AEM Education and Training

A GLOBAL JOURNAL OF EMERGENCY CARE

Elevating the human condition during times of emergency

ORIGINAL CONTRIBUTIONS

Emergency Medicine Resident Efficiency and Emergency Department Crowding
Ryan Kirby, Richard D. Robinson, Sasha Dib et al. 209

Impact of a Mentorship Program on Medical Student Burnout
Jaime Jordan, Daana Watcha, Courtney Cassella et al. 218

The Standardized Video Interview: How Does It Affect the Likelihood to Invite for a Residency Interview?
Abbas Husain, Ida Li, Brahim Ardolic et al. 226

Science Policy Training for a New Physician Leader: Description and Framework of a Novel Climate and Health Science Policy Fellowship
Jay Lemery, Cecilia Sorensen, John Balbus et al. 233

Resident Clinical Experience in the Emergency Department: Patient Encounters by Postgraduate Year
Amy Douglass, Kathleen Yip, Debryna Lumanauw et al. 243

Contents continued inside
Consensus Core Point-of-care Ultrasound Applications for Pediatric Emergency Medicine Training  
Allan Evan Shefrin, Fred Warkentine, Erika Constantine et al.  

Comparison of the Standardized Video Interview and Interview Assessments of Professionalism and Interpersonal Communication Skills in Emergency Medicine  
Laura R. Hopson, Michele L. Dorfsman, Jeremy Branzetti et al.  

Building Emergency Medicine Trainee Competency in Pediatric Musculoskeletal Radiograph Interpretation: A Multicenter Prospective Cohort Study  
Michelle Sin Lee, Martin Pusic, Benoit Carrière et al.  

Comparison of Manikin Versus Porcine Trachea Models When Teaching Emergent Cricothyroidotomy Among Emergency Medicine Residents  
Mark L. Gustafson, Brian Hensley, Marc Dotson et al. 

BRIEF CONTRIBUTIONS

Analysis of the Emergency Medicine Clinical Learning Environment  
V. Ramana Feeser, Zachary Zemore, Nital Appelbaum et al.  

Heart Rate Variability and Acute Stress Among Novice Airway Managers  
Jason M. Mefford, Sarah Kahle, Shikha Gupta et al. 

INNOVATIONS REPORT

Report of a Collaboration Between a Law School and an Emergency Medicine Residency Program for a Full-scale Medical Malpractice Litigation Simulation  
Mark Curato, Adam Shlahet 

COMMENTARY—INVITED

Change Is What You Make It  
Paul S. Auerbach 

CORRESPONDENCE—UNSOLICITED LETTERS TO THE EDITOR

Introducing Medical Student–led Quality Improvement Projects as a Value-added Learning Opportunity in the Emergency Department  
Haridha Sudhakara Pandian, Priyanka Shivaraman Iyer, Jasmine Kaur Bhangra et al. 

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus Core Point-of-care Ultrasound Applications for Pediatric Emergency Medicine Training</td>
<td>251</td>
</tr>
<tr>
<td>Comparison of the Standardized Video Interview and Interview Assessments of Professionalism and Interpersonal Communication Skills in Emergency Medicine</td>
<td>259</td>
</tr>
<tr>
<td>Building Emergency Medicine Trainee Competency in Pediatric Musculoskeletal Radiograph Interpretation: A Multicenter Prospective Cohort Study</td>
<td>269</td>
</tr>
<tr>
<td>Comparison of Manikin Versus Porcine Trachea Models When Teaching Emergent Cricothyroidotomy Among Emergency Medicine Residents</td>
<td>280</td>
</tr>
<tr>
<td>Analysis of the Emergency Medicine Clinical Learning Environment</td>
<td>286</td>
</tr>
<tr>
<td>Heart Rate Variability and Acute Stress Among Novice Airway Managers</td>
<td>291</td>
</tr>
<tr>
<td>Report of a Collaboration Between a Law School and an Emergency Medicine Residency Program for a Full-scale Medical Malpractice Litigation Simulation</td>
<td>295</td>
</tr>
<tr>
<td>Change Is What You Make It</td>
<td>299</td>
</tr>
<tr>
<td>Introducing Medical Student–led Quality Improvement Projects as a Value-added Learning Opportunity in the Emergency Department</td>
<td>301</td>
</tr>
</tbody>
</table>
AEM Education and Training
Editorial Board

A Journal of the Society for Academic Emergency Medicine | saem.org

Editor-in-Chief, Academic Emergency Medicine
Jeffrey A. Kline, MD, Indiana University School of Medicine

Editor-in-Chief, AEM Education and Training
Susan B. Promes, MD, Penn State College of Medicine

DECISION EDITORS
John Burkhardt, MD, PhD
University of Michigan

Teresa Chan, MD, MHPE, FRCP C
McMaster University

Esther H. Chen, MD
University of California
San Francisco Medical Center

Stephen John Cico, MD, MEd
Indiana University School of Medicine

Samuel Clarke, MD, MAS, FACEP
UC Davis Health System

Wendy C. Coates, MD
Harbor-UCLA Emergency Medicine

Nicole M. Deiorio, MD
Virginia Commonwealth University
School of Medicine

Ting Dong, PhD
Uniformed Services University of the Health Sciences

Daniel J. Egan, MD
Columbia University Vagelos
College of Physicians and Surgeons

Michael Gottlieb, MD
Rush University Medical Center

Jaime Jordan, MD
Harbor-UCLA Medical Center

Sorabh Khandelwal, MD
The Ohio State College of Medicine

Anne Messman, MD
Wayne State University School of Medicine

Daniel Runde, MD
University of Iowa Hospitals and Clinics

Sally Santen, MD, PhD
Virginia Commonwealth University
School of Medicine

Jason Wagner, MD
Washington University School of Medicine

Lainie M. Yarris, MD, MCR
Oregon Health & Science University

FELLOW-IN-TRAINING EDITOR
Jessica Fujimoto, MD
Temple University Hospital
Outstanding Peer Reviewers 2018

- William Bond, MD
- Holly Caretta-Weyer, MD
- Joshua Davis, MD
- Douglas Franzen, MD, MEd
- Michael Gottlieb, MD
- Nikhil Goyal, MD
- Phillip Harter, MD
- Ryan Pedigo, MD
- Dina Wallin, MD

The Outstanding Peer Reviewer designation is given annually to peer reviewers who meet the following criteria for excellent performance:

- provided at least 5 high quality reviews (score of 5);
- had no more than 1 late review;
- had a mean score of 85% or higher (≥ r score of 3.4 or higher); and
- accepted 2/3 of all review requests.

Peer reviewers are essential to presenting the high-quality academic contributions that fill the pages of Academic Emergency Medicine Education and Training each quarter. We appreciate all of our AEM E&T peer reviewers and are especially grateful to these five Outstanding Peer Reviewers for their dedicated, conscientious, and exceptional service to AEM E&T in 2018 in contributing timely and thoughtful peer reviews.
Emergency Medicine Resident Efficiency and Emergency Department Crowding

Ryan Kirby, MD, Richard D. Robinson, MD, Sasha Dib, MD, Daisha Mclarty, MD, Sajid Shaikh, MS, Radhika Cheeti, Amy F. Ho, MD, Chet D. Schrader, MD, Nestor R. Zenarosa, MD, and Hao Wang, MD, PhD

ABSTRACT

Objectives: Provider efficiency has been reported in the literature but there is a lack of efficiency analysis among emergency medicine (EM) residents. We aim to compare efficiency of EM residents of different training levels and determine if EM resident efficiency is affected by emergency department (ED) crowding.

Methods: We conducted a single-center retrospective observation study from July 1, 2014, to June 30, 2017. The number of new patients per resident per hour and provider-to-disposition (PTD) time of each patient were used as resident efficiency markers. A crowding score was assigned to each patient upon the patient’s arrival to the ED. We compared efficiency among EM residents of different training levels under different ED crowding statuses. Dynamic efficiency changes were compared monthly through the entire academic year (July to next June).

Results: The study enrolled a total of 150,920 patients. A mean of 1.9 patients/hour was seen by PGY-1 EM residents in comparison to 2.6 patients/hour by PGY-2 and -3 EM residents. Median PTD was 2.8 hours in PGY-1 EM residents versus 2.6 hours in PGY-2 and -3 EM residents. There were no significant differences in acuity across all patients seen by EM residents. When crowded conditions existed, residency efficiency increased, but such changes were minimized when the ED became overcrowded. A linear increase of resident efficiency was observed only in PGY-1 EM residents throughout the entire academic year.

Conclusion: Resident efficiency improved significantly only during their first year of EM training. This efficiency can be affected by ED crowding.

In recent years, studies focusing on various provider efficiencies have been reported including physicians working in community hospitals and academic attending physicians supervising residents, medical students, or advanced practice providers (APPs; e.g., physician assistants, nurse practitioners). Different provider efficiency measurements were used in different studies including new patients per provider per hour, relative value units (RVUs) generated per shift, or mean patient emergency department (ED) length of stay. Study results have variable conclusions on provider efficiency due to the differences of efficiency measurements, ED volumes, interventions, patient acuity, and ED crowding conditions. Such diverse provider efficiency reported in the literature raises the question of whether it is necessary to provide standard training on provider efficiency during EM residency training. Unfortunately, few studies in the current literature...
Although previous studies addressed resident efficiency in general, there were neither efficiency comparisons among residents of different training levels nor dynamic comparisons during the entire academic year, which makes resident efficiency assessment less informative.

ED provider efficiency links closely with patient flow and ED crowding. Patient flow is measured through several time intervals such as the time interval from patient arrival to patient seen by an ED provider (e.g., door-to-provider time), the time interval from patient seen by an ED provider to patient disposition (e.g., provider-to-disposition [PTD] time), or the time interval from patient arrival to patient departure from the ED (e.g., patient total length of stay [LOS]). If an ED runs at its full capacity while no-flow occurs (e.g., unable to move patients out from ED) during a certain time range, this results in crowding. ED crowding can be measured using different crowding estimation tools. The National Emergency Department OverCrowding Score (NEDOCS) is one of the most commonly used scoring systems and is validated widely across the nation. Briefly, NEDOCS are calculated based on the total number of patients in the ED, number of critical care ED patients, and number of admit holds in the ED along with the longest time in department of ED admitted patients and patients in the waiting room. The score has also been adjusted by total number of ED and hospital beds to match for different ED/hospital settings (see 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/aet2.10327/full, for detail descriptions). In general, ED crowding is associated with prolonged patient ED stays. However, such findings neither had a direct link with provider efficiency nor further addressed the association between provider efficiency and severity of ED crowding. More specifically, no literature has reported such an association among residents with different levels of training experience.

Emergency medicine residency programs are generally sponsored in tertiary referral centers with EDs that see a relatively high volume of patients. ED crowding has become more common especially at extremely high-volume EDs. It is important to determine EM resident efficiency during different crowding statuses of an ED, thus providing accurate information to develop future curricula on resident-specific training on ED crowding. We hypothesized that resident efficiency might be affected by ED crowding. Therefore, we aim to 1) compare efficiency and its dynamic changes within an entire academic year among EM residents of different training levels and 2) determine resident efficiency under ED of different crowding statuses.

**METHODS**

**Study Design and Setting**

This is a retrospective single-center observational study. The study hospital is an urban tertiary referral center with an ED that sponsored an ACGME-accredited 3-year emergency medicine (EM) residency program with 12 residents per class. It has a total of 573 licensed beds in the hospital and 54 beds in the ED. The study ED has approximately 120,000 patient visits annually. This study was approved by the John Peter Smith Health Network Institutional Review Board.

**Study Participants**

Study participants were the residents. From July 1, 2014, until June 30, 2017, patients who presented at the study ED and were evaluated by EM residents were included in this study. Residents included EM residents of different training levels (PGY-1, -2, and -3). We excluded 1) patients who were seen and evaluated by APPs (e.g., physician assistants or nurse practitioners), 2) patients seen only by ED attending physicians, 3) patients seen by non-EM residents, and 4) patients who left without being seen or eloped (Figure 1).

**Variables and Outcome Measurements**

Patient general characteristics included age, sex, and ethnicity. ED metrics included patient level of acuity.
using Emergency Severity Index (ESI), patient LOS, and PTD time. ESI is a triage tool to determine patient level of acuity upon patients’ arrival at the ED with ESI-1 being the most emergent and ESI-5 being the least emergent. ED LOS was defined as the time interval in hours beginning at the time of patient arrival at the ED and ending at the time when the patient physically left the ED as documented in the medical record. PTD time was defined as the time interval in hours beginning at the time when the patient was initially seen and evaluated by a provider and ending at the time when the disposition decision was made. The number of new patients per hour was defined as the number of new patients assigned to individual residents within a 1-hour block (e.g., 0800 to 0859). PGY-1 residents have 10-hour clinical shifts while PGY-2 and -3 residents have 9-hour clinical shifts. All residents are required to evaluate any new patients during the first 8 hours of their shifts. Senior EM residents (PGY-2 and -3) have 1 hour of overlap and junior EM residents (PGY-1) have 2 hours of overlap at the end of their clinical shift. Attending physicians have either 8- or 12-hour shifts without any overlap. The study ED has a dedicated “fast-track” area staffed with APPs who saw low-acuity patients (ESI-4 and -5) and higher-acuity patients (ESI-1 to -3) were seen by EM attending physicians and residents. ED crowding status was measured by NEDOCS (see Data Supplement S1). Provider efficiency was based on two outcome measurements: 1) the number of new patients per hour seen by a given resident and 2) PTD time of each patient.

**ED Crowding Measurements**

National Emergency Department OverCrowding Score is one of the most common ED crowding estimation tools and has been externally validated in other studies.\(^\text{12-14}\) NEDOCS is categorized into different levels of crowding, from not crowded to dangerously overcrowded statuses. The study ED has developed a computer program linked to the electronic medical record and is able to calculate NEDOCS score in real-time every hour. To simplify its crowding statuses, NEDOCS categories were modified to three-tier classification schemes in this study, where scores ≤ 100 are classified as “not crowded,” scores > 100 and ≤ 140 are classified as “crowded,” and scores > 140 are classified as “overcrowded.”\(^\text{16}\) Each patient arriving at the ED was assigned a crowding score. ED crowding was measured at the top of each hour throughout the study period. Crowding scores estimated at the beginning of each hour were assigned to all subjects who arrived throughout the given hourly interval. Patients who arrived at the end of the hour were assigned the same crowding score as measured at the start of the hour (e.g., patients who arrived from 0801 to 0859 were assigned to the crowding score measured at 0800).

**Study Protocol**

In this study, we divided patients into three groups based on ED crowding status upon patients’ arrival at the study ED. Provider efficiency outcomes were analyzed and compared under different crowding statuses of the ED (e.g., not crowded, crowded, overcrowded). Additionally, we divided residents into four groups (EM PGY-1, EM PGY-2, EM PGY-3, EM residents with different genders). Provider efficiency outcomes were measured and compared among these four groups.

**Data Analysis**

We used a one-way analysis of variance for continuous data comparisons among different groups and Pearson chi-square test for categorical data comparisons. We calculated both mean and median for ED LOS, PTD, and number of new patients per hour per resident. We used a Wilcoxon rank-sum test for median comparisons between groups. All analyses were performed using STATA 14.2 software with a p-value < 0.05 considered a statistically significant difference.

**RESULTS**

This study included a total of 150,920 patients seen and evaluated initially by 61 EM residents of different training levels. The median age of patients was 46, 53% were male, and 26% were Hispanic. Of these, 30,228 patients were evaluated by PGY-1 residents, 59,520 patients by PGY-2 residents, and 61,172 patients by PGY-3 residents. When NEDOCS was used for ED crowding assessment hourly, 39.7% (59,883/150,920) of patients arrived at the ED under not crowded status, 28.3% (42,716) of patients arrived under crowded status, and 32.0% (48,321) patients arrived under overcrowded status. Table 1 shows the general information of the study patient population.

Using the number of new patients per resident per hour to determine resident efficiencies, we found that senior EM residents (PGY-2 and -3) tended to see more new patients per hour than PGY-1 EM residents. Additionally, senior EM residents (PGY-2 and -3) saw more
new patients per hour when the ED was at higher crowding statuses, indicating resident efficiency increased with advanced EM training (Table 2). When PTD was used for resident efficiency measurement, we found a similar trend among residents of different training levels. Senior EM residents (PGY-2 and -3) had a shorter PTD time than PGY-1 EM residents. However, when PTD was compared under different crowding statuses of the study ED, no significant differences were found among residents regardless of their EM training experience (Table 3). Furthermore, when evaluating patient total ED LOS, we found that increased levels of ED crowding yielded prolonged patient LOS regardless of resident training level (Table 4).

We further evaluate EM residents of different genders. Our results showed male residents saw relatively more new patients per hour (Table 2), with shorter PTD and total patient LOS (Tables 3 and 4) than female residents. Regardless of resident genders, we found that there were more new patients per hour and prolonged PTD and LOS found when the ED was at higher crowding statuses.

Most patients seen by residents were categorized as ESI-1 to ESI-3, with the most being ESI-3 patients (Table 5). No significant differences occurred in terms of median PTD as a function of different levels of acuity; therefore, patient ESI was not considered for resident proficiency measurements (Table 6).

We evaluated the dynamic changes of resident efficiency during the academic year (from July to June of next year) under the different ED crowding statuses. Figure 2 shows the dynamic changes of number of new patients per resident per hour during the academic year. The number of new patients per hour seen by PGY-1 residents increased significantly with an increase in experience regardless of ED crowding statuses (Figure 2). However, such changes did not duplicate among

### Table 1
Study Population Demographics

<table>
<thead>
<tr>
<th>Total Number of Analyzed Patients (N = 150,920)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>46 (32–57)</td>
</tr>
<tr>
<td>Gender, male</td>
</tr>
<tr>
<td>80,291 (53)</td>
</tr>
<tr>
<td>Ethnicity, Hispanic</td>
</tr>
<tr>
<td>38,548 (26)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Number of Residents (N = 61)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Male 37 (61)</td>
</tr>
<tr>
<td>Female 24 (39)</td>
</tr>
</tbody>
</table>

Data are reported as median (IQR) or n (%).

IQR = Interquartile Range.

<table>
<thead>
<tr>
<th>Number of New Patients Per Hour Seen by Residents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Crowded</td>
</tr>
<tr>
<td>Crowded</td>
</tr>
<tr>
<td>Overcrowded</td>
</tr>
<tr>
<td>Overall</td>
</tr>
<tr>
<td>Maximal Changes</td>
</tr>
</tbody>
</table>

| PGY-1 EM residents                             | 2 (1–2) | 2 (1–2) | 2 (1–2) | 2 (1–2) |
| Mean (±SD) 1.75 (±0.86)                       | 1.90 (±0.91) | 1.95 (±0.96) | 1.86 (±0.91) |
| Total number of patients 12,288                | 8,360 | 9,580 | 30,228 |

| PGY-2 EM residents                             | 2 (1–3) | 2 (2–3) | 2 (2–3) | 2 (2–3) |
| Mean (±SD) 2.40 (±1.26)                       | 2.65 (±1.34) | 2.71 (±1.42) | 2.57 (±1.34) |
| Total number of patients 23,330                | 16,950 | 19,240 | 59,520 |

| PGY-3 EM residents                             | 2 (2–3) | 3 (2–3) | 3 (2–4) | 2 (2–3) |
| Mean (±SD) 2.47 (1.31)                        | 2.73 (1.38) | 2.76 (1.36) | 2.64 (1.35) |
| Total number of patients 24,265                | 17,406 | 19,501 | 61,172 |

| All EM male residents                          | 2 (1–3) | 2 (2–3) | 2 (2–3) | 2 (2–3) |
| Mean (±SD) 2.31 (±1.25)                       | 2.56 (±1.33) | 2.61 (±1.36) | 2.48 (±1.32) |
| Total number of patients 37,795                | 27,407 | 31,047 | 96,249 |

| All EM female residents                        | 2 (1–3) | 2 (2–3) | 2 (2–3) | 2 (1–3) |
| Mean (±SD) 2.27 (±1.22)                       | 2.49 (±1.30) | 2.53 (±1.33) | 2.41 (±1.29) |
| Total number of patients 22,088                | 15,309 | 17,274 | 54,671 |

IQR = interquartile range.
Table 3
A Comparison of PTD Time of Patients Seen by Residents of Different Training Levels Under ED of Different Crowding Statuses

<table>
<thead>
<tr>
<th></th>
<th>Not Crowded</th>
<th>Crowded</th>
<th>Overcrowded</th>
<th>Overall</th>
<th>Maximal Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY-1 EM residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2.7 (1.7–3.9)</td>
<td>2.8 (1.8–4.1)</td>
<td>2.9 (1.9–4.1)</td>
<td>2.8 (1.8–4.0)</td>
<td>0.2</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>3.0 (1.8)</td>
<td>3.1 (1.8)</td>
<td>3.2 (1.9)</td>
<td>3.1 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>12,040</td>
<td>8,175</td>
<td>9,373</td>
<td>29,588</td>
<td></td>
</tr>
<tr>
<td>PGY-2 EM residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2.5 (1.6–3.7)</td>
<td>2.6 (1.6–3.8)</td>
<td>2.6 (1.6–3.9)</td>
<td>2.6 (1.6–3.8)</td>
<td>0.1</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>2.8 (2.0)</td>
<td>2.9 (1.8)</td>
<td>2.9 (2.6)</td>
<td>2.9 (2.2)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>22,841</td>
<td>16,617</td>
<td>18,876</td>
<td>58,334</td>
<td></td>
</tr>
<tr>
<td>PGY-3 EM residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2.5 (1.5–3.7)</td>
<td>2.6 (1.6–3.8)</td>
<td>2.6 (1.6–3.8)</td>
<td>2.5 (1.5–3.8)</td>
<td>0.1</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>2.8 (1.8)</td>
<td>2.9 (1.8)</td>
<td>2.9 (1.9)</td>
<td>2.8 (1.6)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>23,875</td>
<td>17,116</td>
<td>19,212</td>
<td>60,203</td>
<td></td>
</tr>
<tr>
<td>All EM male residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2.5 (1.5–3.7)</td>
<td>2.6 (1.6–3.8)</td>
<td>2.6 (1.6–3.8)</td>
<td>2.5 (1.5–3.8)</td>
<td>0.1</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>2.8 (1.8)</td>
<td>2.8 (1.8)</td>
<td>2.9 (1.9)</td>
<td>2.8 (1.6)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>37,136</td>
<td>26,937</td>
<td>30,554</td>
<td>94,627</td>
<td></td>
</tr>
<tr>
<td>All EM female residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2.7 (1.7–3.9)</td>
<td>2.8 (1.8–4.1)</td>
<td>2.9 (1.8–4.1)</td>
<td>2.8 (1.8–4.0)</td>
<td>0.1</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>3.0 (2.0)</td>
<td>3.1 (1.8)</td>
<td>3.1 (2.7)</td>
<td>3.0 (2.2)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>21,620</td>
<td>14,971</td>
<td>16,907</td>
<td>53,498</td>
<td></td>
</tr>
</tbody>
</table>

IQR = interquartile range; PTD time = provider-to-disposition time.

Table 4
A Comparison of Total LOS of Patients Seen by Residents of Different Training Levels Under Emergency Department of Different Crowding Statuses

<table>
<thead>
<tr>
<th></th>
<th>Not Crowded</th>
<th>Crowded</th>
<th>Overcrowded</th>
<th>Overall</th>
<th>Maximal Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY-1 EM residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4.2 (2.8–5.9)</td>
<td>4.6 (3.2–6.5)</td>
<td>5.0 (3.5–7.4)</td>
<td>4.6 (3.1–6.5)</td>
<td>1.4</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>4.7 (±4.0)</td>
<td>5.3 (±4.0)</td>
<td>6.1 (±4.3)</td>
<td>5.3 (±4.1)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>12,284</td>
<td>8,343</td>
<td>9,559</td>
<td>30,186</td>
<td></td>
</tr>
<tr>
<td>PGY-2 EM residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4.1 (2.8–5.8)</td>
<td>4.5 (3.1–6.3)</td>
<td>4.9 (3.3–7.0)</td>
<td>4.4 (3.0–6.3)</td>
<td>1.1</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>4.7 (±3.3)</td>
<td>5.1 (±3.4)</td>
<td>5.8 (±4.5)</td>
<td>5.2 (±3.7)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>23,318</td>
<td>16,910</td>
<td>19,197</td>
<td>59,425</td>
<td></td>
</tr>
<tr>
<td>PGY-3 EM residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4.0 (2.7–5.7)</td>
<td>4.5 (3.1–6.3)</td>
<td>4.7 (3.2–6.8)</td>
<td>4.4 (3.0–6.2)</td>
<td>1.1</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>4.6 (±3.0)</td>
<td>5.2 (±3.6)</td>
<td>5.7 (±4.3)</td>
<td>5.1 (±3.7)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>24,257</td>
<td>17,394</td>
<td>19,493</td>
<td>61,144</td>
<td></td>
</tr>
<tr>
<td>All EM male residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4.1 (2.7–5.7)</td>
<td>4.4 (3.1–6.3)</td>
<td>4.8 (3.2–6.9)</td>
<td>4.4 (3.0–6.3)</td>
<td>1.1</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>4.6 (±3.4)</td>
<td>5.1 (±3.7)</td>
<td>5.7 (±4.1)</td>
<td>5.1 (±3.7)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>37,779</td>
<td>27,357</td>
<td>30,997</td>
<td>96,133</td>
<td></td>
</tr>
<tr>
<td>All EM female residents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4.2 (2.8–5.9)</td>
<td>4.6 (3.2–6.5)</td>
<td>5.0 (3.4–7.2)</td>
<td>4.5 (3.1–6.5)</td>
<td>1.3</td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>4.7 (±3.3)</td>
<td>5.3 (±3.4)</td>
<td>6.0 (±4.7)</td>
<td>5.3 (±3.9)</td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>22,080</td>
<td>15,290</td>
<td>17,252</td>
<td>54,622</td>
<td></td>
</tr>
</tbody>
</table>

IQR = interquartile range; LOS = length of stay.
Table 5
Different ESI Comparisons Among Patients Seen by Residents

<table>
<thead>
<tr>
<th>ESI</th>
<th>Patients Seen by Residents*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESI-1</td>
<td>7,701 (5)</td>
</tr>
<tr>
<td>ESI-2</td>
<td>57,544 (38)</td>
</tr>
<tr>
<td>ESI-3</td>
<td>76,220 (51)</td>
</tr>
<tr>
<td>ESI-4</td>
<td>8,274 (6)</td>
</tr>
<tr>
<td>ESI-5</td>
<td>855 (0.6)</td>
</tr>
</tbody>
</table>

Data are reported as n (%).
ESI = Emergency Severity Index; PTD time = provider-to-disposition time.
*Other 381 patients whose ESIs were unknown.

Senior EM residents (PGY-2 and -3). Overall, the number of new patients seen hourly increased for residents of all training levels when the ED became crowded. Such changes were minimized when ED became overcrowded. When PTD was measured, it showed that PTD times decreased during the academic year regardless of resident training experience. PTD times significantly shortened for patients who were evaluated by PGY-1 EM residents throughout the entire academic year regardless of ED crowding statuses. However, such significant changes also did not duplicate among senior EM residents (Figure 3).

**DISCUSSION**

Provider efficiency has been reported in the literature but mainly focused on the attending physicians. This study focused on EM resident efficiency especially addressing efficiency differences among EM residents under different crowding statuses of an ED. We found that ED crowding did affect resident efficiency, particularly when efficiency was measured by the number of new patients per resident per hour. Such efficiency improved when the ED transitioned from not crowded to crowded conditions. However, no significant changes occurred when ED crowding status changed from crowded to overcrowded conditions. Additionally, although patient total LOS was affected by ED crowding, PTD was not affected significantly. Our study also found that male residents had relatively higher efficiency than female residents. However, there was a similar trend for efficiency at ED of different crowding conditions between the genders. Overall, resident efficiency seemed to be affected during overcrowded ED conditions, but it was little affected when the ED was under crowded status.

More interestingly, both the number of new patients per resident per hour and the PTD time were affected mostly by PGY-1 EM residents whose efficiency improved significantly with more experience gained during their first year of EM training regardless of ED crowding statuses. Since our program does not increase the patient load requirement for EM resident of different training levels, all EM residents are actively seeing new patients for the first 8 hours during their shift. We consider such efficiency changes among PGY-1 EM residents might be related to their steep learning curve of managing patient flow. Unfortunately, in the past, we have not heavily emphasized teaching residents how to properly manage ED flow until reaching their senior level. The results of this study suggest that earlier introduction of an ED administrative curriculum during residency training may be necessary, especially when residents train in an extremely high-volume ED setting. Such administrative curriculum should include teaching EM residents different ED operational strategies to manage patient flow (e.g., provider in triage, vertical flow models, or direct transfer patient to clinic among age patient flow (e.g., provider in triage, vertical flow models, or direct transfer patient to clinic among patients with nonemergency conditions) during their early EM residency training. Therefore, our study findings add some extra value to the current literature pool on how to enhance the administrative management of resident efficiency and provides future guidance on adjusting administrative didactics during EM residency training.

There are many ways to determine provider efficiency in the literature. We chose number of new patients per resident per hour as one of the resident efficiency measurements for the following reasons: 1)
it is a simple and easy way to measure; 2) residents tended to see relatively high-acuity patients in the study ED, and therefore the patient acuity seemed to be less biased across different residents; 3) different shift lengths among residents, where PGY-1 residents had 10-hour shifts with the last 2 hours dedicated to “finishing up” and not picking up new patients and senior residents had 9-hour shifts with the last hour dedicated to “finishing up” and not picking up new patients; and 4) crowding was measured hourly at the study ED, which matched well as the number of new patients per provider per hour, since such crowding statuses change dynamically and patient flow may be affected by different levels of crowding at the ED.16

Under these circumstances, we are also able to analyze differences in resident efficiency under different ED crowding statuses. We measured PTD as another resident proficiency marker since patient LOS can be affected by different ED crowding statuses whereas PTD seemed to not be affected significantly.20,21 Additionally, PTD measured the time interval when the provider was actively engaged in evaluating patients, indicating better control in determining provider efficiency. We believe that using the combinations of new patients per hour and PTD may increase the accuracy of resident efficiency measurements.

More importantly, this study shows the dynamic changes of resident proficiency especially among PGY-1 EM residents, indicating the necessity of earlier training on patient flow/ED administrative management during EM residency training. Such training, including an administrative ED rotation, is usually placed during the PGY-3 academic year.18 However, this might be too late for the current EM residents to train on how to efficiently manage ED patient flow. Additionally, our results showed no significant changes in resident proficiency for both PGY-2 and PGY-3 EM residents. This could be due to: 1) each team at the study ED included one senior EM resident and one or two junior EM/non-EM residents, where the senior EM residents see all high-acuity time-sensitive patients (e.g., ST-elevation myocardial infarction, sepsis, stroke patients requiring tissue plasminogen activator) regardless of whether they were PGY-2 or -3 EM residents; or 2) resident proficiency might reach a plateau by the end of their first year of...
training. However, we are unable to determine such a direct link based on our current findings. Therefore, a large-scale multicenter study specifically designed to externally validate the differences of PGY-2 and -3 resident proficiency is warranted.

LIMITATIONS

Our study has its limitations. First, due to the design of this study, study population selection bias, information loss, and lack of complete variable inclusions occurred that could potentially skew the study findings. Therefore, it is important to have external validation to determine whether our findings can be duplicated in other facilities. Second, resident efficiency can be heavily affected by different levels of patient acuity. EM residents mainly saw high-acuity patients (ESI-1 to -3). Although ESI correlated well with patient severity,22,23 our results could still be deviated by considering patient severity more accurately and comprehensively. Third, resident efficiency could be affected by the attending physicians’ proficiency since attending physicians supervise residents and make final decisions. However, attending pairing with residents is random during different shifts and such risk would have biased results toward the null. Finally, there are other potential provider efficiency measurements not used or compared in this study including patients per shifts (as mentioned above) and RVU. Since RVUs can be affected by documentation, billing, and coding, and residents and attending physicians have different ways to complete/edit their medical record, with some completing records during their shifts while others completing them after the shift, we are unable to accurately know the time spent on each patient chart, making RVU as a measure resident efficiency possibly biased in our study. However, our study was neither able to measure its accuracy nor able to determine the association between RVU and resident efficiency.

