

# AEM Education and Training

A GLOBAL JOURNAL OF EMERGENCY CARE

Elevating the human condition  
during times of emergency



## ORIGINAL CONTRIBUTIONS

### **Resident Supervision and Patient Care: A Comparative Time Study in a Community-Academic Versus a Community Emergency Department**

Ernest E. Wang, Yue Yin, Itai Gurvich et al.

308

### **Does the Emergency Medicine In-training Examination Accurately Reflect Residents' Clinical Experiences?**

Jason J. Bischof, Geremiha Emerson, Jennifer Mitzman et al.

317

### **Feedback With Performance Metric Scorecards Improves Resident Satisfaction but Does Not Impact Clinical Performance**

Mira Mamtani, Frances S. Shofer, Alexander Sackeim et al.

323

### **Navigating Cognitive Dissonance: A Qualitative Content Analysis Exploring Medical Students' Experiences of Moral Distress in the Emergency Department**

Caitlin Schrepel, Joshua Jauregui, Alisha Brown et al.

331

# AEM Education and Training

Contents, continued

Volume 3 • Number 4 • October 2019 • <https://onlinelibrary.wiley.com/journal/24725390>

## **A Retrospective Cohort Study of the Effect of Home Institution on Emergency Medicine Standardized Letters of Evaluation**

Megan Boysen-Osborn, Jessica Andrusaitis, Clelia Clark et al.

340

## **Use of the “Step-back” Method for Education Research Consultation at the National Level: A Pilot Study**

Jaime Jordan, Kaushal Shah, Andrew W Phillips et al.

347

## **Clinical Improvement Interventions for Residents and Practicing Physicians: A Scoping Review of Coaching and Mentoring for Practice Improvement**

Casey MacKenzie, Teresa M. Chan, Shawn Mondoux

353

## **Training in the Management of Psychobehavioral Conditions: A Needs Assessment Survey of Emergency Medicine Residents**

Jason Pickett, Mary Rose Calderone Haas, Megan L. Fix et al.

365

---

### **SOAR REVIEW**

#### **Systematic Online Academic Resource (SOAR) Review: Renal and Genitourinary**

Andrew Grock, Anuja Bhalerao, Teresa M. Chan et al.

375

---

### **BRIEF CONTRIBUTION**

#### **The Revised METRIQ Score: A Quality Evaluation Tool for Online Educational Resources**

Isabelle N. Colmers-Gray, Keeth Krishnan, Teresa M. Chan et al.

387

---

### **NEW IDEAS IN B-E-D-SIDE TEACHING**

#### **Spokes for Our Folks: Public Health Bike Tour**

Christian Rose, Brian Chang, John Brown

393

---

### **COMMENTARY—INVITED**

#### **Embracing FOAM as an Academic Community**

Nicole Munz, Esther H. Chen

396

#### **Thinking Critically About Appraising FOAM**

Teresa M. Chan, Anuja Bhalerao, Brent Thoma et al.

398

---

### **COMMENTARY AND PERSPECTIVE**

#### **Pupil Prose Appraisal: Four Practical Solutions to Medical Student Documentation and Feedback in the Emergency Department**

Mark F. Olaf

403

# AEM Education and Training Editorial Board

A Journal of the Society for Academic Emergency Medicine | [saem.org](http://saem.org)

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, FRCPC**  
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

# EDITORIAL OFFICES

## Peer Review Coordinator

Taylor Bowen  
tbowen@saem.org  
Tel 434.218.3366

## Journal Manager

Stacey Roseen  
sroseen@saem.org  
Tel 847.813.9823 x207

## SAEM

1111 E. Touhy Ave, • Suite 540  
Des Plaines, IL 60018  
Tel 847.813.9823  
Fax 847.813.5450  
saem@saem.org

## Copyright and Copying (in any format)

Copyright © 2019 Society for Academic Emergency Medicine. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA, [www.copyright.com](http://www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising or promotional purposes, for republication, for creating new collective works or for resale. Permissions for such reuse can be obtained using the RightsLink "Request Permissions" link on Wiley Online Library. Special requests should be addressed to [permissions@wiley.com](mailto:permissions@wiley.com)

## Disclaimer

The Publisher, the Society for Academic Emergency Medicine (SAEM) and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, the Society for Academic Emergency Medicine (SAEM) and Editors, neither does the publication of advertisements constitute any endorsement by the Publisher, the Society for Academic Emergency Medicine (SAEM) and Editors of the products advertised.

For submission instructions, subscription and all other information visit [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)2472-5390](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2472-5390)





**Academic Emergency Medicine 1994 – 2019**



# Resident Supervision and Patient Care: A Comparative Time Study in a Community-Academic Versus a Community Emergency Department

Ernest E. Wang, MD,<sup>1</sup> Yue Yin, PhD,<sup>2</sup> Itai Gurvich, PhD,<sup>3</sup> Morris S. Kharasch, MD,<sup>1</sup> Clifford Rice, MD,<sup>1</sup> Jared Novack, MD,<sup>1</sup> Christine Babcock, MD, MSc,<sup>4</sup> James Ahn, MD, MHPE,<sup>4</sup> Steven H. Bowman, MD,<sup>5</sup> and Jan A. Van Mieghem, PhD<sup>2</sup>

## ABSTRACT

**Objective:** The objective was to compare attending emergency physician (EP) time spent on direct and indirect patient care activities in emergency departments (EDs) with and without emergency medicine (EM) residents.

**Methods:** We performed an observational, time-motion study on 25 EPs who worked in a community-academic ED and a nonacademic community ED. Two observations of each EP were performed at each site. Average time spent per 240-minute observation on main-category activities are illustrated in percentages. We report descriptive statistics (median and interquartile ranges) for the number of minutes EPs spent per subcategory activity, in total and per patient. We performed a Wilcoxon two-sample test to assess differences between time spent across two EDs.

**Results:** The 25 observed EPs executed 34,358 tasks in the two EDs. At the community-academic ED, EPs spent 14.2% of their time supervising EM residents. Supervision activities included data presentation, medical decision making, and treatment. The time spent on supervision was offset by a decrease in time spent by EPs on indirect patient care (specifically communication and electronic health record work) at the community academic ED compared to the nonacademic community ED. There was no statistical difference with respect to direct patient care time expenditure between the two EDs. There was a nonstatistically significant difference in attending patient load between sites.

**Conclusions:** EPs in our study spent 14.2% of their time (8.5 minutes/hour) supervising residents. The time spent supervising residents was largely offset by time savings related to indirect patient care activities rather than compromising direct patient care.

From the <sup>1</sup>Division of Emergency Medicine, NorthShore University HealthSystem, Evanston, IL; the <sup>2</sup>Department of Operations, Kellogg School of Management, Northwestern University, Evanston, IL; the <sup>3</sup>Cornell Tech and School of Operations Research and Information Engineering, Cornell University, New York, NY; the <sup>4</sup>Section of Emergency Medicine, University of Chicago Pritzker School of Medicine, Chicago, IL; and the <sup>5</sup>Department of Emergency Medicine, Cook County Health and Hospital System, Rush Medical College, Chicago, IL.

Received October 17, 2018; revision received February 22, 2019; accepted March 11, 2019.

Poster was displayed on Medical Education Day, University of Chicago Pritzker School of Medicine, November 16, 2017.

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

**Author contributions:** EEW and JVM conceived the study and designed the trial; YY, JVM, and IG designed the data collection instrument; MSK and EEW were responsible for coordinating physician scheduling; CR, JN, JA, CB, and SHB were responsible for manuscript review and coordination of resident participation; YY was primarily responsible for observation, data collection, and managing the data; YY and IG provided statistical oversight and analysis as well as manuscript preparation; and EEW, YY, and JVM were primarily responsible for manuscript generation and takes responsibility for the paper as a whole.

Supervising Editor: Sally Santen, MD, PhD.

Address for correspondence and reprints: Ernest E. Wang, MD, FACEP; e-mail: ewang@northshore.org.

AEM EDUCATION AND TRAINING 2019;3:308–316

## BACKGROUND

One of the primary benefits of residency training is the opportunity for residents to observe the work of, and learn from the direct interaction with, the attending emergency physician (EP).<sup>1</sup> Direct attending–resident interaction is also considered one of the best assessment tools for resident competencies.<sup>2</sup>

In 2001, DeBehnke wrote, “... [e]ducating and supervising residents and students while simultaneously providing patient care requires quantifiable faculty time and effort. Academic EDs must identify this time and effort accurately since providing this joint product line has the potential to make our emergency care system inefficient.”<sup>3</sup>

Maintaining effective supervision can be very costly for emergency departments (ED). First, engaging with residents may limit the time that EPs devote to patient care and ED flow. This additional demand on the EP may also increase stress and anxiety, potentially leading to burnout. Second, the EP is responsible for calibrating the level of supervision based on each resident’s knowledge and clinical skill. Residents who lack requisite knowledge with complex cases may struggle and potentially cause patient harm.<sup>4</sup> Third, supervising residents consumes resources. Evidence shows that care at academic hospitals is less cost-effective than care at nonacademic hospitals because of higher frequency of testing and other resource use in the teaching setting.<sup>5</sup>

Resident training takes place in a variety of ED settings. EDs can broadly be categorized into three types: *academic*—defined as university-based, teaching hospitals; *community-academic* (which will be referenced as “CAED” in the remainder of this paper) defined as community EDs with residents and/or students who rotate through and are supervised by EPs; and *nonacademic* EDs (which will be referred as “community ED” in the remainder of this paper) defined as an ED with no learners. In CAEDs, resident supervision is but one of many responsibilities that must be skillfully orchestrated by EPs, alongside other essential tasks such as direct patient care, communication, and documentation.

There are limited data identifying the differences in the time and effort spent by EPs at CAEDs in comparison to community EDs. Chisholm et al.<sup>2,6</sup> tracked EPs’ time expenditures on direct/indirect patient care and personal activities in academic versus community EDs, but did not specify the time spent supervising

residents or performing other care-related activities. Other studies only assessed aggregate effect of residents on departmental throughput.<sup>7–11</sup> To the best of our knowledge, this is the first study to comprehensively quantify and compare the time EPs spend on resident supervision and care-related activities in CAEDs versus community EDs.

## Objective of Investigation

The objective of this study was to compare the time utilization profiles of a group of EPs in a community ED where patients were the only “customers”<sup>3</sup> versus that of the same group of EPs in a CAED where patients and residents generate competing demands for EPs’ time. First, we quantified the time EPs “reallocate” to resident supervision at the CAED. Second, we determined the categories of activities from which EPs shaved time to accommodate supervision.

## METHODS

### Study Site

This study was approved by the first and second author’s respective institutional review boards.

We conducted an observational, time–motion study at a 25-bed CAED versus a 15-bed community ED. Both are Level II trauma centers within a four-hospital health system in northern Illinois. The health system is the primary affiliate and community training site for a university-based emergency medicine (EM) residency program. The CAED trains residents from multiple Accreditation Council for Graduate Medical Education (ACGME)-accredited residency programs. The community ED does not have ACGME trainees. Table 1 shows key performance metrics for both EDs during the fiscal year of 2017.

The EDs have comparable attending physician staffing ratios despite a 10,000 patient census differential between the CAED and community ED. The same group of EPs staff both EDs in our study. Both EDs have double EP coverage from 9 AM to midnight and single coverage overnight. During the double-coverage hours, each EP is supported by a separate team of nurses and is responsible for an evenly divided subset of rooms. Chart documentation is completed via dictation or by typing directly into the electronic health record (EHR). Neither ED has scribes.

Most residents in our study were second- (PGY-2) and third- (PGY-3) year emergency medicine residents. All patients at the CAED evaluated by a resident are

**Table 1**  
Performance Metrics for Community-Academic and Community in FY2017

	Community-Academic	Community
ED visits	37,042	27,555
% Observation admissions	13.3	14.5
% Inpatient admissions	14.7	13.5
LWBS	708	352
Pediatric volume	3,993	4,881
Median turnaround time (minutes)		
Admission	271	242
Discharge	196	178
% Female	55.3	56.0
% Male	44.7	44.0
ESI level (%)		
1	1.0	0.3
2	28.1	20.0
3	52.5	56.4
4	16.3	20.3
5	0.4	0.5
Medicine admissions	8,609	6,363
Surgery admissions	1,831	1,420
Avg. hourly patients seen per shift	2.8	2.2

ESI = Emergency Severity Index; LWBS = leave without being seen.

staffed by an EP as well, which corresponds to practice previously described.<sup>12</sup> In general, a resident takes the patient's history, performs a physical examination, and then presents the findings to the EP. During the presentation, the resident ideally proposes a differential diagnosis and management plan. The EP then provides feedback to the resident (i.e., probing and clarifying questions, nodding, correcting, and using case vignettes from other related cases as supporting evidence). In this study, we consider such care-related, direct EP–resident interaction as the main activity of *supervising residents*. After the discussion, the EP examines the patient independently with or without the resident. After the independent evaluation, the EP continues supervision by evaluating the resident's assessment and modifying the treatment plan together with the resident, if needed. Orders for medications, procedures, and laboratory tests are entered into the EHR and communicated to the nurse, either by the resident or by the EP herself. Residents and EPs both write notes on the patient. The resident notes are generally complete history and physical examination findings, as well as assessments and plans that comprise

the medical decision making. EPs generally write an abbreviated supervisory or physicians at teaching hospitals (PATH) note as stipulated by Centers for Medicare & Medicaid Services. Importantly, the residents at the CAED do not take the EPs role in all these care-related activities. The EP remains the chief provider of care, with the supervision of residents being an added responsibility.

