Received: 28 August 2020 | Revised: 9 December 2020 | Accepted: 4 January 2021

DOI: 10.1002/aet2.10574

ORIGINAL CONTRIBUTION

Longitudinal accuracy analysis of ultrasound performed during a four-year emergency medicine residency

Jessica Schleifer MD[1,2] | Rachel M. Haney MD, MSHPEd[2,3] | Hamid Shokoohi MD, MPH[2] | Calvin K. Huang MD, MPH[2] | Daniel Ratanski BS[2] | Heidi Kimberly MD[4] | Andrew S. Liteplo MD[2]

[1]Department of Anesthesia and Intensive Care Medicine, University Hospital, Bonn, Germany

[2]Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

[3]Department of Emergency Medicine, PeaceHealth Southwest Medical Center, Vancouver, Washington, USA

[4]Department of Emergency Medicine, Newton Wellesley Hospital, Harvard Medical School, Newton, Massachusetts, USA

Correspondence
Andrew S. Liteplo, 326 Cambridge Street, Boston, MA 02114, USA.
Email: andrew.liteplo@mgh.harvard.edu

Funding information
Hamid Shokoohi has received grant money from GE/EMF to perform research unrelated to this project. Calvin K. Huang has received grant money from GE/EMF to perform research unrelated to this project. Andrew S. Liteplo has received grant money from GE/EMF to perform research unrelated to this project.

Supervising Editor: Jason Wagner

Abstract

Background: The objective of this study was to analyze patterns of point-of-care ultrasound (POCUS) performance over 4 years of emergency medicine (EM) residency. Specifically, we aimed to study how accuracy and adherence to standards of scanning changed by postgraduate year (PGY).

Methods: This was a retrospective observational study of resident-performed POCUS at an academic emergency department over 6 years. We reviewed records of POCUS scans performed by PGY-1 to -4 residents that had been collected for quality assurance purposes. Data that were collected about EM residents' performance included the total number and type of scans per year, rate of technically limited scans (TLS), and accuracy on interpreting ultrasound images. Resident performances in each year (PGY-1 to -4) were independently evaluated and reported.

Results: During a 6-year period, 137 different EM residents performed 50,815 ultrasound scans. The median number of scans was 177 for PGY-1, 124 for PGY-2, 118 for PGY-3, and 76 for residents in PGY-4. The accuracy of scan interpretations were high across all PGY levels (>97%), but slight degradation was observed as residents progressed through residency. The TLS rate increased from 4.7% among PGY-1s to 13.6% as PGY-4s.

Conclusions: In this large cohort of POCUS studies by EM residents, POCUS accuracy rates decreased and rates of TLS significantly increased as residents progressed through residency.

INTRODUCTION

Point-of-care ultrasound (POCUS) training is well-established in emergency medicine (EM) residency programs in the United States. Demonstration of competency in POCUS has been required for graduates of United States EM residencies since 2001, yet EM resident POCUS education strategies vary widely.[1–4] In 2013, the American College of Emergency Physicians (ACEP), the Society for Academic Emergency Medicine (SAEM), and the Council of Emergency Medicine Residency Directors (CORD) published general guidelines for EM resident POCUS education and assessment.[1]

Current EM POCUS education literature suggest that POCUS education and assessment should occur longitudinally and through

Presented at the Society for Academic Emergency Medicine New England Regional Meeting, Worcester, MA, March 2019; and the Society for Academic Emergency Medicine Lightning Oral Presentation, Las Vegas, NV, May 2019.

a variety of modalities.[2,3,5–7] These recommendations follow educational theory of the spacing effect, that spreading learning out over time encourages deeper learning and more long-term retention of knowledge.[8] Additionally, it is well known that knowledge and skills degrade over time if not regularly used or reinforced.[9] ACEP POCUS guidelines emphasize that "one rotation without continued learning within the EM residency curriculum is inadequate." Annual POCUS education to reinforce knowledge and skills and/or continuous image review and quality assurance (QA) with direct resident feedback is encouraged.[2,6,7]