CONCLUSIONS

Resident efficiency improved significantly only during their first year of emergency medicine training. In addition, resident efficiency can be affected by ED crowding. Future studies could be focused on resident efficiency changes with the implementations of administrative curriculum and operational strategy to enhance resident performance and optimize patient flow under ED of different crowding conditions.

References


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10327/full

Data Supplement S1. Explanations of NEDOCS Variables for ED Crowding Calculations.
Impact of a Mentorship Program on Medical Student Burnout

Jaime Jordan, MD1,2,5, Daena Watcha, MD, MS3,5, Courtney Cassella, MD4,6, Amy H. Kaji, MD, PhD2,5, and Shefali Trivedi, MD6

ABSTRACT

Background: Burnout can have negative consequences for providers’ health and patient care. Mentorship has positive effects including stress mitigation. We sought to evaluate the impact of a mentorship program on burnout in fourth-year medical students during their 4-week emergency medicine subinternship.

Methods: This was a prospective, quasi-experimental, mixed-methods study at two institutions. We assessed burnout using the Maslach Burnout Inventory, comprising three subscales: Emotional Exhaustion (EE), Depersonalization (DP), and Personal Accomplishment (PA). We compared changes in burnout scores before and after implementation of a resident–student mentorship program. We compared categorical variables using risk ratios and continuous variables using Wilcoxon rank-sum test. To account for potential confounders, we performed multivariable analysis. Students and mentors completed an evaluative survey. We reported descriptive statistics and performed thematic qualitative analysis on free-response data.

Results: A total of 135 students (intervention = 51; control = 84) and 59 mentors participated. Intervention students demonstrated decreased EE and DP and increased PA scores, medians of –2 (–4 to 4), –1 (–3 to 2), and 1 (–1 to 4), respectively, compared to controls, median difference of 0 for all subscales. After adjusting for potential confounders, there was no significant difference in EE (mean difference = –0.2 [–0.5 to 0.2], p = 0.4) or DP scores (mean difference = –0.2 [–1.8 to 1.5], p = 0.9). There was a significant difference in PA scores (mean difference = 2.2 [0.1 to 4.3], p = 0.04). Most students felt the program positively impacted their rotation (39/48) and decreased stress (28/48). Students felt that the program provided career guidance and positively impacted their personal and professional development. The majority (34/37) of mentors enjoyed participating. Qualitative analysis revealed five major themes: relationship building, different perspective, knowledge sharing, personal fulfillment, and self-reflection.

Conclusion: We found an increased sense of personal accomplishment after implementation of a mentorship program. Both mentors and mentees viewed the program positively and perceived multiple benefits.

Burnout, broadly defined as a “state of mental and physical exhaustion related to work or caregiving activities,” is an important problem affecting practicing physicians and medical trainees alike.1–8 Burnout can be detrimental to provider physical and mental health, job satisfaction, physician productivity, and patient care.3,9–13 Because of the great reach of this problem, there has been a call by both individuals and governing bodies to make efforts to mitigate burnout, provide education on well-being to trainees, and treat self-care as an important component of professionalism.14–16

From the 1Department of Emergency Medicine, Ronald Reagan UCLA Medical Center, and the 2David Geffen School of Medicine at UCLA, Los Angeles, CA; the 3Department of Emergency Medicine, Regional Medical Center, San Jose, CA; the 4Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; the 5Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, CA; and 6Mount Sinai Hospital, Icahn School of Medicine at Mount Sinai, New York, NY.

Received February 8, 2019; revision received April 24, 2019; accepted April 25, 2019.

Presented at the Society for Academic Emergency Medicine Annual Meeting, Indianapolis, IN, May 2018.

The authors have no relevant financial information or potential conflicts to disclose.

Supervising Editor: Stephen J. Cico, MD, MEd.

Address for correspondence and reprints: Jaime Jordan, MD; e-mail: jaimejordanmd@gmail.com.

Several groups have undertaken efforts to mitigate burnout and provide the necessary tools to trainees to both recognize and treat burnout and also promote resiliency (the ability to recover from adversity) with mixed results.\textsuperscript{17–24} These attempts have included: 1) changes in the learning environment, such as alteration of grading systems or duty hours; 2) efforts to build a sense of community, such as shared experience sessions, Balint groups, or training in communication skills; 3) instruction on the practice of positive cognitive processes such as narrative medicine, reflection, or journaling; 4) courses to assist in managing stress, such as relaxation or psychotherapeutic technique; and 5) efforts to promote resiliency such as mindfulness, meditation, exercise, and self-compassion.\textsuperscript{17–24} Despite these efforts, best practices to mitigate the complex issue of burnout in medical trainees have yet to be established.

Prior literature has shown many positive effects of mentorship for both the mentor and the mentee.\textsuperscript{25–29} As mentorship provides both psychosocial and career support, it has the potential to decrease burnout.\textsuperscript{30} Additionally, professional development has been identified as a strategy for resilience in medicine.\textsuperscript{31} Limited data have shown a positive impact of mentorship on burnout and stress in both practicing physicians and medical students.\textsuperscript{32,33} It is currently unknown if a structured resident–student mentorship program can decrease burnout in medical students. The objective of this study was to evaluate the impact of such a mentorship program on burnout in fourth-year medical students during their emergency medicine (EM) subinternship.

\section*{METHODS}

\subsection*{Study Setting and Participants}
This study took place at two academic institutions, Harbor-UCLA Medical Center and Mount Sinai Hospital. Study participants were fourth-year (senior) medical students enrolled in the 4-week EM subinternship at the participating sites. A subinternship is a rotation in a specialized area of medicine offered during the fourth year of medical school. A mentor was offered to all subinterns, including those from the home institution as well as external rotating students. Subinterns were allowed to opt out of the mentorship program. Subinterns were informed that participation in the program would not impact their evaluations. Senior EM residents (PGY-3 or PGY-4) were notified of the program and invited to serve as mentors. Resident mentors were provided with a standardized 1-hour training session, which included expectations of the program and tips for successful mentorship. Subinterns and resident mentors were not compensated for their participation. Data were collected between February 2017 and November 2017. This study was approved by the institutional review boards of the David Geffen School of Medicine at UCLA and the Icahn School of Medicine at Mount Sinai.

\subsection*{Study Design}
This was a prospective, quasi-experimental, mixed-methods study. Burnout was assessed in fourth-year medical students at the beginning and end of their EM subinternship using the Maslach Burnout Inventory (MBI) Human Services Survey for Medical Professionals, which consists of three subscales: Emotional Exhaustion (EE), Depersonalization (DP), and Personal Accomplishment (PA).\textsuperscript{34} The MBI has been used extensively to evaluate burnout in practicing physicians, nurses, residents, and medical students.\textsuperscript{19,35} Participants completed a basic demographic questionnaire prior to participation. Changes in burnout scores were assessed prior to (control group) and after the implementation of a big sib/little sib mentorship program (intervention group). The big sib/little sib mentorship program paired senior resident mentors with student mentees in a 1:1 fashion. Because students who opted out did not participate in the mentorship program, they were analyzed as part of the control group. Mentors were purposefully paired with student mentees by the study team based on similar characteristics, such as common interests, medical school, or home town. Minimum requirements of the mentor were to meet in person with their mentee at the beginning and end of the rotation, check in with them by phone or e-mail in the middle of the rotation, and be available by e-mail during the rotation. Mentors were encouraged, however, to meet more frequently as indicated to meet the needs of their mentees. Both mentors and mentees completed an evaluative survey regarding their experience with the program at the end of the rotation which consisted of free-response and Likert-scale items. Study team members with expertise in mentorship, medical education leadership, and questionnaire design developed all surveys, according to established guidelines for survey research, and each of the surveys was piloted on a small group of representative
subjects. The instruments were revised for clarity and readability. The final version of all surveys are available in 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/aet2.10354/full).

### Data Analysis

We performed a sample size calculation and determined that we would achieve >90% power to detect a 5-point difference in change in pre-post burnout scores if we had 60 subjects in the control cohort and seven in the intervention cohort. We calculated and reported descriptive statistics. Risk ratios with 95% confidence intervals (CIs) were used to compare categorical predictor variables. Continuous variables were compared using the Wilcoxon rank-sum test. Multivariable analysis using generalized estimating equations was performed to account for potential confounders identified a priori, including institution, correlations of outcomes within institution (students and mentors at one institution are likely to be similar to one another), number of EM subinternships completed, intended specialty of EM, away rotation, and previously established mentor. We analyzed free-response survey data using a thematic approach. Two analysts, JJ and ST, independently reviewed the data, line by line, to identify recurring concepts and assign codes, which were then further refined into themes using the constant comparative method. After independent review, the two researchers met to establish a final coding scheme that was applied to all data. Overall inter-rater agreement was 86%. We resolved discrepancies by in-depth discussion and negotiated consensus.

### RESULTS

A total of 135 subinterns participated in the study, 84 in the control group and 51 in the intervention group. Fifty-nine mentors participated in the program. Demographic data of student participants are displayed in Table 1. Forty-eight (94.1%) students in the intervention completed the evaluative survey of the program. Thirty-seven (62.7%) mentors completed the evaluative survey.

Students in the intervention group demonstrated decreased scores on both the EE and DP subscales and increased PA scores with medians of $-2$ (–4 to 0), $-1$ (–3 to 2), and 1 (–1 to 4), respectively, compared to control students who did not show any change with a median difference of 0 for all three subscales. This difference between intervention and control groups was not statistically significant (EE $p = 0.2$, DP $p = 0.5$, PA $p = 0.06$). After potential confounders were adjusted for, again there was no significant difference in EE (mean difference $= -0.2$ [–0.5 to 0.2], $p = 0.4$) or DP scores (mean difference $= -0.2$ [–1.8 to 1.5], $p = 0.9$) between groups. There was a significant difference in PA scores (mean difference $= 2.2$ [0.1 to 4.3], $p = 0.04$). Significant independent predictors of change in PA score included number of EM subinternships ($p = 0.01$) and away rotations ($p < 0.0001$).

Results of the evaluative survey showed that student participants generally viewed the program positively (Table 2). The majority of participants felt that the program positively impacted their experience in the subinternship and helped decrease stress with 39 of 48 (81.3%) and 28 of 48 (58.3%) rating these

### Table 1

Demographic Data of Participants

<table>
<thead>
<tr>
<th></th>
<th>Intervention ($n = 51$)</th>
<th>No Intervention ($n = 84$)</th>
<th>Risk Ratios, 95% CI, P-value Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>21 (41.2)</td>
<td>55 (65.5)</td>
<td>0.6, 0.4–0.8, $p = 0.006$</td>
</tr>
<tr>
<td>Male sex</td>
<td>32 (62.8)</td>
<td>49 (58.3)</td>
<td>1.1, 0.7–1.7, $p = 0.6$</td>
</tr>
<tr>
<td>Intended specialty of EM</td>
<td>47 (92.2)</td>
<td>52 (61.9)</td>
<td>4.9, 1.8–12.9, $p &lt; 0.0001$</td>
</tr>
<tr>
<td>Away rotation</td>
<td>40 (78.4)</td>
<td>32 (38.1)</td>
<td>2.9, 1.7–5.0, $p &lt; 0.0001$</td>
</tr>
<tr>
<td>Has a mentor</td>
<td>29 (56.9)</td>
<td>58 (69.1)</td>
<td>0.7, 0.5–1.1, $p = 0.2$</td>
</tr>
<tr>
<td>Number subinternship</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3.2–3</td>
<td>2.1–2</td>
<td>$p &lt; 0.0001$</td>
</tr>
<tr>
<td>Mean</td>
<td>2.7, 2.4–3.0</td>
<td>2.0, 1.8–2.2</td>
<td></td>
</tr>
<tr>
<td>Month of rotation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>9, 9–10</td>
<td>7.6–8</td>
<td>$p &lt; 0.0001$</td>
</tr>
<tr>
<td>Mean</td>
<td>9.3, 9.1–9.6</td>
<td>7.1, 6.7–7.5</td>
<td></td>
</tr>
</tbody>
</table>

*Data are reported as n (%).
statements as a “4” or “5” on a 5-point scale (5 = strongly agree), respectively. The majority of students felt that their big sib demonstrated qualities of good mentors with 35 of 48 (72.9%), 35 of 46 (76.1%), 31 of 48 (64.6%), and 38 of 46 (82.6%) of responding students “strongly agreeing” that their mentors were easily accessible, active listeners, emotionally supportive, and able to answer their questions knowledgeably, respectively. Students also felt that the program provided useful career guidance and positively contributed to their personal and professional development (Table 2). Results of qualitative analysis of free-response data from the mentee survey are displayed in Table 3. Regarding positive impact of the program, two major themes emerged: emotional support and guidance/advice. Major themes for improvement included no change, timing, and a social event. Twenty-nine 46 (63%) students reported meeting with their mentor at both the beginning and end of the rotation.

Table 2
Results of Student Mentee Evaluative Survey

<table>
<thead>
<tr>
<th>Statement</th>
<th>1 - Strongly Disagree</th>
<th>2</th>
<th>3 - Neutral</th>
<th>4</th>
<th>5 - Strongly Agree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>The big sib program positively impacted my experience in the EM subinternship.</td>
<td>1 (2.1)</td>
<td>0 (0)</td>
<td>8 (16.7)</td>
<td>16 (33.3)</td>
<td>23 (47.9)</td>
<td>48</td>
</tr>
<tr>
<td>The BIG SIB program helped decrease my stress during the EM subinternship.</td>
<td>1 (2.1)</td>
<td>4 (8.3)</td>
<td>15 (31.3)</td>
<td>16 (33.3)</td>
<td>12 (25)</td>
<td>48</td>
</tr>
<tr>
<td>The big sib program provided me with useful career guidance.</td>
<td>1 (2.1)</td>
<td>4 (8.3)</td>
<td>8 (16.7)</td>
<td>13 (27.1)</td>
<td>22 (45.8)</td>
<td>48</td>
</tr>
<tr>
<td>The big sib program positively contributed to my personal development.</td>
<td>1 (2.1)</td>
<td>1 (2.1)</td>
<td>18 (37.5)</td>
<td>15 (31.3)</td>
<td>13 (27.1)</td>
<td>48</td>
</tr>
<tr>
<td>The big sib program positively contributed to my professional development.</td>
<td>1 (2.1)</td>
<td>1 (2.1)</td>
<td>12 (25)</td>
<td>21 (43.8)</td>
<td>13 (27.1)</td>
<td>48</td>
</tr>
<tr>
<td>My big sib was easily accessible when I needed them.</td>
<td>1 (2.1)</td>
<td>0 (0)</td>
<td>4 (8.3)</td>
<td>8 (16.7)</td>
<td>35 (72.9)</td>
<td>48</td>
</tr>
<tr>
<td>My big sib was an active listener in our discussions.</td>
<td>1 (2.2)</td>
<td>0 (0)</td>
<td>4 (8.7)</td>
<td>6 (13)</td>
<td>35 (76.1)</td>
<td>46</td>
</tr>
<tr>
<td>My big sib was emotionally supportive.</td>
<td>1 (2.1)</td>
<td>0 (0)</td>
<td>8 (16.7)</td>
<td>8 (16.7)</td>
<td>31 (64.6)</td>
<td>48</td>
</tr>
<tr>
<td>My big sib was able to knowledgably answer my questions.</td>
<td>1 (2.2)</td>
<td>0 (0)</td>
<td>3 (6.5)</td>
<td>4 (8.7)</td>
<td>38 (82.6)</td>
<td>46</td>
</tr>
<tr>
<td>I feel confident that what I discussed with my big sib was kept confidential.</td>
<td>1 (2.2)</td>
<td>0 (0)</td>
<td>7 (15.2)</td>
<td>1 (2.2)</td>
<td>37 (80.4)</td>
<td>46</td>
</tr>
</tbody>
</table>

Data are reported as n (%).

Table 3
Results of Qualitative Analysis of Mentee Survey

<table>
<thead>
<tr>
<th>Question</th>
<th>Major Themes</th>
<th>Exemplar Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>How did the big sib program positively impact your subinternship?</td>
<td>Emotional support Guidance/advice</td>
<td>“I felt like I had an advocate, advisor, and ambassador from day one who was genuinely interested in my well-being and success.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Provided a support system at a new program.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Helped me set goals for my sub-I. Gave me someone to talk to in case I had questions about my rotation.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Gave me career and application info.”</td>
</tr>
<tr>
<td>What is one thing you would like to see changed for the program next year?</td>
<td>No change Timing Communal event</td>
<td>“No changes really. Please keep program around!”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Potentially try to start prior to arrival.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Formal gathering of all big/little sibs”</td>
</tr>
<tr>
<td>Why were you unable to meet with your mentor?</td>
<td>Lack of perceived benefit Time limitations Communication issues</td>
<td>“I’m not going into EM, so a little less applicable for me.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Mentor time limitations, my time limitations.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Mentor has not reached out yet for end of rotation meeting.”</td>
</tr>
</tbody>
</table>
Results of the mentor survey are displayed in Table 4. The majority (34/37; 91.9%) of mentors enjoyed participating in the program and 23 of 37 (62.2%) felt that it positively contributed to their professional development. Participation in the program improved confidence in mentoring ability for 27 of 37 (73%) mentors. Half of the responding mentors stated that it rejuvenated their interest in EM. Qualitative analysis of free response questions from the mentor survey are displayed in Table 5. Regarding positive impact of the program for mentors, five major themes emerged: relationship building, ability to see a different perspective, satisfaction with sharing knowledge/guidance, personal fulfillment, and self-reflection. Regarding suggestions for improvement, two major themes emerged: no change and communication.

**DISCUSSION**

After implementation of a big sib/little sib mentorship program, this study found a significant difference in PA scores and a nonsignificant trend toward decreased EE and DP. This general positive impact is consistent with prior limited data. Given the myriad demonstrated benefits of mentorship, many of which relate to burnout and wellness, we were surprised to not have found a greater effect in this study. It is possible that participants found difficulty establishing and maintaining the mentorship relationship, as has been identified as a barrier. In fact, in our study, 37% of mentor–mentee pairs did not meet the minimum required times. Additionally, the duration of a subinternship is a relatively short amount of time and in this study timing was identified as an opportunity for improvement. A longer mentorship program may have yielded different results. Despite performing a sample size calculation, it is also possible that this study was underpowered to detect smaller but potentially meaningful differences.

Both mentor and mentee participants in this study viewed the program highly positively. This may have implications for resident and student satisfaction as well as additional benefits apart from mitigating burnout. There are additional implications for recruitment as rotating subinterns are often also applicants for residency, and mentorship has been suggested to be a
potential recruitment advantage.\textsuperscript{28} This should be considered for future research. Students felt that the program helped decrease stress and provided emotional support and guidance, which is consistent with prior literature on the value of mentorship and further supports that mentorship can be highly beneficial to trainees.\textsuperscript{25,26,29} Additionally, most students felt that their resident mentors embodied the characteristics of good mentors, including being active listeners, supportive, accessible, and knowledgeable.\textsuperscript{27,38,39} This is likely a reflection of the mentor selection and training process of the program.

Mentors in this study also perceived many benefits consistent with prior literature, such as positive impact on professional development, relationship building, personal fulfillment, knowledge sharing, and self-reflection.\textsuperscript{28} After self-reflection, residents often had a stronger perception of their success, and this has also been shown to be a positive outcome of being a mentor.\textsuperscript{28} This promotion of self-reflection is an important finding as previous literature has shown that reflection can stimulate learning, enhance readiness to apply new knowledge, improve performance, and promote professional development and is a key component of medical education.\textsuperscript{40–42}

Prior literature has found that residents feel underprepared for a career in academics.\textsuperscript{43} A mentorship program such as this may be a means to address this gap as the positive impact on professional development that mentors cited in this study may be particularly useful for those who plan on pursuing an academic career and will likely have future mentor roles. Another benefit cited by participating mentors was the provision of different perspectives. This is important for physicians in training as it helps give them a more global view of their environment and may enhance communication and problem-solving skills. Many of the benefits of the program cited by mentors may help to decrease burnout and promote resiliency congruent with prior literature.\textsuperscript{28} This is another area of potential research as this study did not assess burnout in mentors.

Time constraints, communication barriers, and lack of perceived benefit appear to be the most significant obstacles identified in this study which have been identified previously as barriers to effective mentoring.\textsuperscript{27} Participants suggested improvement strategies center around communal events, timing, and communication themes. Potential means to improve a resident student mentorship program include protecting time for mentoring activities, initiating the relationship earlier to increase continuity and promote more longitudinal relationships, and making communication easier by aligning mentor–mentee schedules and utilizing alternative meeting methods, such as Skype. Additionally supporting group events outside the clinical and classroom environments may help foster relationship building, which is essential for successful mentorship. Implementing these changes may increase the value and impact of such a program.

LIMITATIONS

Since this study took place at two academic centers in the United States, the results may not be generalizable to other settings. The intervention and control groups were not equivalent. However, as we examined change in MBI scores and used participants as their own controls, we do not expect that this impacted our results greatly. Additionally, participants may be sensitized to the construct of burnout and it is possible that our results may have been influenced by a Hawthorne effect. Long-term outcomes and mentor burnout were not assessed and can be an important area of future research. While we controlled for multiple variables, there may be other confounders not accounted for in our analysis that may have influenced our results. Although the MBI has been used extensively to assess burnout in medical professionals including medical students, there are inherent limitations to this instrument including that it was not normed on training physicians and does not account for nonprofessional aspects that can influence burnout.\textsuperscript{44} We also did not have 100% compliance with program activities in the intervention group. Although this reflects real-world implementation, it may have influenced our results. The response rate from the mentors is moderate and there may be viewpoints that were not captured in the data. This study may have not been powered to detect smaller but meaningful differences between the intervention and control groups. Despite these limitations, this study still demonstrates that implementing a resident–student mentorship program is feasible and has benefits for both students and residents.

CONCLUSIONS

This study found an increased sense of personal accomplishment after implementation of a resident–
student mentorship program. There was a nonsignificant trend toward decreased emotional exhaustion and depersonalization scores. Both mentors and mentees viewed the program positively and perceived multiple benefits. Additional research is needed to further evaluate ways to mitigate the complex issue of medical student burnout.

References

30. Pethrick H, Nowell L, Oddone Paolucci E, et al. Psychosocial and career outcomes of peer mentorship in medical


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10354/full

Data Supplement S1. Survey instruments.
The Standardized Video Interview: How Does It Affect the Likelihood to Invite for a Residency Interview?

Abbas Husain, MD, Ida Li, MD, Brahim Ardolic, MD, Michael C. Bond, MD, Jan Shoenberger, MD, Kaushal H. Shah, MD, Arlene S. Chung, MD, Jeffrey Van Dermark, MD, Jonathan M. Bronner, MD, Melissa White, MD, Todd Taylor, MD, Lukasz Cygan, DO, William Caputo, MD, Matthew Silver, MD, William C. Krauss, MD, Daniel J. Egan, MD, and Moshe Weizberg, MD

ABSTRACT

Background: The Association of American Medical Colleges instituted a standardized video interview (SVI) for all applicants to emergency medicine (EM). It is unclear how the SVI affects a faculty reviewer’s decision on likelihood to invite an applicant (LTI) for an interview.

Objectives: The objective was to determine whether the SVI affects the LTI.

Methods: Nine Accreditation Council of Graduate Medication Education (ACGME)-accredited EM residency programs participated in this prospective, observational study. LTI was defined on a 5-point Likert scale as follows: 1 = definitely not invite, 2 = likely not invite, 3 = might invite, 4 = probably invite, 5 = definitely invite. LTI was recorded at three instances during each review: 1) after typical screening (blinded to the SVI), 2) after unblinding to the SVI score, and 3) after viewing the SVI video.

Results: Seventeen reviewers at nine ACGME-accredited residency programs participated. We reviewed 2,219 applications representing 1,424 unique applicants. After unblinding the SVI score, LTI did not change in 2,065 (93.1%), increased in 85 (3.8%) and decreased in 69 (3.1%; p = 0.22). In subgroup analyses, the effect of the SVI on LTI was unchanged by United States Medical Licensing Examination score. However, when examining subgroups of SVI scores, the percentage of applicants in whom the SVI score changed the LTI was significantly different in those that scored in the lower and upper subgroups (p < 0.0001). The SVI video was viewed in 816 (36.8%) applications. Watching the video did not change the LTI in 631 (77.3%); LTI increased in 106 (13.0%) and decreased in 79 (9.7%) applications (p = 0.04).
Conclusions: The SVI score changed the LTI in 7% of applications. In this group, the score was equally likely to increase or decrease the LTI. Lower SVI scores were more likely to decrease the LTI than higher scores were to increase the LTI. Watching the SVI video was more likely to increase the LTI than to decrease it.

In 2017, the Association of American Medical Colleges (AAMC) instituted the standardized video interview (SVI) as a pilot program for all applicants to emergency medicine (EM) residency programs. The SVI is an “online unidirectional video.”1 Applicants are presented six questions via text prompts and have 30 seconds to review the question prior to responding. They then have up to 3 minutes to record an audio and video response to the question. The questions represent a mix of behavioral and situational questions to assess interpersonal and communications skills as well as knowledge of professionalism.1 Applicants record six different 3-minute videos that are subsequently scored by two AAMC-trained expert reviewers. Each video is assigned a score of 1 representing rudimentary to 5 representing exemplary.2 Thus, each applicant receives a total possible score of 6 to 30. Residency programs may view each applicant’s total score and also the entire video response of all six questions.

The SVI was created to provide faculty application reviewers an objective data point that assesses the applicants’ professionalism and interpersonal and communication skills. The intent was that this could widen the pool of applicants invited for an interview.1 However, the decision on the likelihood to invite (LTI) an applicant for an interview for residency is multifactorial and varies among programs.2,3 Although the intent was to widen the applicant pool, the SVI theoretically has the ability to potentially limit an applicant as well. Therefore, in real-world use, it is unclear in which direction the SVI affects faculty application reviewers’ decisions on the LTI. We designed this multicenter study to investigate the impact of the SVI on a faculty reviewer’s decision on whether or not to offer an applicant an interview.

METHODS

Nine Accreditation Council of Graduate Medical Education (ACGME)-accredited EM residency programs participated in this prospective, observational study from September 15 through November 1, 2017. At each study site, faculty members who review residency applications reviewed each Electronic Residency Application Service (ERAS) application. Application reviewers included program directors (PD), assistant PDs, and clerkship directors who are the decision makers involved in application review in their programs.

The LTI decision is multifactorial and varies among programs. Thus, to appropriately control for other factors affecting an institution’s LTI, such as United States Medical Licensing Examination (USMLE) scores, standard letters of evaluation (SLOEs), medical student performance evaluations (MSPEs), and medical school transcripts, we allowed reviewers to view all aspects of the ERAS application but blinded them to the SVI score. Blinding was accomplished by requiring the reviewers to assign an LTI without opening the SVI tab on the ERAS application. The faculty reviewers then assigned an LTI based on the traditional screening process. Faculty application reviewers were subsequently unblinded to the SVI score (without viewing the video) by opening the SVI tab in ERAS and were asked again to determine the LTI. We compared the post-SVI score LTI to the pre-SVI LTI. The LTI was determined on a 5-point Likert scale: 1 = definitely not invite, 2 = likely not invite, 3 = might invite, 4 = probably invite, 5 = definitely invite.4

In a subgroup of the study, faculty reviewers had the option of not only seeing the SVI score but also viewing the actual SVI video. Viewing the video was left to the discretion of the faculty reviewer. In the video group the faculty application reviewers assigned a pre-SVI LTI and a post-SVI video LTI. We then compared post-SVI video LTI to pre-SVI LTI. Ultimately this led to three groups for comparison: pre-SVI LTI, post-SVI score LTI, and post-SVI video LTI.

A predesigned standardized form (Data Supplement S1, Figure S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10331/full) was designed by the principal investigators (face validity) to capture the decision-making process of a faculty application reviewer. It was piloted by four faculty application reviewers to ensure operational and construct validity. All faculty application reviewers were trained in the study protocol via a conference call and
in one-on-one sessions with an investigator to ensure consistency and accuracy.

There were 17 total faculty application reviewers throughout the study period from the nine institutions. Each application was screened by one faculty reviewer per site.

Inclusion criteria were applications received at any one of the participating EM residency programs that were selected to be screened for an interview. We viewed each application as a single data point. Thus, the same applicant could be reviewed at multiple sites and count as separate data points for the purpose of the study. We believe that this was appropriate because each site individually assesses each applicant independent of other sites. Exclusion criteria included applications which had no SVI score reported and applications in which the reviewer prematurely saw the SVI score prior to assigning an LTI.

Applicant demographic data were exported directly from ERAS applications. Reviewer demographic data were entered by the reviewers directly into a Google form. LTI was entered directly into a RedCAP (v 8.2.2) database.

Predetermined subgroup analyses were performed based on applicants’ USMLE scores as follows: ≤200, 201–220, 221–240, 241–260, and >260; and based on SVI scores as follows: 6–11, 12–17, 18–23, and 24–30. These SVI score ranges are described by the AAMC as representing different proficiency levels on the target competencies.

Percentages of LTI were compared using chi-square. Mean SVI scores and LTIs were compared using t-test and ANOVA. Linear regression analysis was used to compare the change in LTI by SVI and USMLE score subgroups. All statistical tests were two-sided. A p-value of ≤0.05 was considered statistically significant. Statistics were calculated using GraphPad InStat.

A sample size analysis was performed prior to initiating the study. AAMC data state that 20% of EM applications result in an invitation to interview. Assuming an alpha of 0.05, to demonstrate that the SVI causes a 10% change in LTI would require a sample size of 1,537. The study was reviewed by the institutional review board at Northwell Health.

RESULTS

Seventeen faculty reviewers at nine ACGME-accredited residency programs participated in the study (Table 1). A total of 2,310 applications were reviewed. We excluded 76 applications that had no SVI score and 15 in which the faculty reviewer was prematurely unblinded to the SVI score, leaving a study population of 2,219 applications (Figure 1).

There were 1,424 unique applicants and demographic data were available for 1,371 (Table 2). The remainder were unavailable as some institutions blocked ERAS demographics. These applications were not excluded from analysis as the demographics would not have affected the LTI.

The mean pre-SVI LTI was 3.0 ± 1.47 and the mean post-SVI score LTI was 3.0 ± 1.4 (p = 0.8). After the SVI score was revealed to faculty application reviewers, LTI was unchanged in 2,065 (93.1%, 95% confidence interval [CI] = 92.4%–94.5%), increased in 85 (3.8%, 95% CI = 3.08%–4.72%), and decreased in 69 (3.1%, 95% CI = 2.42%–3.925) applications (p = 0.22).

Subgroup analyses were performed by USMLE score and by SVI score. For all USMLE subgroups, SVI score did not lead to a change in LTI (Data Supplement S1, Table S1). However, in subgroup analysis by SVI score, the percentage of applications in which the SVI score changed LTI was significantly different.
among the subgroups (p < 0.0001). The largest affect was seen in the 6–11 score range (64.3% decreased; Table 3; Data Supplement S1, Figure S2).

There were 584 unique applicants that were reviewed at more than one site. The intraclass correlation coefficient (ICC) of the effect of SVI score on change in LTI was 0.129. This represents a poor correlation between sites.

The SVI video was viewed in 816 (36.8%) applications (814 viewed part of the video; two viewed the entire video). In the video group, we recorded a pre-SVI LTI while blinded to both the SVI score and the video and then unblinded the SVI score and then the video to compute a post-SVI score LTI and a post-SVI video LTI, respectively. There was no change in post-SVI video LTI in 631 (77.3%, 95% CI = 74.3%–80.2%), it increased in 106 (13.0%, 95% CI = 10.8%–15.5%) and decreased in 79 (9.7%, 95% CI = 7.7%–11.9%; p = 0.04). The percentage of applications in which the SVI video changed the LTI was significantly different for the various SVI subgroups (p < 0.0001; Table 4; Data Supplement S1, Figure S3).

