Variation in Competence of Graduating General Surgery Trainees

Angela E. Thelen, MD, MHPE, *,†, *Kayla M. Marcotte, MS,* *,§,‖ *Sarah Diaz, BSN,* *,§ *Rebecca Gates, MD,* * *Xilin Chen, MPH,* * *Brian C. George, MD, MAEd,* *,‡ *and Andrew E. Krumm, PhD*‡,‖

*Center for Surgical Training and Research, Department of Surgery, University of Michigan, Ann Arbor, Michigan; †MetroHealth Medical Center, Case Western Reserve University School of Medicine, Cleveland, Ohio; ‡Department of Surgery, University of Michigan, Ann Arbor, Michigan; §University of Michigan Medical School, Ann Arbor, Michigan; and ‖Department of Learning Health Sciences, University of Michigan, Ann Arbor, Michigan

OBJECTIVE: To examine the readiness of general surgery residents in their final year of training to perform 5 common surgical procedures based on their documented performance during training.

DESIGN: Intraoperative performance ratings were analyzed using a Bayesian mixed effects approach, adjusting for rater, trainee, procedure, case complexity, and postgraduate year (PGY) as random effects as well as month in academic year and cumulative, procedure-specific performance per trainee as fixed effects. This model was then used to estimate each PGY 5 trainee's final probability of being able to independently perform each procedure. The actual, documented competency rates for individual trainees were then identified across each of the 5 most common general surgery procedures: appendectomy, cholecystectomy, ventral hernia repair, groin hernia repair, and partial colectomy.

SETTING: This study was conducted using data from members of the SIMPL collaborative.

PARTICIPANTS: A total of 17,248 evaluations of 927 PGY5 general surgery residents were analyzed from 2015 to 2021.

RESULTS: The percentage of residents who requested a SIMPL rating during their PGY5 year and achieved a ≥90% probability of being rated as independent, or "Practice-Ready," was 97.4% for appendectomy, 82.4% for cholecystectomy, 43.5% for ventral hernia repair, 24% for groin hernia repair, and 5.3% for partial colectomy.

CONCLUSIONS: There is substantial variation in the demonstrated competency of general surgery residents to perform several common surgical procedures at the end of their training. This variation in readiness calls for careful study of how surgical residents can become more adequately prepared to enter independent practice. (J Surg Ed 81:17−24. © 2023 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

ABBREVIATIONS: PGY, post-graduate year SCORE, Surgical Council on Resident Education SIMPL Society for improving medical professional learning

KEY WORDS: performance variation, competency, operative assessment, workplace-based assessment, surgical resident, resident education, patient care

COMPETENCIES: Patient Care

INTRODUCTION

Patients treated by surgeons with lower operative skill ratings have inferior outcomes.[1] Ensuring the operative competency of surgical residents upon completion of their training is therefore essential for optimizing patient care. A core tenet of graduate medical education is to develop physicians who have the knowledge, skills, and empathy required for autonomous practice.[2] Yet, there are growing concerns that surgical residency programs may struggle to meet this mark.

Many key stakeholders have voiced these concerns over the past several years. For example, practicing surgeons feel new graduates are unprepared to enter independent practice.[3] Fellowship program directors believe most new residency graduates are unable to perform common general surgery procedures without supervision.[4] Even surgical residents themselves attest to feeling

unprepared to enter independent practice.[5] Population studies demonstrate that, on average, senior residents are often provided little autonomy and frequently deemed incompetent to independently perform procedures common to general surgery.[6] These studies—most of which are opinion-based—do not explore the degree to which trainees can complete these procedures to a set standard of performance. Therefore, the competency level of individual graduating general surgery residents and the variation in individual trainees' operative competency remains unclear.

We set out to explore the competency levels of graduating general surgery trainees. This study uses workplace-based assessment data to examine the readiness of general surgery residents in their final year of training to perform 5 common surgical procedures based on their documented performance during training. Identifying where competency levels are subpar will help in targeting improvement efforts, setting more effective performance standards, and ultimately increasing the safety of surgical patients.