## Data Collection

We followed the activity categorization developed by Tipping et al.,<sup>13,14</sup> in which they suggested to include the primary categories of “direct” and “indirect” patient care. We also followed their definition of direct patient care as “those activities involving face-to-face interaction between the [EP] and the patient.”<sup>13</sup> In our study, all the time that EPs spent in patient rooms is quantified as *direct* patient care. Accordingly, bedside charting or teaching is counted toward direct patient care because the EPs are performing these activities in the presence of patients. *Indirect* patient care includes activities “relevant to the patient's care but not performed in the presence of the patient.”<sup>13</sup> Following Tipping et al.,<sup>14</sup> we adjusted to the specifics of our sites by adding customized subcategories. We kept refining the subcategories to ensure that they were “easily observable and identifiable without subjective interpretation from the observer” during a pilot study.<sup>14</sup> Two experienced EPs in our research team helped finalize the categorization (see Table 2). The multitasking activity only involved an EP simultaneously communicating with other nonresident providers while working on EHR. A resident presenting a case while the EP is reviewing the related EHR is categorized as supervision.

## Hypothesis Development

**Hypothesis 1.** A recent study by Hexom et al.<sup>15</sup> reported a mean resident supervision time by EM faculty of 60.8 minutes over 8-hour observation periods. Chisholm et al.<sup>2</sup> also demonstrated that EM faculty devoted 11.9% of their time to resident supervision. Therefore, we expected to see EPs in our study tailoring resident supervision to their workflow at the CAED.

**Hypothesis 2.** In order to fulfill the supervisory responsibility EPs would have to reallocate time spent on other care-related activities. We hypothesized that EPs would delegate portions of *indirect* patient care activities to residents, such as communication and EHR work.

**Table 2**  
Categories of Activity

Primary Category	Main Category	Subcategory	Definition
Direct patient care	Direct patient care		EPs examine, treat, or communicate with patient at bedside or explain treatment to patient's family
Indirect patient care	EHR	Reviewing	EPs collect information from the EHR (e.g., reviewing patient history, test results, radiology images, and ECGs).
		Charting and putting orders	EPs enter information into the EHR (e.g., writing patient note, order test and medicines).
	Communication	Phone calls and consults	EPs communicate with other care providers (e.g., hospitalists and specialists) over the phone for patient care purposes
		Face-to-face communication with other providers	EPs communicate with other nonresident care providers (e.g., nurses and technicians) for patient care purposes
	Multitasking	Communicating with other non-resident providers while working on EHR	EPs communicate with other nonresident care providers for patient care purposes, either over the phone or face to face, while working on EHR
Supervision	Supervising residents		EPs interact with residents for patient care purposes without the presence of patients
Other	Personal		EPs get food, eat, visit the restroom, check private phone/e-mail or requests privacy explicitly; EPs chat with other people for nonclinical purpose
	Transit/travel		EPs wait, move, and wash their hands

ECG = electrocardiogram; EHR = electronic health record.

**Hypothesis 3.** It was unclear how EP direct patient care time would differ between ED settings. On the one hand, patient care by residents might reduce the time that EPs spend on direct patient care. On the other hand, bedside teaching of residents would increase this time. It was also unclear which of these effects would materialize in practice. EPs may constrain the autonomy of residents for reasons of efficiency and safety. Previous studies show limited use of bedside teaching in practice as this also may take more time.<sup>2,16</sup> We hypothesized that time spent by EPs on direct patient care at the CAED would be less than that at the community ED.

### Pilot Observation

In October 2016, we conducted nine pilot observations totaling 1,764 minutes to design the data collection process and to train the observer (a third-year PhD candidate in operations). The observer was trained to be responsive to EP's activity transition and to rapidly track them down. During the pilot study, the observed EPs explained to the observer his/her current activity (e.g., "I am charting" or "I am going to have lunch"). The observer also asked the EPs about ambiguous situations to avoid misinterpreting the activities being performed. Because we only had a single observer in the study, there was no concern for interobserver reliability or consistency in measurement.

### Selection of Participants

We selected a nonrandomized convenience sample from a total of 51 EPs in this health system. The primary inclusion criteria required that the participating EP worked at both EDs during the study. Authors were excluded from the sample. This resulted in 30 eligible EPs, from which 25 EPs verbally consented. They were informed that the study was about ED workflow but were blinded from the specific objective. Data confidentiality was ensured.

### Study Protocol

We observed each EP twice at each ED for a total of 100 observations (50 at the CAED and 50 at the community ED.) The length of a single observation session was 240 minutes,<sup>6,17</sup> totaling 400 observation hours. Importantly, the EPs served as their own control. The observed shifts were a convenience sample but were evenly distributed over the days of the week and AM/PM shifts.

Residents, nurses, and advanced practice providers were not observed directly, but their interactions with the observed EP were recorded. The observer shadowed an individual EP continuously except when the EP was inside the patient room or requested privacy. Tracked information included locations, categorized activity, with start and end times for activity executions. All observational data were recorded using an



iPad app called “Eternity.” To minimize the Hawthorne effect, the observer maintained a “safe distance” from the observed EP and did not engage in conversation during the observation periods. To protect patient privacy, the observer did not enter patient rooms. The iPad was held by the observer and was obscured from the EP during observations.

We calculated patient load and left without being seen (LWBS) rate for each observation using the patient-level turnaround data derived from the health system's data warehouse. Patient load was calculated by adding the number of new patients assigned to the observed EP who received care by that observed EP during the observation, to the number of patients already under the observed EP's care at the beginning of the observation.<sup>13</sup> We also calculated the percentages of patient load in each Emergency Severity Index (ESI) level. ESI provides clinically relevant stratification of patients into five levels from 1 (most urgent) to 5 (least urgent) based on acuity and resource needs.<sup>18</sup> The LWBS rate was defined as the ratio of the number of LWBS new patients (i.e., new patients who arrived during the observation and were assigned to the observed EP, but subsequently left without being seen by this EP),<sup>19</sup> to the number of total new patients who arrived during the observation and were assigned to the observed EP.

The supervised residents were queried confidentially using a survey compiled from previously reported surveys.<sup>20</sup> Residents were asked to rate “How sufficient is the supervision you received from attending in the past four hours” on a three-point Likert scale with the descriptors: “very sufficient,” “sufficient,” and “not at all/slightly sufficient.”<sup>21</sup> The residents were also asked to “Describe your learning outcome in the past four hours” using a three-point scale with the descriptors: “I have learned a significant amount,” “I have learned something,” and “I didn't learn anything.” These surveys were completed shortly after the observations (within 2 hours).

### Measurements and Primary Data Analysis

The time spent on each activity was measured in seconds. Mean time for each main category at community ED versus CAED were presented in percentages (out of a 240-minute observation). We then reported the detailed time spent by EPs on each subcategory across two EDs during observations. We also reported the time an EP spent per patient load on each subcategory during an observation, using the total time the EP

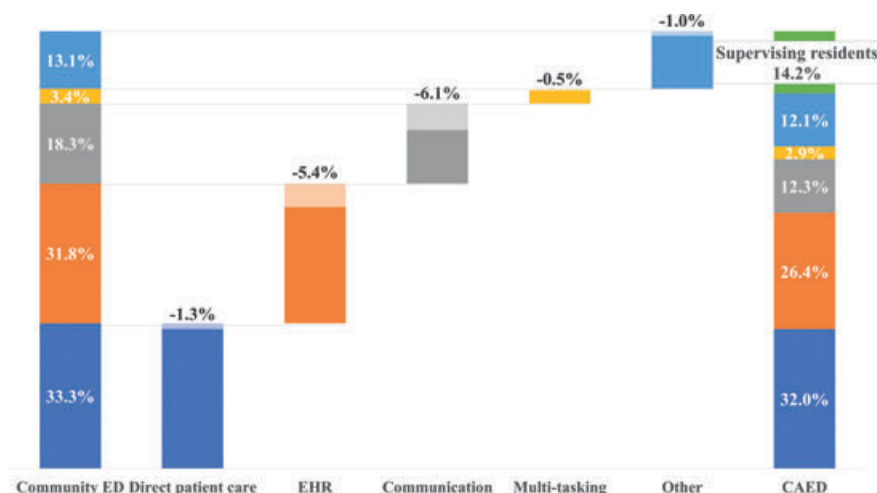
spent on the subcategory divided by the patient load during this observation.<sup>13</sup>

Most of our time-spent data did not follow normal distribution (see Data Supplement S1, Tables A.1 and A.2, available as supporting information in the online version of this paper, which is available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10334/full>); we thereby present the median and interquartile ranges (IQRs) of minutes spent on each subcategory at CAED versus community ED. A Wilcoxon two-sample test was performed. To control the Type I error, we corrected the p-values using the Benjamini-Hochberg procedure with a false discovery rate (FDR) of 0.05.<sup>22,23</sup> Two-sided FDR-adjusted p-values < 0.05 were considered statistically significant. All analyses were run on R version 3.4.3.<sup>24</sup>

## RESULTS

During the formal observational study from March 2017 to August 2017, the 25 observed EPs performed a total of 35,348 executions of the subcategorized activities during the entire 400 observation hours. On average 355 activities were performed per observation at the CAED versus 332 at the community ED ( $p = 0.11$ ). The average patient load across observations was 15 at the CAED versus 13 at the community ED ( $p = 0.40$ ). The ESI scores of these patients were significantly different between the CAED versus community ED for levels 1 to 3: level 1 (1.1% vs. 0.1%,  $p = 0.008$ ), level 2 (37.7% vs. 20.7%,  $p < 0.001$ ), and level 3 (47.9% vs. 62.7%,  $p < 0.001$ ). There was no difference in the average number of patients discharged by the participating EP during the observations (CAED = 7 vs. community = 7,  $p = 0.59$ ). The average LWBS rates were similar (CAED = 1% vs. community = 0.9%,  $p = 0.33$ ).

Figure 1 illustrates the time utilization profile of an average EP per 240-minute observation session. EPs spent 34.2 minutes (14.2%) supervising residents (8.5 minutes/hour). Direct patient care accounted for 76.8 minutes (32%) versus 79.9 minutes (33.3%;  $p = 0.31$ ) at the CAED and community ED, respectively. Indirect patient care accounted for 99.8 minutes (41.6%) and 128.4 minutes (53.5%) at the CAED and community ED, respectively ( $P < 0.001$ ). Non-care-related activities (personal time and travel) accounted for 12.1% versus 13.1% ( $p = 0.96$ ), respectively, and did not statistically differ.



**Figure 1.** Mean task percentages for the 240-minute observations. CAED = community-academic ED; EHR = electronic health record.

Comparing the CAED to the community ED, significant median time decreases were found in EHR review (32.2 minutes vs. 23.9 minutes, FDR-adjusted  $p = 0.003$ ), charting and placing orders (41.6 minutes vs. 36.7 minutes, FDR-adjusted  $p = 0.029$ ), face-to-face communication (25.8 minutes vs. 17.9 minutes, FDR-adjusted  $p = 0.002$ ), phone calls/consult communication (17.1 minutes vs. 8.4 minutes, FDR-adjusted  $p < 0.001$ ), and multitasking (7.2 minutes vs. 4.32 minutes, FDR-adjusted  $p = 0.031$ ). Personal time did not change significantly.

Median minutes spent per patient load on communication, either face to face or by phone, at the CAED decreased by almost 40%. EPs working at the CAED spent 1.14 median minutes less reviewing in EHR (FDR-adjusted  $p = 0.028$ ). No significant change was found with respect to direct patient care, charting and putting orders in EHR, multitasking, and other non-patient care activities.

## Survey Results

Thirty-one residents completed 47 unique session survey responses (response rate is 78.3% from 60 session surveys). One hundred percent of the responses described the supervision as “very sufficient.” Forty-three percent of the responses reported having “learned something” and 57% reported having “learned a lot” during the corresponding sessions.

## DISCUSSION

Building on DeBehnke’s call for a more refined understanding of the time and effort expended on educating and supervising residents, we studied how

EPs adjust their clinical practices when resident supervision is added to their responsibilities. The primary strength of this study is an extensive data set and a subject group that served as their own controls. To our knowledge, ours is the first time-motion study to fully map the time utilization profile of EPs working with and without residents. While this information represents two EDs and mostly upper-level residents, we believe that it is a reasonable starting point for other studies on this important topic.

The key findings are: 1) EPs spent a substantial portion of their clinical time supervising residents, 2) EPs spent the majority of their clinical time in direct patient care, and 3) EPs experienced a significant reduction in indirect patient care when working with residents.

