

ORIGINAL RESEARCH

Factors affecting operative autonomy and performance during otolaryngology training: A multicenter trial

Jenny X. Chen MD^{1,2}  | Lauren E. Miller MD^{1,2}  | Andrey Filimonov MD³  |
Elizabeth A. Shuman MD⁴ | Emily Marchiano MD⁵ | Brian C. George MD⁶ |
Marc Thorne MD⁵ | Steven D. Pletcher MD⁴  | Michael Platt MD⁷ |
Marita Teng MD³ | Elliott D. Kozin MD^{1,2}  | Stacey T. Gray MD^{1,2}

¹Department of Otolaryngology – Head and Neck Surgery, Massachusetts Eye and Ear, Boston, Massachusetts, USA

²Department of Otolaryngology – Head and Neck Surgery, Harvard Medical School, Boston, Massachusetts, USA

³Department of Otolaryngology – Head and Neck Surgery, Mount Sinai Hospital, New York City, New York, USA

⁴Department of Otolaryngology – Head and Neck Surgery, University of California San Francisco, San Francisco, California, USA

⁵Department of Otolaryngology – Head and Neck Surgery, Michigan Medicine, Ann Arbor, Michigan, USA

⁶Center for Surgical Training and Research, Department of Surgery, Michigan Medicine, Ann Arbor, Michigan, USA

⁷Department of Otolaryngology – Head and Neck Surgery, Boston University Medical Center, Boston, Massachusetts, USA

Correspondence

Jenny X. Chen, 243 Charles Street, Boston, MA 02114, USA.

Email: jenny_chen@meei.harvard.edu

Funding information

2019 American Academy of Otolaryngology-Head and Neck Surgery Foundation Women in Otolaryngology Research Grant

Abstract

Objective: Surgical education is changing in an era of new regulations and evolving training cultures. We sought to understand the factors that affect operative experiences during otolaryngology residency.

Methods: From December 2019 to December 2020, five otolaryngology training programs used the SIMPL OR smartphone application to evaluate residents after each operation. Residents and attendings rated the trainee's autonomy on a 4-level Zwisch scale, performance on a 5-level scale, and case complexity on a 3-level scale. We examined associations between ratings of autonomy and performance with variables including postgraduate year (PGY), case complexity, gender, week of the academic year (AY), and whether multiple procedures were logged.

Results: 78 attendings and 92 residents logged 2984 evaluations. PGY level and week of the AY were positively associated with attending ratings of autonomy and performance (PGY3 vs. PGY2: $B = 0.63$, $p < .001$ for autonomy and $B = 1.05$, $p < .001$ for performance; week of the AY: $B = 0.013$, $p = .002$ for autonomy; $B = 0.025$, $p < .001$ for performance). Multiple procedures logged and increasing case complexity were negatively associated with attending ratings (multiple procedures: $B = -0.19$, $p = .04$ for autonomy and $B = -0.48$, $p < .001$ for performance; hardest vs. easiest 1/3 of cases: $B = -1.01$, $p < .001$ for autonomy and $B = -0.59$, $p < .001$ for performance). Attending and trainee genders were not associated with attending ratings of autonomy or performance.

Conclusion: Resident autonomy and performance were positively associated with PGY level and week of the academic year, and negatively associated with case complexity and multiple procedures. These findings highlight the need to align training level with case complexity to promote quality operative experiences.

Level of Evidence: 2.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Laryngoscope Investigative Otolaryngology* published by Wiley Periodicals LLC on behalf of The Triological Society.

KEYWORDS

autonomy, competency, performance, residency training, surgical assessments, surgical education

1 | INTRODUCTION

Modern surgical training has been affected by an increasing emphasis on duty hour restrictions,¹ patient safety,² and hospital productivity.³ As a result, reports of diminished resident surgical volume and operative autonomy have arisen.¹ The range of operative procedures performed has expanded for many specialties, while the volume of operations residents perform has not increased at the same rate.⁴ Across many surgical specialties, concerns have mounted regarding trainee competency and confidence at the end of residency training.^{5,6}

After decades of regulatory and cultural changes to the current system of surgical training, the variables that impact current surgical residency experiences remain unclear. Efforts to quantify, characterize and improve surgical education in otolaryngology have been scarce. A limited number of studies have introduced the use of several instruments to evaluate resident operative experiences and operative skills.⁷⁻¹⁰ However, little is known about potential factors that may cause variability across otolaryngology training experiences. Herein, we conducted a multicenter study to characterize the factors that affect operative skill training in otolaryngology.