MATERIALS AND METHODS

Faculty ratings of the intraoperative performance of general surgery trainees were collected from September 2015 to September 2021 using the Society for Improving Medical Professional Learning (SIMPL) registry. The SIMPL app is a workplace-based assessment tool used to provide feedback to a trainee about a given operative experience. The app allows attending surgeons to rate a trainee's operative autonomy and performance, in addition to the complexity of the case, within 72 hours of case completion. Operative performance was rated using the SIMPL Performance Scale as "Unprepared/Critical Deficiency," "Inexperienced," "Intermediate," "Practice-Ready," or "Exceptional." A comprehensive description of the SIMPL assessment tool has been previously published.[7-10] Specific procedures examined in this study included five of the most common general surgery procedures: appendectomy, cholecystectomy, inguinal and femoral hernia repair, ventral hernia repair, and partial colectomy. These procedure categories were the result of a mapping between the Surgical Council on Resident Education (SCORE) curriculum and SIMPL procedures based on previous work.[11] Supplementary Table 3 presents the mapping between the SCORE-based categories and SIMPL procedures.

We analyzed the SIMPL operative performance ratings aligned to the above categories using a previously developed and evaluated Bayesian mixed-effects model. The model included rater, trainee, procedure, case complexity, and postgraduate year (PGY) as random effects and

month of academic year and cumulative, procedure-specific prior practice-ready (or higher) ratings per trainee as fixed effects.[11] This cumulative variable provided a way to differentiate predicted probabilities based on trainees' individual prior performances. The model was trained on SIMPL evaluations from 2015 to 2021, and for the current analysis, we re-applied the model to the same dataset in order to generate a predicted probability for the likelihood that a trainee could earn a practice-ready (or higher) rating the *next* time the trainee is rated on that procedure using the SIMPL app. We refer to these predicted probabilities as "documented competency scores." The modifier "documented" highlights that not only does a trainee need to perform to practice-ready standard in an operative experience, but also that the trainee needed to have requested that an attending surgeon rate the trainee in the first place.

We applied the model to all PGY 5 trainees within the SIMPL registry and extracted their last documented SIMPL rating for procedures aligned to the above categories. Then, within each procedural category, we averaged the predicted probabilities for each trainee across aligned SIMPL procedures to generate a category average. We then visualized and generated descriptive statistics for each procedural category to characterize trainee's documented readiness to perform these procedures. We further visually explored the variation across SIMPL procedures within a category using boxplots. All analyses were performed using *R* version 4.0.2 with the *brms* package and visualized using *tidybayes*.[12,13] This study was deemed exempt by the University of Michigan Institutional Review Board.

RESULTS

The original model was trained on 63,248 evaluations of 2,605 residents by 1884 faculty at 70 general surgery programs. For this paper, we applied the model to 17,248 evaluations for 927 PGY5 residents. On average, there were 18.9 SIMPL evaluations per resident included in the study. Even though each individual trainee averaged approximately 19 SIMPL evaluations, each SIMPL evaluation contained contextual information that were aligned to robust statistical estimates from the pre-trained model. For example, all predicted probabilities start with a pre-test probability of typical PGY 5 performance that is then adjusted by the specific procedure that has its own probability of success as well as typical month-to-month growth within the PGY 5 academic year. Lastly, a trainee's prior practice-ready rating increases the probability of future success with each successful prior rating. Combining these estimates with each PGY 5 in the dataset allowed us to estimate a

TABLE 1. Cohort Descriptive Statistics

Characteristic	
Evaluations	17,248
Residents	927
Female Gender	385 (42%)
Average Number of Evaluations/ Resident	18.9
Number of Ratings / Practice Ready Ratings (%)	
Appendectomy	1077 / 999 (92.8%)
Cholecystectomy	2926 / 2,614 (89.3%)
Ventral hernia repair	879 / 715 (81.3%)
Groin hernia repair	1305 / 974 (74.6%)
Partial colectomy	1607 / 1,127 (70.1%)

trainee's probability of future success based on their last SIMPL documented SIMPL rating for a given procedure. Table 1 provides descriptive statistics for the cohort analyzed.

