Autonomy in the Operating Room: A Multicenter Study of Gender Disparities During Surgical Training

Jenny X. Chen, MD
Edward H. Chang, PhD
Francis Deng, MD
Shari Meyerson, MD, MEd
Brian George, MD, MAEd
Elliott D. Kozin, MD
Stacey T. Gray, MD

ABSTRACT

Background Gender disparities are prevalent in medicine, but their impact on surgical training is not well studied.

Objective To quantify gender disparities in trainee intraoperative experiences and explore the variables associated with ratings of surgical autonomy and performance.

Methods From September 2015 to May 2019, attending surgeons and trainees from 71 programs assessed trainee autonomy on a 4-level Zwisch scale and performance on a 5-level modified Dreyfus scale after surgical procedures. Multivariable regression models were used to examine the association of trainee gender with autonomy and performance evaluations.

Results A total of 3255 trainees and attending surgeons completed 94 619 evaluations. Attendings gave lower ratings of operative autonomy to female trainees than male trainees when controlling for training level, attending, and surgical procedure (effect size B = –0.0199, $P = .008$). There was no difference in ratings of autonomy at the beginning of training ($P = .32$); the gap emerged as trainees advanced in years (B = –0.0163, $P = .020$). The gender difference in autonomy was largest for the most complex cases (B = –0.0502, $P = .002$). However, there was no difference in attending ratings of surgical performance for female trainees compared to male trainees (B = –0.0124, $P = .066$). Female trainees rated themselves as having less autonomy and worse performance than males when controlling for training level, attending, procedure, case complexity, and attending ratings (autonomy B = –0.0669, $P < .001$; performance B = –0.0704, $P < .001$).

Conclusions While there was no significant difference in ratings of operative performance, a small difference between ratings of operative autonomy for female and male surgical trainees was identified.

Introduction

As of 2019, women represent the majority of US medical students[1] and 41% of general surgery residents.[2] Despite the growing numbers, female trainees continue to face discrimination, harassment, and high rates of burnout.[3,4] In a recent survey of general surgery residents, 65% of female trainees reported gender discrimination and 20% reported sexual harassment during training.[5] In the context of this work environment, gender inequalities have the potential to influence training.[6] Although numerous survey studies document the subjective experiences of female trainees, limited quantitative evidence on gender disparities during training has been reported.[3]

A critical component of surgical training is the graduated levels of autonomy that residents gain in the operating room over years of apprenticeship,[7] and there are growing concerns about modern trainees' readiness for independent practice.[8] Previous small-scale studies have found a significant effect of gender on resident autonomy in general surgery[9,10] and thoracic surgery.[11] However, it is unclear how pervasive these findings are across surgical subspecialties. Herein, we analyzed operative evaluations from general surgery, surgical specialty (including plastic surgery, otolaryngology, orthopedic surgery, and vascular surgery), and surgical fellowship programs to determine if faculty and trainee assessments of trainee intraoperative autonomy and performance are associated with gender.

Methods

We collected postoperative assessments of surgical trainees made by trainees and their supervising attending surgeons between September 2015 and May 2019 on a smartphone application from the Society for Improving Medical Professional Learning (SIMPL, Boston, MA).[12,13] Study programs were members of SIMPL and used the SIMPL OR application as a feedback tool for trainees for a fee. Faculty and trainees at each site underwent standardized rater training sessions.[14] In addition to reporting the operative case complexity ("easiest one-third,"

DOI: http://dx.doi.org/10.4300/JGME-D-21-00217.1

Editor's Note: The online version of this article contains additional data from the study.