**DISCUSSION**

Our study found that the SVI score changed the LTI in 7% of applications. However, there was no difference between the percentage of applications in which the LTI increased or decreased. When comparing SVI scores, lower scores (<11) were more likely to decrease the LTI. Higher SVI scores were not as impactful. This suggests that SVI scores less than 11 could potentially decrease the LTI.

Viewing the video was more likely to increase the LTI than to decrease it. Watching the applicant answer questions and seeing how they communicated and expressed professionalism was more likely to increase the LTI than was the numerical SVI score.

The SVI represents an AAMC initiative to provide objective data for residency application reviewers to evaluate when determining whether or not to interview an applicant. Much of an ERAS application is

<table>
<thead>
<tr>
<th>Table 2 Demographic Characteristics of Applicants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicant Demographics</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Medical school region, n (%)</td>
</tr>
<tr>
<td>Northeast</td>
</tr>
<tr>
<td>Central</td>
</tr>
<tr>
<td>South</td>
</tr>
<tr>
<td>West</td>
</tr>
<tr>
<td>International</td>
</tr>
<tr>
<td>Medical school type, n (%)</td>
</tr>
<tr>
<td>U.S. private</td>
</tr>
<tr>
<td>U.S. public</td>
</tr>
<tr>
<td>Osteopathic</td>
</tr>
<tr>
<td>International</td>
</tr>
<tr>
<td>Licensing examination score</td>
</tr>
<tr>
<td>USMLE 1</td>
</tr>
<tr>
<td>USMLE 2</td>
</tr>
<tr>
<td>USMLE 2 CS</td>
</tr>
<tr>
<td>Passed</td>
</tr>
<tr>
<td>Failed</td>
</tr>
<tr>
<td>COMLEX 1</td>
</tr>
<tr>
<td>COMLEX 2</td>
</tr>
<tr>
<td>COMLEX 2 PE</td>
</tr>
<tr>
<td>Passed</td>
</tr>
<tr>
<td>Failed</td>
</tr>
<tr>
<td>SVI score</td>
</tr>
</tbody>
</table>

COMLEX = Comprehensive Osteopathic Medical Licensing Examination; SVI = standardized video interview; USMLE = United States Medical Licensing Examination.
subjective (i.e., letters of recommendation, curriculum vitae) such that there is little objective information to help differentiate a higher-tier applicant from a lower one. Currently, the USMLE score is the most discerning part of the application as it correlates with board certification pass rates and success on in-training examinations.6

There are currently limited published data available on the SVI or its effects on LTI. The SVI has been shown to not correlate with USMLE scores thus providing an additional piece of data not previously available on ERAS applications.7 One single-site study found no correlation between the SVI score and faculty gestalt scores on communication and professionalism.8

The SVI was designed to quantify professionalism and interpersonal communication skills. As per the AAMC, the purpose of the SVI is to help residency programs widen the pool of applicants.1 This study was designed to see how residency programs would use SVI data to decide whether an applicant warrants an interview invitation. As a new construct, the SVI has not yet been proven to correlate with success or failure in residency training. Therefore, it is unclear how much weight PDs and application reviewers place on the SVI when determining whether to invite an applicant for an interview. Traditional factors outlined previously such as national board and EM testing scores, course clerkship scores, MSPEs, and SLOEs continue to be the major influencers on the LTI for PDs.9 It follows that the value that individual PDs place on the SVI should also be assessed.

Because of the growth of interview season and the numbers of programs applied to by each candidate, residency programs already divert numerous hours and resources away from their primary missions of education, research, and clinical care. A lengthy, time-intensive process of evaluating hundreds of students for each potential PGY-1 residency position has developed. One program in this study received over 1,500 applications which could mean over 450 hours of video. This additional time would add to the cost of

### Table 3
Effect of SVI Score on LTI Subdivided by SVI score

<table>
<thead>
<tr>
<th>SVI SCORE</th>
<th>n</th>
<th>Pre-SVI LTI (Mean ± SD)</th>
<th>Post-SVI Score LTI (Mean ± SD)</th>
<th>p-value</th>
<th>% That Increased (95% CI)</th>
<th>% That Decreased (95% CI)</th>
<th>% Unchanged (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6–11</td>
<td>14</td>
<td>2.43 ± 1.16</td>
<td>1.71 ± 0.99</td>
<td>0.09</td>
<td>64.3 (39.2 to 89.4)</td>
<td>35.7 (10.6 to 60.8)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>12–17</td>
<td>500</td>
<td>2.75 ± 1.45</td>
<td>2.65 ± 1.43</td>
<td>0.25</td>
<td>10.2 (7.68 to 13.3)</td>
<td>89.4 (86.4 to 92.0)</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>18–23</td>
<td>1,483</td>
<td>3.05 ± 1.47</td>
<td>3.07 ± 1.47</td>
<td>0.71</td>
<td>2.9 (2.11 to 3.89)</td>
<td>0.61 (0.28 to 1.15)</td>
<td>96.5 (95.4 to 97.4)</td>
<td></td>
</tr>
<tr>
<td>24–30</td>
<td>222</td>
<td>3.28 ± 1.43</td>
<td>3.47 ± 1.37</td>
<td>0.16</td>
<td>18 (13.2 to 23.7)</td>
<td>0 (76.3 to 86.8)</td>
<td>82.0</td>
<td></td>
</tr>
</tbody>
</table>

Percentage that increased represents the percentage of applications in which the LTI increased after viewing the SVI score.

LTI = likelihood to invite; SVI = standardized video interview;

### Table 4
Effect of SVI Video on LTI Subdivided by SVI score

<table>
<thead>
<tr>
<th>SVI SCORE</th>
<th>n</th>
<th>Pre-SVI LTI (Mean ± SD)</th>
<th>Post-SVI Score LTI (Mean ± SD)</th>
<th>p-value</th>
<th>% That Increased (95% CI)</th>
<th>% That Decreased (95% CI)</th>
<th>% Unchanged (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>6–11</td>
<td>11</td>
<td>2.55 ± 1.21</td>
<td>1.73 ± 1.19</td>
<td>0.12</td>
<td>63.6 (35.2 to 89.2)</td>
<td>36.4 (10.6 to 60.8)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>12–17</td>
<td>179</td>
<td>3.03 ± 1.39</td>
<td>2.80 ± 1.39</td>
<td>0.12</td>
<td>24.0 (18.0 to 31.0)</td>
<td>76.0 (66.6 to 80.1)</td>
<td>73.7</td>
<td></td>
</tr>
<tr>
<td>18–23</td>
<td>521</td>
<td>3.10 ± 1.33</td>
<td>3.18 ± 1.34</td>
<td>0.33</td>
<td>12.9 (10.1 to 16.1)</td>
<td>5.4 (3.60 to 7.66)</td>
<td>81.8</td>
<td></td>
</tr>
<tr>
<td>24–30</td>
<td>105</td>
<td>3.26 ± 1.23</td>
<td>3.63 ± 1.12</td>
<td>0.02</td>
<td>33.3 (24.4 to 43.2)</td>
<td>0.95 (–0.91 to 2.81)</td>
<td>65.7</td>
<td></td>
</tr>
</tbody>
</table>

Percentage that increased represents the percentage of applications in which the LTI increased after viewing the SVI video. p < 0.0001 represents the test of differences in proportions.

LTI = likelihood to invite; SVI = standardized video interview.
screening residency applications and is likely an insurmountable task for most residency programs. EM faculty spend an average of 250 hours per program on their current LTI decisions. In addition, EM residency programs and applicants appear to spend over US$66 million per cycle on the interview process. To view every video for every applicant would represent an insurmountable cost to a residency program. However, a targeted approach to video reviews would be prudent and could be a future subject to investigate. It is important to remember that the AAMC recommends that the SVI not be used in isolation. Rather it should be interpreted in the context of the complete application.

Future research can focus on whether the SVI correlates with real-world residency outcomes including performance on communication and professionalism milestones. Ultimately, it will take several years of data to answer this question.

LIMITATIONS

Each program has a different method of evaluating applicants and places varying values on different elements of the application. Thus, there were limitations in standardization of the application review process. In an attempt to overcome this, we allowed reviewers to view all elements of the ERAS application as they normally would before assigning the LTI. However, they remained blinded to the SVI score. The SVI score was then revealed and they reassessed LTI. Thus, the only variable that changed was viewing the SVI score. In addition, we enrolled several sites in varying geographic locations to capture the true impact of the SVI on an applicant’s LTI.

Each site determined which applications would be screened for an interview based on their preexisting criteria. This was not standardized across the study sites. Thus, there may have been applicants who were not screened in our study due to other factors preventing them from receiving an interview (i.e., low USMLE scores). Also, not all applications at all sites were captured as not all faculty reviewers participated in the study.

The faculty application reviewers were not blinded to the purpose of the study. Thus, there was likely a Hawthorne effect on their decision on the LTI after seeing the SVI score. However, there is no way to blind them to the identity of the applicant because that would prevent them from having access to the personal statement, SLOEs, and MSPEs.

We viewed each application as a single data point. Thus, some applicants were reviewed at multiple sites and counted as separate data points for the purpose of the study. We believe this was appropriate because each site individually assesses each applicant independent of other sites.

We excluded applications in which the reviewer prematurely saw the SVI score prior to assigning an LTI. This was determined by self-reporting of the reviewer. It is possible that some reviews were prematurely unblinded to the SVI but were not excluded.

Our study was limited to nine programs. While we included many applications in the study, it is possible that other programs would value the SVI score differently and it may have a different impact on LTI in each institution. In fact, in our study, the ICC of the effect of the SVI score on change in LTI was poor. This suggests that various sites may use the SVI score in different ways.

Our study occurred during the pilot phase of the SVI. Thus, residency leaders had limited experience with the SVI and may have been unsure how to incorporate it into the application review process. In future years, as comfort with the SVI increases, the SVI may exert stronger influence on application review.

Watching the actual SVI video was optional in our study. We also did not standardize how long reviewers watched the videos. It is possible that by watching more of the video their LTI would have been different.

CONCLUSIONS

In a multicenter prospective observational study, the standardized video interview score changed the likelihood to invite in 7% of applications. In this group, the score was equally likely to increase or decrease the likelihood to invite. Lower standardized video interview scores were more likely to decrease the likelihood to invite than higher scores were to increase the likelihood to invite. The standardized video interview video was more likely to increase the likelihood to invite than to decrease it.

References

1. About the SVI. Association of American Medical Colleges. Available at: https://students-residents.aamc.org/


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10331/full

Data Supplement S1. Supplemental material.
Science Policy Training for a New Physician Leader: Description and Framework of a Novel Climate and Health Science Policy Fellowship

Jay Lemery, MD, Cecilia Sorensen, MD, John Balbus, MD, MPH, Lee Newman, MD, MA, Christopher Davis, MD, Elaine Reno, MD, Renee Salas, MD, MPH, MS, and Emilie Calvello Hynes, MD, MPH

ABSTRACT

The accelerating health impacts of climate change are undermining global health, and the roles of the health sector in addressing the many challenges of climate change are being articulated by governments, multilateral institutions, and professional societies. Given the paucity of physician engagement on this issue to date, there now exists a clear need for health professionals to meet this new challenge with the development and cultivation of new knowledge and skill sets in public health, environmental science, policy, and communication. We describe a novel GME fellowship in climate and health science policy, designed to train a new generation of clinicians to provide the necessary perspective and skills for effective leadership in this field. This fellowship identifies available university resources and leverages external collaborations (government, medical consortiums, affiliate institutions in public health, and environmental science), which we describe as being replicatable to similar training programs of any number of medical specialties and likewise bring meaningful opportunities to their respective training programs and academic departments. The creation of this novel fellowship in climate and health policy provides a roadmap and potential path for similar programs to join us in addressing the defining health issue of this generation and many to follow.

Over the past year, physicians and other health professionals have increasingly borne witness to the effects of climate instability on the health and well-being of people throughout the world. In the United States alone, we have experienced multiple “1000-year” hurricanes in the Atlantic and Gulf of Mexico, resulting in flooded cities and the displacement of millions. Toxic wildfire smoke has blanketed the American west with degraded air quality. Extreme rainfall in California has killed scores from devastating landslides. Historic changes to our ecosystems are undermining healthy living, exacerbating illness, and worsening health outcomes for the most vulnerable of patients.1 Increases in morbidity and mortality presenting as syncope from heat stroke, acute myocardial infarction from extreme heat, dyspnea from degraded air quality and aeroallergens, vomiting and diarrhea from diminished water quality, and depression and anxiety have all been associated with extremes of temperature and precipitation and severe weather events.2–10

The imperative for informed policy and deliberate action is becoming more urgent, and despite the mounting evidence that climate change is negatively affecting human health domestically and worldwide,
the clinical community has been conspicuously absent in developing policy and initiatives to address this threat. Initial surveys have suggested that many American physicians agree that climate change is human caused and is worsening, and most of those responding to the surveys reported that they are at least modestly knowledgeable about the association between climate change and health. Yet a 2016 physician survey by the George Mason Center for Climate Change Communication cited significant barriers to physicians engaging in and taking action on this issue, including a lack of time, an uncertainty on how to communicate about this issue, and a lack of resources and recommendations for patients.

In these multifaceted challenges, faculty at the University of Colorado (CU), Department of Emergency Medicine—with the assistance of many other partners—saw both a need and an opportunity. In early 2017, the Department inaugurated the first graduate medical education (GME) fellowship in climate and health science policy with a purpose to train and equip new physician leaders in climate science, education, and advocacy skills. We envision this GME fellowship to be relevant throughout the entire house of medicine, but starting in emergency medicine (EM) was an intuitive fit. The health impacts from climate change can be best characterized as falling largely on the most vulnerable populations—often in the setting of natural disasters. In this regard, EM is well suited to develop such a program, given its proficiency in disaster response and indefatigable care for the most vulnerable.

In this paper, we describe the specifics of this fellowship and provide a rationale and blueprint for replication in other GME programs. We present this material structured in best practice norms for medical education curriculum development, a six-step approach synthesized by Kern et al.

PROBLEM IDENTIFICATION AND GENERAL NEEDS ASSESSMENT

Despite the accumulation of scientific evidence clarifying climate change and its health impacts, there still remains a conspicuous absence of clinicians in societal discourse on the subject. Much like when the International Physicians for the Prevention of Nuclear War distilled the complex geopolitics of nuclear proliferation to an essential health issue (and won the Nobel Prize in 1985), there exists an analogous opportunity for medicine to dispel myths and to provide accurate, dispassionate risk information on climate change and its related health impacts.

Consider the current state of engagement between known climate science and health care. The science linking current and future changes in climate variability and weather events to increases in morbidity and mortality is strong. The resulting increase in health-related costs, especially for those with acute and chronic medical and psychological needs as well as specific vulnerable populations (i.e., children, women, workers) has also been documented. For example, between 2000-2009, heat waves are estimated to have contributed to over $5 billion in direct health care costs in the United States. The Zika epidemic in the Americas, which may have been facilitated by permissive climatic conditions that increased mosquito vector abundance, led to $1.1 billion of public spending for response and preparedness alone. Another study estimated that just six climate-related events resulted in over $14 billion of health-related costs from provision of medical services and loss of life and productivity. However, the health care costs and needs of those impacted by climate-related events are rarely factored into health policy decisions. Attribution of mortality and morbidity from climate-related events is hindered by a lack of aggregated health data in many regions of the United States (and globally) as well as the challenge in assigning causality of a specific environmental event to climate change. Additionally, there is a dearth of interdisciplinary collaboration and scientific investigation into the direct impacts of climate change on health care cost and utilization, and analyses are almost always retrospective in nature. Compared to other significant environmental health issues, there has been insufficient research investment in the subject to support accurate projections and to form sound policy decisions.

The emerging complexities of climate and health issues necessitate development and dissemination of new knowledge sets. In 2015, this sentiment was supported at the highest levels of the U.S. government when President Obama convened scores of deans and staff from the medical, public health, and nursing colleges and schools to make a public commitment to establish goals, align objectives, and organize actions for education and training on the health impacts of climate change. Unfortunately the follow-through on
these pledges has been inconsistent, and the need remains for health care providers who can effectively identify these risks and provide critical analysis.

Why is there a need for a physician fellowship? *Lancet* has characterized climate change as the greatest global health threat (and opportunity) of the 21st century. Yet there has been a paucity of physician engagement on this topic to date. We believe that a fellowship can fill this gap with physicians capable of assuming leadership, disseminating knowledge, and influencing health care policy—integrating all in a multidisciplinary manner. We also differentiate this novel training from a traditional masters of public health (MPH) program. As there have been increasing calls for physicians to become more engaged in this issue, we propose that formal training in governmental and NGO advocacy, education and curriculum development, and science communication will fulfill a need for physician leadership more precisely than an orthodox, 2-year MPH curriculum.

There is also a business case to be made. For health care policy makers, greater awareness of the magnitude and specifics of climate-driven health events could lead to better care and reduced costs. Physicians trained in blending climate science, policy, and research have the ability to advocate for patient health and to better contribute to the conversation on the systemwide financial risks of climate change on health care expenditure. Physicians knowledgeable in these areas have an opportunity to shape multidisciplinary (i.e., engineering, operations, development) decisions on health care systems’ investment in climate-resilient hospital infrastructure, designed to ensure continuity of services during extreme weather events and prevent the downstream negative health outcomes due to lack of access to care following disasters.

**TARGETED NEEDS ASSESSMENT**

In early 2015, diverse members of the CU came together from its four constituent campuses to form the CU Consortium on Climate Change and Health (CUconsortium.org). Recognizing the complex science behind climate change and the many benefits to address its challenges from a multidisciplinary approach, the Consortium attracted faculty from health research basic science, environmental science, public health, public policy, and disparate clinical disciplines such as nephrology, immunology, occupational and environmental medicine, and emergency medicine. A conference in September 2016 sponsored by the Aspen Global Change Institute codified the mission of the consortium to establish a multidisciplinary approach to climate research, education, and policy.

Through these meetings, new and existing collaborations flourished. Private sector entities and colleagues at government agencies who focused on climate and health issues (CDC, NIEHS) saw strategic synergies in working with academia. Subsequently, the Denver-based non-profit Living Closer Foundation agreed to underwrite a GME fellowship to “train physicians in climate education, meaningful engagement, and effective communication to create leaders in the field of climate change and heath.” Funds were specifically earmarked to 1) support a significant portion of the fellow’s time toward scholarship and research and 2) support travel to off-site preceptorships at the NIH/NIEHS Office of the Director (Bethesda MD) and the CDC (Atlanta, GA). These off-site preceptorships were crafted for the fellow to have access to and training for active policy discussions, such as transagency committees and advisory councils as well as interagency committees and working groups. By establishing official volunteer status within the agencies, the fellows’ unique exposure to governmental meetings facilitated opportunities to participate in authorship and presentations and to play a meaningful role in the review and preparation of science policy documents.

Although other well-established GME science policy fellowship opportunities do exist (American Association for the Advancement of Science, Robert Wood Johnson Foundation), we envisioned our program to better utilize the clinical expertise of the fellow and to be less geographically constricted to Washington, DC. We also understood that climate change and health issues have not been a historical priority for these fellowships, and we were committed to establishing a physician science policy fellowship where this issue had unequivocal primacy. Furthermore, we believe that clinical work confers credibility and a too often absent perspective in policy deliberations for effective leadership in this area.

We have differentiated this fellowship from a traditional course of study through a MPH program and do not believe a concomitant degree is a precondition for a physician to lead on this issue. Yet we do recognize the value of this training to early-career physicians seeking to work within the public health sphere.
Accordingly, we have utilized a close relationship between the CU School of Medicine and the Colorado School of Public Health to offer an optional 2-year fellowship MPH track within our fellowship.

GOALS AND OBJECTIVES

After careful evaluation of existing educational competencies in multiple fields of public health and clinical medicine, we have proposed this set of competencies as a pilot for the Climate Change and Health Science Policy (CCHSP) fellowship—understanding that such competencies may evolve over time. The core list represents essential skills for the fellow given the unique opportunities of partner institutions and the CU. While we believe that these competencies have universal value, other institutions may change these competencies and their relative emphasis given their own strengths, weaknesses, and resources (Table 1).

EDUCATIONAL STRATEGIES AND IMPLEMENTATION

Policy/Advocacy Mentorship

A key facet of the fellowship is to leverage the work of site preceptors and partners to gain experience and exposure to organizations in working on climate and health issues, as a way to consistently reinforce fellowship objectives. Projects the fellow has participated in to date:

- Technical contributor to the U.S. Government Fourth National Climate Assessment.
- National Institute of Environmental Health Science, Global Environmental Health Program: developed and spoke at conference session on the impact of the environment upon women’s health in India, leading to two manuscripts; also participated in meetings with Indian government health officials.
- The World Bank—assessment documents on climate and health – Fellow provided content inputs and critical review.
- Society for Academic Emergency Medicine, leadership within the newly created climate change and health interest group.
- Physicians for Social Responsibility, founding member of Colorado working group.
- Fellow lead collaborative meetings with the Colorado Department of Public Health and Environment to discuss official state reports and the department’s ongoing air and water monitoring activities.
- Citizens climate lobby—clinical advisor.
- Fellow presented to state representatives on climate and health issues in Colorado.
- The Global Consortium on Climate and Health Education at the Columbia University Mailman School of Public Health—fellow contributed to content and review of core educational objectives.
- American Geophysical Union—participant in the Resilience Dialogues, which partners experts with communities to explore their risks from climate variability and change.

Considering future additional fellowships at academic other academic sites, we conclude each subcategory of Educational Strategies and Implementation with “replicable aspects.”

Replicable Aspects for Policy/Advocacy Mentorship.

Many of the aforementioned organizations have abundant opportunities and needs for physician fellows to contribute to their missions and in turn gain valuable training. We envision these groups to be the nidus of a training network for additional fellowships in this field.

Education

A core objective is to develop the fellow’s capacity to be an effective educator—through curriculum development, lecturing (classroom/conference), and social media platforms. The fellow is primarily charged with administering and assistant teaching for the climate and health electives at the Colorado School of Public Health (weekly, fall semester) and at the CU School of Medicine (2-week intensive, spring semester). The fellow is also tasked with other projects, such as spearheading the development of educational products for affiliate partners and constituents (e.g., CME curricula for the Medical Society Consortium on Climate and Health).

Fellow participation in other departmental educational endeavors bolsters their facility with podium presentations and effective science communication (e.g., resident conferences, department/section educational programs for laypeople and undergraduates). Monthly didactic sessions occur during which the fellow meets with identified topic-expert preceptors to discuss key competencies. These meetings occur both in Denver as well as at stakeholder sites (e.g., NIH in greater...
Table 1  
Set of Competencies as a Pilot for the CCHSP Fellowship

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal 1. Fluency with climate and health impacts: understanding how these perturbations in earth science impact human well-being—both pathophysiologic and societal.</strong></td>
<td></td>
</tr>
<tr>
<td>Understand scientific foundations</td>
<td></td>
</tr>
<tr>
<td>Recognize basic science of climate change and identify the sources of greenhouse gases</td>
<td>By 6 months, the fellows will be able to accurately list and describe these objectives during formative individual assessments through structured mentorship meetings with core faculty.</td>
</tr>
<tr>
<td>Describe current climate change projections: U.S. and global</td>
<td></td>
</tr>
<tr>
<td>Illustrate systems’ (physiologic, ecologic, social, etc.) interactions and exposure pathways resulting in health impacts of climate change</td>
<td>By 6 months, the fellows will be able to accurately list and describe two of these objectives during formative individual assessments through structured mentorship meetings with core faculty. At 12 months, the fellows will be able to describe all four of these objectives. Evaluations of scholarship directly incorporating these objectives (at least two of the four) via peer review or editorial process, presentation evaluations, and direct mentorship from core fellowship faculty.</td>
</tr>
<tr>
<td>Recognize policy analysis methods and principles</td>
<td></td>
</tr>
<tr>
<td>Define an understanding of research methods:</td>
<td></td>
</tr>
<tr>
<td>Epidemiologic research methods</td>
<td></td>
</tr>
<tr>
<td>Community and field-based research methods</td>
<td></td>
</tr>
<tr>
<td>Intervention science: dissemination, implementation, and evaluation methods</td>
<td></td>
</tr>
<tr>
<td>Discriminate policies outside of health (urban planning, education, business sustainability/corporate social responsibility, etc.) that could be health-relevant</td>
<td></td>
</tr>
<tr>
<td><strong>Goal 2. Facility with concepts of mitigation and adaptation as actions within public and private entities and to evaluation quality and effectiveness of such actions related to health impacts.</strong></td>
<td></td>
</tr>
<tr>
<td>Interpret the health implications of climate change, including:</td>
<td></td>
</tr>
<tr>
<td>Health impacts in the United States</td>
<td>By 3 months, the fellows will be able to accurately list and describe these objectives during formative individual assessments through structured mentorship meetings with core fellowship faculty.</td>
</tr>
<tr>
<td>Health impacts in other parts of the world</td>
<td></td>
</tr>
<tr>
<td>Articulate the potential health benefits from climate mitigation and adaptation measures</td>
<td>By 3 months, fellows will be able to accurately describe this objective during formative individual assessments through structured mentorship meetings with core fellowship faculty. Evaluation of scholarship directly incorporating this objectives via peer review or editorial process (by the end of fellowship).</td>
</tr>
<tr>
<td>Appreciate sustainable and climate resilient health care facilities</td>
<td>This objective is directly assessed through formative individual assessments through core and affiliate fellowship faculty directly engaged with the fellow’s work with healthcare without harm.</td>
</tr>
<tr>
<td>Understand U.S. government and relevant state policies and institutions</td>
<td></td>
</tr>
<tr>
<td>Differentiate Federal agencies, including facility with the work of the CDC and the NIH</td>
<td>The following objectives are directly assessed through formative individual assessments through core and affiliate faculty directly engaged with the fellow’s work at the NIH and the U.S. Global Change Research Program.</td>
</tr>
<tr>
<td>Explain the clean power plant rule and other executive and legislative measures</td>
<td></td>
</tr>
<tr>
<td>Recognize U.S. Global Change Research Program and other federal interagency activities</td>
<td></td>
</tr>
<tr>
<td>Recognize Executive branch institutions (Office of Science and Technology Policy, Council on Environmental Quality, National Security Council, etc.)</td>
<td></td>
</tr>
<tr>
<td>Explain legislative structures and processes</td>
<td></td>
</tr>
<tr>
<td>Explain U.S. medical societies’ positions on climate change and health</td>
<td>This objective is directly assessed through formative individual assessments through core and affiliate fellowship faculty directly engaged with the fellow’s work at the Medical Society Consortium on Climate and Health.</td>
</tr>
</tbody>
</table>

(Continued)
Washington DC). Other projects the fellow has contributed to: National Institute of Environmental Health Sciences Educational Curriculum on Climate and Health—curriculum development and content contribution; Medical Society Consortium on Climate and Health—lead of continuing medical education; clinical correlates editor of a major textbook on global climate change and human health; and podcast presenter for the Environmental Defense Fund. To further ensure competency with educational objectives, the fellow has regular meetings with fellowship directors and preceptors to discuss key readings on specific climate and health topics.

**Replicable Aspects for Education.** Most GME training programs have deep connectivity with their respective medical schools, undergraduate campuses, and medical campus community outreach programs. There are readily available and accessible curricula on this topic to support a fellowship program (i.e., fellows + faculty) to stand-up educational programs within their affiliated academic communities.
Research/Scholarship
Through collaborations with a broad diversity of actively engaged experts in the fields of public health, clinical medicine, and policy, the fellow has access to mentors and relevant primary data to advance scholarship in this area and to augment proficiency with fellowship educational objectives. The fellow’s successful publications have included primary research, policy papers, conference proceedings, and systematic review articles, including:
- Assessment of excess mortality in Puerto Rico from Hurricane Maria (2017).
- Biomarker assessment of chronic kidney disease from heat stress in Guatemalan field workers.
- Assessment of climate change on vector borne disease in Ecuador.
- An inventory of climate change exacerbations of emergency medicine presentations in the United States.

Replicable Aspects for Research/Scholarship. We discussed that many of these opportunities may be available to a network of fellows through an established training network, many GME programs have access to “local” talent through affiliate academic campuses, in the fields of environmental science, basic medical science, global health, law, human rights, etc. This approach to a multidisciplinary team of mentorship reflects the very nature of climate change and health—that it is a multisector issue well beyond the traditional boundaries of medicine.

Clinical Practice
Clinical practice supports the fellow’s proficiency of the educational objectives by partially offsetting his/her salary by working clinical shifts in the UCHealth Emergency Medicine System (CU Hospital, affiliate UCHealth community hospitals and UCHealth free-standing emergency departments) as an attending physician. Fellow(s) are American Board of Emergency Medicine board-eligible/board-certified. Like many other GME fellowships, this clinical work is an economic necessity to sustain the long-term viability of the fellowship. We also believe that active clinical experience is a positive differentiator in this field, currently dominated by those with traditional training in public policy and public health. We believe that clinical experience helps to add credibility and a necessary perspective for advocacy and effective leadership in this field. By example, the fellow is currently an investigator on a Colorado School of Public Health study in Guatemala performing physical examinations and collecting urine and serologic data to assess at-risk populations under conditions of heat extremes. Other fellowship activities with clinical relevance have included field projects in Ecuador, clarifying climate drivers of the Syrian refugee crisis (Lebanon based), and a posthurricane vulnerability assessment in Puerto Rico. Fellowship funds are specifically earmarked to “buy-down” clinical work, which increases the amount of time the fellow may spend on research and policy scholarship.

Replicable Aspects for Clinical Practice. Many GME fellowships are supported through clinical work, and we believe that this a familiar administrative model to most EM training programs.

Personal/Professional Development
Because of the novelty of this fellowship, regular meetings to assess overall wellness and professional development are integral to ensure success. Because the fellowship affords exposure to so many different organizations and people, much of the mentorship on this topic centers on strategies for procuring opportunities after graduation.

These meetings reassess competency with the fellowship objectives, as well as compliance with minimum expectations, which include:
- Minimum of three peer-reviewed publications each year of fellowship, ideally linking these publications to national or international meetings.
- Presentation at least two national or international meetings.
- Minimum of three editorials or blog posts targeting response to climate related current events.
- Active participation in the educational programs of the department and affiliates.
- Meet all citizenship requirements of department faculty.

Replicable Aspects for Personal/Professional Development. We believe that this approach establishes accountability for all fellowship principals to remain attentive to goals and objectives throughout the year. Such attention will not only ensure the fellowship meets the standards of concomitant GME training programs, but will also allow it to develop, grow, and ultimately thrive in an academic environment.
Duration
The duration of the fellowship is either 1 or 2 years, depending on the specific goals of the fellow, in consultation with the program directors and department leadership. Two or more years’ duration is necessary for completion of the fellowship with concomitant degree-seeking coursework (MPH, MA).

Fellowship Director and Program Administration
The fellowship is administered through the Department of Emergency Medicine, Section of Wilderness & Environmental Medicine (WEM). The fellowship director and an associate fellowship director are clinical faculty at the CU School of Medicine. Site preceptors at the NIH, CDC, and Colorado School of Public Health also play important roles during the fellow’s extended duration and training deployments to Washington, DC, and Atlanta, respectively. The Section of WEM provides administrative support for the fellowship.

As this is an inaugural fellowship, there is no precedent for director qualifications. The director (JL) has a background in wilderness medicine leadership and medical education, which led to over a decade of policy work in climate change and health. The commensurate network development from this work became the effective scaffolding to engage affiliate faculty to support the fellowship. The associate fellowship director (ECH) has a background in global health and is the Director of the Global Emergency Care Initiative at Denver Health/CU. Both the fellowship director and associate director are American Board of Emergency Medicine Diplomates and Fellows of the American College of Emergency Physicians.