First, our EP cohort spent 14.2% of their time supervising residents. This translates to 68 minutes over an 8-hour shift, consistent with the supervision time found by Hexom et al.<sup>15</sup> Time spent interacting with EPs may determine perceived education quality by the residents.<sup>15</sup> This also echoes the positive responses from the residents to our survey.

Second, we confirmed that our EPs spent the majority of their time performing direct patient care.<sup>2,15,17</sup> More importantly, even though the EPs devoted a substantial amount of their time to resident supervision at the CAED, direct patient care time did not change significantly. Direct patient care, the primary priority for our EPs, was preserved despite substantial effort toward supervision. Consistent with the findings by Chisholm et al.,<sup>2</sup> seldom do the EP and the resident at the CAED work simultaneously at the patient’s bedside, except for verification of resident findings on history and physical or procedural supervision.

Third, we observed that time savings from EPs' offloading indirect patient care activities to the residents largely offset the supervision time: EM residents contributed to indirect patient care and expanded EPs capacity in addition to the direct patient care they provided—a “win-win situation” of resident supervision. In our setting, adding an intermediate or upper-level EM resident to the CAED team extended the EPs' ability to evaluate approximately 10,000 more patients per year (who were arguably sicker by our ESI score) with comparable staffing and the same bedside time.

Finally, residents freed EPs from specific indirect patient care activities such as communication and EHR work. At both aggregate and per-patient-load levels, EPs delegated most of the communication to their residents, spending significantly less time making phone calls or face-to-face communication to other health care providers. The EPs providing supervision in exchange for release from the tasks can be viewed as an apprenticeship-type of experience: “The resident will perform these indirect patient care tasks under my supervision, and in return, I will provide the resident with diagnostic feedback on how to perform them better and more effectively.” The presence of residents also endows the EPs with discretion as to how and when to expense their efforts—a “currency of resident apprenticeship.” Several of our EPs suggested that when working at the CAED, they delegated more of the charting to residents when the patient load was low, but would complete more of it on their own when the patient load was high so that residents can help do more direct patient care to accelerate the patient throughput. When working at the community ED, their time spent per patient on charting is relatively stable and independent of the patient load. This anecdotal evidence is consistent with the substantially wider IQR of per-patient-load EHR charting time at the CAED (Data Supplement S1, Table S.4).

## LIMITATIONS

Our results should be interpreted in the light of several limitations. First, some possible confounding factors were beyond our control. For example, the CAED had a larger concentration of high-acuity patients both anecdotally and suggested by their ESI scores. This would bias the observed change in time on direct patient care upward at the CAED, as critical patients typically require more direct patient care. Such bias was difficult to avoid without randomizing patients across EDs and

EPs. Meanwhile, ESI has been criticized for its low accuracy and high variability in clinical practice.<sup>25,26</sup> Observing differences in patient ESI distributions thereby may not suffice to confirm the difference in the actual acuity distributions across two EDs.

Patient load was another potential confounder. With high patient load, EPs probably deferred documentation to make more time for direct patient care. We tried to balance the distribution of patient loads across two EDs by spreading the observations over days of week and AM/PM shifts. The average patient load across observations at the CAED turned out to be higher than at the community ED (15 vs. 13). This difference was not statistically significant ( $p = 0.40$ ), probably because of the large variance in patient loads, but anecdotally, the EPs “feel” the difference of simultaneously carrying more patients at the CAED. To further adjust for the patient load, we reported the average time spent per patient load on each subcategory in Table S.4. Future study with more observations may achieve more consistent estimate and comparison of patient loads.

A potentially stronger method to eliminate these biases introduced by systematic differences in different EDs is to compare the same EP at the same ED on shifts with versus shifts without residents. Although we did not find such opportunity, Salazar et al.<sup>27</sup> achieved this by using a resident strike period as the control and compared quality indicators of patient care during days when residents were on duty versus on strike.

Second, we did not separate the time spent on bedside teaching from direct patient care time. This observation rubric followed prior reported studies.<sup>14</sup> Furthermore, the supervision time we recorded already accounted for a major part of the total EM resident-faculty interaction time as shown in previous study.<sup>2</sup>

Third, we did not capture EPs' time spent after shifts. EPs might have to stay late after shifts to complete patient notes. More data capturing EPs' work after shifts would provide a more holistic analysis of resident effect on their time utilization profile.

Fourth, the accuracy of the single nonclinical observer's interpretation, as well as the Hawthorne effect may limit the results. But we believe that the pilot observations totaling almost 30 hours provided sufficient practice for the observer to achieve reasonable recording accuracy and to remain as unobtrusive as possible during observations.<sup>13</sup> Besides, the inherent bias of the observer's assessment would be carried across both EDs and would not affect the comparison



results. Having multiple observers may make the study more robust and replicable, but that imposes a higher requirement for resources and study design.<sup>28,29</sup>

Fifth, we did not capture overnight shifts in our observations, concerning that the difference in coverage and availability of other medical providers may confound the results. Future research specifically focusing on potential differences of the resident effect on overnight shifts is warranted.

Sixth, these results may not be generalizable to other CAEDs and traditional university-based teaching hospitals, where staffing, supervision, and institutional culture can be quite different. For example, our resident sample consisted entirely of intermediate and upper-level residents. The results may not generalize to institutions that were staffed primarily with junior trainees (i.e., medical students and interns), as senior residents are reported to be more productive than junior trainees.<sup>30–32</sup>

Finally, we had limited number of observations per EP per ED. A larger sample size, with a moderate-large number of observations per EP, per ED, or even under different levels of patient load and acuity, would support a more detailed empirical study that goes beyond comparison of descriptive statistics aggregated at ED level.

## CONCLUSION

Emergency physicians working with emergency medicine residents in a community-academic ED reap significant time savings from the responsibilities of indirect patient care and remunerate those savings in kind to the residents in the form of supervision which accounted for 14.2% of their clinical time. These time savings allow them to foster a clinical learning environment “where residents and fellows can interact with patients under the guidance and supervision of qualified faculty members who give value, context, and meaning to those interactions” and achieve “balance of service and education” in residency training.<sup>33,34</sup> More importantly, it allows community academic emergency physicians to preserve their ability to provide direct high-quality safe patient care, which remains their core mission.

## References

- Merritt C, Shah B, Santen S. Apprenticeship to entrustment: a model for clinical education. *Acad Med* 2017;92:1646.
- Chisholm CD, Whenmouth LF, Daly EA, Cordell WH, Giles BK, Brizendine EJ. An evaluation of emergency medicine resident interaction time with faculty in different teaching venues. *Acad Emerg Med* 2004;11:149–55.
- DeBehnke DJ. Clinical supervision in the emergency department: a costly inefficiency for academic medical centers. *Acad Emerg Med* 2001;8:827–8.
- Atzema C, Bandiera G, Schull MJ, Coon TP, Milling TJ. Emergency department crowding: the effect on resident education. *Ann Emerg Med* 2005;45:276–81.
- Pitts SR, Morgan SR, Schragger JD, Berger TJ. Emergency department resource use by supervised residents vs attending physicians alone. *JAMA* 2014;312:2394.
- Chisholm CD, Weaver CS, Whenmouth L, Giles B. A task analysis of emergency physician activities in academic and community settings. *Ann Emerg Med* 2011;58:117–22.
- Svirsky I, Stoneking LR, Grall K, Berkman M, Stolz U, Shirazi F. Resident-initiated advanced triage effect on emergency department patient flow. *J Emerg Med* 2013;45:746–51.
- Bhat R, Dubin J, Maloy K. Impact of learners on emergency medicine attending physician productivity. *West J Emerg Med* 2014;15:41–4.
- McGarry J, Krall SP, McLaughlin T. Impact of resident physicians on emergency department throughput. *West J Emerg Med* 2010;11:333–5.
- Anderson D, Golden B, Silberholz J, Harrington M, Hirshon JM. An empirical analysis of the effect of residents on emergency department treatment times. *IIE Trans Healthc Syst Eng* 2013;3:171–80.
- Henning DJ, McGillicuddy DC, Sanchez LD. Evaluating the effect of emergency residency training on productivity in the emergency department. *J Emerg Med* 2013;45:414–8.
- Cobb T, Jeanmonod D, Jeanmonod R. The impact of working with medical students on resident productivity in the emergency department. *West J Emerg Med* 2013;14:585–9.
- Tipping MD, Forth VE, O’Leary KJ, et al. Where did the day go?—A time-motion study of hospitalists. *J Hosp Med* 2010;5:323–8.
- Tipping MD, Forth VE, Magill DB, Englert K, Williams MV. Systematic review of time studies evaluating physicians in the hospital setting. *J Hosp Med* 2010;5:353–9.
- Hexom B, Trueger NS, Levene R, Ioannides KL, Cherkas D. The educational value of emergency department teaching: it is about time. *Intern Emerg Med* 2017;12:207–12.
- Aldeen AZ, Gisondi MA. Bedside teaching in the emergency department. *Acad Emerg Med* 2006;13:860–6.
- Füchtbauer LM, Nørgaard B, Mogensen CB. Emergency department physicians spend only 25% of their working time on direct patient care. *Dan Med J* 2013;60:A4558.

18. Gilboy N, Tanabe T, Travers DR. Emergency Severity Index (ESI) a triage tool for emergency department care implementation, version 4. In: Implementation Handbook 2012 Edition. AHRQ Publication No. 12-0014. Rockville, MD: Agency for Healthcare Research and Quality, 2011. <https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/systems/hospital/esi/esihandbk.pdf>. Accessed October 8, 2018.
19. Polevoi SK, Quinn JV, Kramer NR. Factors associated with patients who leave without being seen. *Acad Emerg Med* 2005;12:232–6.
20. Holt KD, Miller RS, Philibert I, Heard JK, Nasca TJ. Residents' perspectives on the learning environment: data from the Accreditation Council for Graduate Medical Education resident survey. *Acad Med* 2010;85:512–8.
21. Jacoby J, Matell MS. Three-point likert scales are good enough. *J Market Res* 1971;8:495.
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
23. Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 2002;15:870–8.
24. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2015, 2018. <https://scholar.google.com/scholar?cluster=4702143145247627313&hl=en&oi=scholar>. Accessed September 6, 2018.
25. Mistry B, Stewart De Ramirez S, Kelen G, et al. Accuracy and reliability of emergency department triage using the Emergency Severity Index: an international multicenter assessment. *Ann Emerg Med* 2018;71:581–7.e3.
26. Buschhorn HM, Strout TD, Sholl JM, Baumann MR. Emergency medical services triage using the Emergency Severity Index: is it reliable and valid? *J Emerg Nurs* 2013;39:55–63.
27. Salazar A, Corbella X, Onaga H, Ramon R, Pallares R, Escarabill J. Impact of a resident strike on emergency department quality indicators at an urban teaching hospital. *Acad Emerg Med* 2001;8:804–8.
28. Lopetegui MA, Bai S, Yen PY, Lai A, Embi P, Payne PR. Inter-observer reliability assessments in time motion studies: the foundation for meaningful clinical workflow analysis. *AMIA Annu Symp Proc* 2013;2013:889–96.
29. Zheng K, Guo MH, Hanauer DA. Using the time and motion method to study clinical work processes and workflow: methodological inconsistencies and a call for standardized research. *J Am Med Inform Assoc* 2011;18:704–10.
30. Jeanmonod R, Jeanmonod D, Ngiam R. Resident productivity: does shift length matter? *Am J Emerg Med* 2008;26:789–91.
31. Lammers RL, Roiger M, Rice L, Overton DT, Cucos D. The effect of a new emergency medicine residency program on patient length of stay in a community hospital emergency department. *Acad Emerg Med* 2003;10:725–30.
32. Brennan DF, Silvestri S, Sun JY, Papa L. Progression of emergency medicine resident productivity. *Acad Emerg Med* 2007;14:790–4.
33. Accreditation Council for Graduate Medical Education. Common Program Requirements. Available at: <http://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements>. Accessed April 25, 2018.
34. Accreditation Council for Graduate Medical Education. ACGME Program Requirements for Graduate Medical Education in Emergency Medicine. Available at: [http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/110\\_emergency\\_medicine\\_2017-07-01.pdf](http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/110_emergency_medicine_2017-07-01.pdf). Accessed September 5, 2018.

## Supporting Information

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10334/full>

**Data Supplement S1.** Supplemental material.

# Does the Emergency Medicine In-training Examination Accurately Reflect Residents' Clinical Experiences?

Jason J. Bischof, MD<sup>1</sup> , Geremiha Emerson, MD<sup>1</sup>, Jennifer Mitzman, MD<sup>1</sup>, Sorabh Khandelwal, MD<sup>1</sup>, David P. Way, MEd<sup>1</sup> , and Lauren T. Southerland, MD<sup>1</sup> 

## ABSTRACT

**Objective:** The American Board of Emergency Medicine Model of the Clinical Practice of Emergency Medicine (ABEM Model) serves as a guide for resident education and the basis for the resident In-training Examination (ITE) and the Emergency Medicine Board Qualification Examinations. The purpose of this study was to determine how closely resident–patient encounters in our emergency departments (EDs) matched the ABEM Model as presented in the specifications of the content outline for the ITE.

