Interpretation of Basic Clinical Images: How Are Surgical Residents Performing Compared to Other Trainees?

Joseph J. Eid, MD,[*,1] Maia I. Reiley, MD,[*,1] Angela L. Miciura, BS,[†,1] Francisco Igor Macedo, MD,[*,1] Edsa Negussie, MD,[‡,1] and Vijay K. Mittal, MD[*,1]

[*]Department of Surgery, Providence-Providence Park Hospital, Michigan State University College of Human Medicine, Southfield, Michigan; [†]Michigan State University College of Human Medicine, East Lansing, Michigan; and [‡]Department of Radiology, Providence-Providence Park Hospitals, Michigan State University College of Human Medicine, Southfield, Michigan

BACKGROUND: During medical training students, residents, and fellows learn how to accurately interpret basic radiographic images. This skill is mostly utilized by physicians in the acute and critical care settings. It is unclear whether surgical residents' interpretation skills differ from that of other trainees.

METHODS: A 30-question online quiz was developed to evaluate trainees' skills in interpreting images using various radiologic modalities. The participating cohort included (1) medical students (MS), (2) general surgery residents (GST), internal medicine residents and fellows (IMT), and radiology trainees (RT). The impact of residency specialty and level of training on performance was evaluated.

RESULTS: A total of 69 postgraduate trainees and 19 MS enrolled in the online quiz. The average score was 67.6% (±16.6). GST scored higher than IMT (74.2% ± 10.7% vs. 67.9% ± 11.3%, $p = 0.038$); however, they were equally proficient to RT. MS had the lowest interpretation accuracy rates compared to postgraduate trainees (57.4% ± 16.8%, $p < 0.001$). On different radiographic modalities, junior GST performance was comparable to MS, JR-IMT, and Junior Radiology Trainees (JR-RT). On computed tomography (CT) body, GST (83.1% ± 15.7%) scored higher than IMT (70.3% ± 17.7%, $p = 0.026$) and MS (61.7% ± 23.4%, $p < 0.001$). Similar findings were demonstrated on ultrasound modality. A difference in performance was not evident for X-rays, CT head, and tubes/lines localization images.

CONCLUSIONS: GST were able to correctly interpret 74.2% of basic clinical images. Although superior in the evaluation of pathologies seen on CT body and ultrasound, GST have comparable performance to other trainees in X-rays, tube/line localization images, and CT head. Integration of radiology education in surgical training may enhance performance and potentially improve patient care. (J Surg Ed 76:1500−1505. © 2019 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: radiology interpretation, surgical education, imaging interpretation, postgraduate training

COMPETENCIES: Practice-Based Learning and Improvement, Patient Care

INTRODUCTION

Medical students (MS), residents, and fellows are expected to analyze various diagnostic modalities as an integral part of delivering optimal patient care. The need for independent radiological interpretation by nonradiology trainees in the acute care setting has become a recent topic of investigation.[1] Providers should be able to enact treatment plans based on their own interpretation of the available imaging. Major deficiencies in identifying normal and abnormal chest X-rays (XR), and renal sonography among different subspecialties has been noted.[2-4] However, some trainees do not receive structured radiology education to overcome these limitations. In fact, they often learn independently from senior

Oral presentation at the Southeastern Surgical Congress, February 2018, Tampa Florida.

Funding: The authors report no external funding.

Correspondence: Inquiries to Joseph J. Eid, MD, Department of Surgery, Providence-Providence Park Hospital, Michigan State University College of Human Medicine, 16001 W 9 Mile Rd, Southfield, MI 48075; fax: (248) 849-5847; e-mail: eidjose1@msu.edu

[1] All authors contributed equally to this research project and manuscript.