At most EM residency programs, a formal ultrasound rotation forms the centerpiece of the resident ultrasound experience. This occurs early in residency so that residents can use ultrasound clinically when caring for patients throughout their training. Subsequent education in later years is varied and less structured. It is not known is how resident ultrasound performance and interpretation changes over time or on how best to structured continued education throughout residency. Some studies have demonstrated how EM resident learning curves change with experience, although they report image acquisition and image interpretation accuracy as a function of number of examinations performed, not as a function of training year or time.[10–14]

The objective of this study was to analyze trends of ultrasound performance as a function of postgraduate year (PGY). We postulated that proximity in time to the ultrasound rotation may portend a higher level of accuracy and adherence to standards of scanning. Our goal was to identify any lapses in ultrasound skill to inform practices and educational models regarding longitudinal teaching of POCUS in an EM residency.

METHODS

Study design

We performed a retrospective analysis of an ultrasound database from a single EM residency training program. Records of POCUS scans performed by EM residents had been collected for QA purposes.

Study setting and population

Our EM residency is a 4-year program composed of approximately 60 EM residents, who rotate at two separate Level 1 trauma centers with a joint volume of over 200,000 patients per year. POCUS education is split between these two hospitals with a coordinated curriculum.

All residents had their ultrasound rotations early in residency. From 2012 to 2014, all PGY-1 residents had a 1-week rotation while PGY-2 residents had a 2-week rotation. From 2014 to 2018, PGY-1 residents completed a 4-week ultrasound rotation. A typical week of an ultrasound rotation would include didactics and hands-on scan shifts with trained POCUS faculty for approximately 18 h a week.

Residents' POCUS skills were assessed via image review and continuous QA of their clinical scans throughout the 4 years of residency. To be deemed competent to perform POCUS independently, residents were required to successfully complete a minimum of 300 expert-reviewed scans prior to graduation and to pass an Observed Structured Clinical Examination (OSCE) as PGY-3s. The OSCE assessed image acquisition, interpretation, and integration skills. Any noted deficiencies were reassessed as PGY-4s.

Throughout the 4 years of residency, residents performed ultrasound scans as part of clinical patient care and for educational purposes. These scans were recorded and saved as video clips and still images and stored in QPath (Telexy). Ultrasound scan interpretation was documented in the electronic medical record, EPIC for clinical scans, or in a structured form for educational scans. During this time period, 100% of residents' clinical and educational scans were reviewed by ultrasound faculty and fellows on a regular basis as part of our QA review. Each scan was assessed for accuracy to assure that images support interpretation. Scans were stratified into one of five QA categories: true positive (TP), true negative (TN), false positive (FP), false negative (FN), or technically limited scan (TLS). A TLS rating was given when sets of images were not adequate enough to support the interpretation, when the image quality was poor, or when minimal scanning criteria were not met (Data Supplement S1 available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10574/full).

Faculty or fellow interpretation of the scan served as the reference standard. At our program, early in the year, scans are reviewed together by fellows and faculty as part of the fellow learning process. Once fellows are trained in QA review and deemed by faculty to consistently perform in the same manner as U.S. faculty, they are allowed to be the only reviewer of a scan. The majority of scans are clearly TN, TP, or TLS (i.e., when minimal scanning criteria are not met). In most of these cases, fellows may have served as the only arbiters of quality. Any scans that were FN or FP or that the fellows were unsure about were flagged and additionally reviewed by a faculty member. Residents received written feedback via email on their scans and interpretations, especially when errors in performance or interpretation were noted.

All scans were entered into QA database (Microsoft SQL). Our database was a Web-based database housed on our hospital's HIPAA-compliant server, developed by the division of ultrasound. Information tracked in the database included patient information, date of scan, indication for imaging, performing physician, reviewing physician, and QA category.