2 | MATERIALS AND METHODS

From December 2019 to December 2020, five independent academic otolaryngology training programs used the SIMPL OR smartphone application from the non-profit Society for Improving Medical Professional Learning (<https://www.simpl.org>, Boston, MA) to evaluate residents after each surgery.^{9,11,12} Faculty and residents participated on a voluntary basis, and participants could opt out at any time without repercussion. This study was approved by the local Institutional Review Board at each study site: Massachusetts Eye and Ear, Boston University, University of Michigan, Mount Sinai Hospital, and University of California San Francisco.

All assessments were completed within or expired at 72 h after the procedures (assessments beyond 72 h were not considered reliable based on prior research).¹³ The smartphone application could record all procedures available in the ACGME case logging taxonomy. Each SIMPL OR evaluation consisted of three questions answered by both the resident and the attending, as well as an option for the attending to dictate audio feedback for the resident.

The first question assessed the surgical autonomy granted by the attending surgeon during the case on a 4-level Zwisch scale: (1) show and tell, (2) active help, (3) passive help, and (4) supervision only.¹¹ In show and tell, the attending surgeon demonstrates and explains the procedural steps to the resident. In active help, the attending surgeon leads the flow of the procedure, while in passive help the attending

surgeon allows the resident to lead the flow of the procedure for the majority of the critical portions. In supervision only, the resident simulates independence while the attending surgeon acts in a supervisory role only. If the resident's logged participation was more than just show and tell, the second question then assessed the resident on a 5-level performance scale: (1) unprepared/critical deficiency, (2) unfamiliar with procedure, (3) intermediate performance, (4) practice ready, and (5) exceptional performance. The third question asked about the tercile of case complexity (easiest 1/3, average 1/3, or hardest 1/3 of cases). All attending surgeon and resident participants underwent standardized training programs prior to the start of the study to improve interrater reliability in the assessment scales.¹⁴ More detailed descriptions of the rating scales are described in prior publications.^{9,15}

2.1 | Statistics

Descriptive statistics were conducted on aggregated data with anonymized participants and programs. Wilcoxon Rank Sum tests were performed to compare the median frequencies of ratings between raters' genders. Two ordinal outcome measures (supervision and performance) were examined to find associations with postgraduate year (PGY), case complexity, week of the academic year (AY, where July 1st is the first week), and whether or not multiple procedures were performed (yes/no). To accomplish this ordinal generalized linear mixed models (GLMM) were conducted using the ordinal package in R version 4.0.2 (Vienna, Austria) by an experienced biostatistician. Reference groups for dummy variables in the regression were PGY = 2, complexity = easy, attending gender = male, resident gender = male, attending autonomy rating = 1, and paired attending performance rating = 2 (because there were no performance ratings of 1). Random effects were included to accommodate clustering effects due to rater, subject, program, and procedure.

3 | RESULTS

Ninety-two residents and 78 attending surgeons from five academic otolaryngology programs participated in the study. Fifty two percent ($n = 48$) of residents and 24% of attending surgeons were female; male and female participants logged a similar number of cases ($p > .05$ for male vs. female residents/attendings). Participants completed a total of 2984 evaluations for 367 different types of procedures. Residents logged a total of 1746 assessments, and attendings logged 1238 assessments. The median number of evaluations submitted per program was 555 (range 332-967). Of all cases logged, 11% were rated as among the easiest third of similar cases ($N = 339$), 68% were