1931-7204/$30.00
https://doi.org/10.1016/j.jsurg.2019.04.011

trainees (SRT) or attending physicians, resulting in major variations in quality and proficiency.[5] Suboptimal radiological training begins during medical school, as only 25% of US allopathic and osteopathic schools require a core radiology clerkship.[6] Studies evaluating MS's competency have demonstrated poor performance in interpreting basic chest radiographs expected for their level of training.[7]

The utilization of radiographic imaging is crucial in the operative planning by surgical residents and general surgeons.[2,5] Analysis of such imaging has become an integral part of trauma and acute care, where immediate interventions are indicated.[11] Variability in surgical residents' accuracy rates in interpreting images has been reported.[12] Assessment of interspecialty performance variations is currently lacking. Herein, the proficiencies and deficiencies in diagnostic imaging interpretation between surgical trainees and other training programs were analyzed. Given their similar limited clinical exposure and focused radiological training, we sought to compare junior GST's (JR-GST) performance to that of MS.

By identifying these deficiencies, GST training may be tailored to meet educational objectives and improve patient care. These benefits may also extend to other training programs.

METHODS

A 30-item web-based test was created on Proprofs.com that included 5 different categories: (1) XR of the chest and abdomen, (2) computed tomography (CT) of the head, (3) CT body, (4) ultrasound (US), and (5) invasive line/tube localizing images. True/false and multiple-choice questions were utilized, which allowed for single or multiple answers. Questions included a radiographic image along with a brief clinical vignette. Five questions were generated for each radiographic modality. An expert panel selected images from a previously established radiology education database. The panel included clinical radiologists and program directors of the participating groups of interest. Items that were considered common conditions encountered by trainees in critical care patients were chosen and then deidentified. Examples of the clinical conditions included pneumonia with consolidation, foreign body in a pediatric patient, ischemic stroke, subdural hematoma, pancreatitis, gastrointestinal perforation, pericardial effusion, deep venous thrombosis, and the location of a central venous catheter and nasogastric tube.

Specific groups of interest were targeted for this study: general surgery residents (GST), internal medicine trainees (IMT), radiology residents (RT), and MS. IMT included internal medicine residents, pulmonary critical care fellows, and cardiovascular (CV) fellows to match the postgraduate training level of GST. MS served as a control group for junior examinees. All groups were recruited from a single tertiary university-affiliated community hospital in Southeastern Michigan. Upon approval from the institutional review board, initial invitation emails were sent out to participants where enrollment was voluntary.

Postgraduate year 1 (PGY1) and PGY2 participants were considered junior trainees (JRT) and PGY3 to PGY6 were considered SRT. PGY2/R1 and PGY3/R2 RT were considered JR-RT, and PGY4/R3 and PGY5/R4 were considered senior residents (SR-RT). Competency in imaging interpretation was established by the percentage of correct answers. GST performance was compared to trainees from different specialties.

Statistical Analysis

Student's t test and analysis of variance (ANOVA) F test followed by post hoc Tukey-Kramer test were used to analyze mean differences in resident accuracy rates between programs, and different levels of training. Categorical proportions were analyzed using chi-square test (χ^2) or Fisher's exact test as appropriate. Data is presented as mean (±standard deviation) and counts (percent, %). The $p < 0.05$ was considered statistically significant. Statistical analysis was performed using Prism 7.0 software (GraphPad Inc., San Diego, CA).

RESULTS

Of the 108 eligible trainees, 88 subjects (81.4%) participated from 2 medical schools (1 allopathic and 1 osteopathic), 3 residency (GS, IM, and radiology), and 2 fellowship programs (pulmonary critical care and cardiovascular). A total of 19/25 MS (76%), 22/23 GST (95.6%), 38/48 IMT (81.5%), and 9/12 RT (75%) participated. Most postgraduate trainees were PGY-3 ($n = 17$, 24.6%) as compared to PGY-1 ($n = 16$, 23.2%), PGY-2 ($n = 16$, 23.2%), PGY-4 ($n = 9$, 13%), PGY-5 ($n = 8$, 11.6%), and PGY-6 ($n = 3$, 4.4%). Male participants ($n = 55$, 62.5%) outnumbered female participants ($n = 33$, 37.5%).