Study protocol

For this study, we reviewed all scans performed from July 2012 to June 2018, for 6 complete academic years of data. Our QA database allowed for searching with specified criteria. We queried the database by each resident separately and filtered by dates so that we could see total numbers and overall performance for each resident by year.

Measurements or key outcome measures

We used two measurements of performance: accuracy and TLS rate. Accuracy was defined as the percentage of adequate scans that were correctly interpreted ((TP + TN)/(TP + TN + FP + FN)). TLS rate reflected what percentage of scans were inadequate for interpretation and was defined as TLS/(TP + TN + FP + FN + TLS). We analyzed both measures based on the PGY level (1 through 4) of the performing physician.

Data analysis

We used RStudio cloud alpha 2019 for the data analysis (R Core Team (2019). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/) and Microsoft Excel for Mac Version 16.32 for statistical analysis. The main outcomes of accuracy and TLS rate were analyzed with a random effects regression model in R using the plm package[15] to determine association between accuracy and PGY and TLS rate and PGY. The data was not normally distributed and therefore median and first to third quartiles were used to describe the data. The study was reviewed by the local institutional review board and which determined a quality improvement initiative and therefore exempt.

RESULTS

A total of 50,815 unique POCUS scans performed by 137 different residents over 6 academic years were analyzed. There was a range of 88 to 90 residents in each PGY (approximately 15 residents in each PGY over 6 years). The number of scans per PGYs varied in our cohort with a continuous reduction in overall number of scans from PGY-1 to -4. The median (interquartile range [IQR]) annual number of POCUS scans performed decreased from 176 (IQR = 105–242) for PGY-1 residents to 76 (IQR = 43–138) for PGY-4 residents (Table 1). Cardiac (41%), lung (21%), and abdominal FAST (12%) were the most common scans performed by EM residents (Figure 1).

Overall, residents showed a very high accuracy in image interpretation with some degradation with higher level of training. Accuracy was highest among PGY-1 scans (98.8%), decreased through PGY-3 (97.1%), and then increased slightly among PGY-4 scans (97.6%). These changes were statistically significant ($p < 0.001$) and are shown in Figure 2A.

Residents also had a higher rate of TLS as they became more senior. TLS rate increased every year, from 4.7% as PGY-1s to 13.6% as PGY-4s. These changes were statistically significant ($p < 0.001$) and are shown in Figure 2B.

DISCUSSION

In one of the largest longitudinal studies to date, we present our findings on resident POCUS performance in an academic emergency department (ED) with a well-established residency and ultrasound program as a function of PGY. Previous reports of longitudinal POCUS education in EM residencies have reported image acquisition and image interpretation accuracy as a function of number of examinations performed, not as a function of training year or time.[10–13] A study by Blehar et al.[10] showed that 50 to 75 scans per application are needed to achieve "excellent interpretation and good image quality." Our QA project was undertaken to identify trends in performance over the course of a 4-year EM residency

TABLE 1 Resident scan numbers and accuracy by PGY

PGY	Residents	Scans	Median	IQR	TLS rate (%)	Accuracy (%)
1	88	15,651	177	105–243	4.7	98.8
2	90	12,968	124	68–196	6.8	97.6
3	89	12,966	118	82–194	11.9	97.1
4	88	9230	76	43–138	13.6	97.6

Abbreviation: TLS, technically limited scan.

