Damned If You Do, Damned If You Don't: Bias in Evaluations of Female Resident Physicians

Esther K. Choo, MD, MPH

Recently, *Modern Healthcare* published a list of the 50 most influential physician leaders of the past year.[1] It included just 7 women (14%), a vivid demonstration of the disparities that have depressed the progress of women in medicine. Numerous studies have documented physician gender disparities across domains of salary,[2–5] promotion,[6] and leadership roles.[7] While the mechanisms are not fully understood, these studies suggest the problem is not easily explained away by women's ambition, selection of specialty, maternity leave, or part-time work; nor is it automatically improving, even 20 years after medical schools began accepting roughly equal proportions of men and women.[8]

Expectations for women's performance are universally lower. Research in the social sciences has demonstrated that changing a fictional person's gender from male to female alone reduces how the individual is evaluated in terms of competence, "hireability," and suggested salary.[9] Further, women are expected to demonstrate communal rather than agentic traits. Thus, even when women have inherently strong leadership skills, there can be pressure to hide them to avoid cognitive dissonance in those with preconceived notions about how women should behave, and the subsequent backlash.[10]

A study by Mueller et al[11] in this issue of the *Journal of Graduate Medical Education* demonstrates how these forces might play out in the day-to-day experiences of physicians in training. In this qualitative analysis of comments in 1317 direct observation evaluations of third-year residents by emergency medicine faculty, the authors found residents were held to a standard of performance delineated mostly by classically "masculine" behaviors and traits (eg, confident, decisive, aggressive), and that women and men received different types of evaluative comments with respect to these behaviors and traits.

Women more often received comments of a polarized and contradictory nature, particularly for these classically male behaviors. For example, female senior residents were chided for lacking autonomy and for demonstrating too much of it, whereas men were almost uniformly praised for behaviors consistent with this trait—even to the extent that 1 male resident's argumentativeness was interpreted positively as asserting confidence. Women in the study, more often than men, received critical evaluations. While quantitative data are not the focus of a qualitative study, there is a stirring narrative in the long strings of zeros in the "strong negative criticism" column for male residents. Male residents also dominated the top half of a rank list based on evaluation scores. The picture that emerges is that there is an unspoken consensus around a standard set of traits desired in emergency medicine residents, and yet women dare not own them. Either way, they suffer in evaluations.

Receiving conflicting information in evaluation is a nightmare from an adult-learning perspective. For example, if a person is uniformly praised for effective communication with families, that person likely will continue to cultivate that component of his or her behavior. It is less clear how a resident might respond, and how clinical behavior might be fine-tuned, if she is told, for example, that she is both receptive to attending guidance and not receptive enough. Might a female physician's performance be compromised because her learning milieu includes this kind of inconsistent and confusing direction?

The subtle messaging about "appropriate" behavior also is important, because individuals tend to rise or sink to the occasion, depending on what expectations are subtly or overtly broadcast to them. Stereotypes about what men and women are good at are established and perpetuated from an early age, often by those responsible for their learning. By elementary school, girls internalize and respond to these stereotypes. For example, girls are shown to perform worse on math tests when they are given gender cues before testing, a phenomenon called "stereotype susceptibility."[12,13] The same phenomenon has been observed among female faculty in medicine.[14] The dilemma faced by women who buck gender stereotypes was captured in a study in which female residents divulged they felt they had to apologize for the authoritative, "counternormative" behaviors used while directing cardiopulmonary resuscitation.[15]

I hope readers of the study by Mueller et al[11] will be struck by how complex the solutions to this

DOI: http://dx.doi.org/10.4300/JGME-D-17-00557.1

problem will be, and how urgent it is that medicine pursue them to allow our female trainees to spend less time learning how to walk the fine line between normative and counternormative behaviors and more time simply learning to be physicians. The evaluation criteria, both explicit (eg, Accreditation Council for Graduate Medical Education competencies) and implicit (ie, the consensus "valued traits" that emerged from this study) must be reevaluated for bias toward typically male traits. This reassessment will need to include the acknowledgement that a wide range of qualities are likely to benefit clinical practice and patient outcomes.[16]

Program leaders must learn to review evaluations with a gender lens, to identify patterns of inconsistent or unduly critical evaluation that may reflect bias rather than actual clinical performance. In the meantime, physician instructors can bring self-awareness to the bedside by performing routine self-assessments of how our opinions, played out in evaluations and feedback, may be systematically biased—and harm our trainees' learning and self-confidence.