The overall average score on the assessment test was 67.6% (±16.6). A difference in performance was identified among the 4 training programs ($p < 0.001$). Table 1 highlights these results. Overall, GST and RT were equally proficient ($p = 0.350$). GST accurately interpreted 74.2% (± 10.7) of the queried images correctly compared to their IM colleagues (67.9% (±11.3), $p = 0.038$). All postgraduate trainees were more competent at interpreting images compared to MS ($p < 0.001$). No difference in accuracy rates was identified based on

TABLE 1. Participant Demographics and Performance

	MS	IMT	GST	RT	p Value
Participants, *N* (%)	19 (21.6%)	38 (43.2%)	22 (25%)	9 (10.2%)	
Level of training, *N* (%)					ns
Junior		21 (55.3%)	13 (59.1%)	5 (55.6%)	
Senior		17 (44.7%)	9 (40.9%)	4 (44.4%)	
Participation rate, *N* (%)	19/25 (76%)	38/44 (79.1%)	22/23 (95.6%)	9/12 (75%)	ns
Gender, *N* (%)					ns
Male	11 (57.9%)	27 (71%)	14 (63.7%)	3 (33.3%)	
Female	8 (42.1%)	11 (29%)	8 (36.4%)	6 (67.7%)	
Exam duration, X min (SD)	24.5 (19.5)	21.8 (14.7)	29.4 (42.9)	32 (30.5)	ns
Work shift, *N* (%)					ns
AM	11 (57.9%)	18 (47.4%)	15 (68.2%)	6 (67.7%)	
PM	8 (42.1%)	20 (52.6%)	7 (31.8%)	3 (33.3%)	
Performance, % (SD)	57.4% (16.8)	67.9% (11.3)	74.2% (10.8)	83% (5.9)	<0.0001*
Junior		64.1% (12.1)	70% (11)	80.7% (6.4)	0.02*
Senior		71% (9.8)	80.4% (7.2)	86% (4.2)	0.003*
Modality performance, % (SD)					
X-rays	46.8% (25)	61.4 % (16.5)	64.1% (18.4)	76.1% (11.8)	0.0013*
CT scan (head)	61.4% (20.1)	72.8% (24.3)	72.7% (19.6)	100% (0)	0.0003*
CT scan (body)	61.7% (23.4)	70.3% (17.7)	83.1% (15.7)	92.1% (10.4)	<0.0001*
Ultrasound	48.3% (22.8)	61% (19.5)	68.2% (21.2)	68.5% (15.5)	0.013*
Invasive lines/tubes	79% (20.5)	82.1% (19.1)	88.2% (15.9)	88.9% (14.5)	ns

ns, not statistically significant.
*p < 0.05, statistically significant.

gender (p = 0.121). Both graduate and postgraduate trainees required the same amount of time to complete the test (p = 0.631; Table 1). The time for completion did not correlate with performance ($r = 0.10$, p = 0.351). Trainees who completed the assessment during working hours scored higher than those who participated after-hours (71.6% ± 13.8% vs. 65% ± 14%, p = 0.029).

A variability in accuracy rate between training categories was seen in the interpretation of different imaging modalities (Table 1). No difference in performance on X-ray images was evident among the postgraduate classes; however, their accuracy rates were superior to MS. GST were less proficient in identifying pathologies seen on CT (head) when compared to RT (p = 0.008). Performance in this modality was similar to IMT and MS. GST and RT were equally correct in their evaluation of CT (body) radiographs (83.1% [±15.7] vs. 92.1% [±10.4], p = 0.596); however, they outperformed their IM (p = 0.046) and MS (p = 0.0016) colleagues. In the US modality, the type of postgraduate training program did not impact accuracy rates; GST's performance was superior to MS's performance (p = 0.021). Examinees from all 4 training programs were equally proficient in evaluating the proper placement of invasive tubes and lines.

The level of training impacted both overall and specific radiographic modality proficiency. Table 2 summarizes the differences between JR participants. JR-GST (70% ± 11%) were not superior in interpreting basic clinical images when compared to JR-IMT (64.1% ± 12.1%, p = 0.181) and JR-RR (80.7% ± 6.4%, p = 0.178). JR-GST outperformed MS (p = 0.023). In assessing radiographs using the 5 queried modalities, the results did not demonstrate JR-GST superiority when compared to their postgraduate colleagues; in fact, they were more likely to misinterpret CT (head) images relative to JR-RT (p = 0.012). Similarly, JR-GST performance did not surpass MS in the XR, US, and localizing images.