The predicted competency rates of graduating general surgery residents for each procedure are included in Table 2. Residents were best prepared to independently perform an appendectomy, with nearly all residents achieving a ≥90% probability of being independent at the time of graduation. The lowest rated resident had an 84% probability of being "Practice-Ready" at the time of graduation for appendectomy. Meanwhile, 5.3% of residents achieved a ≥90% probability of being able to independently perform a partial colectomy. Half of the residents had at least an 81% probability of being "Practice-Ready" to perform a partial colectomy at the time of graduation, and on the low end of the curve some residents had slightly better than a 50-50 chance of being independent to perform this procedure. The distribution of these competency rates is shown in Figure 1. Figures 2−6 illustrate variation within a category by SIMPL procedure. There was a high degree of variation by trainee within the general category of partial

colectomy. There was noticeably less variation within cholecystectomy and appendectomy with overall high documented competency scores. Variation within the inguinal/femoral hernia and ventral hernia is less pronounced than partial colectomy. The variation within the category of ventral hernia demonstrates the potential source of the bimodal distribution, with trainees who form the lower end of the distribution may have only documents ventral hernia repair (laparoscopic).

DISCUSSION

There is substantial variation in the competency of graduating general surgery residents to perform several common surgical procedures. For PGY5 residents, almost all trainees (97.4%) were predicted to be competent (i.e., had greater than 90% predicted probability of being practice-ready) on the next appendectomy that they performed. Additionally, the majority of trainees (82.4%) were predicted to be competent on the next cholecystectomy that they performed. The spread for predicted probability was also very small for these 2 procedures, indicating that there was less variation in the predicted probability of being practice ready for these 2 procedures. In contrast, only 43.5% of trainees were predicted to be competent in performing a ventral hernia repair, 24% were predicted to be competent in performing a groin hernia repair, and 5.3% were predicted to be competent in performing a partial colectomy. Furthermore, the spread of predicted competency was much larger for groin hernia repair and partial colectomy, indicating that there is substantial variation in the competence of residents to independently perform these procedures. This variation raises serious questions: why does such a wide variation in competency exist, and what are the implications of that variation?

One potential source of this variation in competency may be variability in general surgery training. The operative case mix of individual residents can be affected by a

TABLE 2. Distribution of Competency Rates of Graduating General Surgery Residents by Procedure. *N* Reflects the Number of PGY5 Trainees Who Were Rated on That Procedure

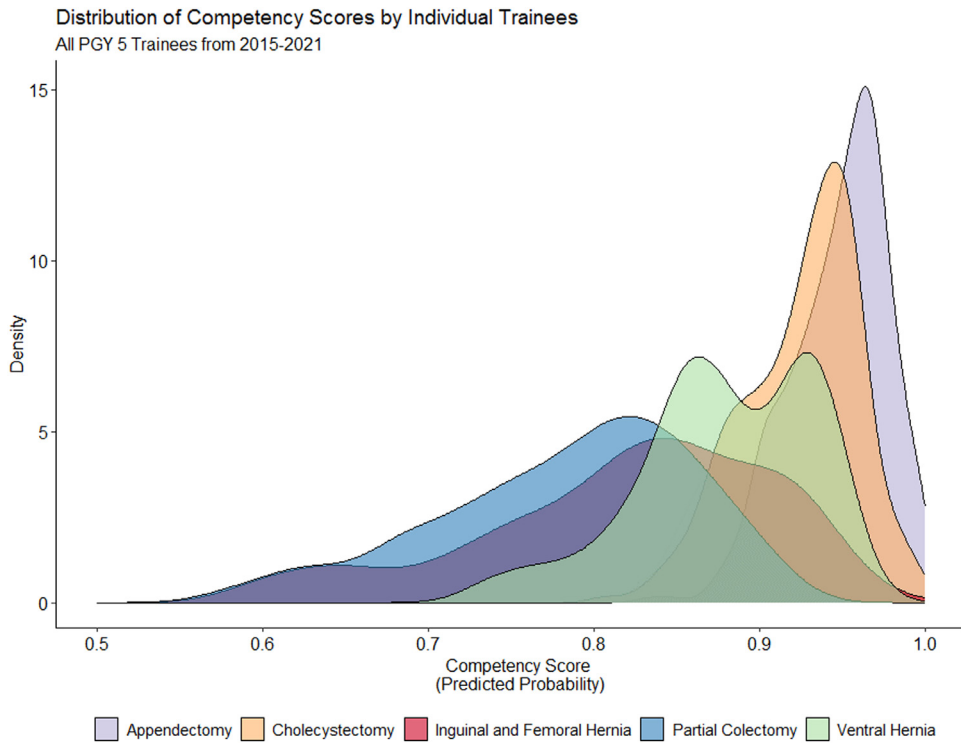|	Trainees		Probability of Being "Practice-Ready"				
	N	Achieved a ≥90% Competence Level *N* (%)	Min.	Q1	Median	Q3	Max.
Appendectomy	378	368 (97.4%)	0.84	0.94	0.96	0.97	>0.99
Cholecystectomy	609	502 (82.4%)	0.82	0.91	0.94	0.96	>0.99
Ventral hernia repair	372	162 (43.5%)	0.72	0.86	0.89	0.93	0.98
Groin hernia repair	434	104 (24%)	0.58	0.77	0.84	0.90	0.99
Partial colectomy	434	23 (5.3%)	0.57	0.74	0.81	0.85	0.93