Fellowship Faculty and Advisory Committee
Fellowship faculty assist the fellow through scholarship and research mentorship, through access to meetings and collaborative policy working groups (government, private, public sector), and through diverse expertise and connectivity. Members of the advisory committee represent basic sciences and clinical departments within the School of Medicine, School of Public Health, and government entities such as the CDC and NIH/NIEHS and affiliate organizations (Medical Society Consortium on Climate Change and Health).

EVALUATION AND FEEDBACK
The fellow has formative monthly meetings with the fellowship director and/or the associate fellowship director. These meetings serve as formal assessments on meeting fellowship goals and on competency with fellowship objectives. Fellowship directors incorporate formative feedback from site preceptors as well as the chief of service of the ED (clinical performance) in these meetings. A summative assessment of both the individual fellow and the overall fellowship is submitted to the department chair and the fellowship advisory committee by the fellowship director at the end of the academic year. The summative assessment incorporates survey data sent to fellowship principals and preceptors, solicited via Likert-scale questions tied to fellowship objectives competencies as well as open-ended questions allowing for qualitative feedback. Formative and summative feedback is also collected on the fellowship itself from all stakeholders, allowing for continuous curriculum maintenance (i.e., deletions and additions) and assessment of off-site projects with goals and objectives.

The fellow is likewise mandated to provide formative and summative assessment on congruence with the fellowship experience and the “boots on the ground” reality with stated goals and objectives. Given fast-paced and dynamic nature of this fellowship, we have also established a culture of less formal, “just-in-time” feedback from the fellow to the fellowship directors.

Although the fellowship is less than 2 years old to date, there are some early lessons learned. One is that of competing priorities amongst the fellowship preceptors. Although the opportunities afforded to the fellows are far and wide, the areas of activity (government policy arena, School of Public Health, national medical consortiums, and school of medicine/department of emergency medicine) often have opposing views of what is of value to the fellow’s training. Such a coalition of divergent opinions in theory offer a rich educational ecosystem for a fellow, but in practice can make for a stressful and chaotic agenda. Early corrections centered on process for the initiation of fellow process and the fellowship director’s better organizing communication among the preceptors. Likewise, such differing points of views obscured concordance in summative assessments among the preceptors.
Other challenges center on recruiting an interested and capable cohort of future fellows who may be otherwise reluctant to apply to a new and unorthodox program different from the EM core competencies of residency training. Finally two persistent challenges are long-term funding for the fellowship and a lingering skepticism from a significant percentage of the medical community on the true risks of climate change on health. Although we hope that a future consortium of climate and health science policy fellowships will emerge—pooling resources to lessen the funding needed for such a fellowship to thrive—the intransigence of leaders (administrative, fundraising) throughout the medical community to promote this effort remains a challenge for dissemination.

DISSEMINATION: APPLICABILITY TO OTHER SPECIALTIES, INSTITUTIONS, AND CONSORTIUM BUILDING

The unique confluence of synergies led to the financial resources and human capital required to launch this initial CCHSP fellowship. We believe that this type of opportunity not only can be duplicated at other institutions, it must be for the growth and advancement of this critical area of expertise. While we anticipate wide applicability of our outlined competencies, they were developed with our specific institution and resources as a framework. Thus, we anticipate that other institutions looking to develop a climate change and health fellowship could tailor their curriculum to highlight their respective internal strengths and unique external collaborations.

Our larger aim is to develop a consortium of institutions dedicated to developing and advancing this type of fellowship education, applicable to all medical specialties. This will allow prospective fellows from across the country access to a wide network of resources and opportunities. For example, relationships with our current collaborating agencies could be utilized (NIEHS, CDC). In addition, we are currently collaborating with climate change and health experts at the Harvard Medical School with intent to stand up a second realization of the fellowship. While the current collaborations are inherently beneficial and fruitful, this site may prove to be a future fellowship site that can duplicate the foundations of the CCHSP curriculum while catering to the unique internal strengths of the larger Harvard community.

CONCLUSION

As the health impacts from climate change become more apparent, the medical community needs to continue to strive to meet the health care needs of its patients. Many of our fellowship partners—academic, private, and public sector—have welcomed participation from motivated, knowledgeable clinicians and have noted that such skill sets and perspectives are rare in policy discussions. The resounding refrain from our external partners has been, “one fellow is good . . . more would be better.” We believe that these external collaborations would be readily available to similar training programs of any number of medical specialties and likewise bring meaningful opportunities to their respective training programs and academic departments. The creation of this novel fellowship in climate and health policy provides a roadmap and potential path for similar programs to join us in addressing the defining health issue of this generation and many to follow.

References

23. Lloyd EA, Oreske N. Climate change attribution: when is it appropriate to accept new methods? Earth’s Future 2018;6:311–25.
Resident Clinical Experience in the Emergency Department: Patient Encounters by Postgraduate Year

Amy Douglass, MD, MPH, Kathleen Yip, MD, Debryna Lumanauw, MD, Ross J. Fleischman, MD, MCR, Jaime Jordan, MD, and David A. Tanen, MD

ABSTRACT

**Background:** During emergency medicine (EM) training, residents are exposed to a wide spectrum of patient complaints. We sought to determine how resident clinical experience changes based on training level in relation to the patient acuity levels, chief complaints, and dispositions.

**Methods:** We performed a retrospective chart review of patients seen at a safety-net, academic hospital in Los Angeles from July 1, 2015, to June 30, 2016. Resident postgraduate year (PGY) level and specialty, patient acuity (based on the Emergency Severity Index), chief complaint (based on one of 30 categories), and disposition were abstracted. Our primary objective was to examine the progression of EM resident experience throughout the course of training. As a secondary objective, we compared the cases seen by EM and off-service PGY-1s.

**Results:** A total of 49,535 visits were examined, and of these, 32,870 (66.4%) were in the adult ED (AED) and 16,665 (33.6%) were in the pediatric ED (PED). The median acuity level was 3, and 27.4% of AED patients and 7.3% of PED patients were admitted. Data from 126 residents were analyzed. This included 94 PGY-1 residents (16 EM and 78 off-service), 16 PGY-2 EM, and 16 PGY-3 EM residents. Residents of different training levels evaluated different types of patients. Senior EM residents were more likely to care for higher-acuity patients than junior EM residents. EM PGY-3s saw higher percentages of acuity level 1 and 2 patients (2.3 and 37.8%, respectively, of their total patients) than EM PGY-1s (0.3 and 18.7%, respectively). Conversely, EM PGY-1s saw higher percentages of acuity level 4 and 5 patients (27.9 and 1.6%, respectively) compared to EM PGY-3s (10.7 and 0.7%, respectively). There was a significant linear trend for increasing acuity with training year among EM residents (p < 0.001). EM PGY-1s saw more patients than off-service PGY-1s with slightly higher acuities and admission rates.

**Conclusion:** The clinical experience of EM residents varies based on their level of training. EM residents show a progression throughout residency and are more likely to encounter higher volumes of patients with higher acuity as they progress in their training. When designing EM residency curriculums, this is a model of an EM residency program.

Residents who work in the emergency department (ED) see a variety of patients with varying acuity levels, chief complaints, and diagnoses.1–9 Emergency medicine (EM) residents often have some control over the number and types of patients they evaluate, since residents in the ED may choose to assign themselves...
Multiple studies have evaluated the clinical experiences of EM residents, typically focusing on factors such as resident productivity, procedural experience, and types of patients seen. Residents see more patients per hour1–3 and generate more RVUs per hour4 as they progress in their training. They also see more patients per hour when they work shorter shifts.5 While productivity generally increases as residents progress in their training,1–3 interestingly, a study by Henning et al.6 found that while EM interns increase productivity significantly over the course of a year, senior EM residents’ productivity did not change significantly. Procedural experience varies among different residency types and settings.7 Residents may not be exposed to all important pathology. In fact, in a study by Langdorf et al.,3 participating residents did not see a large number of core EM diagnoses during a 9-month study period. Another study compared types of patients seen by EM and pediatric residents in a pediatric ED and found that EM residents see more surgical patients, while pediatric residents see more medical patients.8 Finally, two studies compared the training experiences of EM residents in urban versus rural training environments and found that overall experiences were similar with the exception that patients in urban training sites had overall higher acuities, and residents at these sites performed more procedures.9,10 Despite this body of literature evaluating EM training experience, none to our knowledge have quantified differences among the types of patients seen by residents based on their training level.

The aim of our study was to evaluate the clinical experiences of residents in the ED based on their level of training. We hypothesized that residents at different training levels see different types of patients in the ED with varied acuity levels, chief complaints, and resulting dispositions. Our primary objective was to examine the progression of EM resident experience throughout the course of training. As a secondary objective, we compared the cases seen by EM and off-service postgraduate year (PGY)-1s.

**METHODS**

**Study Design**

This was a retrospective chart review conducted at Harbor–UCLA Medical Center, including data from July 1, 2015, to June 30, 2016. Electronic medical records for all patients seen in the adult or pediatric ED who were seen by a resident during this time frame were reviewed. For each patient encounter, a predetermined data collection form was completed including: resident PGY level, gender and specialty, patient acuity level, chief complaint, and disposition. Acuity was classified based on the Emergency Severity Index.11–13 The chief complaint recorded by the triage nurse was categorized into one of 30 categories determined by the study team based on previously published data of the most common chief complaints.14,15 Categorization was reviewed by a second investigator and any discrepancies were reviewed by a third investigator. Resident shift schedules were reviewed and tabulated for the number of adult and pediatric shifts to calculate the number and types of patients seen in each environment. As many patients seen by PGY-1s are supervised by an upper-level EM resident, both the PGY-1 and the upper-level resident were given credit for patients they saw together. Only the first resident of each year to see the patient was considered in the analysis, so residents who assumed care of a patient in sign-out were not credited with the encounter.

**Study Setting and Population**

Harbor–UCLA Medical Center is a large, urban, academic safety-net hospital, located in Los Angeles with nearly approximately 90,000 ED visits per year. Harbor–UCLA Medical Center serves a diverse county population made up of approximately 55% Latino, 20% African American, 18% Caucasian, and 7% other patients with a large percentage of non-English-speaking patients. The hospital includes adult, pediatric, and psychiatric EDs. For this study, patients presenting primarily to the psychiatric ED were not included (i.e., those who were not seen or only had a medical screening exam in adult ED [AED] or pediatric ED [PED]), while patients with psychiatric complaints who were seen in the AED or PED were included in the study. The remaining patient encounters to the AEDs and PEDs were reviewed for the period between July 1, 2015, and June 30, 2016. Patients were excluded from analysis if they were not seen by an intern or ED resident (i.e., patients only seen by attending physicians, nurse practitioners or off-service upper-year rotating residents) since the primary aim of the study was to evaluate the experience of EM residents. Off-service interns were included in the study because they regularly work under the supervision of upper-level
EM residents, while more senior off-service rotators do not.

At the time of our study, Harbor–UCLA’s EM residency program consisted of 3 years of training with 16 residents per year. The AED has off-service rotating residents from multiple departments, including internal medicine, psychiatry, orthopedics, combined internal medicine and pediatrics, and family medicine. The PED is primarily staffed by EM, pediatrics, and family medicine residents. Off-service upper-level residents rotating in the ED were excluded as we expected the heterogeneity of their training experience to make it difficult to draw generalizable conclusions about their ED experience. Shifts in the AED average 9.5 hours, with 6.5 hours devoted to seeing new patients, 2 hours for sign out and teaching and 1 hour for cleanup and note writing time at the end of the shift, while shifts in the PED average 8.5 hours, which is entirely dedicated to patient care with the exception of 30 minutes of cleanup and note writing time at the end of the shift. A typical AED shift has one PGY-2 and -3 resident and often a PGY-1. During signouts PGY-2s receive signouts on more active patients while PGY-3s receive signouts on patients that have been dispositioned or with less active issues. EM and off-service residents from all levels of training are encouraged to see a wide range of types of patients to gain a variety of clinical experiences. Residents are not restricted in the types of patients they see with the exception of trauma activated patients who are only seen by PGY-2s and -3s. At the time of the study, there were no lower-acuity areas of the ED covered by residents; however, there was a fast-track area managed by nurse practitioners. As a result, residents naturally saw a selection of different acuity patients on shifts.

Attending physicians do not see patients primarily except during weekly resident conference, and that data were excluded from analysis.

Confidentiality and Ethics Approval
The study was approved by the Los Angeles Biomedical Research Institute Institutional Review Board. Residents and patients were deidentified to protect confidentiality prior to analysis.

Data Analysis
Data was exported from the Cerner electronic health record into Excel 2013 (Microsoft Corp.). Subsequent deidentification, manipulation, and analysis was done in STATA IC 15 (StataCorp). All data points were exported in an automated fashion by queries created and validated by an investigator who is familiar with the data structure of our system. The queries were compared to the results of manual chart review and modified until no further discrepancies were found. Significant trends in the number of patients seen, acuity levels, and the proportion of patients admitted versus discharge by class year were determined by linear and logistic regression models. Median acuity levels between EM and off-service PGY-1 residents were compared by the Wilcoxon rank-sum test.

RESULTS
Patient Characteristics
A total of 49,535 patient visits were included in the analysis (Figure 1). Crediting both the intern and the supervising upper-level resident who saw a patient together resulted in 56,669 resident–patient encounters for analysis. The demographics and dispositions of the included patient encounters are shown in Table 1.

Resident Characteristics
A total of 126 residents were included in the study: 16 EM PGY-1s, 78 off-service PGY-1s, 16 PGY-2s, and 16 PGY-3s. Data from 13 non-EM interns were excluded because of extremely low numbers of patients seen per shift, suggesting incomplete data from not signing up for patients in the electronic health record. Number of males in each PGY group were as follows: five (31.2%) EM PGY-1s, 47 (60.3%) off-service PGY-1s, 11 (68.9%) EM PGY-2s, and nine (56.3%) EM PGY-3s. EM PGY-1s spent slightly less of their total shift in the AED (47.9%) compared to the PED, while off-service interns, EM PGY-2s and EM PGY-3s spent more of their time in the AED (83.7, 79.3, and 80.2%, respectively).

Main Results
There was a significant linear increase in the number of patients seen by EM residents over the course of their training in the adult but not the pediatric EDs (Table 2). In an average AED shift, EM PGY-1s saw 4.8 patients (0.7 patients/hour), PGY-2s saw 7.8 patients (1.2 patients/hour), and PGY-3 saw 8.2 patients (1.3 patients/hour). In an average PED shift, EM PGY-1s saw 8.6 patients (1.1 patient/hour), PGY-2s saw 11.3 patients (1.4 patients/hour), and PGY-3 saw 9.9 patients (1.2 patients/hour). Patients/hour data only
reflects new patients/hour seen during clinical hours of the shift (excludes education, signout, and end-of-shift note writing/cleanup time) and does not include patients received in signout, which make up a significant portion of resident workload, or any fast track patients, who were only seen by nurse practitioners at the time of the study. The number of patients seen per shift increased significantly from the EM PGY-1 to the PGY-2 years ($p < 0.001$) but not between the PGY-2 and PGY-3 years ($p < 1$).

Patient acuity levels seen by each PGY group in the combined AED and PED settings are shown in Table 3. There was a significant linear trend for higher acuity patients (lower ESI score) over the course of residency among EM residents ($p < 0.001$). This trend in increasing acuity remained significant at $p < 0.001$ when...
comparing just EM PGY-1s to EM PGY-2s and just PGY-2s to PGY-3s. PGY-3s saw the highest percentage of acuity level 1 patients, followed by PGY-2s (p < 0.001). EM PGY-1s saw the highest percentage of acuity level 4 and 5 patients compared with senior EM residents.

EM PGY-1 residents saw significantly more patients per shift than their off-service PGY-1 counterparts in the adult (p < 0.001) but not the pediatric EDs (p < 0.2). On average EM PGY-1s saw 6.6 patients per shift, 4.47 (0.7 patients/hour) in the AED and 8.6 (1.1 patients/hour) in the PED, while off-service PGY-1s saw 3.6 patients per shift, 2.76 (0.4 patients/hour) in the AED and 7.4 (0.9 patients/hour) in the PED. These levels of significance were similar in a sensitivity analysis including the 13 off-service PGY1 residents who had been excluded based on the assumption of a systemic error in signing up for patients. The median acuity of patients seen by both EM and off-service PGY-1 residents was 3, but because of the large sample size, the acuity of patients seen by the EM PGY-1s was significantly greater (p < 0.0001). This difference was so small that it becomes nonsignificant when AED and PED visits were analyzed separately. EM PGY-1s admitted a significantly higher percentage of patients than their off-service PGY-1 counterparts (p < 0.001); however, in a model accounting for the balance of adult versus pediatric shifts, this difference became insignificant (p < 0.5).

The most common chief complaints are shown in Table 4. Over the course of their residency, EM residents saw a significantly greater proportion of trauma patients and significantly lesser proportion of patients with musculoskeletal complaints when accounting for the distribution of adult versus pediatric shifts. The most over- and underrepresented chief complaints (i.e., complaints seen disproportionately more by a given PGY group than the overall prevalence of the complaint) are shown in Figure 2.
Dispositions by PGY levels are summarized in Table 5. In a logistic regression model accounting for the distribution of pediatric and adult shifts, EM residents saw a significantly decreasing proportion of patients who were discharged and significantly increasing proportion of those who were admitted or died. The majority of patients who expired were seen by PGY-2s and PGY-3s. The proportion of patients who were admitted increased significantly from the EM PGY-1 to the PGY-2 years ($p < 0.001$) but not between the PGY-2 and PGY-3 years ($p < 0.5$).

**DISCUSSION**

Residency training is based on experiential growth in the knowledge and breadth of EM. Assuring that residents gain experience to master the knowledge and clinical skills necessary to practice EM is a primary goal of EM residency programs. Residents are encouraged to reach milestones throughout their training with graduated responsibilities and expectations as they advance in training.

Our study demonstrates not only that residents see more patients as they progress from PGY-1 year to upper level years, as has been shown in prior literature,$^{1-3}$ but also that residents at different training levels select different types of patients in the ED in terms of chief complaints and acuity levels. Our data support the desired clinical growth that residencies strive for, that EM residents further in their training see higher-acuity patients and patients with more acute chief complaints. This difference was most pronounced between PGY-1s and upper-level residents, as one might expect. PGY-2s and PGY-3s saw similar numbers of patients per shift but saw higher-acuity patients during the PGY-3 year. While PGY-2s and -3s saw similar numbers of patients, it is important to note that PGY-2s tend to have more supervision, while PGY-3s are more independent in their decision making and take on more of a leadership and teaching role.

### Table 5: ED Disposition by PGY Level

<table>
<thead>
<tr>
<th>Disposition</th>
<th>Total Patients*</th>
<th>PGY-1</th>
<th>Off-service</th>
<th>PGY-2</th>
<th>PGY-3</th>
<th>$p$-value Among EM Residents†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discharged</td>
<td>37,016 (74.7)</td>
<td>6,392 (85.0)</td>
<td>4,419 (77.6)</td>
<td>15,542 (72.9)</td>
<td>15,834 (71.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Admitted</td>
<td>10,230 (20.7)</td>
<td>872 (11.6)</td>
<td>1,034 (18.2)</td>
<td>4,745 (22.3)</td>
<td>5,196 (23.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Interhospital transfer</td>
<td>728 (1.5)</td>
<td>72 (1.0)</td>
<td>87 (1.5)</td>
<td>328 (1.5)</td>
<td>375 (1.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>To other evaluation area at this facility</td>
<td>923 (1.9)</td>
<td>114 (1.5)</td>
<td>107 (1.9)</td>
<td>416 (2.0)</td>
<td>431 (2.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Expired</td>
<td>136 (0.3)</td>
<td>1 (0.0)</td>
<td>3 (0.1)</td>
<td>63 (0.3)</td>
<td>72 (0.3)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Data are reported as $n$ (%).

*Includes total number of patients who had disposition data available.

†$p$-value is for the comparison of EM residents only and refers to the linear trend over the course of training for the proportion with that disposition in a logistic regression model.
role in the ED. In our department, PGY-3s are responsible for helping to manage patient flow for their team (keeping track of all the patients on the team, accepting patient transfers, and triaging patients), which may impact the number of patients they see and contribute to our findings. PGY-2s do not begin to supervise PGY-1s and medical students until the last few months of their PGY-2 year, while PGY-3s are responsible for this supervision throughout the year. This additional teaching and supervision time may impact the number of patients seen by PGY-3s. Finally, PGY-2s and PGY-3s receive different types of patients in signout. PGY-2s receive signouts for more active patients, while PGY-3s receive signout on patients that have been dispositioned with less active issues. While the resident who receives a patient in signout is not credited for that patient in this study, patients that are signed out may impact the numbers of types of patients seen by PGY-2s and -3s. We found that EM PGY-1s see higher volumes of patients than off-service PGY-1s. This difference was only statistically significant in the AED. Additionally, EM PGY-1s tend to see higher-acuity patients and a higher percentage of patients who are admitted; however, these differences are small enough that they are only statistically significant when analyzing AED and PED data as a whole. EM PGY-1s tend to favor patients with chief complaints that are generally thought to be less emergent, such as ear symptoms and skin symptoms, while they select relatively fewer patients with more critical chief complaints, such as syncope and chest pain. These findings appear to be secondary to the fact that EM PGY-1s spend a higher percentage of their total time working in the PED (approximately 50%) where patients tend to present with less emergent chief complaints as compared to off-service PGY-1s (approximately 14%). It is important to note that EM PGY-1s spend a total of 3 months rotating through the ED, while off-service PGY-1s only spend 1 month on average. Additionally, EM PGY-1s typically have experience rotating through the ED as medical students, while off-service PGY-1s may not. Hence, it is likely that EM PGY-1 experience level and comfort with ED logistics impact the numbers of types of patients seen.

Over the past 30 years, studies have shown that EM residents progress in their productivity and patient load over the course of their residencies.1–4 Our study adds to this body of knowledge by demonstrating that the types of patients that residents choose to see also change over the course of their residency training. Educators can use this information to help coach residents at different training levels to expand the types of patients they see.

Future studies are needed to verify the trends observed in our study. It would be beneficial to perform a similar study over a longer time period at additional residency sites. In addition, given that we found that EM PGY-2s and PGY-3s see similar numbers and types of patients, we plan to perform a future study to assess resident autonomy.

LIMITATIONS

While this study is robust in its use of a large sample of an entire year of patient data, it is limited to a single year in a single residency program. We included a total of 126 residents over a 1-year period, of whom 78 were off-service interns who generally only rotate for 1 month in the AED or PED. On each team on a given AED shift, there is only one PGY-1 resident, so off-service PGY-1s are not working during the same shifts as EM PGY-1s and thus should not alter selection of their patients. However, EM PGY-2s only supervise PGY-1s during the last few months of their PGY-2 year, so this could potentially impact types of patients they see. While PGY-1s are advised to see a range of types of patients, it is possible that they could gravitate toward certain types of patients (for example, an orthopedic intern may be interested in seeing patients with orthopedics injuries), which could impact the types of patients seen by the supervising senior residents. Different EM programs have different number of off-service PGY-1 rotators, so this could limit the generalizability of our secondary analysis comparing EM PGY-1s to off-service PGY-1s. Given that each EM class size only consists of 16 residents, it is possible that differences observed between resident classes may be a result of differences in residents within individual classes as opposed to differences in levels of training. Also, the nature of patient flow in our department will be different from other departments. In our program, PGY-3s supervise PGY-1 and medical student encounters in the adult ED, while PGY-2s are only allowed to supervise PGY-1s and medical students in the last few months of their PGY-2 year. Also, PGY-2s take signout on patients whose evaluation is still in progress, while PGY-3s assume care of those whose disposition is settled. These systematic differences in responsibilities between R-2s and R-3s would influence the number of type of patients they would independently see.
de novo. We excluded 56 upper-level off-service residents because we felt that their advanced level of training in different specialties was too heterogenous to categorize as a group and to compare this data to that of upper-level EM residents. Upper-level off-service residents do not see any trauma-activated patients but are otherwise allowed to choose to see any patients they wish to see. Based on our observations, upper-level off-service residents tend to see a range of types of patients. However, it is possible that the patients selected by this group could impact the patients selected by ED residents.

Another limitation of the study is that some data in the medical records were unavailable and may have affected the results if it were included. For example, 597 (1.2%) of patients had unknown acuity levels, and 502 (1.0%) of patients had unknown dispositions. This is an inherent problem with reviewing medical records as not all data are properly recorded and thus was unavoidable in the study. However, Cerner was implemented at our institution in October 2014, and electronic records are ideal for large-volume chart review. In general, resident groups saw similar numbers of these patients with missing data. As 5,983 (12.1%) of patients had chief complaints that did not fall into one of the preset chief complaints, they were reported as “other.” If we had included more categories of chief complaints, we would have reduced the number of other diagnoses. However, our list was based on previously published literature and designed to be a useable length. Finally, this study was conducted at a single county institution with particular challenges such as limited resources and language barriers, so the results may not be generalizable to other residency programs with different curriculums patient populations and nuances in ED flow.

CONCLUSIONS

Our study shows that the clinical experience of emergency medicine residents varies based on their level of training. Emergency medicine residents show a progression throughout residency and are more likely to encounter higher volumes of patients with higher acuity as they progress in their training. Emergency medicine PGY-1s tend to see a higher volume of patients with higher acuities and higher admission rates than off-service interns, which is not surprising given their increased time and experience in the ED. Future multi-institution studies over longer periods of time would help verify this trend. When designing emergency medicine residency curriculums, this a model of an emergency medicine residency program.

References

ORIGINAL CONTRIBUTION

Consensus Core Point-of-care Ultrasound Applications for Pediatric Emergency Medicine Training

Allan Evan Shefrin, MD,1 Fred Warkentine, MD,2 Erika Constantine, MD,3 Amanda Toney, MD,4 Atim Uya, MD,5 Stephanie J. Doniger, MD,6,7 Adam Brand Sivitz, MD,8 Russ Horowitz, MD, RDMS,9 and David Kessler, MD, MSc10

ABSTRACT

Background: Pediatric emergency medicine (PEM) physicians have variably incorporated point-of-care ultrasound (POCUS) into their practice. Prior guidelines describe the scope of POCUS practice for PEM physicians; however, consensus does not yet exist about which applications should be prioritized and taught as fundamental skills for PEM trainees. Members of the PEM POCUS Network (P2Network) conducted a consensus-building process to determine which applications to incorporate into PEM fellowship training.

Methods: A multinational group of experts in PEM POCUS was recruited from the P2Network and greater PEM POCUS community if they met the following criteria: performed over 1,000 POCUS scans and had at least 3 years of experience teaching POCUS to PEM fellows, were a local academic POCUS leader, or completed a formal PEM POCUS fellowship. Experts rated 60 possible PEM POCUS applications for their importance to include as part of a PEM fellowship curriculum using a modified Delphi consensus-building technique.

Results: In round 1, 66 of 92 (72%) participants responded to an e-mail survey of which 48 met expert criteria and completed the survey. Consensus was reached to include 18 items in a PEM fellowship curriculum and to exclude two items. The 40 remaining items and seven additional items were considered in round 2. Thirty-seven of 48 (77%) experts completed round 2 reaching consensus to include three more items and exclude five. The remaining 39 items did not reach consensus for inclusion or exclusion.

Conclusion: Experts reached consensus on 21 core POCUS applications to include in PEM fellowship curricula.

Point-of-care ultrasound (POCUS) has been established in the practice of emergency medicine for over two decades.1,2 More recently, its use has expanded into the practice of pediatric emergency medicine (PEM).3 Most PEM fellowship programs now offer formal ultrasound training to their fellows and incorporate POCUS into the care of the pediatric patient.4,5 In 2015, the American Academy of
Pediatrics (AAP), the American College of Emergency Physicians (ACEP), the Society for Academic Emergency Medicine (SAEM), and the World Interactive Network Focused on Critical Ultrasound (WINFOCUS) issued a joint policy statement on the use of POCUS by PEM physicians thus supporting and further reinforcing its role in the clinical practice of PEM.6,7

In 2001, with revisions in 2008 and 2016, ACEP published POCUS training guidelines in its policy statement for POCUS use by emergency physicians though these were not specific to PEM and do not suggest the applications in which PEM physicians should be competent.8 Further, the Accreditation Council of Graduate Medical Education (ACGME) has made recommendations for ultrasound education of emergency medicine residents through their milestones project.9,10 To date, no similarly comprehensive recommendations exist for PEM fellows.11 Objectives of PEM training include minimal POCUS applications within the American Board of Pediatrics core content for PEM training, certification, and maintenance of certification. Understanding the role of POCUS in the assessment of the trauma victim and sonographic skills related to focused assessment with sonography for trauma, focused cardiac ultrasound, assessment of suspected ectopic pregnancy, and foreign body localization and removal are the only ultrasound-related content included to date.12 Likewise, the Royal College of Physicians and Surgeons of Canada recommends understanding the utility, applications, and limitations of POCUS in its PEM training objectives; however, no specific recommendations exist to guide PEM fellowship directors regarding what POCUS content should be included in fellowship training.13

As PEM physicians have incorporated POCUS into practice the number of described uses for pediatric POCUS has rapidly increased creating a need for detailed educational guidelines for PEM fellowship programs. Consensus guidelines for implementing ultrasound education for PEM fellows have been published by a small group of POCUS leaders, although these are better thought of as a consensus statement on which applications are within the scope of PEM training and not what should form the core content of training.14,15

The P2Network (www.p2network.com) was formed in 2014 by PEM POCUS leaders to further PEM POCUS collaboration. P2Network’s main objectives are to collaborate in the areas of PEM POCUS education, administration, research, and mentorship. The need to arrive at more standardized curricula for PEM fellows and PEM POCUS fellows was recognized as top priorities to better define the skill set, training expectations, and expectations for PEM POCUS performance in practice.

In this study, we set out to establish consensus guidelines for core applications to include in a POCUS curriculum for PEM fellows by leveraging a large diverse group of experts from within and outside the P2Network using a modified Delphi survey.

PATIENTS AND METHODS

The research protocol was reviewed and approved by the Children’s Hospital of Eastern Ontario Research Ethics Board.

Study Design

A modified Delphi method was used to reach consensus core applications. This method involves two or more iterations of a data collection instrument, in this case an electronic survey, to a group of stakeholders to reach consensus.16,17 In designing this study consensus was defined to be reached when ≥80% of respondents agreed on a given item. This level of agreement has been used in previous studies, including ultrasound studies.16–20 Between each stage, data was analyzed and then presented in an anonymized fashion in the subsequent iteration to help influence decision making. The need for future iterations was based on whether there were significant changes in ratings on nonconsensus items between iterations.

Participants

Experts from the P2Network and the greater PEM POCUS community were invited to participate in this study. Participants were considered experts if they had performed at least 1,000 pediatric POCUS scans and met at least one of the following criteria: 3 or more years of experience in teaching POCUS to PEM fellows, PEM POCUS directorship, or leadership at an academic or tertiary pediatric emergency department or PEM POCUS fellowship trained.

At the time the survey was conducted, the P2Network had 87 members. To reach as many experts as possible, we identified PEM POCUS leaders who were not members of the P2Network from previous publications, listservs, and personal relationships. This subject pool was deemed sufficient as the acceptable
number of participants to reach a consensus using the Delphi technique ranges from four to 200.21–25

Survey
The Delphi survey was developed by the investigators, all of whom are local, national, and/or international experts in PEM POCUS. All known PEM POCUS applications were identified by reviewing the current literature and expert opinion of the study team and included in the survey.14,15 All members of the study team reviewed and edited each iteration of the survey and pilot tested it to identify errors and points of confusion and ensure ease of administration. The authors deliberately broke down protocols to their individual applications, as in the case of the extended focused assessment with sonography in trauma (E-FAST), which incorporates abdominal, cardiac, and pulmonary components, with the understanding that educators could teach these together, but each application has value in being able to be performed and understood independent of the other applications.