TABLE 1
Geographic Locations and Specialties of Surgical Programs Studied From September 2015 to May 2019

Program Type	No. of Programs (%)	No. of Ratings (%)
Total programs	71 (100)	94 619 (100)
Northeast programs	24 (34)	39 774 (42)
Midwest programs	24 (34)	31 734 (34)
Southeast programs	9 (13)	12 604 (13)
West programs	9 (13)	4119 (4)
Southwest programs	5 (7)	6388 (7)
Types of surgical programs		
General surgery	39 (55)	78 726 (83)
Advanced fellowships	18 (25)	10 671 (11)
Plastic surgery	4 (6)	531 (1)
Otolaryngology–head and neck surgery	3 (4)	2626 (3)
Orthopedic surgery	2 (3)	397 (< 1)
Vascular surgery	2 (3)	194 (< 1)
Oral and maxillofacial surgery	1 (1)	863 (1)
Neurological surgery	1 (1)	291 (< 1)
Urology	1 (1)	320 (< 1)

Objectives
We aimed to quantify gender differences in trainee operative experiences and explore the variables associated with assessments of surgical autonomy and performance.

Findings
Assessments of 1927 surgical trainees were collected at 71 training programs using a novel smartphone application. There was a small but significant gender gap in attending ratings of resident surgical autonomy, which emerged over time during training and was most significant for complex cases.

Limitations
Assessments were voluntary and susceptible to selection bias.

Bottom Line
As they represent potential harm to young surgeons, these gender differences in operative training must be further studied before designing targeted interventions.

"average one-third," or "hardest one-third" of similar cases), the application asked trainees and faculty to assess the trainee's level of autonomy and performance after each surgery. Assessments were encouraged but not mandatory. Autonomy was rated on the 4-level Zwisch scale describing the autonomy achieved for > 50% of critical portions of the procedure: show and tell, active help, passive help, and supervision only (online supplementary data TABLE 1 describes the levels).[15] If the autonomy level was greater than show and tell, the trainee's performance was then rated on a 5-level modified Dreyfus scale: unprepared/critical deficiency, inexperienced with procedure, intermediate performance, practice ready performance, and exceptional performance (online supplementary data TABLE 2 describes the levels).[12] Faculty/residents could see the other party's evaluation after both assessments were submitted. Participants could review all prior evaluations associated with their accounts. When participants were registered, they reported their genders. Due to limitations in the SIMPL OR application, participants' genders were collected in gender categories (male and female). Assessments

TABLE 2
Demographics of Surgical Trainees Studied From September 2015 to May 2019

Type of Trainee	No. of Participants (%)	No. of Ratings (%)	Ratings Per Participant	
			Mean (SD)	*P* Value
Total residents[a]	1863 (100)	52 241 (100)	28.0 (47.2)	
Male residents	1100 (59)	33 102 (63)	30.1 (52.0)	.024[b]
Female residents	763 (41)	19 139 (37)	25.1 (39.2)	
Total fellows	85 (100)	2747 (100)	32.3 (76.5)	
Male fellows	54 (64)	1841 (67)	34.1 (90.4)	.79
Female fellows	31 (36)	916 (33)	29.5 (43.8)	
Total trainees	1927 (100)	54 998 (100)	28.5 (49.2)	
Male trainees	1139 (59)	34 943 (64)	30.7 (54.9)	.022[b]
Female trainees	788 (41)	20 055 (36)	25.5 (39.5)	
Total attendings	1328 (100)	39 621 (100)	29.8 (53.4)	
Male attendings	943 (71)	30 367 (77)	32.2 (56.1)	.011[b]
Female attendings	385 (29)	9254 (23)	24.0 (45.6)	

[a] The number of total trainees is less than the number of residents and fellows combined as a small number of residents became fellows during the study period.
[b] Indicates significance at the level of $P < .05$.

without user gender were excluded (n = 365, 0.38% of assessments). A pilot study in a single surgical specialty was analyzed in a prior publication.[9]

Statistical Analysis

To determine the effect of trainee gender on attending ratings of autonomy and performance, as well as trainee self-assessments of autonomy and performance, ordinary least squares regressions were used. When analyzing attending ratings of autonomy and performance, the main analyses include fixed effects for attending, trainee postgraduate year (PGY) training level, and surgical procedure, with standard errors clustered by attending. Models with increasing numbers of controls were used to show the robustness of results to different model specifications. When analyzing trainee self-assessments, we included fixed effects for attendings, paired attending ratings, training levels, attendings, case complexity (as rated by attendings), and surgical procedure. Standard errors were clustered by trainee. We also used *t* tests and chi-square tests where indicated. Stata/SE 14.2 (StataCorp LLC, College Station, TX) was used for analyses. An explanation of statistical modeling choices is provided in the online supplementary data. This study was deemed exempt from review by the Massachusetts Eye and Ear Institutional Review Board.