Table 3 depicts the variability in accuracy rates among SR trainees. SR-GST analyzed more basic images correctly in the CT (body) and US modality when compared to SR-IMT. This resulted in higher overall proficiency (80% ± 7.2% vs. 71% ± 9.8%, p = 0.021); however, performance was comparable to SR-RT (p = 0.560). In evaluating XR, CT (body), US. and localizing radiographs, SR-GST demonstrated similar accuracy rates as SR-RT. Conversely, they were outperformed in the CT (head) modality.

DISCUSSION

The proficiencies and deficiencies across multiple specialties and different levels of training were captured. Both medical and surgical trainees were less competent in the interpretation of certain clinical images compared to their radiology counterparts. Since postgraduate trainees circumstantially provide clinical decisions based on their own preliminary evaluation of images, such

TABLE 2. Junior Resident Comparative Performance

	GST	IMT	p_1	RT	p_2	MS	p_3
Participants, *N* (%)	13 (59.1%)	17 (44.7%)		5 (55.6%)		12 (100%)	
Duration, X minutes (SD)	38.6 (54.8)	22.7 (19.8)		37 (35)		24.5 (19.5)	ns
Performance, % (SD)	70.1% (11.1)	64.2% (12.2)	ns	80.7% (6.4)	ns	59.4% (18.1)	0.023
X-rays	59.8% (20.6)	55.5% (17.6)	ns	73.3% (14.9)	ns	46.8% (25)	ns
CT scan (head)	74.4% (20)	70.6% (26)	ns	100% (0)	0.012	61.4% (20.1)	0.04
CT scan (body)	76.9% (16)	66.4% (18.2)	ns	88.6% (12)	ns	61.7% (23.4)	0.05
Ultrasound	60.3% (23.1)	58.8% (20.5)	ns	70% (7.5)	ns	48.3% (22.8)	ns
Invasive line/tube	87.7% (17.4)	80% (21.2)	ns	84% (16.7)	ns	79% (20.5)	ns

$p < 0.05$, statistically significant; ns, not statistically significant.

findings are of high educational value and clinical importance. The occurrence of such instances is highly dependent on the acuity and complexity of care provided by postgraduate trainees, and the 24-hour coverage by in-house radiologists. In the perioperative and trauma setting, surgeons heavily rely on individual interpretation of clinical images to provide immediate interventions. In the evaluation of basic clinic images, an objective assessment test is needed given the limited data comparing surgical trainees' performance to that of other trainees. Postgraduate trainees significantly outperformed MS. These results are expected, as not all medical schools require radiology clerkships, and the emphasis on radiology education is limited. Numerous studies have identified MS as lagging behind residents in several imaging modalities including chest radiographs, spine radiographs, and US.[2,8]

Despite the heavy reliance on plain radiographs by physicians, previously published literature has demonstrated deficiencies in the interpretation of chest X-ray by nonradiologists.[2,9,10] Surgical and nonsurgical trainees utilize chest/abdomen radiographs in the acute and critical care setting where immediate analysis of images is required; however, accuracy rates can be as low as 20%.[10] Although no difference was identified in the overall performance of postgraduate trainees in this modality, increased level of training is associated with improved accuracy rates, consistent with previously published data.[2,9,10] At both the junior and SRT level, GST performance was not statistically different from their RT counterparts. JR-GST correctly interpreted 59.8% of the plain radiographs, while SR-GST correctly interpreted 70.4%. Despite being comparable to RT, an opportunity was identified to ameliorate GST performance in an imaging modality of such importance.[2] Education can be provided through formal didactics, computer-based training, and hands-on application during daily patient-care activities.[2] Such educational strategies may be beneficial for MS, who incorrectly interpreted more than 50% of the queried chest/abdomen radiographs. MS performance in XR was not significantly different from JR-GST, indicating that JR-GST are not better equipped to interpret radiographs. This further emphasizes the necessity of addressing this deficit early on in training. Despite the suboptimal performance by MS, they accurately localized the placement of invasive tubes and lines that are routinely placed in the intensive care setting. Their performance was comparable to surgical and nonsurgical postgraduate trainees.