FIGURE 1. Distribution of competency scores of individual trainees. The distribution of the final predicted probability of each included PGY5 general surgery trainee to independently perform five common general surgery procedures (color).
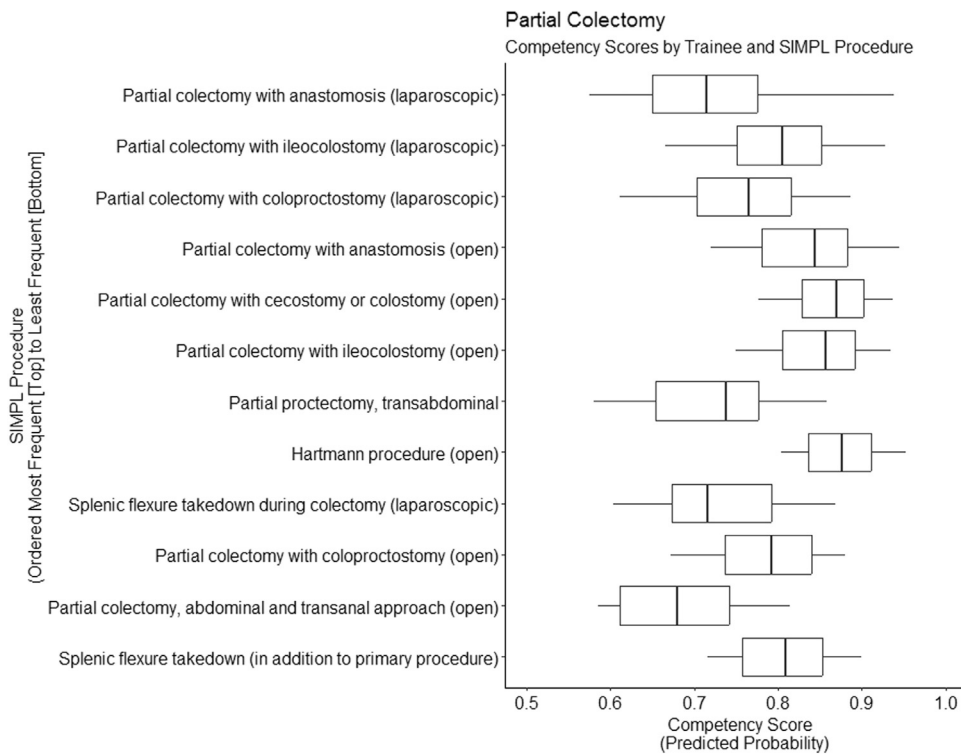