**Methods:** This single-site study of an academic residency program analyzed all documented resident–patient encounters in the ED during a 2.5-year period recorded in the electronic medical record. The chief complaints from these encounters were matched to the 20 categories of the ABEM Model. Chi-square goodness-of-fit tests were performed to compare the proportions of categorized encounters and proportions of patient acuity levels to the proportions of categories as outlined in the content blueprint of the ITE.

**Results:** After the exclusion of encounters with missing data and those not involving EM residents, 125,405 encounters were analyzed. We found a significant difference between the clinical experience of EM residents and the ABEM Model as reflected in the ITE for both case categories ( $p < 0.01$ ) and patient acuity ( $p < 0.01$ ). The following categories were the most overrepresented in clinical care: signs, symptoms, and presentations; psychobehavioral disorders; and abdominal and gastrointestinal disorders. The most underrepresented were procedures and skills, systemic infectious disorders, and thoracic–respiratory disorders.

**Conclusion:** The clinical experience of EM residents differs significantly from the ITE Content Blueprint, which reflects the ABEM Model. This type of inquiry may help to provide custom education reports to residents about their clinical encounters to help identify clinical knowledge gaps that may require supplemental nonclinical training.

The American Board of Emergency Medicine (ABEM) provides an annual residency In-training Examination (ITE) to gauge a resident's educational progress and medical knowledge.<sup>1</sup> Both the ITE and the Board Qualification Examination are constructed from the comprehensive list of emergency medicine (EM) core content contained in the ABEM's Model of Clinical Practice (ABEM Model) and associated table of examination specifications.<sup>2</sup> The ABEM Model is a consensus document that requires periodic updating to reflect the current practice of EM and the educational expectations that serve as the standard

From the <sup>1</sup>Department of Emergency Medicine, The Ohio State University Wexner Medical Center, Columbus, OH.

Received May 30, 2019; revision received July 19, 2019; accepted July 20, 2019.

Presented at the annual meeting of the Society for Academic Emergency Medicine, Las Vegas, NV, May 2019.

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

Author contributions: LTS and JM conceived the study; LTS, GE, JM, SK, and DPW designed the study; LTS, JM, and DPW defined the data fields needed from the EMR and acquired the data through the institution's information warehouse; JJB, DPW, GE, JM, and LTS analyzed and interpreted the data; JJB, GE, LTS, and DPW drafted the original manuscript; and all authors contributed to critical revisions of the manuscript.

Supervising Editor: Margaret Wolff, MD.

Address for correspondence and reprints: Geremiha Emerson, MD; e-mail: Geremiha.Emerson@osumc.edu.

AEM EDUCATION AND TRAINING 2019;3:317–322.

knowledge base against which resident medical knowledge is assessed.<sup>3</sup>

Clinical patient encounters serve as the key component of residency training and in theory should be reflected in the ABEM Model. However, the case mix and content of patient care encounters are difficult to characterize.<sup>4,5</sup> To our knowledge, although the ABEM Model framework was derived from clinical practice experience by expert panels, it has not been formally compared with the clinical experience of EM residents. Consequently, it is unknown whether the current EM resident bedside clinical experiences reflect the broad-based educational requirements as outlined by the ABEM Model. As the ABEM model is also the basis for the questions on the ITE, the clinical environment is critical to the success of trainees in EM. This was a preliminary effort to demonstrate the value of profiling residents' clinical experiences in an academic EM residency program and comparing it to the current ABEM Model to determine if there may be gaps in clinical training that need to be addressed through supplemental, alternative education methods. In other words, this was intended to be a proof-of-concept study.

## METHODS

### Study Setting

The residency program is situated at an academic, tertiary care emergency department (ED) with 86 beds and 80,000 patient visits a year, with a secondary community ED with 26 beds and 52,000 patient visits per year. Only 2% of visits to these EDs are pediatric (age < 18 years), since our residents get their pediatric experience at a freestanding pediatric hospital. Those pediatric visits were not included in this study. At the tertiary care hospital, 90% of patients are insured (56% Medicare and/or Medicaid) and 61% have primary care physicians. At the community ED, 82% are insured (70% Medicare and/or Medicaid) and 39% have primary care doctors.

The residency is a 3-year American Council of Graduate Medical Education–accredited program with 16 EM residents and two EM/internal medicine (EM/IM) residents per year. Residents work 8- to 10-hour shifts without protected time for sign-out or note writing. All patients with psychiatric complaints receive medical screenings by a resident, attending, or advanced practice provider. Both EDs have low-acuity or fast-track sections that are open 10 to 18 hours a

day, Monday through Friday, and are primarily run by advanced practice providers.

### Data Acquisition

This study used deidentified patient data from the electronic health record system (EPIC) and was approved by the local institutional review board. We queried all patient encounters that involved an EM resident at the study sites over a 2.5-year period between July 1, 2015, and December 31, 2017. Data included the Emergency Severity Index (ESI) acuity level, the chief complaint (free-text reason for the visit documented by the triage ED nurse), and unique resident ID code. The initial resident assigned to the patient was considered the resident of record for each encounter. Residents with fewer than 25 encounters were removed from the data set due to concern for incomplete or miscoded data. Records for which no chief complaint was recorded were also removed from the dataset due to an inability to map the encounters to the ABEM Model. ESI and chief complaint information were then mapped to the ABEM ITE Content Blueprint.<sup>1</sup> ED diagnosis was obtained but was unavailable for more than half of the encounters, because the electronic health record system does not require an admission diagnosis.

### Data Analysis

All chief complaints were tabulated and then independently categorized according to the ABEM model by two board-certified EM physicians. Only one category was assigned to each chief complaint. Disagreements between the two reviewers were adjudicated by a third independent board-certified EM physician. For complaints in which all three disagreed, categorization was discussed until consensus was reached. All reviewers are also involved in resident education. The ABEM Model categorized chief complaints and acuity-levels residents experience in the ED were tallied and proportions of the total encounters were calculated to create observed proportions. These observed proportions were then compared to the expected proportions that were derived from the weights provided by the ITE Content Blueprint.

Chi-square goodness-of-fit tests<sup>6,7</sup> ( $\alpha = 0.05$ ) were performed to compare the observed categorized chief complaints to the weights provided by the ITE Content Blueprint. Patient acuity scores were also compared. Effect sizes for the chi-square goodness-of-fit tests were calculated using a formula proposed by Lomax and Hahs-Vaughn ( $ES = \chi^2/N(J - 1)$ ), where



$N$  is the total sample size and  $J$  is the number of categories.<sup>6</sup> These can be interpreted as 0.1 = small effect, 0.3 = medium effect, and 0.5 = large effect. The chi-squares for the content categories were calculated using Motulsky's GraphPad goodness-of-fit calculator.<sup>7</sup> Chi-squares for acuity levels were calculated using VassarStats.<sup>8</sup> Data management was performed through IBM SPSS Statistics for Windows, Version 25.0. Interpreting significant results (post hoc analysis) involved inspection of the standardized residuals. Those that exceeded the absolute value of 1.96 were considered contributors to the significant differences between observed and expected values.<sup>6</sup> To assess the effect of removing encounters without chief complaints, we repeated all analyses by including these encounters in the "other" category.

## RESULTS

The 30-month study period included 160,208 ED encounters. Removing encounters associated with faculty only, fellows, non-EM residents, residents with fewer than 25 encounters, or encounters without a chief complaint resulted in a total number of 125,405 encounters treated by 86 EM or EM/IM residents. There was 66.8% agreement ( $n = 488$  chief complaints) between the initial two reviewers. The remaining 242 chief complaints were reviewed by a third arbitrator. There was three-way disagreement in 6.7% ( $n = 49$ ) of the chief complaints, which were resolved by group consensus.

Tabulation of the list of chief complaints revealed 730 unique chief complaints that were categorized into the 20 categories in the ABEM model. The chi-square goodness-of-fit testing of the observed clinical experience of EM residents and the expected ABEM model categories covered by the ITE are reported in Table 1. Goodness-of-fit tests revealed a statistically significant difference ( $p \leq 0.001$ ), suggesting that the proportion of categories of chief complaints differs from the proportional representation of these categories on the ITE. The effect size is very small; however, post hoc analysis suggests that all but head, ear, eye, nose & throat, and cardiovascular disorders contributed to this significant difference.

Categories that were overrepresented are labeled in Table 1 with a plus sign. Those with large positive standardized residuals ( $>50$ ) included signs, symptoms, and presentations; psychobehavioral disorders; abdominal and gastrointestinal disorders; cutaneous

disorders; and nervous system disorders. Categories that were underrepresented (negative standardized residuals) are labeled in the table with a minus sign. The most underrepresented categories ( $\leq 50$ ) were procedures and skills; systemic infectious disorders; thoracic and respiratory disorders; toxicologic disorders; and environmental disorders. Additionally, 4.8% of clinical encounters were categorized as "other" components, which suggests significant clinical exposure to topics not central to the ABEM model, such as need for preoperative examination, social work consultation, or body-fluid exposure testing (Table 2).

The observed clinical acuity of patient encounters was also significantly different from the ITE Content Blueprint specifications ( $p \leq 0.001$ ) with a moderate effect size, 0.29 (Table 3). Analysis of the standardized residuals suggests that all levels of acuity contributed to this significant difference, as the observed number of encounters were far fewer than expected for the critical (ESI 1) and nonacute acuity levels (ESI 5) and much greater than expected for the lower acuity level (ESI 3 and 4). Results were similar across analyses with and without the data involving missing chief complaints.

## DISCUSSION

The clinical experience of our residents spans the breadth of the ABEM Model categories. The ABEM Model provides the structure and guidance for current EM training, so it is reassuring that all categories were represented, albeit some in lower volumes. Several prior studies have identified challenges with tracking the EM resident clinical experience and assuring adequate clinical exposure.<sup>4,5</sup> New approaches are required to assess the training environment and the current training standards. Chief complaint categorization may be helpful in this regard, because it reflects the diversity of clinical encounters in the ED. Further comparisons of chief complaint, ED visit clinical impression, and eventual diagnosis may be helpful to determine the optimal way to profile the resident clinical experience.

This study capitalizes on the availability of clinical information associated with the increased reliance on electronic medical records (EMRs). As demonstrated by Douglass et al.<sup>5</sup> this type of approach can inform current and future curricular interventions based on the diversity that exists in the current training environment. However, using retrospective data may reduce

**Table 1**  
Number and Percentage of Resident–Patient Encounters for EM or EM/IM Residents

Categories	Observed		ABEM/Expected		Standardized Residual
	Frequency	Percent	Frequency	Percent	
1. Signs, symptoms, and presentations (+)	27,683	22.1	11,286	9.0	154.3
14. Psychobehavioral disorders (+)	11,078	8.8	5,016	4.0	85.6
2. Abdominal and gastrointestinal disorders (+)	18,478	14.7	10,032	8.0	84.3
4. Cutaneous disorders (+)	3,117	2.5	1254	1.0	52.6
12. Nervous system disorders (+)	10,302	8.2	6,270	5.0	50.9
20. Other components** (+)	6,066	4.8	3,762	3.0	37.6
11. Musculoskeletal disorders (nontraumatic) (+)	5,746	4.6	3,762	3.0	32.3
3. Cardiovascular disorders (o)	12,113	9.7	12,541	10.0	−3.8
7. Head, ear, eye, nose, and throat disorders (o)	5,938	4.7	6,270	5.0	−4.2
18. Traumatic disorders (−)	10,937	8.7	12,541	10.0	−14.3
5. Endocrine, metabolic, and nutritional disorder (−)	1,011	.8	2,508	2.0	−29.9
8. Hematologic disorders (−)	974	.8	2,508	2.0	−30.6
15. Renal and urogenital disorders (−)	1,739	1.4	3,762	3.0	−33.0
13. Obstetrics and gynecology (−)	2,639	2.1	5,016	4.0	−33.6
9. Immune system disorders (−)	718	.6	2,508	2.0	−35.7
6. Environmental disorders (−)	352	.3	3,762	3.0	−55.6
17. Toxicologic disorders (−)	1,798	1.4	6,270	5.0	−56.5
16. Thoracic–respiratory disorders (−)	2,937	2.3	10,032	8.0	−70.8
10. Systemic infectious disorders (−)	268	.2	6,270	5.0	−75.8
19. Procedures and skills (−)	1,521	1.2	10,032	8.0	−85.0
<i>Total</i>	125,405	100.0	125,405	100.0	

Encounters are grouped by category of chief complaint and compared to the expected frequencies, which are derived from the 2016 EM Model of Practice content outline for the In-training Examination. These data were used to set up a chi-square goodness-of-fit test for each group. Results are presented in rank order by standardized residuals which are interpreted by evaluating how large or small they are compared to the absolute value of 1.96.  $\chi^2 = 75,870.444$ ;  $df = 19$ ;  $p < 0.0001$ ;  $ES = 0.032^5$

(+) = Have significantly more encounters than expected under this category than recommended from EM Model of Practice.

(−) = Have significantly fewer encounters than expected under this category than recommended from EM Model of Practice.