Results

The 71 training programs included in this study were diverse in geography and specialty (TABLE 1). Fifty-three surgical residency programs were represented, including 39 general surgery programs and 14 surgical specialty programs. Eighteen surgical fellowship programs were included (acute care surgery, bariatric surgery, breast surgery, cardiothoracic surgery, hand surgery, pediatric surgery, surgical oncology, transplant surgery). A total of 3255 surgical trainees and attendings participated. Among these, 1927 were surgical trainees, of whom 59% (n = 1139) were male and 41% (n = 788) were female. A total of 1328 attendings participated, of whom 71% (n = 943) were male and 29% (n = 385) were female.

Trainees and attendings completed 94 619 evaluations. Forty-two percent of evaluations were from attendings (n = 39 621), while 58% were from trainees (n = 54 998; TABLE 2). The mean number of ratings per male trainee was 30.7 (SD 54.9), and the mean number of ratings per female trainee was 25.5 (SD 39.5; unpaired *t* test; $P = .022$). The mean number of ratings per male attending was 32.2 (SD 56.1), and the mean number of ratings per female attending was 24.0 (SD 45.6; unpaired *t* test; $P = .011$). Among attending evaluations, 5324 cases (13%) were rated as the "easiest one-third," 25 107 (63%) were rated as the "average one-third," and 9190 (23%) were the "hardest one-third" of similar cases.

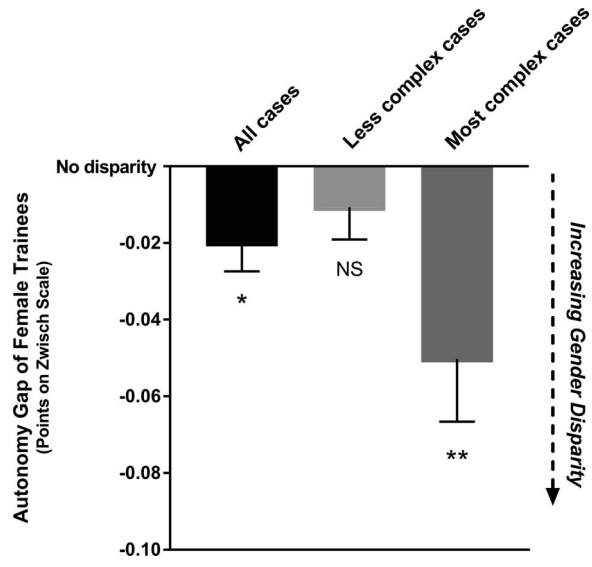

FIGURE 1
Difference in Autonomy Attained by Female Trainees Compared With Male Trainees, Measured on Zwisch Scale of Surgical Autonomy

Abbreviation: NS, no significant difference.

Note: Regressions controlled for attending, level of training, and surgical procedure. Less complex cases represent cases rated as the "easiest one-third" or "average one-third" of similar cases. Most complex cases were rated as the "hardest one-third" of similar cases. Error bars denote standard error of the mean. * indicates $P < .05$, ** indicates $P < .01$.

Average surgical autonomy ratings increased with level of training (online supplementary data TABLE 3; B = 0.130; $P = .002$). After controlling for attending, training level, and surgical procedure, attending ratings of surgical autonomy differed for male and female trainees (B = –0.0199; $P = .008$; FIGURE 1; online supplementary data TABLE 4). Removing fellows did not change the significance of these findings, so they were included in the remainder of these analyses. While there was no initial difference in autonomy afforded to male and female trainees at the start of residency ($P = .32$ for trainee gender; online supplementary data TABLE 5), there were larger gaps in autonomy between genders as training progressed (B = –0.0163; $P = .020$; online supplementary data TABLE 5).