Since the advent of the CT scanner, it has become an integral part of diagnostic medicine. Its utility expands to both surgical and nonsurgical pathologies. While surgeons consistently utilize CT scan images in the trauma

TABLE 3. Senior Resident Comparative Performance

	GST	IMT	p_1	RT	p_2
Participants, *N* (%)	9 (40.9%)	21 (55.3%)		4 (44.4%)	
Duration, X minutes (SD)	16.23 (3.1)	21.1 (9.2)	ns	25.8 (27.4)	ns
Performance, % (SD)	80.4% (7.2)	71% (9.8)	0.021	85.8% (4.2)	ns
X-rays	70.4% (13.6)	66.1% (14.3)	ns	80.6% (5.6)	ns
CT scan (head)	70.4% (20)	74.6% (23.3)	ns	100% (0)	0.02
CT scan (body)	92.1% (10.4)	73.5% (17.1)	0.007	96% (7.1)	ns
Ultrasound	79.6% (11.1)	62.7% (18.9)	0.03	75.8 (8.8)	ns
Invasive line/tube	88.9% (14.5)	83.8% (17.5)	ns	95% (10)	ns

$p < 0.05$, statistically significant; ns, not statistically significant.

evaluations, internists currently employ it for the diagnoses of various medical pathologies such as cerebrovascular events, acute pulmonary emboli, pancreatitis, and evaluation of malignancies. The results of this online quiz demonstrated comparable proficiency and accuracy rates between GST and RT in evaluating CT body images. Concurrently, IMT and MS demonstrated lower proficiency in this modality. This is due to the frequent exposure of GST to CT body images. IMT and MS have limited exposure and heavy reliance on radiologists' interpretation, despite numerous medical pathologies that require CT body imaging for accurate diagnosis. The GST's high level of proficiency in CT body interpretation corroborated prior studies, which demonstrated similar findings specifically in the acute care and trauma surgery patient population.[4,5,12] To the best of our knowledge, this is the first study to investigate IMT's competency in interpreting this radiographic modality. Previous reports have shown equal proficiency between senior surgical residents and RT in the detection of traumatic intracranial pathologies on CT of the head.[11] This was inconsistent with the performance of both surgical and medical trainees who scored 20% and 26% lower than their radiology colleagues. The inclusion of both traumatic and nontraumatic intracranial pathologies in this examination may lead to the suboptimal accuracy rates of GST when compared to RT.

The use of point-of-care US by nonradiologists has expanded to both surgical and medical domains. Currently, practitioners emphasize the use of this modality mainly in the intensive care unit and emergency department for both diagnostic and therapeutic interventions. This modality is utilized for diagnostic purposes such as evaluating cardiac function, pericardial effusions, traumatic intraperitoneal hemorrhage, and acute cholecystitis. It has also been applied during several bedside procedures such as insertion of central venous catheters, directing thoracentesis, pericardiocentesis, and paracentesis. Despite the similar performance between postgraduate trainees in this modality, approximately one-third of the clinical images were incorrectly interpreted. The inability to show a statistically significant difference in performance between general surgery and the remaining trainees may be inherent to the nondiscriminatory nature of the queried images. SR-GST have been able to integrate and utilize US in the trauma setting with high competency.[13,14] The superiority of these trainees in this modality extends beyond the trauma setting, as the queried images included nontraumatic pathologies. This is mainly due to senior surgical resident's participation in advanced US courses.[14,15] A 5-hour training module has been shown to improve IM residents ability to accurately identify key sonographic findings at the level of a radiologist.[4] Additionally, formalized US courses have been shown to increase the level of confidence in the evaluation of US images among surgical residents.[15] This indicates that, given a formal training or course, both surgical and nonsurgical trainees can improve their versatility in the utilization of point-of-care US.

We acknowledge that our study has several limitations. First, although our participation rate was satisfactory (81%), there is a possibility that our findings are not generalizable across other postgraduate training specialties. Second, to create a reasonable time commitment for participants, the expert panel only included 30 imaging items. Although the images selected were best representative of common conditions, they were not inclusive of every possible pathology seen in practice by MS, GST, IMT, and RT. Additional images may have been incorporated into the examination; however, that may have negatively impacted our participation rates. Third, the images used into the exam were static, making the interpretation of dynamic modalities like CT scan and US more difficult. Finally, the quiz was taken under unsupervised conditions. Answer sharing could have impacted the total percentages and affected our final results. Given these potential limitations, the authors still believe that the results of this study show promise for influencing postgraduate radiology training and creating a method for improving patient care.