(o) = Have about the same number of encounters as would be expected under this category as recommended by EM Model of Practice.

complex clinical situations and not fully reflect the resident clinical experience. For example, a patient with an initial chief complaint of “chest pain” and a final diagnosis of “chest pain” could be a 25-year-old with costochondritis requiring minimal evaluation or an 85-year-old with multiple medical comorbidities requiring significant testing and treatment to eliminate other life-threatening causes of chest pain.

Despite these limitations, these data are helpful for demonstrating the value of this type of program evaluation. Results from this study reveal several key points. First, the current clinical encounters experienced by the EM residents at this academic center significantly differ from the ABEM Model in several key areas (i.e., procedures and skills, environmental disorders, toxicology). Given the rarity of certain disease processes

**Table 2**  
List of Chief Complaints Within the “Other” Category

Abnormal CT	Consult	Medication Problem	Referral
Abnormal images	Critical laboratory values	Medication reaction	Routine prenatal visit
Abnormal X-ray	Drain	Medication refill	Second opinion
Admission notification	Drug screen	MRI results	Social work consult
Baclofen pump refill	ED follow-up	ODRC doctor sick call	Surgical follow-up
Body fluid exposure	Exposure to STI	Other	Transportation issues
Cancer	Feeding tube problem	Physician contact	Well child
Chemotherapy	Leaking fluid	Postoperative problem	
Chronic care needs	Letter for school/work	Preoperative examination	
Chronic pain	Medication follow-up	Pump disconnect	

**Table 3**

Number and Percentage of Resident–Patient Encounters Grouped by Level of Patient Acuity for EM/EM-IM Residents With Chi-square Goodness-of-fit Tests Comparing Observed Frequencies to Expected Frequencies Based on the EM In-training Examination Weights

Category	Observed		ABEM/Expected		Standardized Residual
	Frequency	Percent	Frequency	Percent	
Critical	2,658	2.0	37,621.5	30.0	−180.26
Emergent	53,461	43.0	50,162.0	40.0	+14.73
Lower	68,340	54.0	26,335.1	21.0	+258.84
None	946	1.0	11,286.5	9.0	−97.33
Total	125,405	1.00	125,405	1.00	

$$\chi^2 = 109,182.76; df = 3; p < 0.0001; ES = 0.2902.^6$$

this is not entirely unexpected. However, these under-represented content areas may need heavier emphasis through nonclinical education methods such as simulations, conference presentations, or other creative methods. Conversely, areas overrepresented clinically can inform changes in ED flow that may reduce redundant clinical experiences. For example, 8.8% of ED encounters are for psychobehavioral disorders. If by the third year of training these encounters are deemed unlikely to add significantly to the residents' education, a different clinical pathway for these residents might be considered.

Analysis of resident activity can also inform general ED operations. While we expected the signs, symptoms, and presentations category to be overrepresented as we were categorizing by chief complaint, these data suggest that 4.8% of our ED encounters are for “other” problems not categorized in the model. Table 2 demonstrates the breadth of these problems, including the need to expedite the work-up and care of abnormal outpatient testing results, the need for MRI, and care coordination such as medication refills. The ED is the front porch of the entire health system, and care coordination is a necessary part of our practice. Future ABEM models may want to emphasize knowledge of patient flow through health systems, care coordination, and transitions of care.

The observed acuity of EM resident encounters when compared to the ITE distribution differed in the number of critical patients. This is expected due to the ITE's appropriate overemphasis of patients with high acuity. A secondary analysis of acuity including the encounters without a chief complaint (8.7%) did not affect the comparison despite our hypothesis that higher-acuity patients (trauma, stroke, intubated) bypass the traditional triage method. This provides some validity to using chief complaints to profile

resident clinical education, but there is still a chance that missing data impacted study results.

Overall, we found that our resident case mix and acuity vary significantly from the ABEM model, suggesting a need to adjust our clinical workflow or provide alternative education methods. This study uncovered several potential local problems—high volume of patients in the ED for mental health reasons and low volumes of toxicology encounters and critical care. This does not necessarily mean that the ABEM Model does not reflect clinical practice, nor do we suggest that the ITE should be a perfect reflection of clinical care. Residents should be prepared to handle high-acuity, low-frequency illnesses/injuries, and the ITE acuity levels reflect this. However, this analysis could inform updates to the ABEM Model, such as ensuring that residents have an understanding of transitions of care and other uses of the ED than for acute care. Further research into the breadth of clinical practice experienced by residents at other academic institutions could assist other residency programs in identifying surpluses and deficits in their training. Combining resident training information with a fresh national practice analysis would provide comprehensive data to inform future updates to the ABEM Model.

## LIMITATIONS

Our data were limited to the residents' adult patient experience since our residents obtain their pediatric ED experience at a separate children's hospital. Furthermore, this initial database was not broad enough to characterize the complete experience of each resident in our program. For example, 2017 PGY-1 residents would have only 6 months of data in the system, while 2015 PGY-1 residents would have 2.5 years of data. Therefore,

these data cannot be used to characterize patient encounters per resident by PGY level per year. We are also unable to account for the effects of different resident shifts and of advanced practice providers staffing the low-acuity area Monday through Friday. Additional limitations secondary to the retrospective nature of the data capture include missing data or miscoded data that was omitted from the data capture. This, however, would result in an underestimation of the total number of encounters and is likely random across all categories and acuity levels of the encounters given the consistent method of data entry into the EMR that was queried.

Finally, our categorization is limited by the use of triage nurse chief complaint designation. The study team attempted to use natural language processing to assess the medical decision-making section of the notes. However, the notes often have a broad differential diagnosis listed which complicated the identification of the actual diagnosis. Manual chart review would be needed, which was not feasible with 125,000 encounters. A future study option would be to perform manual chart review of a subset of random encounters, blinded to the chief complaint and chief complaint categorization, to assess how often the chief complaint categorization correlates. That could inform confidence intervals on future analyses. We suspect that some conditions such as psychiatric illnesses are likely adequately classified by chief complaint, suggesting that our residents are seeing an overwhelming number of these patients in relation to what they should be learning for their training. Other chief complaints are unlikely to map as well. For example, a chief complaint of nausea could be a gastrointestinal illness, an acute coronary syndrome, or a medication overdose.

## CONCLUSION

In this single residency study, the clinical experiences of emergency medicine residents differed significantly from the 2016 American Board of Emergency Medicine Model. Electronic medical record data on resident clinical experience can be analyzed to identify areas of clinical knowledge requiring additional intervention by residency programs to guarantee adequate training of emergency medicine residents.

## References

1. American Board of Emergency Medicine. In-Training Examination. East Lansing, MI; Available at: <https://www.abem.org/public/for-program-directors/in-training-examination>. Accessed January 22, 2019.
2. Counselman FL, Babu K, Edens MA, et al. The 2016 Model of the Clinical Practice of Emergency Medicine. *J Emerg Med* 2017;52:846–9.
3. Hockberger RS, La Duca A, Orr NA, Sklar DP. Creating the model of a clinical practice: the case of emergency medicine. *Acad Emerg Med* 2003;10:161–8.
4. Langdorf MI, Strange G, Macneil P. Computerized tracking of emergency medicine resident clinical experience. *Ann Emerg Med* 1990;19:764–73.
5. Douglass A, Yip K, Lumanauw D, Fleischman RJ, Jordan J, Tanen DA. Resident clinical experience in the emergency department: patient encounters by postgraduate year. *Acad Emerg Med Educ Train* 2019;3:243–50.
6. Lomax R, Hahs-Vaughn DL. *An Introduction to Statistical Concepts*. 3rd ed. New York, NY: Routledge, Taylor-Francis Group, 2012.
7. Motulsky H. Graphpad Calculator for Chi-Square Goodness of Fit Test. Available at: <https://www.graphpad.com/quickcalcs/chisquared1.cfm>. Accessed April 15, 2019.
8. Lowry R. VassarStats: Website for Statistical Computation. Chi-Square Goodness of Fit Test. Avon, CT. Available at: <http://vassarstats.net/>. Accessed April 15, 2019.



# Feedback With Performance Metric Scorecards Improves Resident Satisfaction but Does Not Impact Clinical Performance

Mira Mamtani, MD, MEd<sup>1</sup>, Frances S. Shofer, PhD<sup>1</sup>, Alexander Sackeim, MD<sup>1</sup>, Lauren Conlon, MD<sup>1</sup>, Kevin Scott, MD, MEd<sup>1</sup>, and Angela M. Mills, MD<sup>2</sup>

## ABSTRACT

**Objectives:** The Emergency Medicine Milestone Project, a framework for assessing competencies, has been used as a method of providing focused resident feedback. However, the emergency medicine milestones do not include specific objective data about resident clinical efficiency and productivity, and studies have shown that milestone-based feedback does not improve resident satisfaction with the feedback process. We examined whether providing performance metric reports to resident physicians improves their satisfaction with the feedback process and their clinical performance.

**Methods:** We conducted a three-phase stepped-wedge randomized pilot study of emergency medicine residents at a single, urban academic site. In phase 1, all residents received traditional feedback; in phase 2, residents were randomized to receive traditional feedback (control group) or traditional feedback with performance metric reports (intervention group); and in phase 3, all residents received monthly performance metric reports and traditional feedback. To assess resident satisfaction with the feedback process, surveys using 6-point Likert scales were administered at each study phase and analyzed using two-sample t-tests. Analysis of variance in repeated measures was performed to compare impact of feedback on resident clinical performance, specifically patient treatment time (PTT) and patient visits per hour.

**Results:** Forty-one residents participated in the trial of which 21 were randomized to the intervention group and 20 in the control group. Ninety percent of residents liked receiving the report and 74% believed that it better prepared them for expectations of becoming an attending physician. Additionally, residents randomized to the intervention group reported higher satisfaction ( $p = 0.01$ ) with the quality of the feedback compared to residents in the control group. However, receiving performance metric reports, regardless of study phase or postgraduate year status, did not affect clinical performance, specifically PTT (183 minutes vs. 177 minutes,  $p = 0.34$ ) or patients visits per hour (0.99 vs. 1.04,  $p = 0.46$ ).

**Conclusions:** While feedback with performance metric reports did not improve resident clinical performance, resident physicians were more satisfied with the feedback process, and a majority of residents expressed liking the reports and felt that it better prepared them to become attending physicians. Residency training programs could consider augmenting feedback with performance metric reports to aide in the transition from resident to attending physician.

In 2013, the Accreditation Council for Graduate Medical Education (ACGME) along with the American Board of Emergency Medicine officially launched the emergency medicine milestones as a component of the Next Accreditation System.<sup>1</sup> These milestones were created as a framework for assessing competencies

From the <sup>1</sup>Department of Emergency Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA; and the <sup>2</sup>Department of Emergency Medicine, Columbia University College of Physician and Surgeons, New York, NY.

Received January 31, 2019; revision received March 27, 2019; accepted March 27, 2019.

Presented at the Society for Academic Emergency Medicine Annual Meeting, Orlando, FL, May 2017; Council of Emergency Medicine Residency Directors Academic Assembly, Ft. Lauderdale, FL, April 2017; and the Council of Emergency Medicine Residency Directors Academic Assembly, Nashville, TN, March 2016.

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

Supervising Editor: Stephen J. CicoMD, MEd.

Address for correspondence and reprints: Mira Mamtani, MD, MEd; e-mail: Mira.Mamtani@uphs.upenn.edu.

AEM EDUCATION AND TRAINING 2019;3:323–330

within several domains of emergency medicine training. Many programs have adapted these milestones as a method to provide focused resident feedback.<sup>2-6</sup>

The annual ACGME survey, a national survey that monitors residency compliance with graduate medical clinical education, includes a question directed to trainees on whether the training program has “provided data about practice habits.”<sup>7</sup> While the emergency medicine milestones include competencies in general patient flow and systems-based management, they do not include specific data about resident practice habits in the emergency department (ED). In addition, studies have shown that milestone-based feedback has not improved resident perception of the quality or satisfaction with the feedback process.<sup>8,9</sup> Objective data on residency efficiency and productivity provide an opportunity to augment feedback on these specific milestones, address the ACGME annual survey question, and improve resident satisfaction with the feedback process.

Satisfaction with feedback has been linked to improvements in staff motivation, satisfaction with level of responsibility and involvement, and perceived support from managers.<sup>10</sup> Providing feedback has also been shown to lower rates of burnout, increase employee engagement, and improve patient safety culture.<sup>11</sup> High-quality feedback is recommended as a tool for program leadership to combat trainee burnout.<sup>12</sup>

Performance metrics obtained from electronic health records (EHRs) are widely used as a way to gauge emergency medicine physician efficiency and quality in the clinical setting with these metrics often tied to physician compensation.<sup>13-15</sup> However, studies on the use of these data as a feedback mechanism to improve resident training and performance have been limited. This study examines whether providing emergency medicine trainees clinical performance metric based feedback improves resident physician 1) satisfaction with the feedback process and 2) performance in the clinical setting, specifically with patient treatment time (PTT) and patient visits per hour (PVHR).