The gender gap in autonomy was disproportionately represented in the cases attendings rated as the hardest one-third (B = –0.0366, $P = .027$, compared with less complex cases controlling for attending, training level, and surgical procedure; online supplementary data TABLE 6). The gender gap for less complex cases (those rated average one-third or easiest one-third of similar cases) was not significant (B = –0.0107; $P = .203$; FIGURE 1; online supplementary data

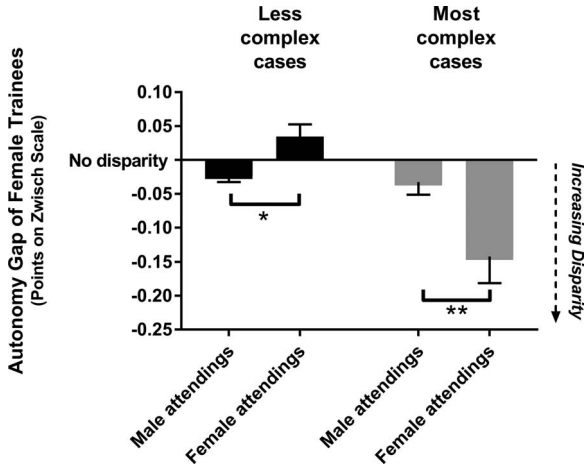

FIGURE 2
Difference in Autonomy Attained by Female Trainees Compared With Male Trainees, Measured on Zwisch Scale of Surgical Autonomy, When Operating With Male or Female Attendings in Cases of Different Complexity

Note: Regressions controlled for attending, level of training, and surgical procedure. Less complex cases were those rated by the attending as the "easiest one-third" or "average one-third" of similar cases. Most complex cases were rated as the "hardest one-third" of similar cases. Error bars denote standard error of the mean. * indicates $P < .05$ and ** indicates $P < .01$ for comparisons.

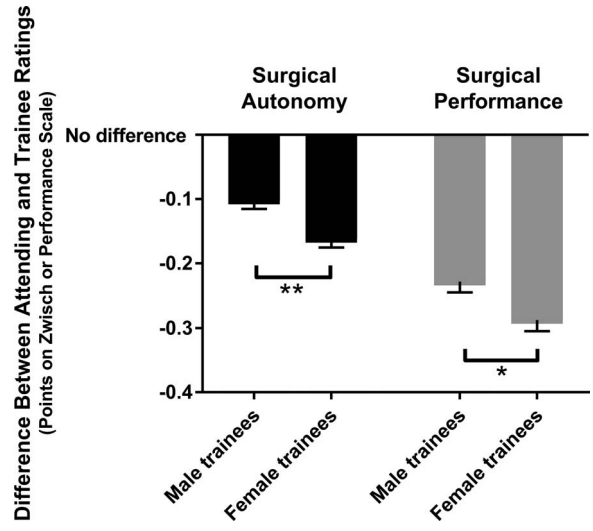

FIGURE 3
Difference in Trainee Self-Assessments and Attending Ratings of Surgical Autonomy and Surgical Performance for Male and Female Trainees

Note: Error bars denote standard error of the mean. * indicates $P < .05$ and ** indicates $P < .01$ for comparisons.

TABLE 7) while the gender gap for the hardest one-third of cases was statistically significant (B = –0.0502; P = .002; FIGURE 1; online supplementary data TABLE 7).