FIGURE 2. Partial colectomy plot by SIMPL procedure. Procedures are organized from most to least frequently done on the y-axis and the predicted probability per trainee is presented on the x-axis.
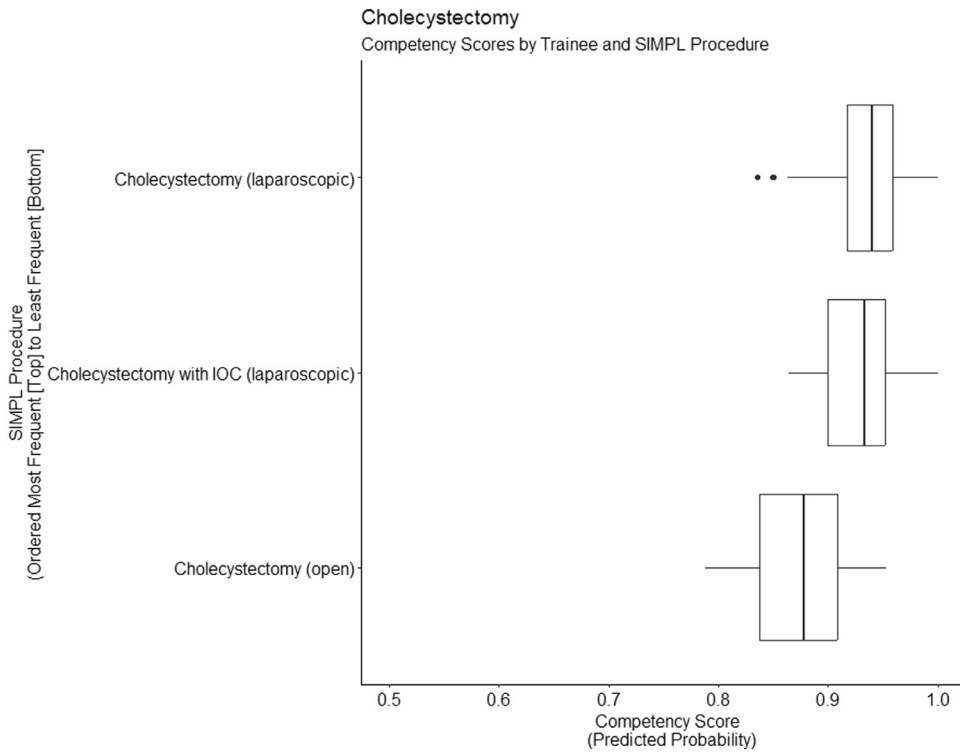
Cholecystectomy
Competency Scores by Trainee and SIMPL Procedure

FIGURE 3. Cholecystectomy plot by SIMPL procedure. Procedures are organized from most to least frequently done on the y-axis and the predicted probability per trainee is presented on the x-axis.