First Iteration
The round 1 survey was sent to all P2Network members by e-mail. Experts who were not part of the P2Network were e-mailed directly to invite them to participate. A unique identifier allowed each participant to access their survey. The first series of questions assessed the respondent’s eligibility and did not allow the participant to proceed if they did not meet the expert eligibility criteria. Baseline demographics including level of training, institution type, and educational structures were collected in this iteration of the survey. Following this, participants were asked to rate each ultrasound application based on their perceived level of importance for PEM fellows. Importance was ranked on a Likert-type scale ranging from 1 to 4 (1 = not important to 4 = extremely important) as has been used in previous Delphi studies.26–29 Participants could add free-text comments including additional applications, justification of choices, and requests for clarification. The round 1 survey is attached included in 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/aet2.10332/full).

Further Iterations
A second iteration was deemed necessary after many items did not reach consensus for inclusion or exclusion. This survey was distributed only to those who completed the first iteration. POCUS applications that did not reach 80% agreement were included in the second survey. Round 1 results for each POCUS application, in the form of means, were provided to the respondents. Respondents rerated each application’s level of importance, as with the first iteration. Additionally, seven applications were added to round 2 based on round 1 free-text comments. In round 2, respondents were asked, in yes or no format, whether they felt that these new items should be included in a future round and if the majority of respondents agreed they should, a third round would be triggered.

Data Collection and Analysis
REDCap (v6.14.1, Vanderbilt University), a secure Web application for building and managing online research surveys and databases, was used to distribute the surveys and collect responses. Ratings were analyzed using SPSS (v22, IBM Corp.), providing a mean frequency for each POCUS application.

RESULTS
In round 1 a total of 62 of 87 (71%) P2Network members and four of five (80%) non-P2Network members responded to the e-mail survey. Of those respondents, 48 met expert criteria and completed round 1 (44 P2Network members, four non-P2Network members). Thirty-seven of the 48 respondents from round 1 (77%) completed round 2 (33 P2Network, four non-P2Network members). The 11 participants who completed round 1 who did not complete round 2 were contacted multiple times by e-mail to solicit completion of the second round but were lost to follow-up.

The majority of responders were working in the United States and Canada at the time of study (Table 1). The expert panel was primarily made up of physicians who completed a pediatric residency followed by a PEM fellowship. More than half of the respondents completed a formal PEM or emergency medicine POCUS fellowship.

Consensus for inclusion was reached if 80% of responses fell within the “extremely important” or “very important” categories. After two Delphi rounds, consensus was reached to include 21 of the 60 possible PEM POCUS applications (Table 2). The recommended applications to include in a POCUS curriculum for PEM fellowship are listed in Table 3. Consensus for exclusion was determined if 80% of
responses fell within the “not important” or “somewhat important” categories. After two Delphi rounds, consensus was reached to exclude seven of the 60 possible PEM POCUS applications (Table 2). The recommended applications to exclude from the PEM fellowship curriculum are listed in Table 4. There was no expert consensus on the remaining 32 applications (Table 5).

In round 1 experts suggested seven applications, via free text, to be added to the Delphi process including: identify severe cardiac valvular disease, assess congenital neck abnormalities, identify corneal foreign bodies, identify ocular drusen, identify bowel obstruction, identify hydrocele, and identify varicocele. Experts were asked in round 2 whether these should be added to future rounds and none received greater than 50% yes votes to trigger a third round. The Delphi process ended after two rounds since no significant changes were observed between the first two rounds on applications that had not achieved consensus.

**DISCUSSION**

In this study we used a modified Delphi technique to determine which POCUS applications should be taught to PEM fellows. PEM POCUS experts identified 21 applications that should be considered core applications and included in all PEM curricula. These core applications include traumatic applications (identification of free peritoneal fluid, pericardial effusion, hemothorax, and pneumothorax), focused cardiac assessment (identification of nontraumatic pericardial effusion and cardiac standstill and evaluation of global cardiac function), soft tissue applications (identification of cellulitis, abscess, and foreign bodies), pulmonary applications (identification of pleural fluid and effusions, lung consolidation, and pulmonary edema), other diagnostic applications (identifying first trimester pregnancy, diagnosing intussusception, and assessing bladder volume), and ultrasound-guided procedures (central and peripheral vascular access, abscess incision and drainage, foreign body localization and removal, and pericardiocentesis). Participants were asked only to rate applications and not items related to ultrasound physics, image generation, common artifacts, and knobology. We recommend that any

| Characteristic                              | No. (%)
|--------------------------------------------|--------
| Country                                    |        
| United States                              | 33 (69)
| Canada                                     | 12 (25)
| Other*                                     | 3 (6)  
| Years of practice outside training, median (IQR) | 5.5 (6.25)
| Postgraduate training (selected all that applied) |        
| Pediatrics                                 | 41 (85)
| Emergency medicine                         | 5 (10) 
| PEM                                        | 43 (90) 
| Other                                      | 1 (2) 
| Ultrasound training (selected all that applied) |        
| PEM POCUS fellowship                       | 17 (35)
| EM POCUS fellowship                        | 13 (27)
| PEM fellowship rotation                    | 12 (25)
| Longitudinal US training within PEM fellowship | 14 (29)
| Longitudinal US training within EM residency | 1 (2) 
| 1 or 2-day course                          | 20 (42)
| Self-trained                               | 15 (31)
| Other                                      | 6 (13) 
| RDMS/RDPS certified                        | 12 (25)
| Involved in ultrasound education of PEM/PEM POCUS fellows | 46 (96)
| Number of years, median (IQR)              | 4.8 (1.8-7.8)
| Fellowships available at expert’s institution |        
| PEM                                        | 42 (88)
| PEM POCUS                                 | 20 (42) 
| EM POCUS                                  | 16 (33) 

IQR = Interquartile range; PEM = pediatric emergency medicine; POCUS = point-of-care ultrasound; US = ultrasound.
*One respondent each from Brazil, Israel, and Jamaica.

| Characteristic                              | No. (%)
|--------------------------------------------|--------
| Country                                    |        
| United States                              | 33 (69)
| Canada                                     | 12 (25)
| Other*                                     | 3 (6)  
| Years of practice outside training, median (IQR) | 5.5 (6.25)
| Postgraduate training (selected all that applied) |        
| Pediatrics                                 | 41 (85)
| Emergency medicine                         | 5 (10) 
| PEM                                        | 43 (90) 
| Other                                      | 1 (2) 
| Ultrasound training (selected all that applied) |        
| PEM POCUS fellowship                       | 17 (35)
| EM POCUS fellowship                        | 13 (27)
| PEM fellowship rotation                    | 12 (25)
| Longitudinal US training within PEM fellowship | 14 (29)
| Longitudinal US training within EM residency | 1 (2) 
| 1 or 2-day course                          | 20 (42)
| Self-trained                               | 15 (31)
| Other                                      | 6 (13) 
| RDMS/RDPS certified                        | 12 (25)
| Involved in ultrasound education of PEM/PEM POCUS fellows | 46 (96)
| Number of years, median (IQR)              | 4.8 (1.8-7.8)
| Fellowships available at expert’s institution |        
| PEM                                        | 42 (88)
| PEM POCUS                                 | 20 (42) 
| EM POCUS                                  | 16 (33) 

IQR = Interquartile range; PEM = pediatric emergency medicine; POCUS = point-of-care ultrasound; US = ultrasound.
*One respondent each from Brazil, Israel, and Jamaica.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. (%)</th>
<th>No. (%)</th>
<th>Final Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include</td>
<td>18 (30)</td>
<td>3 (6.4)</td>
<td>21</td>
</tr>
<tr>
<td>Exclude</td>
<td>2 (3.3)</td>
<td>5 (10.6)</td>
<td>7</td>
</tr>
<tr>
<td>No consensus</td>
<td>40 (66.7)</td>
<td>39* (83)</td>
<td>39*</td>
</tr>
</tbody>
</table>

Data are reported as n (%).
POCUS = point-of-care ultrasound.
*Includes seven additional items added by respondents in round 1.
POCUS educational program include these components. Employing a modified Delphi method, as implemented in this study, to achieve consensus has several benefits, especially when performed electronically. It allows for a large group of people to come together to reach consensus in a cost-effective manner, without geographic limitations. It can be delivered anonymously to allow for pressure-free responses. The Delphi method is a surrogate for a face-to-face meeting where the opinions of others are freely known and can influence the decision making of others. Hence, it is common in the Delphi method to share results from previous iterations.

Reaching consensus on the core applications will hopefully lead to the incorporation of these recommendations by stakeholders such as local training programs, professional organizations, and training accreditation bodies. PEM POCUS educators are encouraged to structure their curricula to, at a minimum, include these applications. Institutions who are unable to meet these core objectives are encouraged to seek out ways to ensure that their fellows learn these applications whether by external training or via recruitment of physicians who are able to teach these applications. This consensus can serve as the baseline for what PEM fellows should learn during their PEM training and what can be expected of new graduates when they enter the workforce. Training programs are encouraged to go beyond this core set of applications based on local expertise and the future practice needs of their trainees. Applications that did not reach consensus might be impactful in the right hands, in the right setting, and at the right time.

This consensus does not direct educators how to teach the applications only what they should include in a POCUS curriculum for PEM fellows. Educational milestones should be created that mirror this consensus much like the Council of Emergency Medicine Residency Directors (CORD) has established for emergency medicine residency. Once these education milestones are developed, curricula can be developed by PEM POCUS educators to meet these objectives. Curricula should include didactic training in ultrasound concepts, recognizing normal and pathological anatomy, indications for use, how to incorporate into medical decision making, and hands-on training to assist in the development of the motor skills necessary for image acquisition. Ultrasound simulators and image data banks can be used to assist learners, especially for less frequent applications, in the development of pattern recognition necessary for POCUS use in PEM.

Other stakeholders that would benefit from this consensus include accreditation bodies and professional organizations. The ACGME, Royal College of Medical Physicians, and the Society of Critical Care Medicine, among others, have worked to develop educational milestones for PEM. This consensus can serve as the basis for ongoing education and training within these organizations.

Table 3
POCUS Applications to Include by Delphi Technique

<table>
<thead>
<tr>
<th>Application</th>
<th>No. of Experts Who Ranked “Extremely/Very Important” (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td></td>
</tr>
<tr>
<td>Identify free peritoneal fluid in trauma</td>
<td>100</td>
</tr>
<tr>
<td>Identify nontraumatic pericardial effusion</td>
<td>100</td>
</tr>
<tr>
<td>Identify pericardial effusion in trauma</td>
<td>100</td>
</tr>
<tr>
<td>Identify hemothorax</td>
<td>98</td>
</tr>
<tr>
<td>Identify pleural fluid/effusion</td>
<td>98</td>
</tr>
<tr>
<td>Identify pneumothorax</td>
<td>98</td>
</tr>
<tr>
<td>Identify cardiac standstill</td>
<td>96</td>
</tr>
<tr>
<td>Abscess incision and drainage</td>
<td>94</td>
</tr>
<tr>
<td>Identify abscess</td>
<td>94</td>
</tr>
<tr>
<td>Central line placement</td>
<td>91</td>
</tr>
<tr>
<td>Evaluate cardiac function</td>
<td>88</td>
</tr>
<tr>
<td>Identify cellulitis</td>
<td>88</td>
</tr>
<tr>
<td>Identify intussusception</td>
<td>87</td>
</tr>
<tr>
<td>Identify intrauterine pregnancy</td>
<td>85</td>
</tr>
<tr>
<td>Identify soft tissue foreign body</td>
<td>85</td>
</tr>
<tr>
<td>Assess bladder volume</td>
<td>83</td>
</tr>
<tr>
<td>Identify lung consolidation</td>
<td>83</td>
</tr>
<tr>
<td>Peripheral IV access</td>
<td>81</td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
</tr>
<tr>
<td>Foreign body localizations and removal</td>
<td>89</td>
</tr>
<tr>
<td>Identify pulmonary edema</td>
<td>84</td>
</tr>
<tr>
<td>Pericardiocentesis</td>
<td>84</td>
</tr>
</tbody>
</table>

POCUS = point-of-care ultrasound.

Table 4
POCUS Applications to Exclude by Delphi Technique

<table>
<thead>
<tr>
<th>Application</th>
<th>No. of Experts Who Ranked “Not/Somewhat Important” (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td></td>
</tr>
<tr>
<td>Identify abdominal aortic aneurysm</td>
<td>90</td>
</tr>
<tr>
<td>Identify myositis</td>
<td>83</td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
</tr>
<tr>
<td>Assess for ovarian torsion</td>
<td>92</td>
</tr>
<tr>
<td>Identify epididymoorchitis</td>
<td>84</td>
</tr>
<tr>
<td>Identify vitreous detachment</td>
<td>84</td>
</tr>
<tr>
<td>Identify vitreous hemorrhage</td>
<td>84</td>
</tr>
<tr>
<td>Identify testicular mass</td>
<td>81</td>
</tr>
</tbody>
</table>

POCUS = point-of-care ultrasound.
Physicians and Surgeons of Canada, and other similar accreditation bodies can include these applications as the POCUS objectives for PEM fellowship. Professional organizations such as the AAP, ACEP, and Canadian Association of Emergency Physicians can also consider incorporating this work into their POCUS position statements.

As the development of PEM POCUS education progresses, dedicated 1-year fellowships in PEM POCUS have been developed to offer higher education in the areas of program oversight, leadership, research, and education. Experts should clearly define via consensus the POCUS objectives for these PEM POCUS fellowships as this would help clearly define the scope of POCUS training required for PEM fellows and PEM POCUS fellows.

We expect our consensus recommendations to change over time as the PEM POCUS field grows and expertise increases. Physicians will be finding new applications for POCUS as ultrasound machine technology, price, and portability improve and these may warrant consideration for inclusion in curricula in the future. Further, POCUS teaching is permeating medical education and we expect the future PEM fellow to begin fellowship with a broader scope of POCUS knowledge and skills than their predecessors. As curriculum are introduced and users achieve greater competence and comfort, we expect some of the applications that at present did not reach consensus for inclusion will in the future. As such, it is our plan to repeat this consensus-building project in 5 to 7 years.

LIMITATIONS

Our study has several limitations. We did our best to identify leaders both inside and outside of the P2Network, and it is possible that we missed some. We expect future iterations to include more participants as the P2Network continues to grow and the amount of PEM POCUS users who meet expert criteria will grow. We will also continue to identify experts outside of the P2Network by surveying worldwide POCUS networks, scanning publications, reviewing conference speaker lists, and leveraging personal relationships. Our project only surveyed PEM POCUS experts using expert criteria to best standardize POCUS training for PEM fellows. We did not survey non-POCUS using PEM physicians or POCUS users in the general emergency medicine community and it is possible that our results would differ if they participated.

CONCLUSIONS

A modified Delphi method was used by the ultrasound collaborative P2Network to reach a consensus on 21 core applications for inclusion in a point-of-care ultrasound curriculum for pediatric emergency medicine fellowship. This consensus establishes a baseline for education and expectations of graduating fellows.
The authors acknowledge Dr. Nick Barrowman, Senior Statistician, Clinical Research Unit at the Children’s Hospital of Eastern Ontario, for his assistance with survey design and statistical analysis and Ms. Kendra Sikes, Research Coordinator, University of Louisville, for her assistance with the manuscript preparation.

References


Supporting Information
The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10332/full
Data Supplement S1. Delphi Round 1 PEM POCUS Curriculum Consensus Survey.
Comparison of the Standardized Video Interview and Interview Assessments of Professionalism and Interpersonal Communication Skills in Emergency Medicine

Laura R. Hopson, MD1, Michele L. Dorfsman, MD2, Jeremy Branzetti, MD3, Michael A. Gisondi, MD4, Danielle Hart, MD5, Jaime Jordan, MD6, James A. Cranford, PhD7, Sarah R. Williams, MD8, and Linda Regan, MD9

ABSTRACT

Objectives: The Association of American Medical Colleges Standardized Video Interview (SVI) was recently added as a component of emergency medicine (EM) residency applications to provide additional information about interpersonal communication skills (ICS) and knowledge of professionalism (PROF) behaviors. Our objective was to ascertain the correlation between the SVI and residency interviewer assessments of PROF and ICS. Secondary objectives included examination of 1) inter- and intrainstitutional assessments of ICS and PROF, 2) correlation of SVI scores with rank order list (ROL) positions, and 3) the potential influence of gender on interview day assessments.

Methods: We conducted an observational study using prospectively collected data from seven EM residency programs during 2017 and 2018 using a standardized instrument. Correlations between interview day PROF/ICS scores and the SVI were tested. A one-way analysis of variance was used to analyze the association of SVI and ROL position. Gender differences were assessed with independent-groups t-tests.

Results: A total of 1,264 interview-day encounters from 773 unique applicants resulted in 4,854 interviews conducted by 151 interviewers. Both PROF and ICS demonstrated a small positive correlation with the SVI score ($r = 0.16$ and $r = 0.17$, respectively). ROL position was associated with SVI score ($p < 0.001$), with mean SVI scores for top-, middle-, and bottom-third applicants being 20.9, 20.5, and 19.8, respectively. No group differences with gender were identified on assessments of PROF or ICS.

Conclusions: Interview assessments of PROF and ICS have a small, positive correlation with SVI scores. These residency selection tools may be measuring related, but not redundant, applicant characteristics. We did not identify gender differences in interview assessments.

From the 1Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI; 2Department of Emergency Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3Ronald O. Perelman Department of Emergency Medicine, New York University School of Medicine, New York, NY; 4Department of Emergency Medicine, Stanford University, Stanford, CA; 5Department of Emergency Medicine, University of Minnesota Medical School, St. Paul, MN; 6Department of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; 7Department of Psychiatry, University of Michigan, Ann Arbor, MI; 8Department of Emergency Medicine, Stanford University School of Medicine, Stanford, CA; and the 9Department of Emergency Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

Received January 4, 2019; revision received March 16, 2019; accepted March 23, 2019.

The authors have no relevant financial information or potential conflicts to disclose.

Supervising Editor: Anne Messman MD, FACEP.

Address for correspondence and reprints: Laura R. Hopson, MD; e-mail: Lhopson@med.umich.edu.

AEM EDUCATION AND TRAINING 2019;3:259–268

© 2019 by the Society for Academic Emergency Medicine
doi: 10.1002/aet2.10346 ISSN 2472-5390 259
Residency programs invest substantial resources into the high-stakes process of resident selection. These efforts aim to identify individuals with a high likelihood of success during training. Studies of residency applications show mixed results for predicting future performance including medical school grades, United States Medical Licensing Examination (USMLE) performance, and letters of recommendation.\textsuperscript{1–3} Success in residency training and beyond likely requires a mixture of cognitive and nontechnical skills. As defined by the Accreditation Council for Graduate Medical Education (ACGME), professionalism (PROF) requires a commitment to carrying out professional responsibilities and an adherence to ethical principles; interpersonal communication skills (ICS) require the effective exchange of information and collaboration with patients, their families, and health professionals.\textsuperscript{4} These elements are poorly represented in standard application materials; however, these may be crucial determinants of professional success for future trainees.\textsuperscript{5}

Despite the mixed data surrounding their predictive value for training outcomes, conclusions drawn from residency interviews are heavily weighted in the selection process.\textsuperscript{1,2} In emergency medicine (EM), assessments of communication skills and interactions with program personnel during interviews are among the most heavily weighted considerations when ranking applicants.\textsuperscript{5} There are important reasons for this focus. Deficits in PROF and ICS are exceedingly difficult and time-consuming to remediate during graduate medical education training.\textsuperscript{6–9} PROF and ICS impact patient care and have substantial real-world implications beyond the challenges of remediation including risk of future state medical board disciplinary action, medical errors, and malpractice suits.\textsuperscript{10–14} There is also a strong correlation between burnout and low PROF, especially in residents and early-career faculty.\textsuperscript{15} For all these reasons, PROF and ICS are critically important characteristics to measure in residency applicants.

In response to feedback from program directors about the lack of PROF and ICS data in residency applications, the Association of American Medical Colleges (AAMC) developed the Standardized Video Interview (SVI). Using definitions identical to those of the ACGME, the SVI attempts to provide information about PROF and ICS, through the use of behavioral and situational interview questions, allowing programs to utilize these data when deciding who to invite for an interview.\textsuperscript{16} The AAMC SVI became required for all applicants in EM for the 2017 to 2018 residency application season (2018 Match), and expansion to other specialties is under consideration.\textsuperscript{17} The SVI consists of six interactions scored by trained raters on a 5-point scale resulting in a numerical score between 6 and 30. The AAMC has ongoing research on the predictive value of the SVI for residency performance. There remain practical questions about how residency programs should utilize the SVI score for applicant invitation and selection purposes as the predictive value of the SVI for future performance is still an area of active research. Additionally, SVI field testing has been met with some controversy, so understanding its potential role in candidate selection is of crucial importance.\textsuperscript{18}

Through a multisite study with seven EM residency programs, we studied the correlation of the SVI and interviewer assessments of PROF and ICS conducted during standard residency interviews. The primary objective of this study was to assess the degree of correlation between the SVI and interviewer assessments of PROF and ICS, in an effort to determine whether these assessments contribute similar or different data to the residency selection process. A high correlation (generally defined as $r > 0.4$) would indicate that the SVI and interview assessments are measuring similar candidate attributes, while a low correlation (commonly defined as $r < 0.19$) would indicate that they are measuring different attributes. We hypothesized that we would observe a strong correlation. Secondary objectives included an examination of 1) inter- and intrainsitutional assessments of ICS and PROF, 2) correlation of SVI scores with rank order list (ROL) positions, and 3) finally, due to recent research, we wanted to define the potential influence of gender on interview day assessments for ICS and PROF.\textsuperscript{19–21}

**METHODS**

We conducted an observational study using data collected prospectively from seven EM residency programs and the applicants to these programs during the 2018 Match. The participating programs represent diverse geography and training settings, with representation from both the 36-and the 48-month training formats. Site characteristics are summarized in Table 1. Institutional review boards at each of the participating sites approved the study or deemed it exempt.

All residency applicants who completed an interview at any one of the seven programs during the 2018 Match season and had an SVI score available in the
electronic residency application service (ERAS) were eligible for inclusion. Applicants without a SVI score were excluded from the analysis. Study approval and manuscript review were obtained from the AAMC for use of the SVI data; however, authors retained control over the data and final manuscript.

Two assessment tools were utilized in this study: 1) applicant SVI scores available through ERAS and 2) interviewer assessments of applicant PROF and ICS behaviors including interview discussions and other interactions during on-site interviews. All seven participating sites utilized the same five-point scoring rubric in order to standardize interviewer assessments (Figure 1). Use of a previously validated tool was not possible, as there is no single validated assessment tool widely used. However, many programs are attempting to measure these domains during interviews. Similarly, we did not attempt to completely standardize the student interviews across all seven sites, as we wanted to reflect the usual process conducted by most residency programs. Our assessment scale was developed through an iterative approach that included review of existing assessments from participating programs, the generic SVI scoring algorithm, and group consensus from the study authors.17 Content validity was established through review of this assessment tool by expert educators who have extensive residency leadership and applicant interview experience as well as advanced medical education training. The assessment items were determined to represent the construct being evaluated, with behavioral characteristics that all program directors believed were critical to assess (Figure 1), and were considered similar constructs of PROF and ICS that are intended to be assessed with the SVI.17,22 Response process validity was addressed by 1) ensuring that each interviewer completed their applicant ratings prior to discussion of the applicants with other interviewers; 2) attempting to blind interviewers to the applicants’ SVI scores during the immediate preinterview file reviews by providing the files as PDF documents, blocking access to the SVI scores through ERAS, or requesting that program leadership (who may still have been able to access SVI scores through ERAS) not look at the SVI scores proximate to interview day; and 3) training interviewers to utilize the anchors to inform their scoring through the use of brief site-specific in-person as well as written methods prior to any interviews being conducted. Internal structure validity evidence was provided by assessing interinstitutional correlations between ICS and PROF scores. All sites assessed ICS and PROF during interview days prior to this study, and no other aspects of the interview day were changed during this process (e.g., number of interviews, timing of interviews, etc).

Data were merged using the AAMC identification number to allow cross-referencing of otherwise blinded

<table>
<thead>
<tr>
<th>Site</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training venue</td>
<td>Midwest suburban university</td>
<td>West urban university</td>
<td>Mid-Atlantic urban university</td>
<td>Northeast urban university</td>
<td>Midwest urban university</td>
<td>West suburban county</td>
<td>Midwest urban county</td>
<td>Total</td>
</tr>
<tr>
<td>Training duration</td>
<td>48 months</td>
<td>48 months</td>
<td>48 months</td>
<td>48 months</td>
<td>36 months</td>
<td>48 months</td>
<td>36 months</td>
<td>36 months</td>
</tr>
<tr>
<td>No. of PGY-1 positions</td>
<td>16</td>
<td>16</td>
<td>12</td>
<td>16</td>
<td>16</td>
<td>15</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>No. of applicants interviewed</td>
<td>203</td>
<td>132</td>
<td>196</td>
<td>171</td>
<td>177</td>
<td>229</td>
<td>156</td>
<td>156</td>
</tr>
<tr>
<td>% Female</td>
<td>43.3%</td>
<td>43.2%</td>
<td>40.3%</td>
<td>49.1%</td>
<td>33.3%</td>
<td>38.9%</td>
<td>41.0%</td>
<td>41.1%</td>
</tr>
<tr>
<td>Interviews/applicant</td>
<td>5</td>
<td>2–3</td>
<td>3</td>
<td>5</td>
<td>4–5</td>
<td>3–7</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Interview duration</td>
<td>15 and 30 min</td>
<td>20 min</td>
<td>20 min</td>
<td>20 min</td>
<td>20 min</td>
<td>10 and 20 min</td>
<td>20 min</td>
<td>20 min</td>
</tr>
<tr>
<td>Individual interviews conducted</td>
<td>1,015</td>
<td>360</td>
<td>588</td>
<td>839</td>
<td>787</td>
<td>849</td>
<td>416</td>
<td>4,854</td>
</tr>
<tr>
<td>Individual interviews/interviewer, mean (±SD)</td>
<td>31.7 (±41.8)</td>
<td>10.0 (±10.9)</td>
<td>23.5 (±12.3)</td>
<td>83.3 (±50.4)</td>
<td>52.5 (±53.1)</td>
<td>77.2 (±65.0)</td>
<td>18.9 (±33.4)</td>
<td>18.9 (±33.4)</td>
</tr>
<tr>
<td>Structured interviews</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>No. missing SVI (%)</td>
<td>2 (1.0%)</td>
<td>2 (1.5%)</td>
<td>3 (1.5%)</td>
<td>2 (1.2%)</td>
<td>5 (2.8%)</td>
<td>1 (2.2%)</td>
<td>1 (0.6%)</td>
<td>1 (0.6%)</td>
</tr>
</tbody>
</table>

SVI = Standardized Video Interview
data across sites. Apart from gender, all other individually identifying candidate information was removed. Collected information about each interviewer included administrative title, years of interview experience, and gender. We utilized data routinely collected as a part of residency applications including the applicant’s SVI score (possible range = 6–30), applicant interview day PROF and ICS scores (each scored on a separate five-point scale), and the applicant’s final position on the ROL (top third, middle third, lower third, or do not rank [DNR]).

All data were maintained in Excel 2016. Statistical analysis was performed with IBM SPSS Statistics for Windows, Version 25.0. We estimated zero-order Pearson’s correlations between interview day assessments of PROF and ICS, each using five-point anchored scales and the SVI score. Subject data for PROF and ICS were aggregated into a single rating by using the mean of all ratings. We conducted a multiple regression analysis of PROF and ICS as predictors of SVI to gain perspective on the degree of overlap between these assessments.

Analysis of the relationship between SVI score and ROL position used a one-way analysis of variance (ANOVA), with ROL position as the independent variable. Follow-up pairwise comparisons were performed with the Tukey HSD test when analyzing SVI scores for ROL position. Using the applicant and the interviewer as the unit of analysis, applicant gender differences, and interviewer assessment differences in PROF, ICS, and SVI scores were assessed with a series of independent-groups t-tests. We did not calculate inter-rater reliability among interviewers, as each interview is a distinct interaction and applicants may perform differently in each interview. For testing our primary objective, the unit of analysis was the applicant (n = 1,264). Tests for secondary objectives used the interviewer (n = 151) or the

---

**Figure 1.** ICS and PROF scoring tool for interview interactions, which was developed through group consensus after review of the literature. Expected standards would include local-level program judgment as well as performance at Level 1 for the communication and PROF ACGME milestones. ACGME = Accreditation Council for Graduate Medical Education; ICS = interpersonal communication skills; PROF = professionalism.

---

<table>
<thead>
<tr>
<th>Major deficits – 1</th>
<th>Minor deficits – 2</th>
<th>Appropriate – 3</th>
<th>Excellent – 4</th>
<th>Outstanding – 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not meet standards/ Abrasive, unclear, poor focus</td>
<td>Meets minimum standards/ May fail to be clear or focused at times, requiring redirection</td>
<td>Meets expected standards/ Effective in non-complex or stressful situations</td>
<td>Above expected standards/ Articulate and effective in both simple and most complex/stressful situations</td>
<td>Significantly above expected standards/ Clear, concise, poised with constant success across the spectrum of situations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Major deficits – 1</th>
<th>Minor deficits – 2</th>
<th>Appropriate – 3</th>
<th>Excellent – 4</th>
<th>Outstanding – 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not meet standards/ Concerning deficits</td>
<td>Meets minimum/ May have minor concerns, but can develop</td>
<td>Meets expected standards/ No concerns</td>
<td>Above expected standards/ Noted to have some advanced qualities</td>
<td>Significantly above expected standards/ Wise beyond expected experience</td>
</tr>
</tbody>
</table>

---

I. **Verbal /Communication Skills**

II. **Professionalism/Maturity**
interview (n = 4,854) as the unit of analysis where appropriate.

RESULTS

Tables 1 and 2 summarize the descriptive statistics by sites. A total of 1,264 applicants were interviewed by 151 interviewers across the seven participating EM programs. This resulted in a total of 4,854 interviews, across 773 unique applicants. Missing data points were very infrequent. In descending order of frequency, administrative positions included general faculty and fellows (49.7%), residents (16.6%), assistant/associate program directors (15.9%), student clerkship leadership (5.3%), other leadership including deans and chairs (5.3%), program directors (4.6%), and non-physicians (2.6%). Scores for interview day assessments of ICS and PROF ranged from a low of 1 to a high of 5. However, very few students (less than 0.2%) had a score of 1 on ICS or PROF, and less than 2.1% had a score of 5 on ICS or PROF. Interviewer role was statistically significantly associated with PROF scores (F(6,143) = 2.2, p = 0.04) and with ICS scores (F(6,143) = 2.5, p = 0.02). Generally, lower PROF and ICS scores were provided by program directors (mean = 3.6), and higher PROF and communication scores were provided by nonphysicians, residents, and “others” (means = 3.9–4.0). There were strong interinstitutional correlations between ICS and PROF scores, all of which were statistically significant (r range = 0.38 to 0.68, all p < 0.05). Full data are provided in 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/aem.210346/full). We calculated interinstitutional correlations using data from students who completed interviews at more than one institution. To achieve 80% power to detect a medium-to-large interinstitutional correlation of 0.40 at an alpha of 0.05, pairwise correlations with a sample size of 40 or higher were examined. Of the 42 possible interinstitutional correlations, 14 met this inclusion criterion. As shown below, the 14 interinstitutional correlations ranged from 0.38 to 0.68, and all were statistically significant at 0.05 (Table 3).