Overall, 24% (5870 of 24 751) of all cases with male trainees were rated by attendings as complex cases compared with 22% (3320 of 14 870) of cases with female trainees ($\chi^2(2n = 39\ 621) = 10.065$, P = .002). Both junior (PGY-4 and lower) and senior (PGY-5 and higher) female trainees logged a lower proportion of complex cases than male trainees of similar training levels. For PGY-4 and lower, 22% (3894 of 18 036) of cases with male trainees were rated by attendings as hardest one-third compared with 21% (2261 of 10 982) of cases with female trainees ($\chi^2(2n = 29\ 018) = 4.10$, $P = .043$). For PGY-5 and higher, 29% (1976 of 6715) of cases with male trainees were rated by attendings as hardest one-third compared with 27% (1059 of 3888) of cases with female trainees ($\chi^2(2n = 10\ 603) = 5.78$, P = .016). The interaction between case complexity and trainee gender in modeling autonomy was complicated by attending gender. In cases rated less complex (average one-third or easiest one-third of similar cases), female attendings gave female trainees and male trainees similar ratings of autonomy (B = 0.0345; P = .058, not statistically significant; FIGURE 2; online supplementary data TABLE 8), while male attendings gave female trainees lower ratings of autonomy than male trainees (B = –0.0229; P = .016; FIGURE 2; online supplementary data TABLE 8). There was a significant difference in autonomy ratings given to male and female trainees by male and female attendings for less complex cases (B = –0.0481; P = .014; FIGURE 2). For the hardest one-third of cases, female attendings gave lower ratings of autonomy to female trainees than male trainees (female attendings B = –0.142; $P < .001$; FIGURE 2; online supplementary data TABLE 9), while male attendings gave similar ratings of autonomy to male and female trainees (B = –0.0327; P = .08, not statistically significant; FIGURE 2; online supplementary data TABLE 9). There was a significant difference in autonomy ratings given to male and female trainees by male and female attendings for the most complex cases (B = –0.109; P = .007; FIGURE 2).

Average surgical performance ratings also increased with level of training (B = 0.141; $P < .001$; online supplementary data TABLE 10). When controlling for attending, training level, and surgical procedure, there was no statistically significant difference in attending ratings of performance for male and female trainees (B = –0.0124; P = .07; online supplementary data TABLE 11). There was similarly no gap for the most complex cases (B = –0.0110, P = .48).

There were significant differences in trainee perceptions of their own operative experiences by gender. Female trainees consistently rated themselves as less autonomous than male trainees when controlling for training level, attending, surgical procedure, case complexity, and paired attending ratings of autonomy (B = –0.0669; $P < .001$; FIGURE 3; online supplementary data TABLE 12). Similarly,

female trainees rated their performance worse than male trainees when controlling for training level, training program, surgical procedure, case complexity, and paired attending ratings of performance (B = –0.0704; $P < .001$; FIGURE 3; online supplementary data TABLE 13). All trainees rated themselves worse than paired attending ratings for both autonomy (male: B = –0.102, $P < .001$; female: B = –0.162, $P < .001$) and performance (male: B = –0.228, $P < .001$; female: B = –0.287, $P < .001$). However, the differences between attending ratings and trainee ratings were significantly larger for female trainees than male trainees (autonomy: B = –0.0602, $P = .001$; performance: B = –0.0596, $P = .014$; FIGURE 3).

Discussion

In this study of the operative experiences of surgical trainees, 3 primary gender differences were identified. First, there was a small but significant gender gap in surgical autonomy ratings, which emerged over time during training and was most significant for complex cases in which female trainees were rated to have had less autonomy. Despite the difference in autonomy ratings, there was no gender difference for ratings of surgical performance. Second, female trainees' self-assessments of operative experiences were significantly lower than male trainees' self-assessments, even when controlling for factors including attending ratings. Lastly, female trainees logged fewer complex cases than male trainees of similar training levels.