In conclusion, physicians continue to make clinical decisions based on their own interpretation of imaging studies. It is imperative to provide trainees with necessary training to improve their diagnostic skills. Our findings show that although progress is seen throughout the length of training, both GST and IMT underperformed compared to RT. Due to higher exposure of GST to CT body, they had comparable accuracy rate in this imaging modality. As expected, GST outperformed MS in this examination. JR-GST had similar false interpretation rates in X-ray, invasive tube/line placement, and US images when compared to MS. We advocate that additional educational resources and opportunities should be implemented early during surgical residency training to improve radiological skills. Future studies should evaluate how dedicated radiology training in nonradiology residency programs would impact such skills and improve patient care.

AUTHORS' CONTRIBUTIONS

JJE, MIR, ALM, FIM, EN, and VM equally contributed to the conception and design of the study, or acquisition of data, or analysis and interpretation of data, drafting the article or revising it critically for important intellectual content, and final approval of the version to be submitted.

REFERENCES

1. Butler KL, Chang Y, DeMoya M, et al. Needs assessment for a focused radiology curriculum in surgical residency: a multicenter study. *Am J Surg*. 2016;211: 279-287.

2. Eisen LA, Berger JS, Hegde A, Schneider RF. Competency in chest radiography: a comparison of medical students, residents, and fellows. *J Gen Intern Med*. 2006;21:460-465.

3. Caronia J, Panagopoulos G, Devita M, et al. Focused renal sonography performed and interpreted by internal medicine residents. *J Ultrasound Med*. 2013;32:2007-2012.

4. Wong C, Taubman K, Hoehne F, et al. The quality of surgical resident interpretation of after-hour CT scans of the head and abdomen in trauma. *Am Surg*. 2005;71:772-775.

5. Khan SA, Sleem R, Schiralli M, et al. Assessment of surgical residents' ability to predict appendicitis with computed tomography. *Am Surg*. 2013;79:1308-1310.

6. Poot JD, Hartman MS, Daffner RH. Understanding the US medical school requirements and medical students' attitudes about radiology rotations. *Acad Radiol*. 2012;19:369-373.

7. Jeffrey DR, Goddard PR, Callaway MP, Greenwood R. Chest radiograph interpretation by medical students. *Clin Rad*. 2003;58:478-481.

8. Taylor J, Clopton PE. Interpretation of abnormal lumbosacral spine radiographs: a test comparing students, clinicians, radiology residents, and radiologists. *Spine*. 1995;20:1147-1154.

9. Kaufman B, Dhar P, O'Neill DK, et al. Chest radiograph interpretation skills of anesthesiologists. *J Cardiothorac Vasc Anesth*. 2001;15:680-683.

10. Gatt ME, Spectre G, Paltiel O, et al. Chest radiographs in the emergency department: is the radiologist really necessary? *Postgrad Med J*. 2003;79: 214-217.

11. Vorhies RW, Harrison PB, Smith RS, Helmer SD. Senior surgical residents can accurately interpret trauma radiographs. *Am Surg*. 2002;68:221-225.

12. Arentz C, Griswold JA, Halldorsson A, et al. Accuracy of surgery residents' interpretation of computed tomography scan in trauma. *Am J Surg*. 2008;196:809-812.

13. Kern SJ, Smith RS, Fry WR, et al. Sonographic examination of abdominal trauma by senior surgical residents. *Am Surg*. 1997;63:669-674.

14. Smith R, Kern SJ, Fry WR, Helmer SD. Institutional learning curve of surgeon-performed trauma ultrasound. *Arch Surg*. 1998;133:530-536.

15. Kotagal M, Quiroga E, Ruffatto BJ, et al. Impact of point-of-care ultrasound training on surgical residents' confidence. *J Surg Educ*. 2015;72:82-88.

SUPPLEMENTARY INFORMATION

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jsurg.2019.04.011.