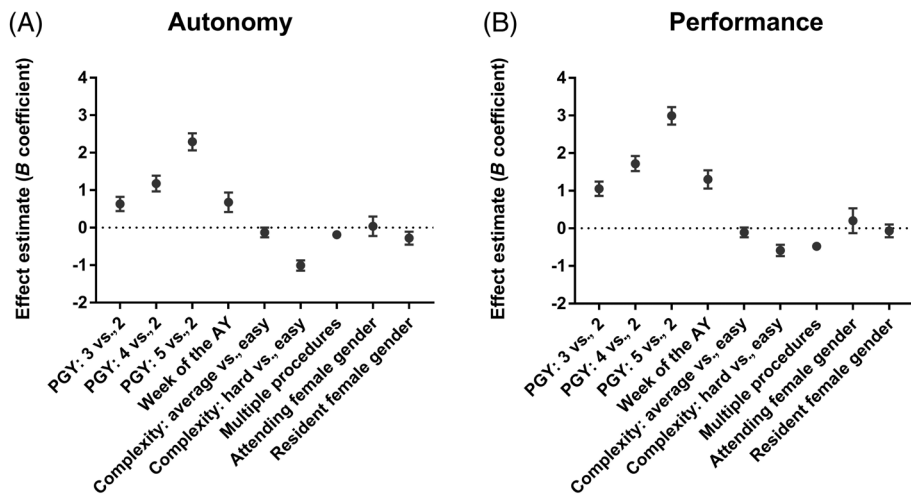


FIGURE 1 The fixed effect estimates for the model used to predict attending ratings of trainee autonomy (A) and performance (B). Week of the academic year (AY) is scaled to represent the effect over 1 year or 52 weeks

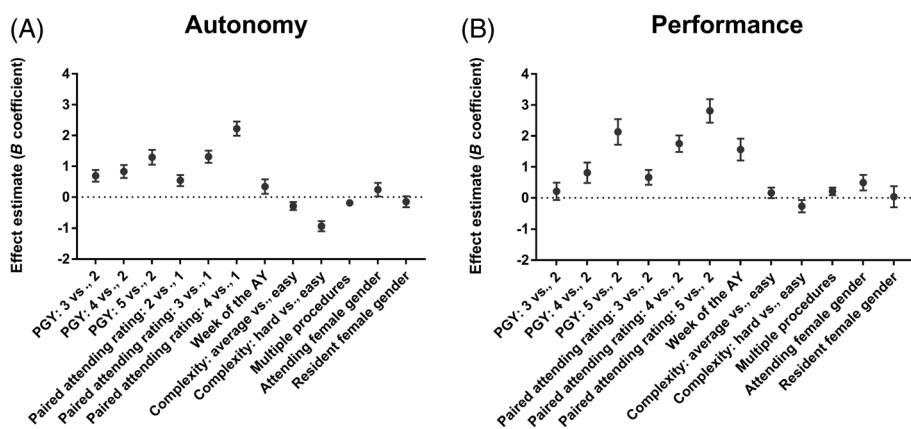


FIGURE 2 The fixed effect estimates for the model used to predict resident self-assessments of trainee autonomy (A) and performance (B). Week of the academic year (AY) is scaled to represent the effect over 1 year or 52 weeks

rated as average ($N = 2020$), and 21% were rated as among the hardest third of similar cases ($N = 625$).

Resident participation was well distributed across PGY2–5: 556 (19%) assessments for PGY5 residents, 792 (27%) assessments for PGY4 residents, 700 (24%) assessments for PGY3 residents and 903 (31%) assessments for PGY2 residents. As there were very few assessments of PGY1 and PGY6 residents ($N = 32$ and $N = 1$, respectively), they were excluded from the subsequent analyses. Ratings of performance/autonomy/case complexity for assessments at each PGY level of training were previously reported.¹⁰

3.1 | Factors affecting attending ratings of operative experiences

In multivariable modeling, attending ratings of surgical autonomy were associated with resident level of training, case complexity, week of the AY, and the inclusion of multiple procedures in a single evaluation. PGY level was positively associated with increased autonomy; for example, increasing training level from PGY2 to PGY3 was associated with a $B = 0.63$ increase in the Zwisch level of autonomy (standard error, $SE = 0.19$, $p < .001$, Figure 1A, full model in Table S1). Each

new week of the AY was also associated with a small increase in autonomy ratings ($B = 0.013$, $SE = 0.0045$, $p = .002$). Variables associated with decreases in autonomy included case complexity and the presence of multiple procedures logged. Increasing the complexity from the easiest third to the hardest third of cases was associated with a $B = -1.01$ reduction in the Zwisch level ($SE = 0.14$, $p < .001$). Similarly, the presence of multiple procedures logged for a single evaluation was associated with a $B = -0.19$ reduction on the Zwisch scale ($SE = 0.091$, $p = .04$). Attending gender was not associated with ratings of autonomy ($B = 0.038$, $p = .9$); trainee female gender had a small negative impact on attending ratings of autonomy but this finding did not reach statistical significance ($B = -0.28$, $p = .1$).