The gender difference in ratings of operative autonomy identified in this study is small in magnitude, representing approximately 0.05 points on the 4-point Zwisch scale. Coupled with the finding that attending ratings of surgical performance were not different for male and female trainees, some would celebrate the small size of this gender gap in autonomy as a triumph of decades of social progress. However, the cumulative impact of small differences in autonomy over thousands of procedures over years may affect trainee competency or confidence at the end of training and must be studied further.[10,16]

Notably, similar to research in other sectors,[17] this study determined that working with male or female supervisors did not change the direction of gender differences in operative autonomy. In fact, female attending ratings of trainee autonomy for the most complex cases appeared to drive the difference in ratings between male and female trainees for these cases, suggesting that they must be included in any future interventions to improve gender equity in surgical training. Further qualitative studies should explore the gendered interactions in the operating room that might contribute to this finding, similar to the ethological observations of social behaviors in the operating room that highlighted how the gender composition of surgical teams affected the distribution of cooperative and conflict behaviors.[18]

This study found that male and female trainees rated their own operative autonomy and performance lower than their attending surgeons, suggesting a level of humility among surgical trainees. The difference in male and female trainee self-assessments of autonomy and performance found in this study may be explained by imposter syndrome, a psychological phenomenon that disproportionately affects female trainees in fields with imbalanced gender representation, which causes one to feel inferior to their peers.[19,20] Alternatively, as we cannot assume that attending ratings are universally accurate, female trainee self-assessments may be more realistic than male trainee self-assessments, which might be more likely to overestimate capabilities.[21] Whatever the etiology of this difference, this gender gap in self-assessments may translate to less surgical confidence among female trainees. For example, if female trainees feel they have lower levels of operative autonomy during residency or fellowship, they might feel less prepared to perform more difficult procedures after graduation, a hypothesis supported by a recent study in which female attending general surgeons performed less complex cases than male colleagues even when controlling for subspecialty and seniority.[22]

Lastly, we found that female trainees logged fewer complex cases than their male colleagues, which warrants further exploration and possible intervention. If female trainees were actually performing fewer difficult cases because they were assigned to or volunteered for them, programs should increase oversight to ensure even allocation of complex cases. An alternative hypothesis is that female trainees were less inclined to log more complicated cases, perhaps because they preferred to ask for in-person feedback after difficult cases instead of using the smartphone application. Further investigation at the institutional level should be conducted to determine if there is a need to oversee equal distribution of cases, potentially using additional metrics like surgical case logs[16] or electronic medical record documentation of operative assistants for validation.

Despite significant social progress over recent decades, gender inequalities persist in medicine with the potential to affect women's advancement in the workplace throughout training and beyond.[5,23–26] As they represent potential harm to young surgeons, the gender differences identified in this study must be studied qualitatively to elucidate the mechanisms through which these differences are perpetuated before designing targeted interventions for trainees and attending surgeons of all genders. Possible etiologies

of gender differences in operating room experiences include gender stereotypes, gender norms, and unconscious biases.[27,28] The effects of any future interventions must be evaluated carefully, potentially with the use of novel assessment instruments like the mobile application in this study that can longitudinally track operative experiences. The identification and correction of inequities during surgical training may improve the way we recruit and train future leaders in surgery.

Several study limitations exist. First, factors beyond the scope of this study may shape operative experiences, such as the number of times a trainee has operated with an attending or geographic differences in surgical teaching behaviors. Qualitative details of interactions in the operating room were also unable to be assessed and warrant further investigation. The possibility that subjective perceptions of performance may affect the autonomy allotted to trainees was considered. However, performance was not used as a covariate in the analysis of the gender gap in autonomy because trainee gender may affect both performance and autonomy in the same way and therefore may confound estimates of the effect of gender on either outcome. Second, most institutions studied were university-affiliated centers and may not reflect the experience of all programs. Third, the data included surgical specialties that may have differing training practices. Fourth, SIMPL evaluations were voluntary and therefore susceptible to selection bias. The overall response rate is unknown as it is unclear what proportion of all cases performed was logged and whether logged cases are representative. It is also unknown whether study participants are representative of the general population of surgical trainees. This study is limited by its retrospective nature, and future prospective studies will be needed to confirm and further elucidate the mechanisms through which gender shapes training experiences.