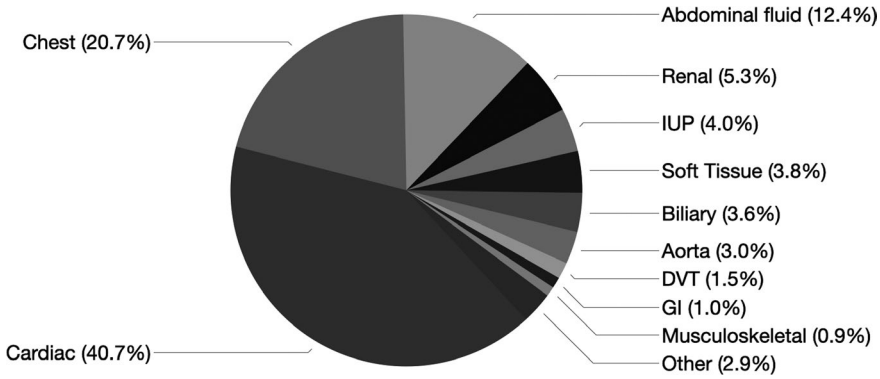

FIGURE 1 Distribution of types of ultrasound applications performed by EM residents. DVT, deep vein thrombosis; GI, gastrointestinal; IUP, intrauterine pregnancy

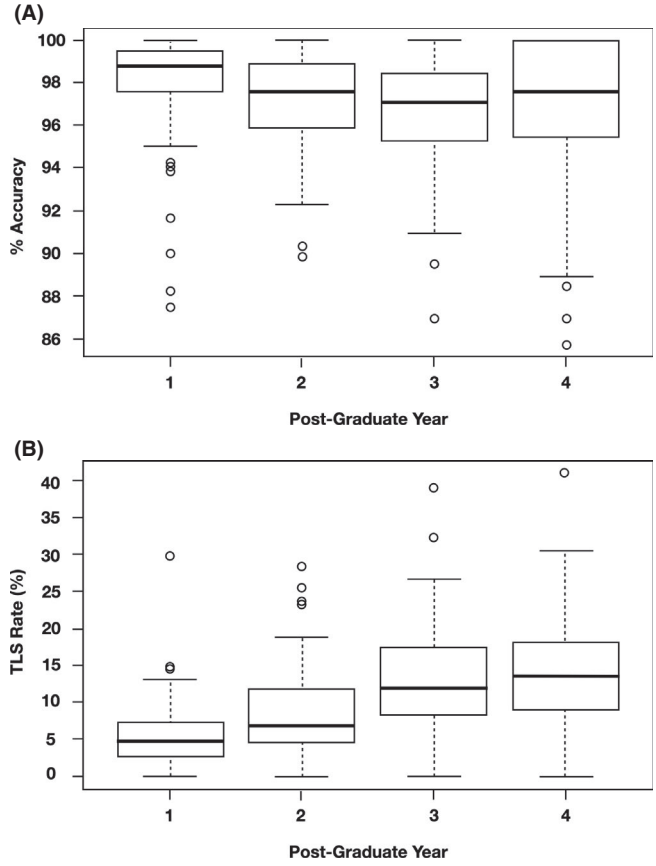

FIGURE 2 (A) Accuracy as a function of PGY. (B) TLS rate as a function of PGY. PGY, postgraduate year; TLS, technically limited scan

training in an effort to glean actionable data to inform additional educational initiatives. We noted several interesting trends.

First, the number of scans performed per year as residents progressed through their training dropped significantly. Many factors could have contributed to this finding. The bulk of the structured ultrasound education occurs in the PGY-1 year. Additionally, it is possible that once residents meet competency requirements (a minimum of 30 scans for each POCUS application) they cease to document or perform scans for educational purposes. In the PGY-4 year we saw a major drop in documented scans. Some part of this may be that in their supervisory role in the ED, senior residents might have delegated some POCUS examinations to junior residents. While we have a mechanism in place by which senior residents could be credited for these scans as well, it is not automated and requires manual entry and as such leads to intermittent compliance. Regardless, the finding that residents perform fewer scans each year is an important one. As a remedy, additional educational initiatives or incentives could be considered.

Second, the rate of TLS scans increased every year. There are a number of reasons that a scan may be technically limited. Most commonly, this is dependent on the operator. As residents progress through their training, each year they are further removed from their ultrasound rotation, and they may forget the minimal criteria required for each examination. It is also possible that lack of constant supervision of their ultrasound studies leads to an apathy of sorts, and their sense of importance of correctly performing or saving complete scans wanes. To be fair, sometimes the factors are specific to patients (i.e., difficult habitus) or situations (i.e., completing a study would have interfered with patient flow). However, it is unlikely that this proportion of patients or situations would change over time, unless residents became more comfortable attempting scans in more challenging patients rather than selecting easier patients to scan when first learning.