Appendectomy
Competency Scores by Trainee and SIMPL Procedure

FIGURE 4. Appendectomy plot by SIMPL procedure. Procedures are organized from most to least frequently done on the y-axis and the predicted probability per trainee is presented on the x-axis.
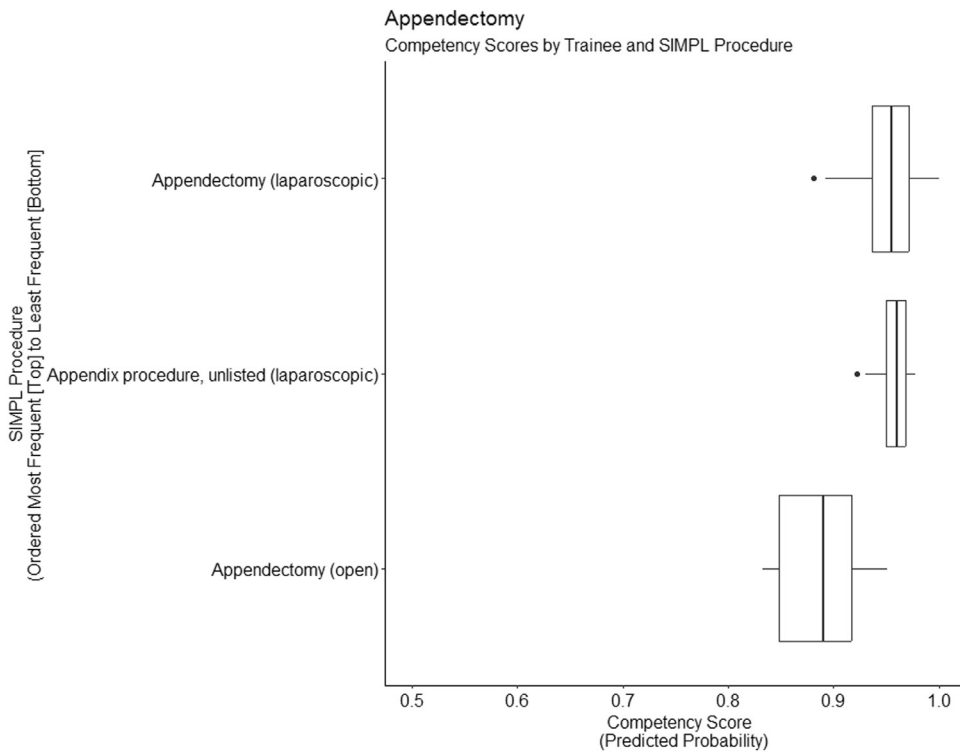
FIGURE 5. Inguinal and femoral hernia plot by SIMPL procedure. Procedures are organized from most to least frequently done on the y-axis and the predicted probability per trainee is presented on the x-axis.
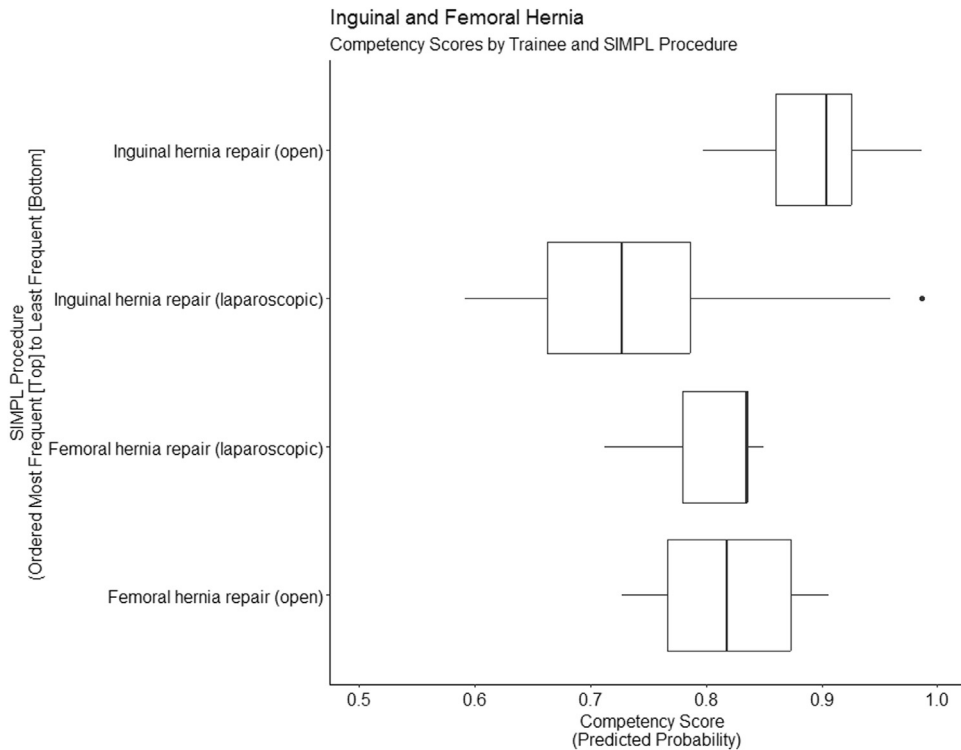
multitude of factors including training program, prevalence of disease, time spent on each rotation, competition from other residents, or personal learning plan.[14] After the implementation of duty hour restrictions in the United States, resident case numbers initially declined. While these case numbers have since rebounded, the heterogeneity of individual resident operative experiences has
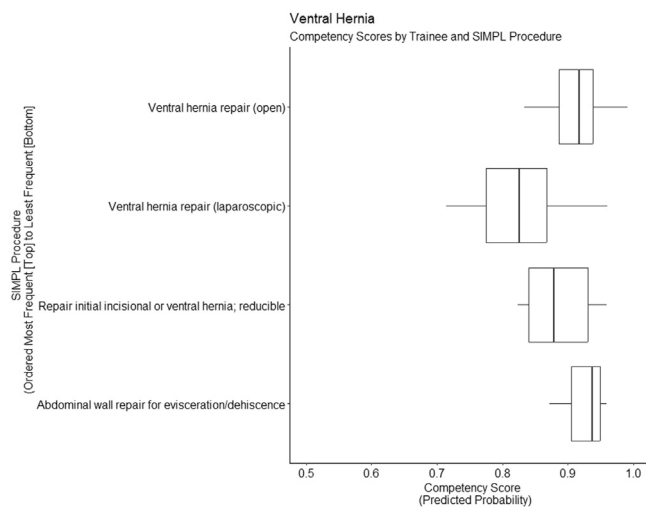

FIGURE 6. Ventral hernia plot by SIMPL procedure. Procedures are organized from most to least frequently done on the y-axis and the predicted probability per trainee is presented on the x-axis.

increased.[15,16] In addition, it is well known that the amount of autonomy granted to residents has decreased over recent decades.[17] The reduction in operative autonomy almost certainly influences the variability in competency. Operative autonomy itself can be inconsistent and affected by individual resident learning behaviors, prior experience with the operation, availability of a qualified assistant, productivity pressures on faculty, and faculty entrustment behaviors.[18,19] Reducing some of the variability in general surgery training may subsequently reduce the variability of the graduates of this system.