Conclusions

We found no gender difference in ratings of operative performance, yet small differences exist in the ratings of operative autonomy and the proportion of complex cases performed by female and male surgical trainees.

References

1. Association of American Medical Colleges. The majority of U.S. medical students are women, New Data Show. https://www.aamc.org/news-insights/press-releases/majority-us-medical-students-are-women-new-data-show. Accessed June 28, 2021.
2. Association of American Medical Colleges. Table B3. Number of active residents, by type of medical school, GME specialty, and sex. https://www.aamc.org/data-reports/students-residents/interactive-data/report-residents/2019/table-b3-number-active-residents-type-medical-school-gme-specialty-and-sex. Accessed June 28, 2021.
3. Bruce AN, Battista A, Plankey MW, Johnson LB, Marshall MB. Perceptions of gender-based discrimination during surgical training and practice. *Med Educ Online*. 2015;20:25923. doi:10.3402/meo.v20.25923
4. Lebares CC, Braun HJ, Guvva EV, Epel ES, Hecht FM. Burnout and gender in surgical training: a call to re-evaluate coping and dysfunction. *Am J Surg*. 2018;216(4):800–804. doi:10.1016/j.amjsurg.2018.07.058
5. Hu Y-Y, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. *New Engl J Med*. 2019;381(18):1741–1752. doi:10.1056/NEJMsa1903759
6. Bergen PC, Turnage RH, Carrico CJ. Gender-related attrition in a general surgery training program. *J Surg Res*. 1998;77(1):59–62. doi:10.1006/jsre.1998.5335
7. Meyerson SL, Teitelbaum EN, George BC, Schuller MC, DaRosa DA, Fryer JP. Defining the autonomy gap: when expectations do not meet reality in the operating room. *J Surg Educ*. 2014;71(6):e64–e72. doi:10.1016/j.jsurg.2014.05.002
8. Hashimoto DA, Bynum WE, Lillemoe KD, Sachdeva AK. See more, do more, teach more: surgical resident autonomy and the transition to independent practice. *Acad Med*. 2016;91(6):757–760. doi:10.1097/ACM.0000000000001142
9. Meyerson SL, Odell DD, Zwischenberger JB, et al. The effect of gender on operative autonomy in general surgery residents. *Surgery*. 2019;166(5):738–743. doi:10.1016/j.surg.2019.06.006
10. Hoops H, Heston A, Dewey E, Spight D, Brasel K, Kiraly L. Resident autonomy in the operating room: does gender matter? *Am J Surg*. 2019;217(2):301–305. doi:10.1016/j.amjsurg.2018.12.023
11. Meyerson SL, Sternbach JM, Zwischenberger JB, Bender EM. The effect of gender on resident autonomy in the operating room. *J Surg Educ*. 2017;74(6):e111–e118. doi:10.1016/j.jsurg.2017.06.014
12. Chen JX, Kozin ED, Bohnen JD, et al. Assessments of otolaryngology resident operative experiences using mobile technology: a pilot study. *Otolaryngol Head Neck Surg*. 2019;161(5):939–945. doi:10.1177/0194599819868165
13. Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intraoperative performance assessment with SIMPL (System for Improving and Measuring Procedural Learning): early experience from a multi-institutional trial. *J Surg Educ*.