## METHODS

From July 1, 2015, to June 30, 2016, we conducted a three-phase stepped-wedge randomized pilot study of emergency medicine residents at a single, urban academic site with an annual census of approximately 61,000. Informed consent was obtained from resident physicians who were randomized within each postgraduate year (PGY) level to receive either traditional

feedback (control group) or monthly performance metric reports in addition to traditional feedback (intervention group). Traditional feedback at this institution included end-of-shift milestone-based assessment and qualitative comments from faculty, peers, and medical students, as well as semiannual data regarding attendance and involvement in educational experiences, performance on in-service examination, number of procedures completed, and compliance with other requirements of the residency program. Notably, aside from the monthly performance metric report, no additional feedback on resident efficiency in the clinical learning environment was provided during the study period.

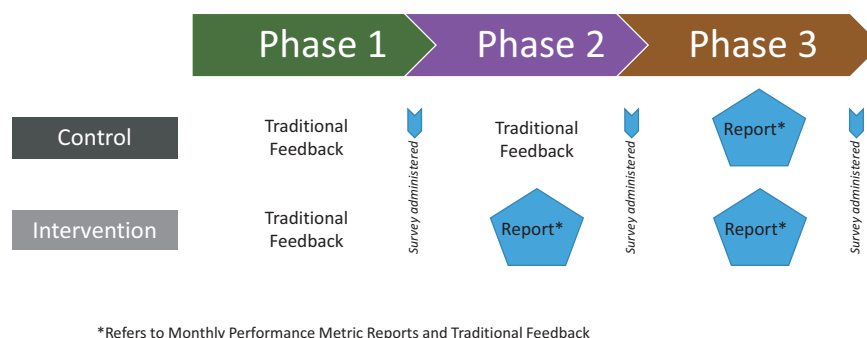
A stepped-wedge model was employed with each phase lasting 4 months. During phase 1 of the study, all residents received traditional feedback. During phase 2 of the study, residents were randomized to receive traditional feedback (control group) or traditional feedback and performance metric reports (intervention group). During the final phase of the study, all residents received monthly performance metric reports and traditional feedback. Off-service residents were not included in the study. Surveys were administered to the residents during each phase of the study (Figure 1). This study was approved by our institutional review board.

## Survey

The three surveys were electronically administered through an online survey program, Survey Monkey, to all study participants at key junctures of the study phase (Figure 1). Likert scale (1–6 = strongly disagree to strongly agree) questions inquired into satisfaction with the feedback process, subjective experience of receiving data on performance in relation to peers, and perceived accuracy of the reports. A free-text component was also included, allowing participants to provide general observations and recommend additional performance metrics that should be included in the scorecard (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/aet2.10348/full>). The survey was piloted through an iterative process with the resident on the study team. Results were anonymous.

## Performance Metric Report

The performance metric report was created using data queried from our EHR (Epic System) and our



**Figure 1.** Three-phase stepped-wedge randomization model with control and intervention group. \*Refers to monthly performance metric reports and traditional feedback.

shift scheduling software (ShiftAdmin). Provider-specific metrics extracted from the EHR included total visits, total PVHR, acuity level of patients seen as measured by the Emergency Severity Index (ESI), median treatment times, and the disposition of patients. Over the course of the year, all resident physicians had worked an equal number and a similar distribution of shifts compared to their colleagues in the same PGY of training. Individual performance metrics were displayed in relation to the mean values

of similar PGY level (Table 1). Reports were delivered monthly to the resident physicians via the hospital e-mail system.

### Data Analysis

To examine resident clinical performance, median PTT and mean PVHR were calculated for each resident during the three study phases. The median PTT was chosen over the mean PTT due to nonnormality of the data. In contrast, the mean PVHRs were

**Table 1**  
Sample Monthly Performance Scorecard

Metrics	July	August	September	October	Your Class	
					Median	25%–75%
Efficiency						
Total visits	41	188	44	152	107	49–145
Visits per hour	1.3	1.41	1.22	1.22	1.08	0.97–1.22
Median treatment time (minutes)	187	167	161	163	195	118–300
Admitted patients	187	149	134.5	161	171	91–280
Discharged patients	196	174	161	161	205	129–309
Median time from bed request to CRC approval (minutes)	25	40.5	44	49	25	14–47
Median time from bed request to doc-to-doc (minutes)	120	145	94	130	102	63–159
Median length of stay (minutes)	468	324	285	342	390	258–558
ESI level (%)						
1	2.5	1	0	2.6	0.9	0.0–1.9
2	57.5	35.1	29.6	38.8	44.7	39.2–49.4
3	35	55.9	50	50	45.7	41.7–50.4
4	2.5	7.4	18.2	7.2	6.6	4.5–8.7
5	2.5	0.5	2.3	1.3	1	0.0–2.0
Disposition rates (%)						
Inpatient	46.3	25	27.3	30.3	31.8	28.7–36.2
Discharge	51.2	72.9	70.5	67.1	66.4	61.5–70.2
Left against medical advice	2.4	2.1	2.3	2	1.5	0.4–2.6

Definitions: total visits = number of visits each month in which you were the first resident assigned to a patient; visits per hour = total visits divided by hours worked in ShiftAdmin; treatment time = time in minutes from when a patient was placed in a room until a disposition was documented in Epic; time from bed request to CRC approval = time in minutes from bed order placement until the CRC approved the bed for admission; time from bed request to doc-to-doc = time in minutes from when a bed order was placed until doc-to-doc was completed in Epic; length of stay = time in minutes from registration in triage until the patient left the emergency department.

chosen over median PVHR as that data set had a normal distribution. Mean PVHR was restricted to the PGY-1 and PGY-2 residents as the study site had PGY-3 and PGY-4 resident physicians engaged in supervisory roles and, as such, could not accurately attribute patients to the senior level residents during these shifts. Median PTT was calculated for all levels of trainees across the three study phases. The senior-level residents occasionally saw patients primarily, and as such, the median PTT could be calculated in those instances. If a handoff occurred, the initial assigned resident was the trainee attributed to that patient.

To determine differences between the control group and the intervention group with regard to PTT and PVPH, an analysis of variance in repeated measures was performed, with study phase ( $n = 3$ ) as the repeated measure and feedback group ( $n = 2$ ) and PGY status ( $n = 2$  [for PGY-1 and PGY-2] and  $n = 4$ , respectively) as grouping variables. To reduce variability in the data set, mean patient ESI level at triage and percentage of patients admitted during each study phase was added as time-varying covariates.

Summary statistics (frequencies and percentages) were used to describe feedback on the survey reports. To assess survey data by feedback group, two-sample *t*-tests were performed, as not all participants completed the surveys at all study points. All analyses were performed using SAS statistical software (version 9.4, SAS Institute).

## RESULTS

Of the 43 emergency medicine residents at our academic institution, 42 were eligible to receive performance metric scorecards (one resident was excluded due to involvement in creation of the study) with 41 (98%) electing to participate. During phase 2 of the study, 20 residents were randomized to the control group and 21 residents to the intervention group, with equal representation per PGY of training.

### Survey

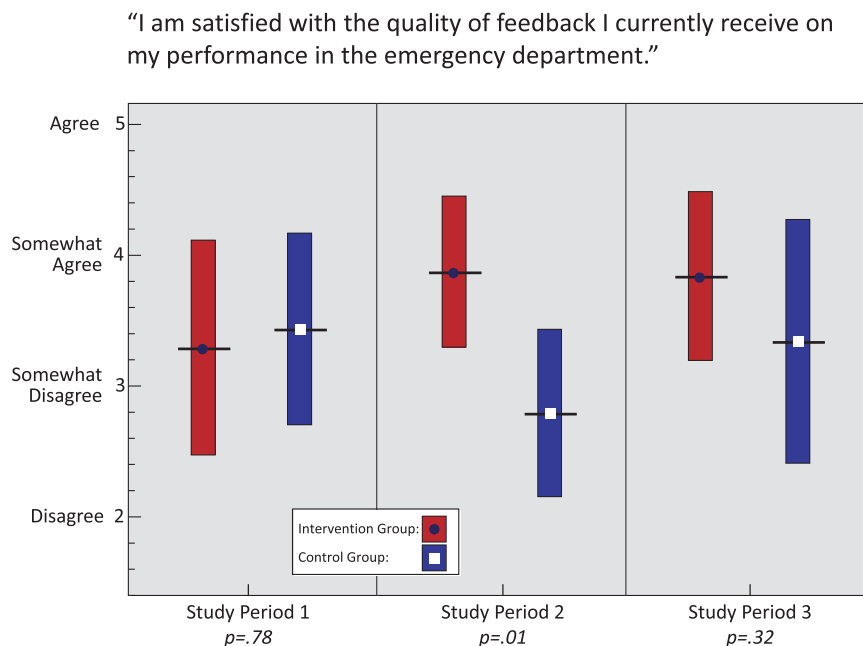
Of the 41 residents participating in the study, 28 (68%) completed the survey during study phase 1, 14 (70%) control group residents and 15 (71%) intervention group residents completed the subsequent survey at the end of study phase 2 (randomization), and finally nine (45%) control group and 12 (57%) intervention group residents completed the survey at the end of the study (phase 3).

All of the residents checked their reports. Over 90% of the residents liked receiving the report with 50% believing that the reports helped them identify areas of improvement. A total of 74% believed that the reports better prepared them for understanding the expectations of becoming an attending physician. Several residents noted that the reports were “really helpful [and of] great benefit.” Forty percent of respondents noted that they had some increase in anxiety about their performance compared with their peers because of the reports.

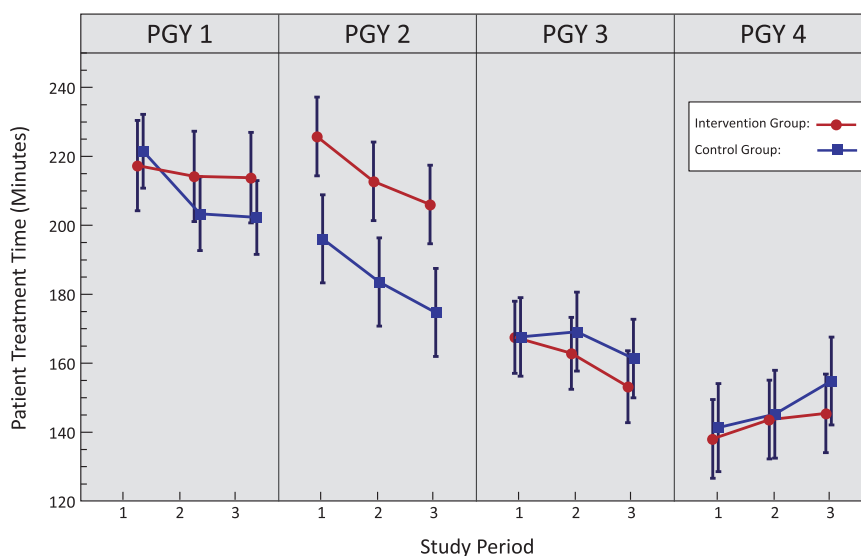
Prior to receiving the performance metric report, the resident physicians reported no difference in their satisfaction with the quality of the feedback they received (intervention vs. control group 3.3 vs. 3.4,  $p = 0.78$ ; Figure 2). After the intervention, residents who received performance metric monthly score cards reported a statistically significant increase (3.9 vs 2.8,  $p = 0.01$ ) in their satisfaction with the quality of the feedback compared to the control group during phase 2 of the study. During phase 3 of the study, when both groups were receiving the performance scorecards, there was no statistically significant difference in satisfaction with the quality of feedback (3.8 vs. 3.3,  $p = 0.32$ ; Figure 2). Additional metrics that the residents believed would be of value as a feedback mechanism included patient satisfaction, new patients per hour, hospital capacity at time of patient visit, and the comparable results for the attending physicians to use as a target.