There are multiple improvements that may be considered by training programs to increase the readiness of their graduating surgical residents for autonomous practice. To start, increasing the amount of workplace-based assessment and formative feedback could help accelerate performance improvement and competency levels. These tools provide residents with a consistent measure of their progress and provide direct, actionable feedback they can use to improve their performance.[20] Many residents are dissatisfied with the amount of feedback they receive, and both resident and faculty development to optimize feedback quantity and quality may be important avenues to improve competency.[20-23] The upcoming implementation of Entrustable Professional Activities is an important step in promoting workplace-based assessment and feedback.

Alternative educational methods may also serve an important role in improving resident competency. In a recent study by Collings et al.,[24] a porcine simulation lab was used to assess chief surgical resident operative performance. Not only was the porcine lab found to be highly feasible, the simulation also helped identify residents with significant technical deficiencies that were able to be remediated prior to graduation. In a randomized controlled trial, Bonrath and colleagues[25] demonstrated that comprehensive surgical coaching is superior to conventional training for the development of technical skills in residents. Surgical coaching is gaining popularity and implementing these techniques in training programs may be another important adjunct to enhance the education and competency of surgical residents.

This study has several limitations. First, competency rates were predicted based on individual resident's cumulative procedure-specific performance in SIMPL. If the bulk of a resident's evaluations of a particular procedure were early in their training, their growth may not be fully accounted for in our model. While we did only include residents who requested a SIMPL evaluation during their PGY 5 year and controlled for time in the academic year to help circumvent this, predicted competency by the end of training may be underestimated for some residents. Additionally, sampling bias may be introduced as residents and faculty choose which cases to evaluate in the SIMPL app. Finally, while our sample size was large, all programs included were members of SIMPL and evaluations may not accurately represent the entire population of graduating US surgical residents.

The variation in the competency of graduating general surgery residents is vast even for the most common general surgery procedures. The transition to a more competency-based medical education system may be a promising solution to decrease this variability and ensure the readiness of all trainees for independent practice,[23] yet that evolution will likely take decades to be fully realized. More incremental improvements to surgical education could certainly be implemented now. These efforts could immediately improve surgical education, and by extension, the outcomes of surgical patients.

REFERENCES

1. Fecso AB, Szasz P, Kerezov G, Grantcharov TP. The effect of technical performance on patient outcomes in surgery: a systematic review. *Ann Surg*. 2017;265(3):492. https://doi.org/10.1097/SLA.0000000000001959.

2. Common Program Requirements (Residency). 2021. Accreditation Council for Graduate Medical Education (ACGME). 2021. https://www.acgme.org/globalassets/PFAssets/ProgramRequirements/CPRResidency2021.pdf. Accessed November 30, 2021.

3. Napolitano LM, Savarise M, Paramo JC, et al. Are general surgery residents ready to practice? A survey of the American College of Surgeons Board of Governors and Young Fellows Association. *J Am Coll Surg*. 2014;218(5):1063–1072.e31. https://doi.org/10.1016/j.jamcollsurg.2014.02.001.

4. Mattar SG, Alseidi AA, Jones DB, et al. General surgery residency inadequately prepares trainees for fellowship: results of a survey of fellowship program directors. *Ann Surg*. 2013;258(3):440–449. https://doi.org/10.1097/SLA.0b013e3182a191ca.

5. Anderson TN, Payne DH, Dent DL, Kearse LE, Schmiederer IS, Korndorffer JR. Defining the deficit in US surgical training: the trainee's perspective. *J Am Coll Surg*. 2020;232(4):623–627. https://doi.org/10.1016/j.jamcollsurg.2020.11.029.

6. George BC, Bohnen JD, Williams RG, et al. Readiness of US general surgery residents for independent practice. *Ann Surg*. 2017;266(4):582–594. https://doi.org/10.1097/SLA.0000000000002414.

7. 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. https://doi.org/10.1016/j.jsurg.2016.08.010.