Finally, there was a trend toward a decreasing accuracy as residents progressed through training. Accuracy of scans fell from first to third PGY and then increased slightly in the PGY-4 year. While overall sensitivities and specificities remained very high, the decrease was statistically significant and warrants pause. As above, the ultrasound rotation occurs early in residency and is the bulk of the formal education. It is possible that the great length of time since this rotation resulted in a decrease in skill and accuracy. It is also likely that there is an association between the decreased number of scans performed by senior residents and their quality. This is in line with prior studies that support competency being directly proportional with experience of scan numbers. Finally, there may be an element of the Dunning-Kruger effect. Residents might have grown overconfident in their ultrasound skills and not have been fully aware of their limitations and therefore lost some proficiency in the acquisition and interpretation of their scans.[16] Despite this finding of a statistically significant decrease in accuracy, it is important to note that resident accuracy over the 4 years of training remained very high, and it is questionable whether or not this decrease is clinically significant. However, in light of the concordant data about an increase in TLS rate, an overall decrease in performance over time warrants consideration.

Our findings were surprising and informed interactive changes to our educational POCUS curriculum. PGY-1 residents demonstrated a high accuracy, high number of scans, and low TLS rate. Our data show that the ultrasound education provided during the intern ultrasound rotation is of high quality and effective in conveying necessary skills. Yet over time, there is atrophy of skills. We have since introduced additional longitudinal educational initiatives, such as in-person meetings for senior residents to review their scan numbers and quality and a 1-week refresher rotation for PGY-4 residents. We have also introduced a POCUS mentorship program, with each resident assigned to an ultrasound faculty mentor who can track progress longitudinally. The goal is to prevent attrition of skills so that residents are competent, proficient, and comfortable with the use of POCUS for clinical care of patients upon graduation. Future prospective studies may allow educational researchers to investigate additional factors that impact POCUS training, including residents' individual characteristics and real-time supervision by faculty.

LIMITATIONS

Given that our study was only at a single academic residency, it has limited generalizability to other residency programs. Additionally,

our study is limited by the bias inherent to any retrospective study. We cannot account for unknown confounding factors that may have contributed to our findings. Another limitation of this study is that while we believe that our QA database captured the vast majority of scans, it is possible that a minority of scans were missed and not entered into the database. It is possible that actual resident performance could be different than what our data showed, although we think that with our large sample size, this is unlikely. We also acknowledge the potential confounder of faculty presence and supervision of image acquisition during the ultrasound rotation. This could partially account for some of the increase in accuracy and decrease in TLS rate by PGY-1 residents that we observed. Finally, our methodology was limited because we did not pursue an inter-rater reliability assessment among image reviewers.

CONCLUSIONS

In this large cohort of point-of-care ultrasound scans in an emergency medicine residency, the majority of the residents achieved competency with very high accuracy in their interpretation. However, point-of-care ultrasound accuracy rates decrease, and rates of technically limited scan increase as emergency medicine resident progress through residency.

ACKNOWLEDGMENTS

The authors thank Anshul Kumar for his statistical expertise and help in this project.

CONFLICT OF INTEREST

Andrew S. Liteplo has consulted for Philips Healthcare and is involved in ultrasound-related research sponsored by Fujifilm/Sonosite, unrelated to this project. The other authors have no potential conflicts to disclose.

AUTHOR CONTRIBUTIONS

Jessica Schleifer—study concept and design, acquisition of data, analysis and interpretation, statistical expertise, drafting of the manuscript, critical revision of the manuscript. Rachel M Haney—study concept and design, acquisition of data, analysis and interpretation, drafting of the manuscript. Hamid Shokoohi—critical revision of the manuscript, statistical expertise. Calvin K. Huang—study concept and design, acquisition of data, analysis and interpretation critical revision of the manuscript, statistical expertise. Daniel Ratanski—acquisition of data. Heidi Kimberly— study concept and design, critical revision of the manuscript. Andrew S. Liteplo—study concept and design, analysis and interpretation, drafting of the manuscript, critical revision of the manuscript.