Results presented in Table 4 indicate that, for our sample of applicants, SVI scores ranged from 12 to 28 and the mean (±SD) AAMC SVI score was 20.4 (±2.8), which was statistically significantly higher than the mean (±SD) score of 19.1 (±3.1) reported for the AAMC SVI performance data in the 2018 Match.23 Table 5 also shows a strong positive correlation between interview day assessments of PROF and ICS (r = 0.81, p < 0.001), which was consistent across all interview sites (r range = 0.71 to 0.89, median r = 0.81). Both PROF and ICS were positively and statistically significantly correlated with SVI score, but the correlations were small in magnitude (r = 0.16; r values s by site ranged from 0.11 to 0.26, median r = 0.18) and 0.17 (r values by site ranged from 0.05 to 0.29, median r = 0.19), respectively. These patterns of correlation did not vary when examined by institution or by applicant gender (Data Supplement S1). Only two of the participating programs utilized

<table>
<thead>
<tr>
<th>Site</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of interviewers</td>
<td>32</td>
<td>36</td>
<td>25</td>
<td>10</td>
<td>15</td>
<td>11</td>
<td>22</td>
<td>151</td>
</tr>
<tr>
<td>Interviewer years experience, mean (±SD)</td>
<td>5.5 (±5.5)</td>
<td>7.7 (±9.5)</td>
<td>8.3 (±8.2)</td>
<td>5.2 (±8.1)</td>
<td>6.7 (±7.3)</td>
<td>10.4 (±10.2)</td>
<td>9.6 (±9.4)</td>
<td>7.5 (±8.3)</td>
</tr>
<tr>
<td>% Female</td>
<td>40.6%</td>
<td>44.4%</td>
<td>40.0%</td>
<td>60.0%</td>
<td>26.7%</td>
<td>36.4%</td>
<td>40.9%</td>
<td>41.1%</td>
</tr>
</tbody>
</table>
interviews that included predefined behaviorally based interview questions; in the other programs, questions asked were at the discretion of the individual interviewer. When examined by interview method, the two programs utilizing structured interviews did not differ from those with unstructured interviews. Interestingly, despite the small correlation between interview day PROF and ICS scores and SVI scores, there was very little statistical overlap between PROF and ICS scores and AAMC SVI scores when multiple regression analyses were conducted. Taken together, PROF and ICS scores were associated with a statistically significant squared multiple correlation (adjusted $R^2 = 0.028$, $p < 0.001$), indicating that both variables together were associated with 2.8% of the variance in SVI scores. Although statistically significant, these results indicate that over 97% of the variance in SVI scores was not associated with PROF and ICS scores from the interview day assessments.

Results from a one-way ANOVA ($n = 1,264$ applicants) showed that there was a statistically significant association between ROL position and SVI score ($F(2,1176) = 16.5$, $p < 0.001$). Follow-up pairwise comparisons found that mean SVI scores for top-, middle-, and bottom-third and DNR applicants were 20.9, 20.5, 19.8, and 19.8, respectively. All means were statistically significantly different from each ($p < 0.05$), except for the bottom-third and DNR groups ($p = 0.90$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>1. SVI</th>
<th>2. ICS</th>
<th>3. PROF</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SVI</td>
<td>—</td>
<td>20.4</td>
<td>2.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ICS score</td>
<td>0.17*</td>
<td>—</td>
<td>3.7</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>3. PROF score</td>
<td>0.16*</td>
<td>0.81*</td>
<td>—</td>
<td>3.7</td>
<td>0.6</td>
</tr>
</tbody>
</table>

ICS = interpersonal communication skills; PROF = professionalism; SVI = Standardized Video Interview.

Based on the applicant ($n = 1,264$) gender, a series of independent-groups t-tests showed no statistically significant differences between female and male applicants on assessments of 1) PROF (male mean $= 3.7$ vs. female mean $= 3.8$, $t(1262) = 0.6$, $p = 0.57$); 2) ICS (male mean $= 3.7$ vs. female mean $= 3.7$, $t(1262) = 0.8$, $p = 0.41$); or 3) SVI score (male mean $= 20.5$ vs. female mean $= 20.3$, $t(1243) = 1.1$, $p = 0.25$).

Results for interviewer ($n = 151$) gender influences showed that male interviewers gave statistically significantly higher scores than female interviewers on 1) PROF (male mean $= 3.9$ vs. female mean $= 3.8$, $t(147) = 2.1$, $p = 0.04$); and 2) ICS (male mean $= 3.9$ vs. female mean $= 3.7$, $t(147) = 2.1$, $p = 0.03$) across all of the institutions, but these gender differences were small in magnitude. Male interviewers had significantly more years of experience than female interviewers (male mean $= 9.3$ years vs. female mean $= 5.1$ years, $t(148) = 3.1$, $p = 0.003$). However, interviewers’ years of experience was not statistically significantly correlated with PROF scores ($r = 0.01$, $p = 0.85$) or ICS scores ($r = -0.06$, $p = 0.45$). No statistically significant effects of applicant–interviewer gender concordance on PROF or ICS scores were observed ($F < 1.0$, $p > 0.50$).

**DISCUSSION**

Our study presents data on comparisons between interview day assessments of residency applicants and the new SVI that demonstrates a small positive correlation between the SVI and interview day assessments of PROF and ICS. While SVI and interviewer assessments of PROF and ICS have some overlap, they are also potentially measuring separate domains. There are a number of potential reasons that could account for only finding a small degree of correlation between the two assessments. First, it is possible that residency
Interviewers’ assessments of PROF and ICS are fundamentally different than the manner in which the AAMC assesses these domains as measured by the SVI. Since interview day assessments often rate applicants in multiple domains in addition to PROF and ICS, it is possible that interviewers are concurrently accounting for some of these SVI subcompetencies in other domains or possibly not at all. This explanation would also be concordant with observations by Schnapp et al. who in a single-institution study showed no correlation between faculty global gestalt scores of PROF and ICS with the SVI. Second, the AAMC SVI utilizes behavioral and situational questions mapped to their clear definition of PROF and ICS. SVI examples include: “Imagine you are leading a multidisciplinary team composed of professionals with different areas of expertise. How should you make sure everyone works together effectively?” (ICS) and “One of your patients refuses treatment because it is incompatible with the patient’s religious beliefs. What should you do in this situation?” (PROF). Only two of our participating sites used structured or semistructured, interviews, and all incorporated broad topics into questions beyond just PROF and ICS questions, which may have contributed to the small degree of correlation with the SVI. Our anchors for interview day assessments of ICS and PROF encompass a broader skill set than what is measured in the AAMC SVI. For example, our ICS rating anchors focus on the applicant having focused, articulate, effective, and poised communication, while our PROF ratings are less specific and include aspects of maturity and wisdom. Third, while AAMC SVI scoring relies on raters trained to emphasize content more than delivery, our interviewers are likely measuring components of both, including fluency of communication and nonverbal behaviors. Thus, it seems that these two assessments are measuring different, but related, aspects of ICS and PROF. However, further study may be required to truly understand what interviewers are measuring when asked to assess these domains.

It is also likely that factors such personality, similarities with the interviewer, and consideration of previous interactions, such as by e-mail or during clerkships, may influence interview day scores despite the presence of anchors for scoring. In addition, interviewers are not blinded to other elements of the residency application, which may create halo or horn effects on PROF and ICS assessments from the interview itself. Thus, interviewer PROF and ICS scores may represent additional candidate factors that are not assessed in the SVI score. PROF can be a difficult domain to assess and measure, and our ability to do this effectively within a brief interview encounter may also be limited.

Of interest is the strong correlation ($r = 0.81$) between PROF and ICS assessments on our scoring instrument. This could suggest that raters, despite use of an anchored rating scale, have difficulty discriminating between behaviors that represent each domain and thus conceptualize them as a single entity. This observation is consistent with literature suggesting that ICS skills positively affect scoring in many other domains of the ACGME Core Competencies. It is also possible that this correlation is an accurate representation of high covariance between the domains. Longitudinal research will be needed to clarify the discriminant validity of the PROF and ICS scores.

Interinstitutional correlations of scores from applicants completing interviews at more than one institution were significantly correlated and provide validity evidence for the interview day assessment tool. We chose not to place weight on the comparison of PROF and ICS scores of the same applicant who interviewed at more than one site in our study. We felt that the same applicant might perform very differently at different interview sites on different days due to a variety of factors, including interest in the program, physical factors such as illness or sleep adequacy, and external factors or stressors. As such, we viewed each interview as a discreet encounter that is potentially not comparable across sites.

The statistically significant association between AAMC SVI scores and ROL position is not surprising given that the SVI is intended to evaluate skills that are highly valued by EM program directors. However, the scores for top-, middle-, and bottom-third applicants only revealed small absolute differences (0.4 and 0.7, respectively) and this small variation may not be meaningful in practice and is significant due to the large sample size. Interestingly, the SVI scores of the DNR applicants were identical to the lower-third group. We postulate that the DNR group is heterogeneous due to technical disqualifiers (e.g., lack of USMLE scores by rank list submission deadlines as required by the institutional selection policies, withdrawal by the applicant) as well as individuals with behavioral or academic concerns. In addition, ROL positioning, while it may emphasize academic traits, is subject to many influences.
the SVI does not appear to be a useful discriminator for DNR positioning, and the practical significance of the SVI score differences by ROL position may be an avenue for future research.

Recent research has demonstrated gender discrepancies in ACGME EM Milestone proficiency levels assessments, as well as within letters of recommendation and the medical school performance evaluation. We felt that it was important to understand whether gender bias could be a factor affecting interview day assessments, which, if present, could affect correlations with the SVI. We did not find any statistically significant differences in interview day assessment scoring related to applicant-interviewer gender concordance. We did find that male interviewers gave slightly higher ICS and PROF scores than female interviewers. However, this small difference in scores of 0.21 points and 0.12 points, respectively, while statistically significant, is likely not meaningful in a practical sense. It is possible that use of an anchored rating scale may provide one mechanism for residency programs to minimize gender bias within their applicant assessment processes. This process has been used successfully by the AAMC, which has found no evidence of gender bias in the AAMC SVI scores.

In our results, program directors generally gave lower scores on both scales and higher scores were provided by nonphysicians, residents, and interviewers in the “other” category. This association is not entirely surprising as prior medical literature has shown differences in ratings between assessor groups and that assessors’ interpretations are framed within their discipline, experience, and level of involvement with the learner. Program directors may be more critical of applicants, knowing that they will have the responsibility of managing any remediation issues. They are also more likely to interview the largest number of candidates and thus may have a broader sample of applicants across which to calibrate their ratings. These differences in assessments, however, do not necessarily reflect bias or mean that one is more or less accurate than another. Literature suggests that differing assessments, as long as raters possess the skills and expertise to accurately judge the construct of interest, represent distinct but equally valid perspectives.

We did not find any statistically significant association between interviewer years of experience and PROF and ICS scores, suggesting that duration of experience is not the sole factor required to accurately assess these skills.

LIMITATIONS

There are many potential sources of bias inherent to the interview process for which we could not control. Each site was free to conduct interviews per their normal process. The study protocol did not include scripted questions or formats; the only commonality was the PROF and ICS assessment tool. This design was utilized so that our study would represent “real-world” assessments rather than an artificial idealized interview day state, with the goal of increasing the generalizability and external validity of the results as opposed to limiting them. This does mean that applicants were not all necessarily put in complex or stressful situations during all interviews. We also did not provide extensive rater training, as was done for the SVI. While additional training of interviewers could further standardize this process, we again elected not to do this in an attempt to represent real-world interview day assessments. Terms such as “minimum standards” and “expected standards” were not explicitly defined for the interviewers during the training process and were left up to individual interviewers’ discretion when rating applicants although reference to Level 1 of the ACGME milestones was an implicit part of the concept. Further standardization would require a substantial investment of resources and could include standard setting exercises with a variety of in-person or recorded interview interactions.

Although the scoring instrument was created iteratively by the group in an effort to enhance validity, it did not undergo any formal piloting prior to implementation in actual interviews. In the experience of the authors, this is in line with standard practice of residency interview scoring; the tools used are generally based on content validity without further validity testing.

Our study population was preselected from review of ERAS application materials; this cohort may have different abilities than those not invited to interview. This may have impacted our overall data and its resultant correlations. For example, we observed a restricted score range of PROF and ICS assessments which likely attenuated our ability to identify true correlations that may be present in a study population that was not preselected.

The participating sites included five 4-year programs and two 3-year programs. While we found no differences between sites, it is possible that our skewed sample may make these findings more applicable to 4-year programs. Only two of the seven programs used semistructured interviews. It is unclear whether this
percentage is reflective of the overall EM community, and these results may be more applicable to programs using unstructured interviews.

CONCLUSIONS

In this multicenter study aiming to determine whether the Standardized Video Interview and usual interview day assessments of professionalism and interpersonal communication skills contribute similar or different data to the residency selection process, we found that interview day assessments using a novel tool have only a small, positive correlation with Association of American Medical Colleges Standardized Video Interview scores. It is therefore likely that both assessments provide meaningful, distinct information. For secondary objectives, there was strong interinstitutional correlation between interview day professionalism and interpersonal communication skills scores; a small but statistically significant correlation between Standardized Video Interview and rank order list position across institutions and no gender influences on interview day scores. However, the difference between a top-third candidate and a do-not-rank candidate was minimal, with only a 1.1 point Standardized Video Interview score difference. Similarly, the Standardized Video Interview could not distinguish between a bottom-third and a do-not-rank candidate. Further study is required to examine the predictive ability of both the Standardized Video Interview and well-designed interview day assessments on future clinical performance.

References


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10346/full

Data Supplement S1. Pearson correlations between SVI, ICS, and PROF scores across sites.
Building Emergency Medicine Trainee Competency in Pediatric Musculoskeletal Radiograph Interpretation: A Multicenter Prospective Cohort Study

Michelle Sin Lee, MDCM,1 Martin Pusic, MD, PhD,2 Benoit Carrière, MD, MHPE,3 Andrew Dixon, MD,4 Jennifer Stimec, MD,5 and Kathy Boutis, MD, MSc1

ABSTRACT

Objectives: As residency programs transition from time- to performance-based competency standards, validated tools are needed to measure performance-based learning outcomes and studies are required to characterize the learning experience for residents. Since pediatric musculoskeletal (MSK) radiograph interpretation can be challenging for emergency medicine trainees, we introduced Web-based pediatric MSK radiograph learning system with performance endpoints into pediatric emergency medicine (PEM) fellowships and determined the feasibility and effectiveness of implementing this intervention.

Methods: This was a multicenter prospective cohort study conducted over 12 months. The course offered 2,100 pediatric MSK radiographs organized into seven body regions. PEM fellows diagnosed each case and received feedback after each interpretation. Participants completed cases until they achieved a performance benchmark of at least 80% accuracy, sensitivity, and specificity. The main outcome measure was the median number of cases completed by participants to achieve the performance benchmark.

Results: Fifty PEM fellows from nine programs in the US and Canada participated. There were 301 of 350 (86%) modules started and 250 of 350 (71%) completed to the predefined performance benchmark during the study period. The median (interquartile range [IQR]) number of cases to performance benchmark per participant was 78 (60–104; min = 56, max = 1,333). Between modules, the median number of cases to achieve the performance benchmark was different for the ankle versus other modules (ankle 366 vs. other 76; difference = 290, 95% confidence interval [CI] = 245 to 335). The performance benchmark was achieved for 90.7% of participants in all modules except the ankle/foot, where 34.9% achieved this goal (difference = 55.8%, 95% CI = 45.3 to 66.3). The mean (95% CI) change in accuracy, sensitivity, and specificity from baseline to performance benchmark was +14.6% (13.4 to 15.8), +16.5% (14.8 to 18.1), and +12.6% (10.7 to 14.5), respectively. Median (IQR) time on each case was 31.0 (21.0–45.3) seconds.

Conclusions: Most participants completed the modules to the performance benchmark within 1 hour and demonstrated significant skill improvement. Further, there was a large variation in the number of cases completed to achieve the performance endpoint in any given module, and this impacted the feasibility of completing specific modules.

From the 1Division of Pediatric Emergency Medicine, Department of Pediatrics, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada; the 2Department of Emergency Medicine and Division of Learning Analytics at the NYU School of Medicine, New York, NY; 3Division of Emergency Medicine, CHU Sainte-Justine and Université de Montréal, Montreal, Quebec, Canada; the 4Division of Pediatric Emergency Medicine, Department of Pediatrics, Stollery Children’s Hospital and University of Alberta, Edmonton, Alberta, Canada; and the 5Department of Diagnostic Imaging, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada.

Received October 14, 2018; revision received January 11, 2019; accepted February 4, 2019.

Funded by the Royal College of Physicians and Surgeons: There was no role of the funder in the study design, data collection, analysis, interpretation of the data, writing of the report, or decision to submit the article for publication.

The course implemented in this study operates as a nonprofit under the academic umbrellas of the Hospital for Sick Children and University of Toronto. Although there was a fee for the course ($150 per participant), none of the authors nor content advisors get financial compensation at any point for their work related to this course. The fee is used to pay for operational expenses for the course as these funds are not available from...
Over 20 million diagnostic films of pediatric extremities are done annually in the United States in emergency departments (EDs).1 Emergency medicine physicians are tasked with real-time interpretations of these images and need to incorporate their diagnostic impressions into immediate decision making at the bedside. An appropriate initial diagnosis decides management, minimizing morbidity and long-term dysfunction. Thus, it is imperative for training programs to ensure that emergency medicine trainees develop competence in the skill of pediatric musculoskeletal (MSK) radiograph interpretation before they make high-stakes decisions with patients in independent practice.2,3

Unfortunately, the interpretation of pediatric MSK radiographs has been shown to be relatively deficient among graduating pediatric and emergency medicine residents who will be working in EDs.4–8 These training results echo what has been noted in clinical practice, where misinterpretation of pediatric MSK radiographs in general and pediatric EDs has been estimated to be up to 19%,1,9–11 and pediatric extremity radiographs have been found to be one of the most common causes of discrepant radiograph reports.12,13 As many as 86% of these misinterpretations require a change in management,14 and in the United States, misdiagnoses of pediatric fractures has accounted for the third largest amount of dollars paid out to settle malpractice claims.15 Research in this area has called for effective education solutions to reduce radiograph interpretation errors at the bedside.9,16–18

To bridge the knowledge–practice gap in pediatric MSK radiograph interpretation among graduating emergency medicine physicians, we developed an online education system that allows active learning of this skill using 2,100 authentic pediatric MSK images that represent seven regions of the pediatric MSK system (https://imagesim.research.sickkids.ca/demo/msk/).19–24 With this platform, the presentation of images is simulated to mirror how clinicians interpret them in the clinical field and presented in large numbers so that learners can learn similarities and differences between diagnoses, identify weaknesses, and build up a global representation of possible diagnoses.19,21,22 Specifically, cases are presented with a brief clinical stem, standard images and views, and juxtaposition of normal and abnormal cases.19,23 There are also hundreds of cases to review and after each case the system provides visual and text feedback, which allows for deliberate practice25 and an ongoing measure of performance as part of the instructional strategy.26,27 Furthermore, in line with the unfolding competency-based residency framework, which promotes greater accountability and documentation of actual capability,28–31 this platform also requires learners to complete cases until they reach a predefined performance level.

Despite these aforementioned strengths, the feasibility of embedding this type of learning system into a postgraduate training program needs to be studied rather than assumed. That is, it is equally important to ensure that the course will be completed when it is embedded longitudinally within a training program that has multiple competing learning requirements. Using Bowen’s framework of how to conduct feasibility studies,32 we focused on the elements of integration, adaptation, and expansion.

We introduced a Web-based pediatric MSK radiograph learning system with a performance benchmark into pediatric emergency medicine (PEM) fellowships and aimed to establish the feasibility and effectiveness of implementing this intervention. Specifically, we examined the number of cases needed to achieve the performance threshold, the proportion of participants who successfully achieved the performance standard, and skill gains from baseline to performance standard.

METHODS

Study Design and Setting
This was a prospective cohort multicenter study. The education intervention was developed at Hospital for Sick Children (Toronto, Ontario, Canada) and New York University (New York, NY). The education intervention was implemented July 2015 to June 2016 at nine credentialed PEM fellowship programs in Canada (n = 7) and United States (n = 2).

Selection of Participants
The participants were first-year PEM fellows that completed a pediatrics residency. Sites that embedded the education intervention into their core PEM fellowship the institutions where this course was developed. Three of the authors were involved in the development of the education intervention (JS, MP, KB) and one author (KB) is the academic director of the course (www.imagesim.com).

Supervising Editor: Jaime Jordan, MD.
Address for correspondence and reprints: Dr. Kathy Boutis; e-mail kathy.boutis@sickkids.ca.
AEM EDUCATION AND TRAINING 2019;3:269–279
learning goals included the Hospital for Sick Children, New York University, Columbia University, British Columbia Children’s Hospital, Stollery Children’s Hospital, Alberta Children’s Hospital, CHU Sainte Justine, Children’s Hospital of Eastern Ontario, and Montreal Children’s Hospital. Once the respective program director committed to the education experience, the respective PEM fellows were provided unique username/password access to the learning system. The website outlined that engaging in the education intervention included the possibility that de-identified participant data may be used for research purposes. This study was approved by the institutional review boards of the Hospital for Sick Children and New York University.

Education Intervention Radiograph Collection. We collected the radiographs by purposefully sampling from a pediatric emergency clinical setting. Specifically, images were identified by reviewing a tertiary care pediatric emergency diagnostic imaging database for pediatric MSK radiographs ordered from the ED from January 1, 2012, to December 31, 2014, in which the indication for the radiograph was exclusion of a fracture/dislocation. Radiographs were then downloaded by research assistants from the institutional picture archiving and communication system in JPEG format along with the respective final attending pediatric radiologist’s report. One study team member (KB) reviewed all the images and excluded cases that had markers like casting material or embedded arrows that would suggest a diagnosis, very poor quality films such that radiograph findings were obscured, or those with an incomplete number of views. Cases with an uncertain final diagnosis were additionally reviewed with the collaborating pediatric MSK radiologist (JS) to establish a final diagnosis, and those that remained uncertain were excluded. In a separate diagnosis specific search of the archiving system that included the years 2010 to 2014, rarer but clinically important diagnoses (e.g., Monteggia or scaphoid fractures) were identified to ensure that there were a few examples of these injuries. For each case, a brief clinical history was written based on available clinical data and radiographs were categorized depending on whether there was the presence or absence of a fracture/dislocation. Cases with a fracture/dislocation were further subclassified by diagnosis and the location of the abnormality on the image. This resulted in a pool of 2,985 radiographs from which 885 normal cases were excluded to result in 2,100 radiographs with a case-mix frequency of 50% normal and 50% abnormal cases.19,23

Online Software Application for Presentation of Radiograph Cases. An online learning platform was previously developed using HTML, PHP, and Flash and a description of this can be found elsewhere.20 In brief, the case experience is organized into seven regions of the pediatric MSK system, each of which contains 200 to 400 case examples: 1) skull, 2) shoulder/clavicle/humerus, 3) elbow, 4) wrist/forearm/hand, 5) pelvis/femur, 6) knee/tibia-fibula, and 7) ankle/foot. The participant reviews all relevant images of a case and commits to a response (fracture/dislocation absent/present) and is provided with immediate text and visual feedback on their interpretation (Figure 1). Once the participant has considered this information, they moved on to the next case and continued doing cases until a performance-based standard of 80% accuracy, sensitivity, and specificity was achieved on the most recently completed 25 cases. In our prior research,22 we found that 25 cases provided stable estimates of learner performance in these metrics. The software tracked participant progress through the cases and recorded responses to a mySQL database.

Performance Benchmark. In the absence of an evidence based performance standard that confirms competency, we considered several factors to select the 80% performance benchmark. The goal was not necessarily to achieve mastery during a fellowship program, which often requires several years of practice with feedback postgraduation. We also considered that beside exposure in the ED, radiology, and orthopedic rotations still had an important role in increasing image interpretation skills for a common clinical problem. Importantly, while diagnostic errors do occur, the vast majority are subtle fractures and are not of high clinical consequences.11 Further, most EDs have radiology review of images available within 24 hours, and as such errors are captured and reported. Finally, we reviewed performance achieved for this participant group in our prior studies to understand what was feasible for most fellows.20,22 Based on these principles, our aim was to expose participants to a large number of cases that would likely take years to acquire via bedside emergency medicine practice alone toward the goal of an “acceptable and feasible” performance benchmark in a setting where additional education on this topic was ongoing at the bedside and quality...
assurance interventions would capture diagnostic errors.

**Course Fees.** There was a $150 Canadian fee per participant for 1 year of access to the cases (www.imagesim.com). The education system operates as a non-profit under the academic umbrellas of the Hospital for Sick Children and University of Toronto. Fees were used to pay for operational expenses. Neither the authors nor the content advisors were paid for their work on this education intervention or research.

**Study Protocol**

**Recruitment.** This learning platform was introduced at a national Canadian PEM meeting (Pediatric Emergency Research Canada) in January 2015. Program directors from PEM fellowships interested in participating enrolled their fellow participants in the education intervention.

**Participant Engagement.** Upon enrollment, PEM fellow participants received a brief 10-slide presentation on the course reviewing the educational theory and goals of course participation. Fellows participated by using a computer of their choice and having online 24/7 access to the cases. Secure entry was ensured via a participant name and password given to each participant. After the fellow accessed the learning system, each participant was given some general information, which included assurance of confidentiality, the purpose of the exercise, and some information on how to use the system. The following participant demographics were captured: country of participation (Canada or United States) and sex (male or female). Fellows were not provided with any information about the proportion of normal to abnormal cases or types of pathology. No time limitation per case interpretation was imposed and each participant selected the order in which to complete a given module. As described, the participant then completed cases until the performance benchmark of at least 80% accuracy, sensitivity, and specificity was achieved.

**Course Integration Into PEM Fellowship.** Prior to course initiation at a given site, each program

---

**Figure 1.** Visual and text feedback after case interpretation.
director met with the academic director of the pediatric MSK course to review strategies for integrating this course into the curriculum. Program directors were asked to consider if there were specific rotations (e.g., orthopedics or radiology) during which their fellows should be focusing on these modules. Program directors were also encouraged to schedule group sessions at a computer lab every 2 months to provide structure and timelines to complete the modules over the 12-month period. Finally, each PEM fellowship director was provided access to a dashboard that tracked each participant’s progress and was encouraged to review the status of completion of modules with the PEM fellows at routine semiannual reviews. To our knowledge, programs did not incentivize fellows to complete the modules nor punish them for lack of completion.

**Adaptation and Course Feedback.** Feedback was formally solicited from program directors and PEM fellow participants. Every 2 months, the study team contacted the PEM program directors and highlighted their site’s participant progress and asked for feedback from the perspective of a program director or PEM fellow participant. When PEM fellow participant challenges were identified, the study team worked with the program director to focus on strategies that may be more successful at their site. Further, the online learning system allowed for each participant to comment on every case they reviewed and provide general comments. Any errors in case details were corrected every 2 months. Participants also had access to technical support 15 hours per day, 7 days per week. Any challenges were typically resolved within 24 hours.

**Outcome Measures**
The primary outcome was the median number of cases completed to reach the predefined performance benchmark overall and per module. Secondary outcomes included the proportion of participating fellows that started the education intervention and completed at least 25 cases and the proportion who achieved the performance benchmark in a given module. We also measured the median time in seconds required to complete a case and compared this between modules. Further, we determined the mean change in accuracy, sensitivity, and specificity from baseline to performance benchmark overall and independently for each of the seven modules and compared performance gains across different modules. Finally, we reviewed feedback provided by program directors and participants and used this information to better understand outcomes and consider how we could improve the experience in real time and for future implementations.

**Data Analyses**

**Unit of Analysis.** Each case completed by a participant was considered one item. Normal items were scored dichotomously depending on the match between the participant’s response and the reference standard diagnosis. Abnormal items were scored correct if the participant had both classified it as abnormal and indicated the correct region of abnormality on at least one of the images of the case. Participant data were included only if they completed a minimum of 25 cases in a given module.

**Number of Cases Completed.** We calculated the median number of cases to the performance benchmark with respective interquartile ranges (IQRs) for all modules and for each module (primary analysis). We also presented graphic data demonstrating cases completed overall, distinguishing which participants achieved the performance benchmark and which did not. The Kruskal-Wallis test was used to compare for differences between participants.

**Participants Who Completed Modules.** Proportions were reported with respective 95% confidence intervals (CIs). The Kruskal-Wallis test was used to compare for differences between modules.

**Median Time on Case.** We calculated the median time on case in seconds for all participants who started the intervention overall and per module. The Kruskal-Wallis test was used to compare for differences between modules.

**Performance Increase.** Baseline accuracy, sensitivity, and specificity scores were calculated after the initial 25 cases were completed, while final score was determined based on an average of the terminal 25 cases. Comparisons of within-subject performance changes were performed using the paired Student’s t-test. Analysis of variance was used to compare performance differences between multiple groups and post hoc analyses were performed adjusted using the Bonferroni’s test.

Significance was set at $p < 0.05$, but for post hoc analyses was set at $<0.01$ to adjust for multiple testing.
All analyses were conducted using SPSS software analysis package (version 23).

RESULTS

The program was provided to 59 first-year PEM fellows. Of these, 50 (84.7%) PEM fellows started the assigned modules, 45 (90.0%) of whom were in a Canadian fellowship and 31 (62.0%) were female.

Across all modules, the median (IQR) number of cases required to achieve the performance benchmark (primary outcome) for participants was 78 (60–104), with respective minimum and maximum values of 56 and 1,333. Of the participants, 70.8% completed 100 cases or less to achieve the performance target, while 92.7% achieved this goal in 200 cases or less (Figure 2). Between participants there was a significant variation in number of cases required to achieve the performance benchmark for a given module (p = 0.03). Between modules, the median (IQR) number of cases required to achieve the performance benchmark differed between modules (p < 0.0001), Table 1. This comparison between the ankle/foot module and other modules demonstrated a difference of 290 cases (95% CI = 245 to 335) completed.

Of the 301 of 350 (86.0%) modules that were started, participants reached the performance benchmark in 249 to 350 (71.1%) modules. Four of the nine (44.4%) sites scheduled sessions at a computer laboratory at regular intervals during the study period. At these sites, 79 of 97 (81.4%) of modules were completed to the performance benchmark, while this was true for 170 of 204 (83.3%) of modules at sites that did not schedule group computer sessions (difference = –1.9%, 95% CI = –6.6 to +11.8). The performance benchmark was achieved for 90.7% of participants in all modules except the ankle/foot, where 34.9% achieved this skill level (difference = 55.8, 95% CI = 45.3 to 66.3). Median (IQR) time on each case was 31.0 (21.0–45.3) seconds. Median time on case differed between modules (p < 0.0001; Table 1). Post hoc analyses demonstrated that time on case was greater for the ankle relative to all other modules (difference = +16.9, 95% CI = +10.1 to +23.6).

The overall change in accuracy from baseline to the performance benchmark was +14.6% (95% CI = 13.4 to 15.8), with the corresponding Cohen’s effect size of 1.8 (95% CI = 1.1 to 2.5). The respective overall change in sensitivity and specificity were +16.5% (95% CI = 14.8 to 18.2) and +12.6% (95% CI = 10.7 to 14.5); increase in sensitivity relative to specificity was 2.8% (95% CI = 0.4 to 5.2). The per module mean changes in accuracy, sensitivity, and specificity are detailed in Table 2. Overall, the pre–post changes in accuracy and sensitivity were different between modules (p < 0.0001) but not for specificity (p = 0.01).

DISCUSSION

This study demonstrated that the implementation of an online pediatric MSK image interpretation learning system with a performance benchmark in PEM fellowships resulted in significant increases in PEM fellow interpretation skill. Further, most of these PEM fellow participants completed the seven modules to a preset performance target. However, there was a large variation in the number of cases required to achieve the performance goal between individual participants and between different types of modules, and this impacted module completion for some participants, particularly in the ankle/foot case set.

The finding that the number of cases required to achieve a performance endpoint for a given participant was variable demonstrates that individual learners reach their milestones at varying speeds, and this learning model affords them the flexibility of how they learn while providing guidance on when they have mastered a particular skill.29 Despite this variation, about 70% of participants were still able to achieve performance benchmark in a feasible time frame, about 1 hour. In contrast, some participants needed to complete several hundred cases over several hours before they reached the performance target. This was particularly true for the ankle/foot case set, whereby only about one-third of the participants achieved the 80% performance goal. It is difficult to know why this set posed additional challenges relative to other modules. Feedback identified from comments left on the system stated that many of the fractures were small and difficult to see (e.g., distal fibular fractures, subtle foot fractures), resulting in low performance and the need to do more cases to improve scores. The latter may have led to fatigue and decreased motivation for many participants. Future iterations of this module may consider reducing the number of minor foot/ankle fractures of little clinical significance or incorporating adaptive learning algorithms to refine the case presentation such that the system selectively presents cases most relevant to an individual participant’s
Figure 2. Participants are ranked left to right in ascending order of how many cases they completed. Each bar is one individual. Gray bars are participants who successfully achieved the performance threshold. White bars are participants who did not achieve the performance threshold. The x-scales differ as not all body regions have the same number of participants. The y-scale is the same for all graphs except the bottom-most where it is adjusted for the substantially greater number of cases completed for the ankle/foot module.
weaknesses. As emergency medicine transitions to competency-based residency training programs, educators will have to consider how to balance ideal performance thresholds in specific areas that demonstrate a high degree of individual variability with ensuring completion of an entire emergency medicine curriculum in a reasonable time frame for all their trainees. Ultimately, it is likely that while individual participants make take longer in one task, they may achieve competency more quickly in other tasks. Hence the need to compromise on standards for a given task will not likely be necessary.