### Performance Metric Report

Residents receiving performance metric reports and traditional feedback (intervention group) compared to those receiving traditional feedback (control group), regardless of time period or PGY status did not differ by median PTT (183 minutes vs. 177 minutes,  $p = 0.34$ ; Figure 3) or mean PVHR (0.99 vs. 1.04,  $p = 0.46$ ; Figures 4). For the PGY-2 level, the control and intervention groups had started off at different points prior to the intervention, but improved their median PTTs at the same rate despite the intervention. For median PTT, there were significant improvements in treatment times as PGY status increased (means = 212, 200, 164, and 145 minutes, for PGY-1 to PGY-4, respectively;  $p < 0.0001$ ) as well as a decrease over the three study phases (means = 184, 179, and 177 minutes,  $p = 0.10$  for PGY-1 vs. PGY-2 and  $p = 0.01$  for periods 1 vs. 3; Figure 3). For PVHR, again, the same pattern was seen. PGY-2's treated more patients per hour (0.92 vs. 1.1,  $p = 0.01$ )



**Figure 2.** Resident satisfaction with feedback by feedback group.



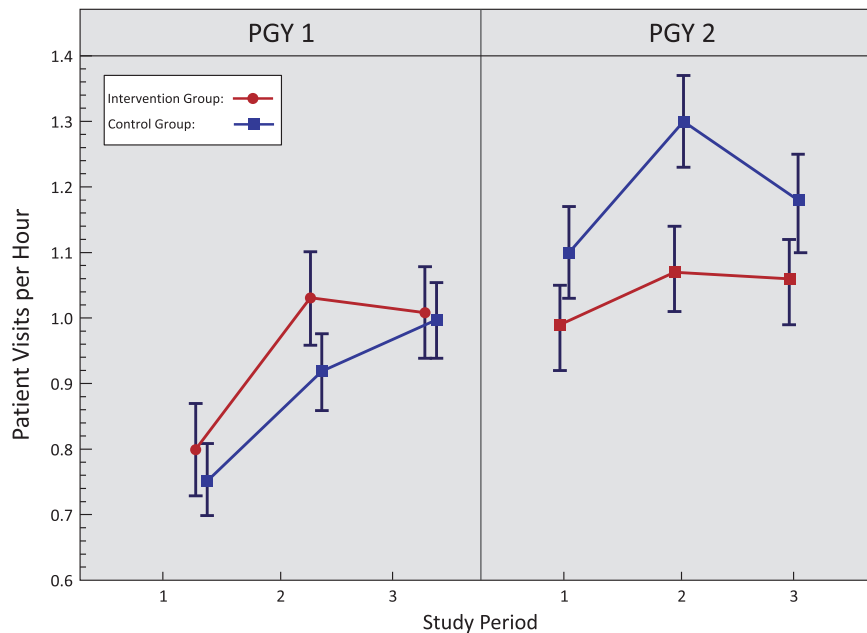
**Figure 3.** Median patient treatment time in minutes by PGY status and feedback group.

compared to PGY-1's, and both PGY groups increased the number of patients by the second study phase but leveled off by the third (0.91 vs 1.08 vs. 1.06 PVHR,  $p < 0.0001$ ; PGY-1 vs. PGY-2), again regardless of feedback group (Figure 4). Including ESI and admission percentage to control for variability did not alter the above associations.

## DISCUSSION

To our knowledge, this is one of the first studies examining the effect of performance metric reports on

emergency medicine resident satisfaction with the feedback process and performance in the clinical setting. Emergency medicine attending physicians are measured on their clinical productivity with their performance on these metrics often tied to financial incentives and compensation. However, there is a dearth of literature on resident physician engagement and awareness of specific clinical productivity metrics.<sup>13–15</sup> Emergency medicine residents may benefit from access to their clinical performance metrics to better inform them of the current expectations of attending clinical practice. We found that



**Figure 4.** Mean patient visits per hour by PGY status and feedback group.

approximately three-quarters of the resident physicians felt that receiving these metrics made them more likely to understand the expectations of becoming an attending physician.

While the vast majority of resident physicians reported that they liked receiving the report and it better prepared them for the expectations of becoming an attending, 40% of residents noted that they had some increase in anxiety about their performance compared with their peers. These results are important when designing feedback opportunities for learners. A prior study examining feedback in medical students demonstrated that institutional culture was most effective in combating anxiety and resistance through the use of engagement throughout the institution and meaningful feedback to support professional development.<sup>16</sup> Standardization of the feedback process has also been shown to decrease the anxiety of feedback delivery to residents.<sup>17</sup> As incorporation of performance metric reports in the feedback process has not yet become standardized at our institution, it is not surprising that resident physicians felt anxiety about the process. Prior studies have demonstrated a need for improved quantitative feedback provided to residents beyond the milestone scores and a recent study found that feedback can improve patient safety culture and lower burnout rates.<sup>8–11</sup> Performance metric reports provide an opportunity to integrate data from the EHR and augment feedback from the competencies patient flow and system-based management to improve feedback opportunities for our learners.

Providing feedback with clinical performance reports did not significantly impact resident clinical productivity and efficiency. A recent Cochrane review revealed that feedback to health care professionals may be more effective when it is offered in multiple methods over multiple sessions and should include explicit targets and action plans.<sup>18</sup> We designed our study to provide monthly performance scorecards in a single method without providing additional in-person feedback regarding resident efficiency in the clinical setting as a method of standardizing the study to measure efficacy. Learners would benefit coupling performance metric scorecards with in-person feedback with specific action plans to improve clinical performance. As the goals and objectives for resident physicians differ based on their level of training, resident physicians may also benefit from different types of feedback: senior-level learners would likely benefit from performance metric scorecards coupled with explicit action plans, junior-level learners may benefit from awareness of specific clinical productivity metrics and its role within the field of emergency medicine.

Interestingly, we also showed a statistically significant improvement in resident clinical performance with PTTs throughout the year, independent of type of feedback given. This is reassuring given that we would expect resident physicians to improve in clinical performance with experience throughout the year. We also showed a statistically significant improvement in resident productivity with PVHR and PTTs as they progress through each year of residency training, which



has been found in prior studies modeling resident productivity in the emergency department.<sup>19</sup>

## LIMITATIONS

While we were able to provide treatment time metrics for residents at all levels of training, we were not able to accurately provide PVHR metrics for the senior residents when their shifts included a supervisory role. Arguably, these metrics would be most helpful for the senior residents as they are preparing to become independent practitioners and may have compensation or other incentives tied to their clinical performance. In addition, as a method of standardizing the study, no additional feedback on resident clinical efficiency and productivity was provided during the study period. Providing feedback in a single format may have limited the ability for performance metric scorecards to be as effective as possible in changing resident behavior. Future studies including more robust metrics for senior-level residents and feedback with explicit action plans to target deficiencies should be performed to determine impact on clinical productivity and satisfaction.

The survey used in our study was not a validated instrument. We created it through an iterative process in our study committee. It is possible that the structure of the survey or wording of questions introduced bias. We attempted to limit this by writing simply worded questions as affirmative statements with agreement assessed on a 6-point Likert scale. Future studies on resident perceptions of feedback would benefit from a validated instrument.

This pilot study was performed at a single institution and, as such, other institutions may have different ranges for patients' visits per hour and PTTs. In addition, learners in other settings may find additional metrics to be of greater value for their specific clinical learning environment.

## CONCLUSIONS

Our study reveals that residents are more satisfied with the feedback process when receiving monthly performance metric reports. Furthermore, a majority of the residents liked receiving reports and felt that it better prepared them for the expectations of being an attending physician. Residency training programs could consider including standardized clinical performance reports as a way of augmenting feedback to aid in the transition from resident to attending physician.

Feedback inclusive of performance metrics, however, did not show statistically significant improvement in resident productivity in the clinical setting. Future studies would need to be conducted to determine which metrics would be of most value to medical trainees and inclusive of a larger group of participants to better understand how to impact learners' productivity in the clinical learning environment.

## References

1. Swing SR, Beeson MS, Carraccio C, et al. Educational milestone development in the first 7 specialties to enter the next accreditation system. *J Grad Med Educ* 2013;5:98–106.
2. Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Acad Med* 2009;84:301–9.
3. Rosenbaum ME, Ferguson KJ, Kreiter CD, Johnson CA. Using a peer evaluation system to assess faculty performance and competence. *Fam Med* 2005;37:429–33.
4. Brasel KJ, Bragg D, Simpson DE, Weigelt JA. Meeting the Accreditation Council for Graduate Medical Education competencies using established residency training program assessment tools. *Am J Surg* 2004;188:9–12.
5. Higgins RS, Bridges J, Burke JM, O'Donnell MA, Cohen NM, Wilkes SB. Implementing the ACGME general competencies in a cardiothoracic surgery residency program using 360-degree feedback. *Ann Thorac Surg* 2004;77:12–7.
6. Reisdorff EJ, Hayes OW, Carlson DJ, Walker GL. Assessing the new general competencies for resident education: a model from an emergency medicine program. *Acad Med* 2001;76:753–7.
7. The Accreditation Council of Graduate Medical Education. Resident Survey Question Content Areas. 2019. Available at: <https://www.acgme.org/Data-Collection-Systems/Resident-Fellow-and-Faculty-Surveys>. Accessed March 25, 2019.
8. Angus S, Moriarty J, Nardino RJ, Chmielewski A, Rosenblum MJ. Internal medicine residents' perspectives on receiving feedback in milestone format. *J Grad Med Educ* 2015;7:220–4.
9. Yarris LM, Jones D, Kornegay JG, Hansen M. The milestones passport: a learner centered application of the Milestone framework to prompt real-time feedback in the emergency department. *J Grad Med Educ* 2014;6:555–60.
10. Frampton A, Fox F, Hollowood A, et al. Using real-time anonymous staff feedback to improve staff experience and engagement. *BMJ Qual Improv Rep* 2017;6(1).
11. Sexton JB, Adair KC, Leonard MW, et al. Providing feedback following Leadership Walk Rounds is associated with better patient safety culture, higher employee engagement and lower burnout. *BMJ Qual Saf* 2018;27:261–70.

12. Gordon EK, Baranaov DY, Fleisher L. The role of feedback in ameliorating burnout. *Curr Opin Anaesthesiol* 2018;31:361–5.
13. Pines JM, Ladhania R, Black BS, Corbit CK, Carlson JN, Venkat A. Changes in reimbursement to emergency physicians after Medicaid expansion under the Patient Protection and Affordable Care Act. *Ann Emerg Med* 2019;73:213–24.
14. Scholle SH, Roski J, Adams JL, et al. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manage Care* 2008;14:833.
15. Scholle SH, Roski J, Dunn DL, et al. Availability of data for measuring physician quality performance. *Am J Manage Care* 2009;15:67.
16. Nofziger AC, Naumburg EH, Davis BJ, Mooney CJ, Epstein RM. Impact of peer assessment on the professional development of medical students: a qualitative study. *Acad Med* 2010;85:140–7.
17. Bing-You RG. Internal medicine residents' attitudes toward giving feedback to medical students. *Acad Med* 1993;68:388.
18. Joseph JW, Henning DJ, Strouse CS, Chiu DT, Nathanson LA, Sanchez LD. Modeling resident productivity in the emergency department. *Ann Emerg Med* 2017;70:185–90.
19. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;(6):CD000259.

### Supporting Information

---

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10348/full>

**Data Supplement S1.** Intervention group survey.



# Navigating Cognitive Dissonance: A Qualitative Content Analysis Exploring Medical Students' Experiences of Moral Distress in the Emergency Department

Caitlin Schrepel, MD, Joshua Jauregui, MD , Alisha Brown, MD, Jamie Shandro, MD, MPH, and Jared Strote, MD, MS

## ABSTRACT

**Background:** As undergraduate medical students are acculturated into clinical practice, they develop a set of refined professional values that impact their decision making. We aimed to use students' reflective narratives on ethical dilemmas to identify how students experience moral distress while working in the emergency department (ED) to better understand how to support them in the development of their own agency to act ethically.

**Methods:** Students rotating in our emergency medicine clerkship are required to submit an essay describing an ethical dilemma they encountered. We selected a random sample of these reflective pieces from the 2015 and 2016 academic years and used an exploratory qualitative thematic analytic approach to identify frequently recurring themes. This process was continued until thematic sufficiency was reached.

**Results:** Two-hundred essays were coded, and seven unique themes were identified. The moral distress students described in reflective writing narratives stemmed from patient-provider discord, uncertainty, and social injustices. In each case, students were expressing the cognitive dissonance they experienced as they began to reconcile the difference between their perceptions of optimal patient care and the actual care delivered to the patient.

**Conclusion:** Understanding medical students' cognitive dissonance in the ED will help educators support their students as they negotiate the differences between preferences and principles while being acculturated into clinical practice. Future work should develop specific interventions to promote educator understanding of learners' moral distress and to develop novel models of support for learners.

As undergraduate medical students are acculturated into clinical practice, they develop a set of refined professional values that impact their decision making. This transformative process toward internalization of "the way of being a physician," or professional identity formation (PIF), is influenced by multiple factors in a student's professional and personal life.<sup>1</sup> Clinical experience is a particularly influential factor in a students'

PIF and introduces students to new realities that may conflict with their previously formed ethical ideals and have the potential to lead to burnout and diminished empathy.<sup>2-4</sup> As such, it is important that educators understand how students are experiencing the clinical environment to best guide them in their PIF.

We relied on student reflective narratives to better understand how ethical challenges in the emergency

From the Department of Emergency Medicine, University of Washington, Seattle, WA.

Received April 4, 2019; revision received July 25, 2019; accepted July 27, 2019.

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

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

Author contributions: all authors contributed to the study concept and design, the acquisition of data, analysis and interpretation of the data, drafting of the manuscript, and critical revision of the manuscript for important intellectual content.

Supervising Editor: Sally Santen, MD, PhD.

Address for correspondence and reprints: Joshua Jauregui, MD; e-mail: joshjaur@uw.edu.