Factors associated with attending ratings of resident surgical performance were the same as for surgical autonomy. PGY level was positively associated with autonomy; increasing training level from PGY2 to PGY3 was associated with a $B = 1.05$ increase in the performance rating (standard error, $SE = 0.19$, $p < .001$, Figure 1B, full model in Table S2). Each subsequent week of the AY was associated with a small increase in ratings ($B = 0.025$, $SE = 0.0047$, $p < .001$). Increasing the complexity from the easiest third to the hardest third of cases was associated with a $B = -0.59$ reduction ($SE = 0.15$, $p < .001$) and the presence of multiple procedures logged was associated with a

$B = -0.48$ reduction ($SE = 0.094$, $p < .001$) in performance ratings. Attending gender and trainee gender were not associated with attending ratings of performance.

3.2 | Factors affecting trainee self-assessments of operative experiences

Resident self-assessments of surgical autonomy were positively associated with increasing level of training (PGY2 as compared with PGY3-5), and paired attending ratings of autonomy (Figure 2A, full model in Table S3). Resident ratings of autonomy were negatively associated with case complexity, as well as with multiple procedures logged. Week of the AY, attending gender and resident gender were not associated with resident ratings of autonomy.

Resident self-assessments of surgical performance were similarly positively associated with senior levels of training (PGY4 and 5 but not PGY3), paired attending ratings of performance, and the week of the AY (Figure 2B, full model in Table S4). Attending female gender was also positively associated with resident self-assessments of performance ($B = 0.49$, $SE = 0.25$, $p = .04$). Resident ratings of performance were not associated with resident gender, case complexity or whether multiple procedures were logged.

4 | DISCUSSION

This is the largest study to date studying the factors that shape resident operative experiences in otolaryngology training. A number of these findings were expected: attending surgeon and resident self-ratings of resident operative autonomy and performance were positively associated with level of training and each passing week of the AY, while they were negatively associated with case complexity and multiple concurrent procedures performed.

These findings reinforce the idea that resident case assignments can be very important in determining the quality of the operative experiences. For example, this study suggests that junior residents assigned to more complex cases, particularly earlier in the AY, are less likely to have meaningful participation. Although this finding may appear self-evident to most readers, it may be challenging for training programs to optimize case assignments, influenced by a multitude of factors such as resident availability across different sites, the identity of the specific person who makes the case assignments and their priorities (i.e., faculty vs. fellows vs. residents), and the distribution of inpatient floor/consult duties among residents of different levels. The structure of resident rotations may also impact a program's ability to optimize pairings between cases and residents. Based on this study's findings, to promote progressive resident autonomy and performance, careful design and oversight of level-appropriate case assignments should be a priority for programs.

In this study, resident gender was not associated with either attending evaluations or resident self-assessments. This is particularly interesting because in previous studies pooling SIMPL OR data across

surgical specialties, trainee gender was found to be a source of disparities in attending ratings of surgical autonomy and performance.¹⁶ Many other studies using a range of self-assessment instruments have also found that female residents are more self-critical than male residents.¹⁶⁻¹⁸ One possible explanation for the difference in findings between this multicenter study of otolaryngology residents compared to studies that pooled data across surgery programs nationwide could be the nearly equal numbers of male and female trainees in the present study. In this study of large academic training programs where female trainees were no longer in the minority, female residents did not exhibit historically gender-based psychological phenomena like imposter syndrome. Alternatively, this study may be too underpowered to discern small gender disparities compared to the larger pooled studies. Interestingly, operating with female attendings was associated with a small increase in resident self-assessments of performance. The etiology of this difference is unclear, and further qualitative studies are needed to discern gender-specific interactions between faculty and residents that may explain this finding.

5 | LIMITATIONS

A number of limitations can be found in this study's design. First, the study was conducted at large academic training centers in urban locations across the United States; thus, findings may not be generalizable to training programs with other attributes (size, location, case mix, gender distribution of residents, etc.). Second, evaluations were voluntary and susceptible to selection bias. It is also unknown what proportion of all cases performed was logged, and whether the logged cases were representative of overall case volume. Third, this study was unable to determine which specific procedures at differing levels of case complexity are best suited for different levels of trainees; heterogeneity in rotation schedules across training years at the small number of training programs studied limited this analysis. This study was also unable to answer the question raised in the discussion as to whether case assignment by residents or by attendings/fellows was more conducive to optimizing surgical experiences. Lastly, insufficient data was collected in the last few months of training for PGY5 residents to do predictive analyses to project the level of surgical autonomy and performance they can expect to achieve for key indicator procedures, as has been done in the general surgery literature. Finally, the study year was the first year of the SARS-CoV-2/COVID-19 pandemic, which may have had a significant impact on resident training.