To be clear, both male and female instructors are guilty of these biases. Even so-called gender bias experts are susceptible. In a feedback moment I now deeply regret, I once told a female resident that she spoke too softly when she led resuscitations. "I have a vocal cord dysfunction," she told me bluntly. "That's as loud as I get." Over the next year, I witnessed her command the room many times effortlessly with her soft voice. It turns out that her incredible knowledge base, her clarity of thought, and her decisiveness around clinical management—not her inability to live up to my fixed notions of what a resuscitation team leader looked and sounded like—were the key elements of her effectiveness as a budding physician. My notions changed, and so should those of others across the house of medicine.

References

1. 50 most influential physician executives and leaders 2016. *Modern Healthcare*. http://www.modernhealthcare.com/community/50-most-influential/ 2016. Accessed August 3, 2017.
2. Jena AB, Olenski AR, Blumenthal DM. Sex differences in physician salary in US public medical schools. *JAMA Intern Med*. 2016;176(9):1294–1304.
3. Seabury SA, Chandra A, Jena AB. Trends in the earnings of male and female health care professionals in the United States, 1987 to 2010. *JAMA Intern Med*. 2013;173(18):1748–1750.
4. Lo Sasso AT, Richards MR, Chou CF, et al. The $16,819 pay gap for newly trained physicians: the unexplained trend of men earning more than women. *Health Aff (Millwood)*. 2011;30(2):193–201.
5. Ly DP, Seabury SA, Jena AB. Differences in incomes of physicians in the United States by race and sex: observational study. *BMJ*. 2016;353:i2923.
6. Jena AB, Khullar D, Ho O, et al. Sex differences in academic rank in US medical schools in 2014. *JAMA*. 2015;314(11):1149–1158.
7. Wehner MR, Nead KT, Linos K, et al. Plenty of moustaches but not enough women: cross sectional study of medical leaders. *BMJ*. 2015;351:h6311.
8. Association of American Medical Colleges. Table 1: medical students, selected years, 1965–2010. https://www.aamc.org/download/170248/data/2010_table1.pdf. 2010. Accessed August 3, 2017.
9. Moss-Racusin CA, Dovidio JF, Brescoll VL, et al. Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci U S A*. 2012;109(41):16474–16479.
10. Rudman LA, Glick P. Prescriptive gender stereotypes and backlash toward agentic women. *J Soc Issues*. 2001;57(4):743–762.
11. Mueller AS, Jenkins TM, Osborne M, et al. Gender differences in attending physicians' feedback to residents: a qualitative analysis. *J Grad Med Educ*. 2017;9(5):xx–xx.
12. Ambady N, Shih M, Kim A, et al. Stereotype susceptibility in children: effects of identity activation on quantitative performance. *Psychol Sci*. 2001;12(5):385–390.
13. Spencer SJ, Steele CM, Quinn DM. Stereotype threat and women's math performance. *J Exp Soc Psychol*. 1999;35(1):4–28.
14. Fassiotto M, Hamel EO, Ku M, et al. Women in academic medicine: measuring stereotype threat among junior faculty. *J Womens Health (Larchmt)*. 2016;25(3):292–298.
15. Kolehmainen C, Brennan M, Filut A, et al. 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.
16. Tsugawa Y, Jena AB, Figueroa JF, et al. Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians. *JAMA Intern Med*. 2017;177(2):206–213.

Esther K. Choo, MD, MPH, is Associate Professor, Center for Policy and Research in Emergency Medicine, Department of Emergency Medicine, Oregon Health & Science University.

Corresponding author: Esther K. Choo, MD, MPH, Oregon Health & Science University, CDW-EM, 3181 SW Sam Jackson Park Road, Portland, OR 97239, 503.494.1440, chooe@ohsu.edu