ORCID

Rachel M. Haney https://orcid.org/0000-0002-6145-6111
Andrew S. Liteplo https://orcid.org/0000-0002-3439-6651

REFERENCES

1. Lewiss RE, Pearl M, Nomura JT, et al. CORD-AEUS: consensus document for the Emergency Ultrasound Milestone Project. *Acad Emerg Med*. 2013;20(7):740-745.
2. Damewood SC, Leo M, Bailitz J, et al. Tools for measuring clinical ultrasound competency: recommendations from the ultrasound competency work group. *AEM Educ Train*. 2019;4(Suppl 1):S106-S112.
3. Amini R, Adhikari S, Fiorello A. Ultrasound competency assessment in emergency medicine residency programs. *Acad Emerg Med*. 2014;21(7):799-801.
4. Ahern M, Mallin MP, Weitzel S, Madsen T, Hunt P. Variability in ultrasound education among emergency medicine residencies. *West J Emerg Med*. 2010;11(4):314-318.
5. Lewiss RE, Hoffmann B, Beaulieu Y, Phelan MB. Point-of-care ultrasound education. *J Ultrasound Med*. 2014;33(1):27-32.
6. American College of Emergency Physicians Policy Statement: Ultrasound Guidelines: Emergency, Point-of-Care, and Clinical Ultraosund Guidelines in Medicine. ACEP website. Published 2016. Accessed November 20, 2019. https://www.acep.org/globalassets/new-pdfs/policy-statements/ultrasound-guidelines---emergency-point-of-care-and-clinical-ultrasound-guidelines-in-medicine.pdf.
7. Atkinson P, Bowra J, Lambert M, Lamprecht H, Noble V, Jarman B. International Federation for Emergency Medicine point of care ultrasound curriculum. *Can J Emerg Med*. 2015;17(2):161-170.
8. Gooding HC, Mann K, Armstrong E. Twelve tips for applying the science of learning to health professions education. *Med Teach*. 2017;39(1):26-31.
9. Pusic MV, Kessler D, Szyld D, Kalet A, Pecaric M, Boutis K. Experience curves as an organizing framework for deliberate practice in emergency medicine learning. *Acad Emerg Med*. 2012;19(12):1476-1480.
10. Blehar DJ, Barton B, Gaspari RJ. Learning curves in emergency ultrasound education. *Acad Emerg Med*. 2015;22(5):574-582.
11. Jang TB, Ruggeri W, Dyne P, Kaji AH. Learning curve of emergency physicians using emergency bedside sonography for symptomatic first-trimester pregnancy. *J Ultrasound Med*. 2010;29(10):1423-1428.
12. Peyrony O, Legay L, Morra I, et al. Monitoring personalized learning curves for emergency ultrasound with risk-adjusted learning-curve cumulative summation method. *AEM Educ Train*. 2018;2(1):10-14.
13. Gaspari RJ, Dickman E, Blehar D. Learning curve of bedside ultrasound of the gallbladder. *J Emerg Med*. 2009;37(1):51-56.
14. King A, Tyransky A, Coffman A, et al. A formalized three year emergency medicine residency ultrasound education curriculum. *JETem*. 2016;1(2).
15. Croissant Y, Millo G. Panel data econometrics in R: the plm package. *J Stat Softw*. 2008;27(2):1-43.
16. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999;77(6):1121-1134.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Schleifer J, Haney RM, Shokoohi H, et al. Longitudinal accuracy analysis of ultrasound performed during a four-year emergency medicine residency. *AEM Education and Training*. 2021;5:1–5. https://doi.org/10.1002/aet2.10574