2016;73(6):e118–e130. doi:10.1016/j.jsurg.2016.08.010

14. George BC, Teitelbaum EN, Darosa DA, et al. Duration of faculty training needed to ensure reliable OR performance ratings. *J Surg Educ*. 2013;70(6):703–708. doi:10.1016/j.jsurg.2013.06.015
15. George BC, Teitelbaum EN, Meyerson SL, et al. Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. *J Surg Educ*. 2014;71(6):e90–e96. doi:10.1016/j.jsurg.2014.06.018
16. Gurgel RK, Cardon BR, Allen CM, et al. Evaluating gender parity in operative experience for otolaryngology residencies in the United States. *Laryngoscope*. 2020;130(7):1651–1656. doi:https://doi.org/10.1002/lary.28306
17. Faniko K, Ellemers N, Derks B, Lorenzi-Cioldi F. Nothing changes, really: why women who break through the glass ceiling end up reinforcing it. *Pers Soc Psychol Bull*. 2017;43(5):638–651. doi:10.1177/0146167217695551
18. Jones LK, Jennings BM, Higgins MK, Waal FBM de. Ethological observations of social behavior in the operating room. *Proc Natl Acad Sci U S A*. 2018;115(29):7575–7580. doi:10.1073/pnas.1716883115
19. Villwock JA, Sobin LB, Koester LA, Harris TM. Impostor syndrome and burnout among American medical students: a pilot study. *Int J Med Educ*. 2016;7:364–369. doi:10.5116/ijme.5801.eac4
20. Mullangi S, Jagsi R. Imposter syndrome: treat the cause, not the symptom. *JAMA*. 2019;322(5):403–404. doi:10.1001/jama.2019.9788
21. Hautz WE, Schubert S, Schauber SK, et al. Accuracy of self-monitoring: does experience, ability or case difficulty matter? *Med Educ*. 2019;53(7):735–744. doi:10.1111/medu.13801
22. Chen Y-W, Westfal ML, Chang DC, Kelleher CM. Under-employment of female surgeons? *Ann Surg*. 2021;273(2):197–201. doi:10.1097/SLA.0000000000004497
23. 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(43):13201–13206. doi:10.1073/pnas.1510649112
24. 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
25. Carr PL, Ash AS, Friedman RH, et al. Faculty perceptions of gender discrimination and sexual harassment in academic medicine. *Ann Intern Med*. 2000;132(11):889–896. doi:10.7326/0003-4819-132-11-200006060-00007
26. Martell RF, Lane DM, Emrich C. Male-female differences: a computer simulation. *Am Psychol*. 1996;51(2):157–158. doi:10.1037/0003-066X.51.2.157
27. 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. doi:10.4300/JGME-D-17-00126.1
28. Myers SP, Dasari M, Brown JB, et al. Effects of gender bias and stereotypes in surgical training: a randomized clinical trial. *JAMA Surg*. 2020;155(7):552–560. doi:10.1001/jamasurg.2020.1127

Jenny X. Chen, MD, is a Clinical Fellow, Department of Otolaryngology–Head and Neck Surgery, Massachusetts Eye and Ear, Harvard Medical School; **Edward H. Chang, PhD,** is Assistant Professor of Business Administration, Negotiation, Organizations and Markets Unit, Harvard Business School; **Francis Deng, MD,** is a Clinical Fellow, Department of Radiology, Massachusetts General Hospital, Harvard Medical School; **Shari Meyerson, MD, MEd,** is Professor of Medicine, Department of Thoracic Surgery, University of Kentucky; **Brian George, MD, MAEd,** is Assistant Professor of Surgery, Department of Surgery, Center for Surgical Training and Research, Michigan Medicine; **Elliott D. Kozin, MD*,** is Assistant Professor, Department of Otolaryngology–Head and Neck Surgery, Massachusetts Eye and Ear, Harvard Medical School; and **Stacey T. Gray, MD*,** is Associate Professor, Department of Otolaryngology–Head and Neck Surgery, Massachusetts Eye and Ear, Harvard Medical School.

* Contributed equally as authors.

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

Conflict of interest: The authors declare they have no competing interests. Drs Shari Meyerson, Brian George, and Elliott Kozin are members of the Society for Improving Medical Professional Learning's Steering Committee. Dr George is also the executive director of the organization. These positions are unpaid and without vested financial interest.

The authors would like to thank the program directors, coordinators, faculty, and trainees from the participating programs for making this study possible. They also wish to thank all the institutional members of the Society for Improving Medical Professional Learning for supporting this study.

Corresponding author: Stacey T. Gray, MD, Massachusetts Eye and Ear, stacey_gray@meei.harvard.edu

Received February 22, 2021; revision received May 31, 2021; accepted June 22, 2021.