6 | CONCLUSION

In this multicenter study, attending surgeon assessments of operative autonomy and performance were associated with increasing PGY experience, case complexity, week of the AY, and whether multiple procedures were logged. Assessments were not associated with attending or resident genders.

ACKNOWLEDGMENTS

The authors are grateful to all the residents and faculty who participated in this study at each institution, and to the faculty, trainees, and institutional members who collectively sustain the work of the Society for Improving Medical Professional Learning (SIMPL).

CONFLICT OF INTEREST

None.

ORCID

Jenny X. Chen  <https://orcid.org/0000-0002-2992-4771>

Lauren E. Miller  <https://orcid.org/0000-0002-5458-0544>

Andrey Filimonov  <https://orcid.org/0000-0002-4285-5862>

Steven D. Pletcher  <https://orcid.org/0000-0003-0127-927X>

Elliott D. Kozin  <https://orcid.org/0000-0002-0305-0682>

REFERENCES

- Damadi A, Davis AT, Saxe A, Apelgren K. ACGME duty-hour restrictions decrease resident operative volume: a 5-year comparison at an ACGME-accredited university general surgery residency. *J Surg Educ.* 2007;64(5):256-259. doi:10.1016/j.jsurg.2007.07.008
- Dickinson KJ, Bass BL, Nguyen DT, Graviss EA, Pei KY. Public perception of general surgery resident autonomy and supervision. *J Am Coll Surg.* 2021;232(1):8-15.e1. doi:10.1016/j.jamcollsurg.2020.08.764
- Puram SV, Kozin ED, Sethi R, et al. Impact of resident surgeons on procedure length based on common pediatric otolaryngology cases. *Laryngoscope.* 2015;125(4):991-997. doi:10.1002/lary.24912
- Malangoni MA, Biester TW, Jones AT, Klingensmith ME, Lewis FR. Operative experience of surgery residents: trends and challenges. *J Surg Educ.* 2013;70(6):783-788. doi:10.1016/j.jsurg.2013.09.015
- Fonseca AL, Reddy V, Longo WE, Udelsman R, Gusberg RJ. Operative confidence of graduating surgery residents: a training challenge in a changing environment. *Am J Surg.* 2014;207(5):797-805. doi:10.1016/j.amjsurg.2013.09.033
- George BC, Bohnen JD, Williams RG, et al. Readiness of US general surgery residents for independent practice. *Ann Surg.* 2017;266(4):582-594. doi:10.1097/SLA.0000000000002414
- van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg.* 2010;97(7):972-987. doi:10.1002/bjs.7115
- Zhou NJ, Kamil RJ, Zhu J, et al. Preoperative briefings and postoperative debriefings to increase resident operative autonomy and performance. *J Surg Educ.* 2021;S1931-7204(21):00055-00056. doi:10.1016/j.jsurg.2021.03.002
- 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
- Chen JX, Deng F, Filimonov A, et al. Multi-institutional study of otolaryngology resident intraoperative experiences for key indicator procedures. *Otolaryngol Head Neck Surg.* 2021;01945998211050350. doi:10.1177/01945998211050350
- 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
- 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
- Williams RG, Chen XP, Sanfey H, Markwell SJ, Mellinger JD, Dunnington GL. The measured effect of delay in completing operative performance ratings on clarity and detail of ratings assigned. *J Surg Educ.* 2014;71(6):e132-e138. doi:10.1016/j.jsurg.2014.06.015
- 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
- Chen JX, Kozin E, Bohnen J, et al. Tracking operative autonomy and performance in otolaryngology training using smartphone technology: a single institution pilot study. *Laryngoscope Investig Otolaryngol.* 2019;4(6):578-586. doi:10.1002/lio2.323
- Chen JX, Chang EH, Deng F, et al. Autonomy in the operating room: a multicenter study of gender disparities during surgical training. *J Grad Med Educ.* 2021;13(5):666-672. doi:10.4300/JGME-D-21-00217.1
- Cooney CM, Aravind P, Lifchez SD, et al. Differences in operative self-assessment between male and female plastic surgery residents: a survey of 8,149 cases. *Am J Surg.* 2021;221(4):799-803. doi:10.1016/j.amjsurg.2020.04.009
- 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:10.1002/lary.28306

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Chen JX, Miller LE, Filimonov A, et al. Factors affecting operative autonomy and performance during otolaryngology training: A multicenter trial. *Laryngoscope Investigative Otolaryngology.* 2022;7(2):404-408. doi:10.1002/lio2.750