AR, Blumenthal DM.** Sex differences in physician salary in US public medical schools. *JAMA Intern Med.* 2016;176(9):1294.
13. **Barnes KL, McGuire L, Dunivan G, Sussman AL, McKee R.** Gender bias experiences of female surgical trainees. *J Surg Educ.* 2019;76(6):e1-e14.
14. **Cochran A, Hauschild T, Elder WB, Neumayer LA, Brasel KJ, Crandall ML.** Perceived gender-based barriers to careers in academic surgery. *Am J Surg.* 2013;206(2):263-268.

15. ACGME Data Resource Book Academic Year 2021-2022. Chicago, IL: Accreditation Council for Graduate Medical Education; 2022. https://www.acgme.org/globalassets/pfassets/publicationsbooks/2021-2022_acgme_databook_document.pdf. Accessed October 26, 2023.
16. Wheeler M, de Bourmont S, Paul-Emile K, et al. Physician and trainee experiences with patient bias. *JAMA Intern Med.* 2019;179(12):1678.
17. Klein R, Julian KA, Synder ED, et al. From the Gender Equity in Medicine (GEM) workgroup. Gender bias in resident assessment in graduate medical education: review of the literature. *J Gen Intern Med.* 2019;34(5):712-719.
18. Kolehmainen C, Brennan M, Filut A, Isaac C, Carnes M. Afraid of being "witchy with a B": a qualitative study of how gender influences residents' experiences leading cardiopulmonary resuscitation. *Acad Med.* 2014;89(9):1276-1281.
19. Sharma M. Applying feminist theory to medical education. *The Lancet.* 2019;393(10171):570-578.
20. Crenshaw K. *On Intersectionality: Essential Writings.* New York: The New Press; 2017.
21. Spencer SJ, Steele CM, Quinn DM. Stereotype threat and women's math performance. *J Exp Soc Psychol.* 1999;35(1):4-28.
22. Burgess DJ, Joseph A, van Ryn M, Carnes M. Does stereotype threat affect women in academic medicine?. *Acad Med.* 2012;87(4):506-512.
23. Creswell JW, Clark VL. *Designing and Conducting Mixed Methods Research.* Thousand Oaks (CA): Sage Publications; 2017.
24. Bunniss S, Kelly DR. Research paradigms in medical education research. *Med Educ.* 2010;44(4):358-366.
25. Sandelowski M. What's in a name? Qualitative description revisited. *Res Nurs Health.* 2010;33(1):77-84.
26. Bullock JL, Lockspeiser T, Del Pino-Jones A, Richards R, Teherani A, Hauer KE. They don't see a lot of people my color: a mixed methods study of racial/ethnic stereotype threat among medical students on core clerkships. *Acad Med.* 2020;95(11S):S58-S66.
27. Spencer SJ. *The Effect of Stereotype Vulnerability on Women's Math Performance.* [Doctoral thesis]. Ann Arbor (MI): University of Michigan; 1993.
28. Morgan, DL. *The Focus Group Guidebook.* London: Sage Publications Limited; 1997.
29. Santen SA, Yamazaki K, Holmboe ES, Yarris LM, Hamstra SJ. Comparison of male and female resident milestone assessments during emergency medicine residency training: a national study. *Acad Med.* 2020;95(2):263-268.
30. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. *Med Teach.* 2020;42(8):846-854.
31. Dedoose Version 9.0.62. cloud application for managing, analyzing, and presenting qualitative and mixed method research data (2022). Los Angeles (CA): SocioCultural Research Consultants, LLC www.dedoose.com.
32. Nelson J. Using conceptual depth criteria: addressing the challenge of reaching saturation in qualitative research. *Qual Res.* 2017;17(5):554-570.
33. Barry CA, Britten N, Barber N, Bradley C, Stevenson F. Using reflexivity to optimize teamwork in qualitative research. *Qual Health Res.* 1999;9(1):26-44.
34. Schoonenboom J. *Routledge Handbook for Advancing Integration in Mixed Methods Research.* New York: Routledge; 2022. Chapter 4, Developing the Meta-Inference in Mixed Methods Research through Successive Integration of Claims. p. 55-70.
35. Bartels C, Goetz S, Ward E, Carnes M. Internal medicine residents' perceived ability to direct patient care: impact of gender and experience. *J Womens Health.* 2008;17(10):1615-1621.
36. Mueller AS, Jenkins TM, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender differences in attending physicians' feedback to residents: a qualitative analysis. *J Grad Med Educ.* 2017;9(5):577-585.
37. Dayal A, O'Connor DM, Gadhri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med.* 2017;177(5):651-657.
38. Ju M, van Schaik SM. Effect of professional background and gender on residents' perceptions of leadership: *Acad Med.* 2019;94:S42-S47.
39. Pennington CR, Heim D, Levy AR, Larkin DT. Twenty years of stereotype threat research: a review of psychological mediators. *Pavlova MA, ed. PloS one.* 2016;11(1):e0146487.
40. Gartner RE, Sterzing PR. Gender microaggressions as a gateway to sexual harassment and sexual assault: expanding the conceptualization of youth sexual violence. *Affilia.* 2016;31(4):491-503.
41. de Bourmont SS, Burra A, Nouri SS, et al. Resident physician experiences with and responses to biased patients. *JAMA Netw Open.* 2020;3(11):e2021769.
42. McIntyre RB, Paulson RM, Lord CG. Alleviating women's mathematics stereotype threat through salience of group achievements. *J Exp Soc Psychol.* 2003;39(1):83-90.
43. Marx DM, Ko SJ. Superstars "like" me: the effect of role model similarity on performance under threat. *Eur J Soc Psychol.* 2012;42(7):807-812.
44. Marx DM, Roman JS. Female role models: protecting women's math test performance. *Pers Soc Psychol Bull.* 2002;28(9):1183-1193.
45. Carnes M, Devine PG, Manwell LB, et al. Effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Acad Med.* 2015;90(2):221-230.
46. Girod S, Fassiotto M, Grewal D, et al. Reducing implicit gender leadership bias in academic medicine with an educational intervention. *Acad Med.* 2016;91(8):1143-1150.
47. Aronson J, Fried CB, Good C. Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *J Exp Soc Psychol.* 2002;38(2):113-125.
48. Blackwell LS, Trzesniewski KH, Dweck CS. Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Dev.* 2007;78(1):246-263.
49. Johns M, Schmader T, Martens A. Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychol Sci.* 2005;16(3):175-179.
50. Martens A, Johns M, Greenberg J, Schimel J. Combating stereotype threat: the effect of self-affirmation on women's intellectual performance. *J Exp Soc Psychol.* 2006;42(2):236-243.
51. Woodcock A, Hernandez PR, Estrada M, Schultz PW. The consequences of chronic stereotype threat: domain disidentification and abandonment. *J Pers Soc Psychol.* 2012;103(4):635-646.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.