The differential experience between modules to reach a performance standard highlights the fact that, from an educational standpoint, the number of cases to a performance threshold in one module cannot be assumed to be similar for even within the same domain of pediatric MSK images, let alone a different image type entirely (e.g., point-of-care ultrasound images). These findings are consistent with the educational phenomenon of "case specificity," where the diagnostic performance on one case does not predict performance on another. While the specific reasons for this differential performance were not elucidated in this research, this does map to what has been noted at the bedside for this clinical area. Mounts et al. reviewed 220 cases and found that some pediatric MSK radiograph interpretations were more prone to missing fractures than others; the most frequently missed fractures were of the hand phalanges (26.4%) followed by metatarsus (9.5%), distal radius (7.7%),

Table 1
Per-module Data for Achieving Performance Threshold

<table>
<thead>
<tr>
<th>Pediatric MSK Module</th>
<th>Cases Completed to Reach Performance Standard, Median Number* (IQR)</th>
<th>Proportion of Participants Completed Module To Performance Standard, n/N (%)</th>
<th>Time on Case (Seconds), Median (Min, Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skull</td>
<td>79 (68–113) n = 37</td>
<td>37/42 (88.1)</td>
<td>36.3 (15.4, 88.6) n = 42</td>
</tr>
<tr>
<td>Shoulder/clavicle/humerus</td>
<td>90 (76–119) n = 36</td>
<td>36/39 (92.3)</td>
<td>32.3 (13.4, 58.1) n = 39</td>
</tr>
<tr>
<td>Elbow</td>
<td>73 (68–190) n = 42</td>
<td>42/48 (87.5)</td>
<td>28.9 (8.0, 63.2) n = 48</td>
</tr>
<tr>
<td>Forearm/hand</td>
<td>72 (66–105) n = 46</td>
<td>46/50 (92.0)</td>
<td>37.4 (10.8, 77.9) n = 50</td>
</tr>
<tr>
<td>Pelvis/femur</td>
<td>69 (57–111) n = 38</td>
<td>38/40 (95.0)</td>
<td>36.0 (10.4, 80.0) n = 40</td>
</tr>
<tr>
<td>Knee/tibia-fibula</td>
<td>89 (78–103) n = 33</td>
<td>33/39 (84.6)</td>
<td>28.9 (7.6, 79.5) n = 39</td>
</tr>
<tr>
<td>Ankle/foot</td>
<td>366 (150–548) n = 15</td>
<td>15/43 (34.9)</td>
<td>41.0 (19.8, 89.5) n = 43</td>
</tr>
</tbody>
</table>

MSK = musculoskeletal.
*Performance standard defined as achieving ≥ 80% accuracy, sensitivity, and specificity.

Table 2
Changes in Accuracy, Sensitivity, and Specificity from Baseline to Performance Standard of at Least 80% Accuracy, Sensitivity, and Specificity

<table>
<thead>
<tr>
<th>Pediatric MSK Module</th>
<th>Percent Change in Accuracy (95% CI)</th>
<th>Percent Change in Sensitivity (95% CI)</th>
<th>Percent Change in Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skull (n = 37)</td>
<td>17.6 (15.3–19.9)</td>
<td>18.4 (14.5–22.4)</td>
<td>15.6 (11.3–19.9)</td>
</tr>
<tr>
<td>Shoulder/clavicle/humerus</td>
<td>15.9 (12.4–19.4)</td>
<td>13.5 (9.4–17.6)</td>
<td>18.9 (13.6–24.2)</td>
</tr>
<tr>
<td>Elbow (n = 42)</td>
<td>15.4 (12.1–18.8)</td>
<td>18.7 (15.5–21.9)</td>
<td>14.2 (9.5–18.9)</td>
</tr>
<tr>
<td>Forearm/hand (n = 46)</td>
<td>10.4 (8.2–12.6)</td>
<td>10.6 (7.3–13.9)</td>
<td>10.6 (6.7–14.5)</td>
</tr>
<tr>
<td>Pelvis/femur (n = 38)</td>
<td>8.7 (6.1–11.3)</td>
<td>8.7 (6.2–11.2)</td>
<td>7.1 (2.6–11.7)</td>
</tr>
<tr>
<td>Knee/tibia-fibula (n = 33)</td>
<td>16.4 (13.2–19.7)</td>
<td>23.8 (19.1–28.5)</td>
<td>7.1 (3.0–11.2)</td>
</tr>
<tr>
<td>Ankle/foot (n = 15)</td>
<td>26.1 (20.2–32.0)</td>
<td>32.7 (24.7–40.7)</td>
<td>17.9 (10.3–25.5)</td>
</tr>
</tbody>
</table>

MSK = musculoskeletal.
tibia (7.3%), and phalanges of the foot (5.5%). While some of these identified clinical challenges do map to our findings in the setting of an educational intervention, it is important to note that in this other publication, the study focused only on identifying missed fractures. In contrast, our intervention also identified overcalling of image interpretation, and this may explain some of the differences. Nevertheless, overcalling pathology also carries harm, often in the form of incorrect treatment and unnecessary follow-up. Future research should explore both types of pediatric MSK interpretation errors to allow a more informed view of diagnostic interpretation needs, which will consider data from stand-alone education interventions as well as errors made during clinical practice.

The findings of our study have implications for practice. Our results reinforce the motivations for competency-by-design. A “one-size-fits-all” curriculum applied to postgraduate trainees would result in a variable level of skill as residents prepare to enter the world as independent clinicians. Thus, one of the key benefits of the type of educational platform embedded in this research is that these learning tools allow trainees to automatically get an experience weighted toward evening out case exposure and performance to a specified standard. As such, we encourage the implementation of learning assessment platforms for emergency medicine skills amenable to cognitive simulation and intensive deliberate practice. Specific to the learning intervention described in this study, it is available to any postgraduate trainee or attending-level physician to participate as an individual or as part of a program engagement (https://imagesim.com/). Further, at the end of this study period, the duration of access to the course was expanded from 12 to 24 months to be more in keeping with the duration of academic fellowships. Importantly, the pairing of these learning platforms with learning analytics can provide a granular look at individual- and group-level performance. This will in turn allow educators to identify learners who face difficulties early such that additional learning interventions could be implemented in a timely manner or the development of quality assurance programs in areas identified as common deficiencies. Finally, as the shift toward competency-based medical education progresses, there is a need to explore the new technologies that will help feasibly establish and assess competency. The learning platform presented in this research demonstrated effectiveness for training on clinical tasks with dichotomous outcomes. Where the assessment is more complex, our model could be adapted to address unique training needs for different types of clinical tasks.

LIMITATIONS

This research has limitations that warrant consideration. Although all participants were PEM fellows, there could have been heterogeneity between the participants (e.g., number of radiology or orthopedic rotations), which may have impacted performance outcomes. Since we did not collect information on these variables, we were not able control for these potential confounders. Performance on this education platform may not necessarily translate into performance in a clinical setting where real-time patient information and higher resolution monitors may impact interpretation skills. This intervention did not explore skill retention from this education intervention, which is the subject of another study (Boutis et al., manuscript submitted for publication). This education intervention requires a fee for participation and may pose a barrier to participation for some programs. Finally, while we have previously shown that PEM fellows and emergency medicine residents have similar skill acquisition using this system, this study enrolled a convenience sample of PEM fellows, and therefore our results may have limited generalizability to the broader group of postgraduate trainees.

CONCLUSIONS

There was significant variation between participants and between modules in the number of cases required to achieve a predefined performance standard. Nevertheless, the performance target was achieved for a high percentage of participants on average in about 1 hour per module, which speaks to the feasibility of implementing similar programs. The learning outcomes for the ankle/foot case set were different than the other case sets, highlighting an area that might require additional training. Overall, there were significant gains made in accuracy, sensitivity, and specificity across all modules. Future studies are required to explore the best ways to establish defensible evidence-based competency thresholds and how similar strategies can be applied to the training of other skill sets.

The authors acknowledge Dr. Martin Pecaric of Contrail Consulting Services for providing technical development and support for the education intervention.
References


Comparison of Manikin Versus Porcine Trachea Models When Teaching Emergent Cricothyroidotomy Among Emergency Medicine Residents

Mark L. Gustafson, DO,1,2 Brian Hensley, MD,1 Marc Dotson, DO,1 Mike Broce,3 and Alfred Tager, MD1,3

ABSTRACT

Objectives: Emergent cricothyroidotomy (EC) is a rarely used yet lifesaving procedure that is important for an emergency physician to master throughout his or her training. We evaluated the difference in utilization of a manikin or porcine model among emergency medicine (EM) residents when teaching EC. We also evaluated the difference in the models using two different commonly utilized and taught techniques: "scalpel-finger-bougie" (SFB) technique and the Melker technique (MT).

Methods: This was a prospective crossover design. Instructions about the procedure were provided; study participants were randomly assigned to one of two groups and performed cricothyrotomy on both manikin and porcine simulators using both the Melker and the SFB techniques. Each group was started with the technique on the second simulator opposite what they started with on the first simulator. After the procedures, study participants completed a questionnaire that used the same format for both groups. All survey questions required a 7-point Likert scale response. Confidence, difficulty, reality of the anatomy, and landmarks were compared.

Results: Fifteen EM residents participated in the study. Compared to the porcine-first group, the manikin-first group rated the anatomy more realistic (question 5) after their first attempt (6.29 vs. 5.87) than their second attempt (using the porcine model; 4.29 vs. 5.62; main effect for attempts [first vs. second] p = 0.027; interaction p = 0.074). Interestingly, the porcine model–first group rated the landmarks (question 6) significantly easier to find in both attempts (6.87 vs. 5.86 on the first attempt and 6.50 vs. 5.57 on the second attempt; p = 0.012). Twelve participants (80%) chose SFB as their preferred technique to use in real-life scenario.

Conclusion: The Manikin model tends to be more realistic in cricothyrotomy simulation than the porcine model. The influence of fresh pig skin overlying the porcine model may require further research. Landmark identification during the procedure was easier on both models when participants start with the porcine model. Participants prefer the SFB over MT if faced with a real-life scenario.

Emergent cricothyroidotomy (EC) is a lifesaving procedure that is performed when endotracheal intubation has failed. This most commonly occurs when a patient has a difficult airway and cannot be intubated or ventilated. This procedure has been described using several different techniques. EC has traditionally been described in an open technique using a tracheal hook and dilator.1 This open technique is traumatic and may be difficult to perform in the emergency department (ED). A modified open technique utilizing a...
scalpel, gum elastic bougie, and finger as a dilator has been described. This procedure has been coined the "scalpel-finger-bougie" (SFB) technique and is frequently taught as it can be performed with minimal equipment. Also, multiple manufacturers make kits utilizing the Seldinger technique to perform this same procedure in efforts to reduce trauma. One of the more commonly used Seldinger techniques is referred to as the Melker technique (MT). Previous research evaluating which technique is superior has been limited and has been unable to show a significant difference in one EC technique over another.2

Emergent cricothyroidotomy has become very difficult to teach novice learners such as emergency medicine (EM) residents as they are sparsely exposed to this procedure in clinical practice. The majority of resident training on this procedure is simulation focused.3 The EC procedure is traditionally taught and practiced utilizing a manikin. To make this simulation more realistic it is also taught in cadaver labs; however, they are expensive and logistically difficult.4 Due to these limitations the use of porcine tracheas can be used to simulate this procedure in an effort to make it more realistic.5 Thus far, research comparing the use of different simulators has been limited.

The goal of our study was to evaluate the difference in utilization of a Manikin or porcine model among EM resident when teaching EC. We also evaluated the difference in the models using two different commonly utilized and taught techniques.

METHODS

Study Design
This was a prospective crossover design. Instructions about the procedure were provided at a didactic session explaining indications, contraindications, anatomy, and complications of each technique using a 20-minute video. Participation in the training experience and study was voluntary, and verbal consent was obtained prior to training. Participants were offered and could have had the training without study participation. Following the video session, study participants were randomly assigned to one of two groups and performed cricothyrotomy on both manikin and porcine simulators using both the Melker and the SFB techniques. To control for bias with respect to technique each group was started with the opposite technique on the second simulator than they started with on the first simulator.

Larynxes from slaughtered pigs were provided for the porcine simulator. Prelaryngeal tissue was eliminated from the larynx. Pig skin was not provided. All porcine models were provided utilizing preserved pig tracheas procured from a healthcare simulation wholesaler (Figure 1). The Simulab TraumaMan was used as the Manikin simulator. Intact synthetic skin was provided for each resident (Figure 1). Study participants were monitored during the entire procedure by the primary investigator to ensure proper technique.

After the procedures, each group was verbally consented for participation in the study. They then completed a questionnaire that used the same format for both groups. All survey questions required a 7-point Likert scale response. This study was approved by Charleston Area Medical Center Institutional Board Review.

Data Analysis
Data are reported as a number (%) or mean. The survey score of each model were compared with a repeated-measures analysis of variance test using SPSS (IBM Corp., Released 2010, IBM Statistics for Windows, Version 19.0). A p-value of <0.05 was considered statistically significant. The current study resulted in a 2-by-2 factorial design where the first factor was the type of model (manikin vs. porcine) and the second factor was the order of the attempts (first vs. second). Therefore, the results were tested for main effects for each factor (type of model and order of attempt). The main effect for each factor was the effect of one independent variable on the dependent variable, ignoring the effects of other independent variable. In addition, the results were tested for a possible interaction between the two independent variables (type of model and order of attempt). A statistical interaction could occur if the effect of one independent variable on the dependent variable is influenced by the level of the other independent variable.

RESULTS
Study participants consisted of 15 EM resident physicians, four PGY-1 (26.7%), four PGY-2 (26.7%), four PGY-3 (26.7%), and three PGY-4 (20%). Eight (53.3%) participants had previous cricothyrotomy training experience with manikin model only; five (33.3%) participants had previous cricothyrotomy training experience with both manikin and porcine models one (6.7%) participant had previous cricothyrotomy training experience with manikin, porcine model, and actual patient; and
one (6.7%) did not any previous cricothyrotomy training experience. Eight (53.3%) candidates started with the porcine model first and then switched to the manikin; the other seven (46.7%) started with manikin first then switched to the porcine model.

We found no significant differences between the two groups (manikin model first or porcine model first) in terms of the questions that related to confidence and difficulty (the first four questions of the survey; Table 1). All participants performed both techniques (SFB and MT) on both types of models (manikin and porcine; see Figure 2).

The porcine model–first group rated the landmarks (question 6) significantly easier to find (6.69 vs. 5.71, p = 0.012; see Table 2). There was a significant reduction for both groups for rating the models as being realistic to the human anatomy (question 5) after performing two attempts (p = 0.027; see Table 3). The manikin-first group rated the anatomy more realistic (question 5) after the first attempt (6.29) compared to...
(4.29) after the second attempt (i.e., with the porcine model). In comparison, to a lesser degree, the porcine-first group rated their first attempt higher (5.87) than their second attempt (5.62; interaction p = 0.074; see Table 4). The delta differences between the first and second attempts for the two groups are illustrated in Figure 3.

The majority of the participants (12 t of 15, 80%) chose SFB as their preferred technique to use in a real-life scenario. Technique preference did not differ based on whether the candidate started on the porcine model (6/8, 75%) or the manikin model (6/7, 85.7%) first (p = 1.00).

**DISCUSSION**

The study survey found no significant difference between utilization of the Manikin or porcine simulators with respect to questions related to confidence and difficulty. This finding is somewhat similar to the only previous study to compare utilization of manikin versus porcine simulators. Cho et al.5 previous study compared the simulators using only the Poretex cricothyroidotomy kit technique while our study utilized SFB and MT.

A previous study by Klock6 found that while a manikin can simulate the likeness of a real airway anatomically, the compressibility of the tissues and the frictional component of the manikin materials need to be taken into consideration. There is a difference between the actual physical fidelity versus how functionally realistic these manikin simulators are with respect to the realism of the tissues. In the previously mentioned study by Cho et al., they found that the porcine model had a higher visual analog scale for overall reality and preference among the participants.5 In our study, when participants started on the Manikin first they felt that the manikin was more realistic than the porcine model. When participants started on the porcine model first there was less of a difference in realism between the two models (interaction p = 0.074; Figure 3). The differences in realism of the porcine model in our study may be due to utilization of fresh pig skin by Cho et al. overlying their porcine model. We believe that it may be logistically difficult for most simulation centers to obtain and dispose of fresh pig skin.

Identification of landmarks when performing EC is important for proper technique and success. Our study showed that participants who started with the porcine model first found the anatomical landmarks easier to identify on both the porcine and the manikin models than participants who started on the manikin model first (Figures 1 and 2). This finding is also different than that of Cho et al. where porcine landmarks were rated more difficult to find. However, the disparity in findings could also be due to the prior experience of participants. In our study, the majority had previous training (14/15 = 93.3%), with

### Table 2

Mean Survey Scale Scores by Type of First Model, Manikin Versus Pig (Main Effect for Model)

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manikin First</td>
<td>6.00</td>
<td>0.96</td>
<td>5.81</td>
<td>1.11</td>
<td>5.90</td>
<td>1.03</td>
<td>0.737</td>
</tr>
<tr>
<td>Porcine Model First</td>
<td>6.29</td>
<td>0.61</td>
<td>5.94</td>
<td>1.06</td>
<td>6.10</td>
<td>0.88</td>
<td>0.460</td>
</tr>
<tr>
<td>Overall</td>
<td>5.90</td>
<td>1.03</td>
<td>5.90</td>
<td>1.03</td>
<td>5.90</td>
<td>1.03</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Mean scores are presented without accounting for other independent variables.

### Table 3

Mean Survey Scale Scores by Order of Attempt First Versus Second (Main Effect for Attempt)

<table>
<thead>
<tr>
<th>Attempt</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Attempt</td>
<td>5.93</td>
<td>0.96</td>
<td>5.87</td>
<td>1.13</td>
<td>5.90</td>
<td>1.03</td>
<td>0.616</td>
</tr>
<tr>
<td>Second Attempt</td>
<td>6.13</td>
<td>0.83</td>
<td>6.07</td>
<td>0.96</td>
<td>6.10</td>
<td>0.88</td>
<td>0.616</td>
</tr>
<tr>
<td>Overall</td>
<td>6.06</td>
<td>0.99</td>
<td>5.94</td>
<td>1.00</td>
<td>6.00</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Mean scores are presented without accounting for other independent variables.

### Table 4

Mean Survey Scale Scores by Type of First Model and by Order of Attempt

<table>
<thead>
<tr>
<th>Model</th>
<th>First</th>
<th>Second</th>
<th>Delta</th>
<th>First</th>
<th>Second</th>
<th>Delta</th>
<th>Interaction p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manikin First</td>
<td>6.00</td>
<td>6.00</td>
<td>0.00</td>
<td>5.87</td>
<td>5.75</td>
<td>0.12</td>
<td>0.616</td>
</tr>
<tr>
<td>Porcine Model First</td>
<td>6.29</td>
<td>6.29</td>
<td>0.00</td>
<td>6.00</td>
<td>5.87</td>
<td>0.13</td>
<td>0.616</td>
</tr>
<tr>
<td>Overall</td>
<td>6.15</td>
<td>6.15</td>
<td>0.00</td>
<td>5.93</td>
<td>5.75</td>
<td>0.18</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Mean scores are presented after accounting for both independent variables and test for interaction.
only five (33.3%) having experience with porcine and manikin models, and only one participant (6.7%) had experience with an actual patient, while all participants had previous training and nearly half (46.7%) had actual patient experience in the study by Cho et al. Our survey results also showed that after completing the training on both models 12 of 15 participants stated that they would prefer the SFB technique over the MT in a real-life scenario.

Although our main interest was to evaluate the types of training models (manikin vs. porcine) and techniques (SFB and MT), it is possible that participants could have perceived that their competence or skill level was also being measured. Our survey instrument did not contain identifiers; participants used both models in one-setting and were randomized to the order of the first model encountered. These safeguards were in place to reduce the apprehension surrounding participant responses to the survey questions. However, to be fair, this type of simulation training is based not only on participant preference, but self-perception and/or self-assessment. Effects of training can be short-lived.\(^7\) We were not specifically measuring the degree of learning or skill, but at the same time we must mention that other factors could have influenced participants’ perceptions. It has been reported that “self” rating objectivity is difficult because it is an emotional response.\(^8\) Sargeant and colleagues\(^9\) suggest that there are four phases of self-assessment: 1) receiving feedback from external sources, 2) reflecting on feedback to reconcile with self-assessment, 3) plan to use feedback, and 4) use feedback for improvement and learning. Often during simulation training feedback from instructors can be important, and researchers have suggested that debriefing is beneficial.\(^10\) Students can perceive peer and faculty debriefing as equally effective.\(^11\) Recently researchers have put forth “a toolbox for simulation educators” for difficult debriefing situations.\(^12\) The authors list types of learners and provide information on how to deal with different types of debriefing situations. Anyone or a multitude of these factors, intrinsic or extrinsic, could have influenced the perceptions of the participants in our research study. Perhaps a more complete and sensitive survey instrument along with debriefing could have allowed for some of the factors to have been identified and controlled for during statistical testing.

**LIMITATIONS**

Besides the self-perception issues mentioned, this study is limited by a small sample size from a single simulation center. In an attempt to overcome this limitation, we used a repeated-measures statistical test, so each study participant served as his or her own control to reduce error variance. The study used an unvalidated survey instrument that has been developed by our study team. However, we were only asking questions concerning confidence and difficulties with each technique, and we believe that these questions and responses to be straightforward. We also applied the crossover effect to decrease the bias of taking the same survey questions twice (one for each model). Finally, in the porcine model, fresh pig skin was not provided, which might affect the realistic anatomy of the procedure.

**CONCLUSION**

Our study results, which are based on self-perception, suggest that the manikin model tends to be more
realistic in cricothyrotomy simulation than the porcine model. Influence of fresh pig skin overlying the porcine model may require further research. Landmark identification during the procedure was easier on both models when participants start with the porcine model. Participants prefer the scalpel-finger-bougie technique over the Melker technique if faced with a real-life scenario.

References

Analysis of the Emergency Medicine Clinical Learning Environment

V. Ramana Feerer, MD1, Zachary Zemore, MD1, Nital Appelbaum, PhD2, Sally A. Santen, MD, PhD1, Joel Moll, MD1, Brian Aboff, MD3, and Robin R. Hemphill, MD, MPH4

ABSTRACT

Background: Residencies are grappling with ways to identify methods to internally monitor and improve their learning environments. Building on prior work, the objective of this study was to determine emergency medicine (EM) internal evaluations of perceived organizational support and psychological safety and compare to the results from the Accreditation Council for Graduate Medical Education (ACGME) Resident Survey for the purpose of program improvement and to explore factors affecting residents’ perception of their learning environment.

Methods: In 2017, the Virginia Commonwealth University School of Medicine Office of Graduate Medical Education and Office of Quality and Safety conducted an in-person, anonymous safety survey of the residents across 19 residency programs on the Short Survey of Perceived Organizational Support (SPOS) and Psychological Safety Scale (PSS). These were compared to the ACGME Resident Survey for 19 programs. Resident interviews and open response evaluation data informed content analysis on program experiences.

Results: Institutional response rates were 63% for the internal learning environment survey and 96% for ACGME Resident Safety Survey. EM residents responded positively on the SPOS and PSS compared to other programs (ranked second highest scores on both scales). One-hundred percent of respondents agreed or strongly agreed on SPOS items: “Help is available from my department when I have a problem.” “My department really cares about my well-being.” “My department values my contribution to its well-being.” Furthermore, EM had the highest overall training experience score (mean = 4.83) on the ACGME survey compared to the 18 other training programs. Qualitative responses suggest program strengths included supportive program leadership, positive working relationships with faculty, and emphasis on trainee wellness.

Conclusions: Compared to other programs, EM has created a positive environment of safety and support as perceived by their residents. Internal surveys of the learning environment can help programs understand their culture for purposes of improvement and align with the ACGME survey.
organization values their contribution and cares about their well-being.\textsuperscript{1} This perception is based on fairness, supervisor support, organizational rewards, and job conditions. The importance of POS is that it is a reciprocal relationship whereby if the resident perceives a favorable organizational environment, the resident has an increased commitment to give back positively to the organization.\textsuperscript{1} With increased commitment and performance to the organization, there is less individual withdrawal and burnout.\textsuperscript{1}

Psychological safety, the belief that one can speak up without fear on a team, has been identified as another critical factor in enabling performance and collaboration to accomplish shared work because it creates a climate that mitigates interpersonal risks when trying to raise concerns in learning hierarchies.\textsuperscript{2} High levels of psychological safety allow for increased creativity, increased collaboration, admitting mistakes, asking for help, providing feedback, challenging the status quo, identifying problems or opportunities for improvement, and offering ideas to improve their organization’s well-being.\textsuperscript{2}

Goals of This Investigation
Building on prior work,\textsuperscript{3} the objective of this study was to evaluate emergency medicine (EM) internal assessments of perceived organizational support and psychological safety and compare with results from the Accreditation Council for Graduate Medical Education (ACGME) Resident Survey and then focus on factors that may affect residents’ perception for the purpose of program improvement. By identifying how these relate to overall program experience score on the ACGME resident survey, we may then be able to target interventions to the items that positively or negatively influence the clinical learning environment for EM residents.

METHODS
Our health system internally surveyed residents across 19 residency training programs between May 2017 and June 2017.\textsuperscript{3} A non-EM research faculty member administered the paper survey during resident specific conferences to allow confidentiality in responses. The 19 ACGME Resident Survey results for FY17 were downloaded from ACGME’s online reporting system.

The Short Survey of Perceived Organizational Support\textsuperscript{4} (SPOS) and the Psychological Safety Scale\textsuperscript{5} (PSS) were selected to operationalize the clinical learning environment from a cultural lens. SPOS and PSS, validated in prior studies, have been utilized in studies across numerous industries to explain the dynamics between people and the organizations in which they work and learn. The two scales were embedded into a longer survey regarding resident handoffs and event reporting. Our institutional review board approved the resident survey.

The SPOS consists of 16 items measured on a 5-point agreement Likert scale (1 = strongly disagree, 5 = strongly agree). The PSS has seven items similarly measured. Items were adapted so the words “team” and “organization” were replaced with “department” considering that residents primarily associate their clinical learning environment with their department. The ACGME Resident Survey\textsuperscript{6,7} for FY17 provided an overall experience item to evaluate each training program on a 5-point Likert scale (1 = very negative, 5 = very positive).

Data Analyses
Department-level frequency statistics and mean scores were calculated for SPOS and PSS (SPSS Version 24) were compared against the overall training experience score at the departmental level (ACGME survey). For the ACGME Resident Survey, we converted the percentages into an overall mean experience score for each department based on response rate information.\textsuperscript{3} Inter-rater agreement ($r_{wg}$) was calculated to support aggregation to the departmental level for SPOS and PSS.\textsuperscript{8}

To understand the results and give context to the survey findings, we reviewed with content analysis recent internal annual EM program evaluation where residents provided free-text responses. In addition, we discussed with residents in small group settings their impressions of the program.

RESULTS
Across 19 programs 322 residents (63% response rate) completed the learning environment survey and 496 residents (96% response rate) completed the FY17 ACGME Resident Survey. Within EM, there was a 63% response rate (n = 19) for the learning environment survey and a 100% response rate (n = 30) for the ACGME Resident Survey. There was support to aggregate SPOS ($r_{wg} = 0.80$) and PSS ($r_{wg} = 0.79$) scores to the departmental level based on inter-rater reliability calculations.\textsuperscript{8}
Within our institutional sample, EM had the highest overall training experience score on the ACGME Resident Survey (mean = 4.83 out of 5) compared to the other 18 programs (Table 1). In addition, EM was rated second highest on perceived organizational support (mean = 4.01 out of 5) and psychological safety (mean = 3.93 out of 5).

The perceived organizational support items (Table 2) revealed that 100% (n = 19) of the EM residents agreed/strongly agreed that their department values residents’ contributions to its well-being and cares about residents' well-being. All (n = 19; 100%) of the EM respondents also agreed/strongly agreed that help is available from the department when they have a problem. For psychological safety (Table 3), the majority of EM respondents disagreed/strongly disagreed that if they made a mistake in the department, it is often held against them (n = 15, 79%) or that it is difficult to ask other members of the department for help (n = 19, 100%). Upon review of the annual ACGME EM residency program evaluation, there were 105 comments three themes that emerged (Table 4) in relation to the learning environment: program leadership, teamwork, and work–life balance.