8. George BC, Bohnen JD, Schuller MC, Fryer JP. Using smartphones for trainee performance assessment: a SIMPL case study. *Surgery*. 2020;167(6):903–906. https://doi.org/10.1016/j.surg.2019.09.011.

9. 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):90–96. https://doi.org/10.1016/j.jsurg.2014.06.018.

10. DaRosa DA, Zwischenberger JB, Meyerson SL, et al. A theory-based model for teaching and assessing residents in the operating room. *J Surg Educ*. 2013;70(1):24–30. https://doi.org/10.1016/j.jsurg.2012.07.007.

11. Krumm AE, Marcotte KM, George BC. Model-based operative performance expectations: a

pathway for quantifying competency. *JAMA Surg*. 2023;158(5):515–521. https://doi.org/10.1001/jamasurg.2023.0014.

12. Buerkner PC. BRMS: an R package for Bayesian multilevel models using Stan. *J Stat Softw*. 2016;80 (1):1–28. https://doi.org/10.18637/jss.v080.i01.

13. Kay M. tidybayes: tidy data and geoms for Bayesian Models. 2023. doi:10.5281/ZENODO.1308151Please provide missing details for ref 13.

14. Malangoni MA, Biester TW, Jones AT, Klingensmith ME, Lewis FR Jr. Operative experience of surgery residents: trends and challenges. *J Surg Educ*. 2013;70(6):783–788. https://doi.org/10.1016/j.jsurg.2013.09.015.

15. Quillin RC, Cortez AR, Pritts TA, Hanseman DJ, Edwards MJ, Davis BR. Operative variability among residents has increased since implementation of the 80-hour workweek. *J Am Coll Surg*. 2016;222 (6):1201–1210. https://doi.org/10.1016/j.jamcollsurg.2016.03.004.

16. Bell RH, Biester TW, Tabuenca A, et al. Operative experience of residents in us general surgery programs: a gap between expectation and experience. *Ann Surg*. 2009;249(5):719–724. https://doi.org/10.1097/SLA.0b013e3181a38e59.

17. Anjaria DJ, Kunac A, McFarlane JL, Oliver JB. A 15-year analysis of surgical resident operative autonomy across all surgical specialties in veterans affairs hospitals. *JAMA Surg*. 2021;157(1):76–78. https://doi.org/10.1001/jamasurg.2021.5840.

18. Cortez AR, Winer LK, Kassam AF, et al. See none, do some, teach none: an analysis of the contemporary operative experience as nonprimary surgeon. *J Surg Educ*. 2019;76(6):e92–e101. https://doi.org/10.1016/j.jsurg.2019.05.007.

19. Sandhu G, Thompson-Burdine J, Nikolian VC, et al. Association of faculty entrustment with resident autonomy in the operating room. *JAMA Surg*. 2018;153(6):518–524. https://doi.org/10.1001/jamasurg.2017.6117.

20. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007;29(9-10):855–871. https://doi.org/10.1080/01421590701775453.

21. Jensen AR, Wright AS, Kim S, Horvath KD, Calhoun KE. Educational feedback in the operating room: a gap between resident and faculty perceptions. *Am J Surg*. 2012;204(2):248–255. https://doi.org/10.1016/j.amjsurg.2011.08.019.

22. Marcotte KM, Holmstrom AL, Thelen AE, Gates R, George BC, Krumm AE. Opportunities to increase workplace feedback: a multi-institutional examination. *J Surg Educ*. 2022;79(6):e124–e129. https://doi.org/10.1016/j.jsurg.2022.08.012.

23. Holmboe ES. Realizing the promise of competency-based medical education. *Acad Med*. 2015;90(4):411–413. https://doi.org/10.1097/ACM.0000000000000515.

24. Collings AT, Stefanidis D, Doster D, et al. Assessment of chief resident practice readiness in a porcine lab: a 4-year experience. *J Surg Educ*. 2021. https://doi.org/10.1016/j.jsurg.2021.11.009. S1931-7204(21)00329-9.

25. Bonrath EM, Dedy NJ, Gordon LE, Grantcharov TP. Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. *Ann Surg*. 2015;262(2):205–212. https://doi.org/10.1097/SLA.0000000000001214.

SUPPLEMENTARY INFORMATION

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jsurg.2023.11.005.