**DISCUSSION**

It is important for residency programs to focus on their learning environments with internal surveys but also in response to external ACGME surveys. We found that the POS, PS, and ACGME responses were high for our EM program. The items on the POS and

**Table 1**
Mean Scores of Residents’ Perceived Support, Psychological Safety, and Overall Training Evaluation by Department

<table>
<thead>
<tr>
<th>Department</th>
<th>Perceived Organizational Support</th>
<th>Psychological Safety</th>
<th>ACGME Overall Training Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM</td>
<td>4.01</td>
<td>3.93</td>
<td>4.83</td>
</tr>
<tr>
<td>Department B</td>
<td>3.99</td>
<td>3.75</td>
<td>4.76</td>
</tr>
<tr>
<td>Department C</td>
<td>3.92</td>
<td>3.53</td>
<td>4.75</td>
</tr>
<tr>
<td>Department D</td>
<td>4.24</td>
<td>4.09</td>
<td>4.67</td>
</tr>
<tr>
<td>Department E</td>
<td>3.89</td>
<td>3.56</td>
<td>4.63</td>
</tr>
<tr>
<td>Department F</td>
<td>3.51</td>
<td>3.22</td>
<td>4.50</td>
</tr>
<tr>
<td>Department G</td>
<td>3.60</td>
<td>3.66</td>
<td>4.47</td>
</tr>
<tr>
<td>Department H</td>
<td>3.75</td>
<td>3.65</td>
<td>4.36</td>
</tr>
<tr>
<td>Department I</td>
<td>3.14</td>
<td>3.22</td>
<td>4.35</td>
</tr>
<tr>
<td>Department J</td>
<td>3.23</td>
<td>3.35</td>
<td>4.32</td>
</tr>
<tr>
<td>Department K</td>
<td>3.63</td>
<td>3.51</td>
<td>4.30</td>
</tr>
<tr>
<td>Department L</td>
<td>3.83</td>
<td>3.83</td>
<td>4.28</td>
</tr>
<tr>
<td>Department M</td>
<td>3.50</td>
<td>3.44</td>
<td>4.20</td>
</tr>
<tr>
<td>Department N</td>
<td>3.12</td>
<td>3.01</td>
<td>4.18</td>
</tr>
<tr>
<td>Department O</td>
<td>3.64</td>
<td>3.40</td>
<td>4.07</td>
</tr>
<tr>
<td>Department P</td>
<td>3.37</td>
<td>3.46</td>
<td>3.98</td>
</tr>
<tr>
<td>Department Q</td>
<td>3.18</td>
<td>3.38</td>
<td>3.85</td>
</tr>
<tr>
<td>Department R</td>
<td>3.11</td>
<td>3.21</td>
<td>3.82</td>
</tr>
<tr>
<td>Department S</td>
<td>3.44</td>
<td>3.41</td>
<td>3.79</td>
</tr>
</tbody>
</table>

**Table 2**
Perceived Organizational Support

<table>
<thead>
<tr>
<th>Item</th>
<th>Strongly Disagree/ Disagree</th>
<th>Neutral</th>
<th>Agree/ Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>My department values my contribution to its well-being.</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>19 (100)</td>
</tr>
<tr>
<td>*If my department could hire someone to replace me at a lower salary it would do so.</td>
<td>11 (58)</td>
<td>4 (21)</td>
<td>4 (31)</td>
</tr>
<tr>
<td>*My department fails to appreciate my extra effort from me.</td>
<td>12 (63)</td>
<td>3 (16)</td>
<td>4 (21)</td>
</tr>
<tr>
<td>My department strongly considers my goals and values.</td>
<td>1 (5)</td>
<td>2 (11)</td>
<td>16 (84)</td>
</tr>
<tr>
<td>*My department would ignore any complaint from me.</td>
<td>17 (90)</td>
<td>1 (5)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>*My department disregards my best interests when it makes decisions that affect me.</td>
<td>15 (79)</td>
<td>4 (21)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Help is available from my department when I have a problem.</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>19 (100)</td>
</tr>
<tr>
<td>My department really cares about my well-being.</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>10 (100)</td>
</tr>
<tr>
<td>*Even if I did the best job possible, my department would fail to notice.</td>
<td>14 (74)</td>
<td>4 (21)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>My department is willing to help me when I need a special favor.</td>
<td>0 (0)</td>
<td>5 (26)</td>
<td>14 (74)</td>
</tr>
<tr>
<td>My department cares about my general satisfaction at work.</td>
<td>1 (5)</td>
<td>4 (21)</td>
<td>14 (74)</td>
</tr>
<tr>
<td>*If given the opportunity, my department would take advantage of me.</td>
<td>12 (63)</td>
<td>4 (21)</td>
<td>3 (16)</td>
</tr>
<tr>
<td>*My department shows very little concern for me.</td>
<td>19 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>My department cares about my opinions.</td>
<td>0 (0)</td>
<td>5 (26)</td>
<td>14 (74)</td>
</tr>
<tr>
<td>My department takes pride in my accomplishments at work.</td>
<td>2 (11)</td>
<td>2 (11)</td>
<td>15 (79)</td>
</tr>
<tr>
<td>My department tries to make my job as interesting as possible.</td>
<td>0 (0)</td>
<td>3 (16)</td>
<td>16 (84)</td>
</tr>
</tbody>
</table>

Data are reported as n (%). *Item is reverse coded.
PS also help us identify areas of strengths and improvements of our residency. High POS and PS may also contribute to enhanced work performance, job satisfaction, positive mood, and commitment.1,2 With the ACGME showing a heightened emphasis on resident burnout and well-being, there are opportunities for shared learning between those programs that are performing well in the domains of POS and PS, such as EM, and those that are less successful. However, for such shared learning to be successful, greater investigation is needed to learn what features of the program create a more positive learning environment.

There are lessons learned of how to improve the learning environment from the qualitative results. Our preliminary qualitative analysis suggests focusing on behaviors of program leadership, teamwork dynamics between residents and attendings, and approach to wellness. Residents have taken note of changes with rotations, conference activities, teaching time during shifts, and residency retreats. Many of these revisions to the program stemmed from active feedback sessions done by the leadership team. This not only addresses the root problem voiced but also enhances the perception that a resident’s opinion is valued by the leadership team and empowers them to speak up about other concerns.9

A second theme that arose related to the EM environment was the strong emphasis on teamwork. The resident–attending team interactions encourage discussion of patient care, interpersonal relationships, and teaching. This positive close-knit interaction also allows for a nonthreatening environment in which questions can be posed and potential mistakes can be addressed.

A final factor that may relate to the scores on the ACGME survey is a healthy work–life balance.10 The ability to control one’s work schedule and have the necessary time off for significant life events is a priority.11 The perception of control over one’s life is a basic tenet of POS. We implemented a wellness committee, annual residency retreat, and wellness strategies such as “anchor sleep” and “clockwise shift rotation” have also been maintained.12,13 All13 residents are

### Table 3
Psychological Safety

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree/ Disagree</th>
<th>Neutral</th>
<th>Agree/ Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>If you make a mistake in my department, it is often held against you.</em></td>
<td>15 (79)</td>
<td>1 (5)</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Members of my department are able to bring up problems and tough issues.</td>
<td>0 (0)</td>
<td>3 (16)</td>
<td>15 (84)</td>
</tr>
<tr>
<td><em>People in my department sometimes reject others for being different.</em></td>
<td>18 (95)</td>
<td>1 (5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>It is safe to take a risk in my department.</td>
<td>0 (0)</td>
<td>7 (37)</td>
<td>12 (63)</td>
</tr>
<tr>
<td><em>It is difficult to ask other members of my department for help.</em></td>
<td>19 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No one in my department would deliberately act in a way that undermines my efforts.</td>
<td>4 (21)</td>
<td>2 (11)</td>
<td>13 (69)</td>
</tr>
<tr>
<td>Working with members of my department, my unique skills and talents are valued and utilized.</td>
<td>0 (0)</td>
<td>5 (26)</td>
<td>14 (74)</td>
</tr>
</tbody>
</table>

Data are reported as n (%). *Item is reverse coded.

### Table 4
Example Resident Responses on the Annual EM Residency Program Evaluation

<table>
<thead>
<tr>
<th>Theme</th>
<th>Example Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program leadership</td>
<td>“I feel like our residency leadership advocated for us as residents fiercely and consistently.”</td>
</tr>
<tr>
<td></td>
<td>“EM residency administration is focused and adaptable to change.”</td>
</tr>
<tr>
<td></td>
<td>“Appropriate autonomy that progresses as time goes on. Leadership is a huge advocate for us.”</td>
</tr>
<tr>
<td></td>
<td>“Program leadership heavily invested in resident educations and needs.”</td>
</tr>
<tr>
<td>Teamwork</td>
<td>“Faculty are always happy to supervise and answer questions when needed.”</td>
</tr>
<tr>
<td></td>
<td>“Faculty being great teachers and very supportive.”</td>
</tr>
<tr>
<td></td>
<td>“Faculty involvement and dedication to our education.”</td>
</tr>
<tr>
<td>Work–life balance</td>
<td>“After seeing multiple residencies, VCU offers the best work/education/life balance.”</td>
</tr>
<tr>
<td></td>
<td>“Emphasis on wellness, fun place to work.”</td>
</tr>
</tbody>
</table>
given a Dictaphone and senior residents are able to use a scribe. This benefit also allows residents to complete their shifts in a timely fashion.

**LIMITATIONS**

There are several limitations to this study. First, survey methodology is subject to potential response biases, in particular social desirability bias. The steps taken to protect respondent identity during data collection efforts were meant to reduce this bias. Furthermore, while our data suggest a favorable clinical learning environment in EM, we cannot know whether our residents’ responses were driven by features of our program or whether other aspects impacted their perceptions of support and safety in speaking up. Finally, the study was at a single institution, limiting generalizability. Future research would tackle prospective evaluation of interventions aimed to increase residents’ perceptions of support and ability to speak up and extend the study to other programs.

**CONCLUSIONS**

With increased focus on the clinical learning environment, residency programs need to understand the inputs into perceptions of that environment. The ACGME, SPOS, and PSS survey tools may be used to quantify that environment, and this study has attempted to describe some strategies that may help other residency programs consider ways to improve their own learning environment.

**References**

Heart Rate Variability and Acute Stress Among Novice Airway Managers

Jason M. Mefford, MD, MAS1, Sarah Kahle, PhD2, Shikha Gupta, BS1, Daniel Tancredi, PhD3, Aaron R. Danielson, MD, MAS1, and Samuel O. Clarke, MD, MAS1

ABSTRACT

Background: The nature of medical emergencies places emergency physicians at risk for high levels of acute psychological stress (APS). Stress-modifying techniques like visualization, breath control, and mental practice may help mitigate APS, but objective markers of stress are difficult to measure in the clinical setting. We explored the relationship between heart rate variability (HRV), a real-time measure of autonomic arousal, and self-reported APS among emergency medicine (EM) residents learning to intubate on actual patients.

Methods: This was a prospective study of postgraduate year 1 (PGY-1) EM residents at a single academic medical center during their 1-month anesthesia rotation. We obtained repeated measures of HRV immediately before and during the first intubation attempt each day. Participants completed the modified Spielberger State-Trait Anxiety Inventory (STAI-6) before intubation attempts and scored intubation difficulty using the Intubation Difficulty Scale. We analyzed HRV using root mean square of successive differences and analyzed data using clustered data methods and Pearson correlation coefficients.

Results: We enrolled eight PGY-1 residents and recorded 64 intubations. Mean HRV in the 2 minutes before intubation (17.88 ± 9.22) and during intubation (21.17 ± 13.46) was significantly lower than resting baseline (32.09 ± 15.23; adjusted mean difference [95% CI] = −13.90 [−20.35 to −7.45], p < 0.001; and −10.77 [−17.65 to −3.88], p = 0.02). Preintubation anxiety was negatively correlated with HRV (r = −0.39 [−0.58 to −0.16], p = 0.001). Intubation difficulty was not significantly correlated with HRV (r = −0.12 [−0.36 to 0.13], p = 0.35).

Conclusions: HRV shows promise as a real-time index of autonomic arousal and may serve as an outcome measure in the evaluation of stress-modifying interventions.

The practice of emergency medicine (EM) demands the ability to lead teams, simultaneously diagnose and treat disease, and perform lifesaving procedures. The unpredictable and high-stakes nature of EM places physicians at risk for suboptimal performance and burnout.1,2 While some degree of autonomic arousal may facilitate performance (known as the Yerkes-Dodson law), high degrees of acute psychological stress (APS) increase the risk of medical and procedural errors.3–6 Interest is growing in psychological skills training (PST) as a means to combat APS and improve resuscitation performance.3,4 Originally derived from sports psychology, PST is a set of techniques including visualization, positive self-talk, and breath control. Due to a lack of robust metrics of APS that are measurable in the clinical setting, the efficacy of formalized PST is unknown. To address this need, we assessed an objective measure of autonomic arousal known as heart rate...
variability (HRV). HRV is the beat-to-beat variation in heart rate and reflects the real-time activity of the autonomic nervous system. In general, highly stressed people exhibit reduced HRV as a result of the overactivity of the sympathetic nervous system and reduced activity of parasympathetic nervous system. Understanding the physiologic changes among trainees learning a high-stakes procedure would provide an objective means for assessing APS as well as the efficacy of stress modifying interventions such as PST. As we enter the era of low-cost wearable technology, measuring physiologic data in the clinical setting becomes a realistic possibility.

We sought to measure HRV among postgraduate year 1 (PGY-1) EM trainees learning to intubate in a real-life clinical setting and to examine the relationship between HRV and self-reported anxiety and intubation difficulty. We chose to study this phenomenon in an authentic clinical context to elicit a high degree of psychological engagement from participants. We likewise chose novice (PGY-1) airway managers learning to intubate as they are less likely to have well-rehearsed strategies for mitigating acute stress compared to more experienced providers. Endotracheal intubation was chosen as the focus for this study due to its central importance to the practice of EM, its time-dependent and high-stakes nature, and its integration of cognitive and psychomotor skills. We hypothesized that residents would exhibit low HRV before and during intubation compared to baseline, indicating heightened autonomic arousal. We further hypothesized that HRV would be inversely correlated with self-reported anxiety and with procedural difficulty.

METHODS

This was a prospective observational study conducted at a U.S. EM residency program. PGY-1 EM residents were eligible for participation in the study. Exclusion criteria were previous anesthesia or EM residency, nurse anesthetist or paramedic training, preexisting cardiac disease, or use of medications affecting heart rate. The study site was an affiliated community hospital where our PGY-1 EM residents spend 1 month on an anesthesia rotation. During this rotation they learn endotracheal intubation on surgical patients under the direct supervision of an attending anesthesiologist. The study was approved by the University of California Davis Institutional Review Board, and all participants provided written consent. Please see Data Supplement S1 (available as supporting information) for a detailed description of our data collection and analysis.

Our primary outcome measure was HRV, which we assessed by comparing mean changes, relative to resting baseline, during the 2 minutes before and during intubation. Our secondary outcome measures were self-rated anxiety before intubating as measured by the modified Spielberger State-Trait Anxiety Inventory (STAI-6) and intubation difficulty as measured using the Intubation Difficulty Scale (IDS). This was a pilot study using a convenience sample to provide proof of concept and establish a foundation for future studies and as such did not rely on a predetermined sample size. Eight of nine eligible PGY-1 EM residents were enrolled: five men (62%) and three women (38%) with mean ± SD age of 28 ± 1.92 years (range = 26–31 years); one eligible resident declined to participate. All were healthy subjects with no history of heart disease or medication use.

RESULTS

We collected HRV recordings on a total of 64 intubation attempts, with a mean of eight attempts collected per participant (range = 6–10). We failed to obtain HRV recording on two intubation attempts. Mean time to intubate was 2.5 minutes (range = 1–6 minutes) with a mean of 1.19 attempts (range = 1–3 attempts). Mean ± SD self-rated anxiety score before intubating on the STAI-6 instrument was 34.3 ± 8.9, and mean ± SD intubation difficulty score was 1.8 ± 2.4.

At baseline, mean HRV measured in the time domain of mean square of successive differences (RMSSD) was 32.1 ± 15.2, with an intraclass correlation coefficient of 0.50, indicating a mild level of test–retest reliability. Mean HRV was not significantly lower at baseline 2 compared to baseline 1: mean difference = −1.7 (95% CI = −9.3 to 5.9), p = 0.66. Mean HRV measured in the time domain of RMSSD during the 2-minute period immediately before and during intubation was significantly decreased compared to resting baseline (Figure 1). Adjusted mean differences based on our regression model were −13.90 milliseconds (95% confidence interval [CI] = −20.35 to −7.45) prior to intubation and −10.77 milliseconds (95% CI = −17.65 to −3.88) during the intubation period. Univariate correlation analysis of preintubation STAI-6 score and preintubation HRV showed a significant negative correlation (r = −0.39, p = 0.001), whereas
scored intubation difficulty did not correlate significantly with HRV \(r = -0.12, p = 0.35\).

**DISCUSSION**

In this study, we demonstrated HRV to be significantly lower (indicating increased autonomic arousal) in residents just before and during intubation compared to resting baseline. Self-rated anxiety correlated significantly with lower measures of HRV, consistent with our hypothesis. While this was a limited sample, our baseline HRV measurements were consistent with published norms for resting HRV and the direction of correlation was consistent with prior studies using RMSSD as a measure of HRV under stressful conditions.\(^8\,10\)

Notably, in 36 of the 62 measurements, preintubation HRV was lower than intraintubation HRV with the proclivity for a higher measurement associated with learner (Fisher’s Exact test \(p\)-values = 0.03). This finding suggests that, for some learners, anticipatory APS appears to be higher than that observed during the procedure itself. We found no relationship between intubation difficulty and HRV during intubation. The majority of intubations in our study sample were determined to be relatively easy per the IDS (in which an intubation with a score > 5 is considered difficult),\(^8\) and this may have attenuated the relationship between intubation difficulty and HRV.

Higher resting HRV is associated with greater emotional regulation,\(^5,11\) and successful regulation of emotion can likewise increase HRV.\(^12\) EM residents who encounter fear with procedures may benefit from interventions such as breath training and biofeedback that help bring HRV under conscious control.\(^13,14\) Future directions of this research will include describing the expected ranges of HRV while intubating in comparison to rest, as well as recovery times. Understanding the physiological profile of trainees performing stressful procedures may help identify those with exaggerated or prolonged stress responses. In this way, HRV may prove to be an essential tool for both shaping and exploring the efficacy of PST interventions.

**LIMITATIONS**

This was a small pilot trial conducted at a single institution, and as such our findings are of limited generalizability. While we found evidence of test–retest reliability within our sample, three or more baseline measurements would be desirable to achieve a mean baseline score with suitable reliability (>70%) for research purposes.\(^15\) Additionally, a number of potentially confounding factors were not measured, including participants’ caffeine consumption, concurrent stressful life events, and the social interaction between study participants and anesthesia attendings on the airway rotation. We also chose to conduct our study in a controlled (operating room) setting rather than in the ED to standardize our testing conditions and minimize potential confounders, but this was done as a knowing trade-off in terms of generalizability to the ED setting. Our use of the IDS scale also relied on the subjective assessment of Cormack-Lehane grade by an individual novice observer. Future studies would benefit from the use of video laryngoscopy and multiple observers to address this limitation. Finally, we intentionally limited our investigation to PGY-1 residents early on in their airway training, and as such our findings may not generalize to more experienced practitioners.

**CONCLUSIONS**

Our study demonstrates that heart rate variability is lower among emergency medicine trainees just before and while intubating compared to rest and that lower
heart rate variability ratings correspond to higher degrees of self-reported anxiety. These findings suggest that emergency medicine trainees experience acute psychological stress while learning to perform the procedure of endotracheal intubation.

REFERENCES


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10335/full Data Supplement S1. Detailed description of research methods.
ABSTRACT
Background: Medical malpractice litigation is a prevalent challenge for emergency physicians, but there is a dearth of dedicated training in residency programs on this subject. As a result, when emergency physicians become the subject of a medical malpractice action they often find themselves ill-equipped to successfully navigate the process.

Objectives: We sought to create an in-depth medical malpractice simulation encompassing all key aspects of medical practice litigation. We collaborated with a law school for a semester-long simulation of a fabricated medical malpractice case brought against an emergency medicine (EM) resident.

Discussion: In harnessing the legal expertise of law students and faculty we were able to deliver a months-long medical malpractice simulation to our EM residency program. Similarly, in lending the medical expertise of our residents and faculty to the project, the law students were afforded a unique practical experience in learning to litigate medical malpractice.

Conclusions: The project resulted in a rich educational experience for both the EM residents and the law students. We present a guide for replication by any residency program in collaboration with a law school.

NEED FOR INNOVATION
More than 75% of emergency physicians will be named in a medical malpractice action at least once in their career, but there is a dearth of training in emergency medicine (EM) residency programs on the subject of medical malpractice litigation. Issues such as understanding the components and natural course of a malpractice suit, coping with the anxiety and stress associated with being a defendant, the importance of clinical documentation, giving an effective deposition, working with one’s attorney, and the range of possible outcomes are often opaque to physicians until such time as they must experience them firsthand.

BACKGROUND
Reports of structured medical malpractice litigation education in the literature are scant. Lectures and mock depositions and trials have been reported2,4.
although such exercises have been limited in scope and lacked the detail and experiential learning that a full-scale mock litigation provides. There has been reported collaboration between an EM residency program and a law school, although the purpose was limited to teaching deposition skills. To our knowledge, this is the first collaboration between a law school and EM residency in which all major aspects of medical malpractice litigation have been simulated during a months-long experience.

**OBJECTIVE OF THE INNOVATION**

We sought to deliver a realistic, immersive experience encompassing all key aspects of medical malpractice litigation. Our collaboration allowed EM faculty and residents to contribute medical realism to the law students’ learning experience and the law professors and students to contribute legal realism to the EM residents experience; the goal was a mutually beneficial learning experience. The project was institutional review board exempt.

**DEVELOPMENT PROCESS**

The project required a collaborator from both the EM residency and the law school. We developed a course at the law school called “Medical Malpractice Litigation Practicum” and six law students enrolled. Prior coursework in evidence and trial advocacy were prerequisites so as to allow for specific focus on medical malpractice and not on the basics of trial law.

As the law professor was designing the law school course, the EM faculty member created the fact pattern for the case. Realism, clinical uncertainty, and comprehensiveness were paramount in fabricating the medical record. Realism was achieved by beginning with actual medical records from our emergency department and then carefully deidentifying them and altering the content and data to suit our needs. The result was that the fonts, layout, and appearance were authentic. Meticulous attention was paid to every facet of the visit (notes, electrocardiograms [ECGs], orders, etc.) because the medical record would be perused by all parties (especially the law students) and any irregularity would destroy the realism. Clinical uncertainty was the next major consideration, such that a reasonable argument (as judged by the facilitators) could be made by both the plaintiff and the defendant. In our scenario we devised a case of a middle-aged man with chest pain, risk factors for coronary artery disease, an equivocal ECG, and a minimization of symptoms by the patient. He was discharged only to return hours later with an obvious ST-elevation myocardial infarction. Comprehensiveness was important because it was impossible to predict which data would become relevant during the course of the litigation, so an unabridged medical record was necessary.

The participants were then recruited. An actor was hired by the law school to serve as the plaintiff, two senior faculty members from the EM residency program were to be expert witnesses (one for the plaintiff and one for the defendant), and an EM resident was to be the defendant.

**IMPLEMENTATION**

At the law school, each class meeting was divided into lecture and skills performance for each component of a medical malpractice case. Practitioner lecturers from both the defense and the plaintiff’s bar were guest instructors. Following the initial plaintiff/client interview simulation, the students drafted the complaint that was subsequently served to the defendant as he was in the midst of a clinical shift. Word soon spread through the department that a resident was being sued in a mock malpractice case and interest in the process among his colleagues peaked (evidenced by informal but numerous questions posed to the faculty and the “defendant” resident). The next step would be an initial defense/client meeting between the defendant and his lawyers (i.e., the law students assigned to the defense team). The client meetings were in the law school classroom setting and mentored by the law professor.

As the mock litigation progressed we engaged the EM residency at large, so that the proceedings involving their colleague could be translated vicariously into an immersive learning experience for all. We utilized lectures and discussions in weekly conference, reflective writing, and “individualized interactive instruction” (for example, each resident was tasked with drafting an expert witness report). All documents (i.e., complaint, medical record, deposition transcripts) were made available online. Through these various didactic activities (Table 1) the residents learned the components malpractice litigation, followed the progress of the case, and tracked the impact it had on their colleague. The law students, meanwhile, learned and performed the various steps of pretrial litigation with the benefit of actual physician participants.
The culmination was a trial before a judge and jury in a mock courtroom at the law school (held in lieu of our regularly scheduled EM conference). The residents in the gallery, although acquainted with the details of the case, self-reported being unsettled by the sight of their friend and colleague seated with his attorneys at the defense table. As the morning progressed, they appeared captivated by the direct and cross-examinations of the plaintiff, defendant, and both expert witnesses.

**OUTCOMES**

The desired outcome was that residents, if ever involved in a medical malpractice action, will be prepared to participate in their defense with a knowledge of the process and an awareness of common mistakes and also that they will react to a malpractice action with emotional temperance. While those outcomes are not measurable, feedback from residents indicated that this immersive experience was successful in demystifying the malpractice milieu and demonstrating the factors involved in successfully navigating it. They commented specifically on empathy with the “defendant” and an increased appreciation for importance of careful clinical documentation. Course evaluations from the law students demonstrated appreciation for the challenge of litigating a mock case with actual medical experts.

**REFLECTIVE DISCUSSION**

We identify three potential limitations involved in replicating this project:

1. The law school may wish to offer the course every year, but repeating the full exercise annually would be of limited benefit to the EM residents. To address this, we have maintained the relationship between the law school and the residency by supplying faculty and resident volunteers to participate in the law school class annually, but to repeat the full simulation (for the residency program) every 3 to 4 years.

2. There were financial costs. We utilized a professional actor to portray the plaintiff and a stenographer to record the depositions; the law school paid their fees. Where this funding is unavailable, it is possible to depend on a volunteer plaintiff and to record/transcribe depositions.

3. While an argument could be made against physicians participating in the training of future plaintiff malpractice litigators, we believe that the civil justice system benefits from competent advocacy for all parties and also that the transfer of knowledge should not be limited by the hypothetical possibility that a learner might later use that knowledge to litigate a frivolous suit against a physician.

The collaboration described here is transferrable to any EM residency program in which a law school exists within reasonable geographical proximity. Interdisciplinary cooperation can add value to both legal and medical education, and knowledge sharing among EM learners and law learners (and their respective teachers) is as natural as it is mutually beneficial. The legal expertise rendered by the law school participants and the medical expertise rendered by the EM residency participants yielded a rich learning experience for all parties. The project was well adapted to the time constraints involved for an EM residency curriculum and also to the requirements and schedule for a law school elective.

The authors wish to acknowledge Drs. Howard Greller, Daniel Murphy, Rikin Shah, and Mr. James Kainen for their significant contributions to the project.

**References**


2. Guerges M, Slama E, Zayadin Y, Kieninger A. Use of a mock deposition program to improve resident


Climate change puts humanity in the crosshairs of nature, which doesn’t often miss. If you’re a doctor who cares about this issue, you know it’s time for a host of changes beyond climate. To halt human-caused climate change, we need to change a lot of things about ourselves as individuals and societies.

On this topic, Lemery and his colleagues at the University of Colorado are more than emergency physicians. They’ve gone beyond the usual, “find a problem, make an intervention, see some improvement, and move to the next situation.” Were that climate change was that easy. It’s not a single problem with one solution, because so much of the world’s existence is predicated on burning fossil fuels. We saw it coming, but we didn’t appreciate the force multipliers and how fast it would happen, and here we are. The Living Closer Foundation Fellowship in Climate & Health Policy Science is a big step in the right direction, so much so that our specialty should refine it and then do everything possible to share this educational advance across all of medicine and perhaps beyond.

Climate change is a phenomenon that will kill people—if current predictions hold, then certainly hundreds of thousands and perhaps millions. The debate is not about what happens to people who are caught in floods (they drown), are trapped by fires (they burn), inhale polluted air (they suffer lung disease), are overwhelmed by the heat (they die), or forced to migrate due to drought-famine-conflict (they lose everything). It’s about whether we’re to blame—whether we contribute to this or it’s simply evolution and fate.

What we do next depends on where we line up. If we don’t believe that climate change is causing global warming, then we do little or nothing. If we do believe that we are witnessing global warming, then we have to do many things simultaneously in two main categories—slow it down and eventually stop it, while dealing with the effects that cannot be avoided. We can’t change evolution, but we can change human behavior. Before we convinced people to stop smoking tobacco, we needed to treat their emphysema and lung cancer. Before Earth’s atmospheric temperature stabilizes, we need to be better prepared to cool people. Many of us believe that we are playing catch-up now, so time is of the essence.

Doctors may not wish to become involved. We work long hours, have electronic medical records to complete, can’t fully keep up with the issues, and feel helpless to effect change in the face of economic and political forces mightier than ours. I would argue that if there is any profession that can step up to this challenge, it’s medicine. We know how to work hard, we’re used to wading into misery, we can handle the truth, and we’re brave.

Emergency medicine has been typecast (I’m guilty for having done this) as being on the front lines. That role is important, but it’s not the one that’s going to protect humans from global warming. The truth on this matter is that as medical professionals, we are also the last line of defense. Politicians have changing motivations and may not have an enduring commitment to preventing global warming. Likewise, industry is often more concerned with profits than preservation of environmental resources. So, the medical profession must dig deep trenches and put up a line of resistance. We’ve responded to enough disasters—now we need to prevent them. Let emergency medicine lead the way.
If every day isn’t a “spare the air” day, then they will all become “shorten your lifespan” days. If we are truly in the midst of an environmental crisis, how will future generations of doctors define themselves? Will they cling to their usual safety zone and remain predominately analysts and educators, or will they find the energy and creativity to push science forward? Doctors need to do more than treat and educate. They need to not only report the news. They need to make the news. Furthermore, they need to be in the headlines—what they are doing, how they are doing it, and why it’s being done. They must be balanced but relentless—ever seeking new knowledge with the understanding that what might be at stake is not just feeling a bit better or living a few years longer, but rather, survival, not of the fittest, but of the least fit.

Within its sphere of influence, emergency medicine can gather data, train its spokespersons, treat patients, and tell tales of strength and redemption. But that sphere needs to grow quickly and forever. This fellowship should be for every specialty, not just ours. Every branch of medicine needs to glean whatever can be learned by investigating climate change as a whole and the impact of catastrophic weather events, food and water insecurity, migration of infectious disease vectors, and so forth. There is no part of anatomy or physiology untouched by the events occurring all around us. Where we suspect or observe medical implications, we must discern, confirm, and publicize.

The debate about climate change is different than the debate about nuclear war. The entire world understands the causes and effects of a nuclear explosion. Climate change is more complicated. Atomic bombs get people’s attention. Icebergs breaking off and permafrost melting do not. Doctors who coalesce to face climate change will need to learn ways to garner support for their cause, which will very importantly include funding for essential research. Doctors also need to practice what they preach. The medical profession should demonstrate that it has learned lessons from prior catastrophes and is better prepared for the next one. If we continue to see hospitals in flood zones with generators in the basement and clinics adjacent to wildland forests that burn like tinder, we’ll struggle with credibility.

We have another malignant climate change in the United States—the toxic climate of politics. The inability of our elected leaders to be reliable and responsible has removed our trust and confidence in them. From where will our leaders come? The medical profession needs to stay above the fray and remain true to its sacred commitment to better humankind by diagnosing, curing, and preventing illnesses and injuries in all their manifestations. To this mix of healers now must be added specialists in climate change and human health. We need real experts who become the go-to resources for everyone who cares about the impacts of climate change on human health. These experts will interact with scientists, educators, legislators, economists, and others who need to collaborate to conquer the greatest challenge the world has ever faced. This calls for our profession’s best minds. We need to train, nurture, and fund medical experts who look upon this challenge as their life’s calling. Our future will be in their hands.

Reference

To the Editor:

We read with great interest the commentary by Griffith et al.¹ that poses an important question: “can faculty partner with medical students to add value in the emergency department (ED), in a way that students can learn and patients benefit in the process?” As medical students we have realized that a busy ED is not always conducive to learning. Our experiences are in agreement with the authors’ statement that there can be a blurred line between clinical learning opportunities and “scutwork.”

In addition to the value-added learning opportunities outlined in the article, we would like to propose medical student–led quality improvement (QI) projects as another method of experiential learning in the ED. As part of our fourth-year curriculum, we are required to design and manage a 6-month QI project in a U.K. hospital.

Manning et al.² have previously emphasized the value and feasibility of adding QI in the medical student curriculum. Our experiences have reinforced this, as undertaking the project enabled us to integrate into the clinical team and provided an insight into the clinical workflow. It offered the opportunity to identify areas requiring improvement and proactively seek solutions. Griffith et al.¹ mention that brief student rotations in the ED do not allow for clinical educators to develop a relationship with their students. Managing a project from its inception to completion has allowed us to partner with a senior clinician and the ED staff for an extended period of time. For the first time, we have been able to work with clinicians as colleagues in contrast to the traditional “student–teacher” relationship. Instead of feeling like a burden in a hectic department, undertaking a QI project has given us a sense of empowerment to make positive change. We have developed key research skills including data collection, critical analysis and academic writing. Many students have gone on to present their QI project at conferences and submit their work for publication. This has served as a foundation of knowledge and confidence for future research.

However, running a QI project within the ED is not without its difficulties. Issues encountered include poor engagement and high turnover of staff affecting the uptake of interventions. Therefore, our first recommendation for introducing value-added QI projects in the ED is to have a dedicated supervisor per student group. This allows for clear direction for the project and successful integration into the ED. We also recommend structured QI implementation tutorials, in the form of online modules to provide students with adequate training. Our final recommendation is to allow medical students to indicate an area of interest in which to undertake the project, increasing student engagement and motivation.

In conclusion, we strongly believe that adding QI projects to the medical student curriculum echoes the stance of Griffith et al.¹ stance on medical student value-based learning opportunities. QI projects do not qualify as “scutwork” as they provide a plethora of learning opportunities for medical students and clinicians, enabling students to directly contribute to patient care and safety within the ED.
Haridha Sudhakara Pandian, BSc (Hons)
(haridha.pandian@kcl.ac.uk)
Priyanka Shivaraman Iyer, BSc (Hons)
Jasmine Kaur Bhangra, BSc (Hons)
Aditi Nijhawan, BSc (Hons)
Faculty of Life Sciences and Medicine, King’s College
London, Guy’s Campus, London, UK

All authors have contributed equally to this manuscript.

Supervising Editor: Susan B. Promes, MD, MBA, FACEP

References