AEM EDUCATION AND TRAINING 2019;3:331-339.

department (ED) might impact PIF. As part of the required emergency medicine (EM) clerkship at our institution, students must write reflective narratives regarding an ethical dilemma they have faced in the ED. Reflective narratives<sup>5,6</sup> can positively contribute to PIF by providing students with a way to better characterize their own beliefs in a safe environment with formative feedback.<sup>1,7</sup> Narratives also provide educators with insight into the professional and ethical challenges students face while working in the clinical environment.<sup>8-10</sup>

The ED provides students with a rich clinical environment from which to learn, but also exposes them to complex clinical and ethical scenarios that may lead to moral distress. Moral distress is defined as the negative feelings that arise when one knows the morally correct thing to do but they feel compelled to act in a way that contradicts with their values.<sup>11</sup> Moral distress can have a significant impact on health care providers and has previously been linked to burnout,<sup>12,13</sup> compassion fatigue,<sup>14</sup> and suboptimal patient care.<sup>15</sup>

While several recently published studies have focused on how medical students are affected by moral distress,<sup>11,16,17</sup> they have not focused on the unique learning environment of the ED. On the other hand, previous studies which used reflective narratives during EM clerkships have considered broad ethical and professional challenges students face<sup>8,18</sup> and identified modern struggles students encounter as they encounter ethical dilemmas<sup>10</sup> but did not focus on the impact of moral distress from these scenarios. We aim to add to this research by situating what we learn within a theoretical framework<sup>19-21</sup> to better understand the ethical dilemmas students experience in the ED and how they might create moral distress. We hope to provide educators the needed context to stimulate ideas for mitigating any negative impact these scenarios might have on PIF.<sup>19-21</sup>

## METHODS

### Study Design

We performed a qualitative content analysis using a conventional exploratory approach.<sup>22,23</sup> Because such analysis requires researchers to be iteratively engaged with the data, we were reflexive about our roles.<sup>24,25</sup> In the interest of transparency, our team was made up of two EM clerkship directors (JSh, JSt), an EM chief resident (CS), an EM assistant residency program director (AB), and an EM faculty member with

advanced training in qualitative research methods (JJ). Our experience in graduate and undergraduate medical education helped us better understand the professional lives and responsibilities of our students. The chief resident, as a near peer and learner in the clinical environment, added unique understandings to our interpretations. We met regularly to discuss the impact of our personal perspectives on our data interpretations and adjudicate these interpretations.

### Sample and Data

All medical students in our required fourth-year EM clerkship rotate either at our primary academic clerkship site, consisting of a tertiary referral university medical center and a level I trauma center county hospital, or at a community ED clerkship site. All students must write a reflective writing (RW) essay about an ethical dilemma they encountered in the ED. Clerkship directors provide students with the following prompt: "There are many ethical dilemmas faced in the emergency department on a daily basis, such as in this clerkship bioethical reading.<sup>26</sup> Pick one such situation you encountered during your EM clerkship and describe what you learned from it." Although clerkship directors would occasionally discuss the content of certain reflective writing essays with individual students, there was no formal use of, or response to, the reflective writing essays save for ensuring completion of the assignment. All RW samples submitted in the 2015 to 2016 and 2016 to 2017 academic years were subject to analysis. A research assistant anonymized RW samples submitted from this time period onto a standardized Word document (Microsoft Corp., 2013) and assigned each sample a unique, nonsequential numeric identifier. The University of Washington Institutional Review Board deemed that this study was not human subjects research.

### Data Analysis

We conducted an exploratory qualitative thematic analysis of the data, analyzing a random sample of the RWs from the study period.<sup>22,23</sup> Three researchers (CS, JJ, and JSt) independently read and coded each RW sample, 25 samples at a time. After each round of coding, the analysis team met to iteratively discuss their codes, explore differing perspectives, and resolve any differences in their thematic findings. After six rounds of coding (150 RW samples), no new significant themes were emerging from the data. To ensure that we had reached thematic sufficiency, we coded 50 more RW samples.<sup>27</sup> The research team then met to

confirm the thematic findings and reach consensus on a final thematic scheme and the relationships between the themes, thematic categories, and the emerging conceptual framework of cognitive dissonance and its relationship to student moral distress. We used Dedoose for data organization and analysis (Sociocultural Research Consultants, Version 7.5 software).

To ensure trustworthiness of the data we fostered reflexive dialogue among the authors about our interpretations of the data and maintained an audit trail of all our findings. Triangulation of themes from all rounds of coding helped ensure credibility of the findings.<sup>20</sup>

## RESULTS

We analyzed a total of 200 RW samples. The majority (164, 82%) of the RWs were from students rotating at our primary academic site while the remaining RWs (36, 18%) were from students rotating at 15 different community clerkship sites. We identified three major thematic categories causing moral distress, encompassing seven unique themes (Table 1). Below we provide several examples of quotations that best exemplify a particular theme. Each quote is followed by parentheses with a code identifier. Of note, while pronouns are used in discussing the quotes, the sex of the students was deidentified prior to analysis.

### Discord: Patient–Provider Conflict

Students identified several themes related to discord evoked by patient–provider conflict that they perceived

to hinder ideal patient care. These discordances involved patient autonomy in clinical decision making, consent for procedural learning, and challenging communication.

### Patient Autonomy in Clinical Decision Making.

Students identified patient autonomy as hindrance to providing what they perceived to be the most appropriate clinical care. One student described frustration with the decision made by a patient who “became angry and left without treatment” (student 230) despite presenting with a gunshot wound to the leg. Several students also discussed the added challenge that substance use or altered mental status play in determining patients’ ability to make decisions. They expressed both a fear of the patient deteriorating clinically a fear of limiting autonomy. One student wrote,

I wonder if this gentleman were not addicted to IV methamphetamine, would he make the same autonomous decision? Is he acting as a competent agent who can exercise true autonomy if he is struggling with severe addiction? I don’t have a clear answer without starting to sound paternalistic” (student 75).

Several students also noted instances where providers chose to limit patient autonomy, leading to important clinical consequences. One student discussed the extra stress and cost imposed on patients when we don’t give them an “opinion on the decision to admit” (student 75), while other students expressed

Table 1

Major Themes Identified in Reflective Writing Samples, Organized by the Type of Cognitive Dissonance Experienced by Students When Reconciling the Difference Between Their Perceptions of Idealized Patient Care and Delivered Patient Care

Theme	Definition
<i>Discord: Patient–Provider Conflict</i>	
Patient autonomy	Student struggled with the balance of providing appropriate medical care and patient autonomy
Patient consent	Student struggled with reconciling their need for procedural learning, while ensuring proper consent
<i>Indecision: Working in the Face of Uncertainty</i>	
Anxiety related to clinical uncertainty	Students discussed a personal and emotional struggle with uncertainty
<i>Disillusion: The Impact of Social Injustices</i>	
System limitations: caring for vulnerable populations	Students identified scenarios in which the health care system and social safety net failed to provide appropriate care for vulnerable populations
System limitations: providing psychiatric care	Students noted frustration with delays or inadequate psychiatric care
Clinical consequences of overcrowding	Students noted changes to patient comfort or clinical treatment because of overcrowding in the ED
Clinical challenges related to the opioid epidemic	Students identified challenges in pain management, treatment of withdrawal, or enrollment in detox programs for patients with opioid dependency

the stress they felt personally when the team made the decision to limit autonomy. For example, one student discussed a patient who was not permitted to leave the ED to smoke a cigarette. The student wrote that by not allowing the patient to make this decision “it was clear to me that he would leave putting himself at serious risk and we could have avoided this by letting him go outside for 15 minutes . . . I was left wondering if we could have done more. It tore at me for the rest of the day after the shift” (student 617).

**Consent for Procedural Learning.** Students discussed the difficulty they faced in reconciling their need to prioritize procedural learning, while ensuring that the consent process was being respected. One student expressed this sentiment stating, “I was struck by the ethical dilemma posed by training new doctors while still giving patients standard of care. Parents and patients don’t want to be guinea pigs, but medical students and residents must learn by doing, and everyone will have a ‘first-time’ patient. Do the wishes of the parents or patients override our need to learn in a teaching hospital? How much information is required for informed consent?” (student 541). Students understood the importance of procedural learning but struggled with determining what constituted proper consent and appropriate patient care. One student wrote, “While I admit, I appreciated the hands-on learning, and US (ultrasound) is a non-invasive and generally harmless study to have a number of students repeat, it still felt a little funny” (student 484). Another student described the emotional struggle she felt stating, “I sometimes feel like I am taking advantage of patients who don’t think they have the option to refuse a student or who don’t fully understand my role on the team” (student 19).

It was also evident to many students that the patient population of the ED can add another level of complexity to this ethical dilemma, as they worried about whether they were taking advantage of vulnerable populations for the benefit of their learning. One student wrote that she was told to place an IV on an intoxicated patient, but wondered “if this patient would have agreed to have a medical student practice starting an IV on him if he was lucid and knew that this was only my third time placing an IV” (student 11). Another student wrote that she wondered, as she was performing an incision and drainage on a man who came from jail, “if we disproportionately work on those of lower socioeconomic status” (student 417).

## **Indecision: Working in the Face of Uncertainty**

Students wrote about ethical dilemmas that caused emotional stress secondary to the inherent clinical uncertainty in the ED, leaving them with a sense of indecision.

**Anxiety Related to Clinical Uncertainty.** Many reflective narratives focused on anxieties students felt as they reflected on the uncertainties in the ED amid treating critically ill patients or those with “uncertain code status” (student 84). They felt powerless in these scenarios and puzzled by the decision-making process. One student wrote that she was confused by “the relatively subjective way that medical futility was determined” (student 672). When comparing this to an inpatient rotation she noted, “This case highlighted for me the challenges of treating patients in the ED when medical histories are unknown and there is no way to determine the patient’s wishes” (student 672).

Difficult resuscitations were not the only clinical scenarios that evoked this sense of helplessness. Students noted that “emergency physicians are in a difficult position when it comes to the ethical mandate of ‘first, do no harm.’ They have to make rapid decisions without all of the necessary information and sometimes even without help from the patient” (student 309). Scenarios involving clinical decisions around aggressive patients were some of the most impactful. One student wrote about the decision to sedate an aggressive patient stating, “The decision about what to do next had impacts on this patient’s autonomy, this patient’s health, the safety of other patients, the safety of staff (especially after they left the hospital), and the legal position of EM physicians and the hospital. It had to be made with very little information as to the patient’s health, which had not been fully evaluated” (student 331).

They also highlighted some emotions that are common and reasonable among new providers, including a lack of confidence in making decisions. One student wrote, “I just don’t yet have the confidence and experience to have made such a decision on my own” (student 383), while another described the discomfort she felt when discharging a patient with an ambiguous prognosis, stating that it was “difficult to deal with given the acute nature of this fatal disease” (student 334).

## **Disillusion: The Impact of Social Injustices**

Many students discussed concerns of social justice in their narratives. Students highlighted the clinical



challenges and disillusion they faced as it related to limitations of our health care system, definitive psychiatric care, hospital overcrowding, and opiate dependence.

**System Limitations.** Caring for vulnerable populations. Students' narratives often discussed the frustration they felt when our system failed to care for vulnerable populations. One student wrote about a patient with a fractured hand who had presented to the ED four times for new splints. While the student was initially upset with the patient for destroying the splints, she later states that she realized that "navigating the streets homeless is trying enough with a functional dominant hand" (student 643). Other students also discussed how homelessness had impacted a patient's ability to heal. For example, a student wrote about a patient with infected wound stating, "it isn't the care that he received but the social situation he was in that predisposed him to infection . . . The limited availability of respite for the injured and sick puts the patient in a bad situation where their health outcome is at risk, leaves a bad taste in the mouth of the provider who disposes their patient to the street . . ." (student 539).

Students described the anxieties they felt when they perceived that they had to ration health care resources. Whether they realized that "sandwiches are a resource that we have to decide how to allocate" (student 502) or that they could not admit a patient with withdrawal because of a bed shortage, they described the frustration of feeling as though they "only have the time and the resources to put a Band-Aid on a bullet wound" (student 50).

**System Limitations: Providing Psychiatric Care.** Students were particularly frustrated about the failure of our system to care for psychiatric patients. One student expressed, "The all-too-often ugly truth for many psychiatric patients is that they use the ED like a revolving door. However, our coolness and ease with discharging patients, sometimes, seemingly brushing off their complaints in light of known psychiatric disease does not feel ethical. It certainly sets a stage up to question the justice and beneficence of the case. I cannot imagine repeatedly discharging a patient to self-care when there is clear evidence that it will not solve the problem . . ." (student 258).

While this student is upset that her patient is discharged without psychiatric care, other students were

equally bothered by holding patients with psychiatric complaints against their will. One student wrote,

How can we justify essentially imprisoning someone for an extended period of time in an environment that is not at all therapeutic except in that we can stop the patient (hopefully) from hurting him/herself, especially if the only reason the patient is still there is because no one can find placement for her? It is important to keep staff members safe from harm, but for a psychotic patient who has not been violent with staff, keeping her cooped up in a small room for more than 4 days with staff members discussing putting her in restraints is hard to stomach (student 347).

**Consequences of Hospital Overcrowding.** Over the course of their rotation, students noted that overcrowding has led to important clinical concerns related to safety and privacy, especially for vulnerable populations who use the safety-net care of the ED. Students were particularly bothered by the increasing lack of privacy. One student wrote, "I have noticed on my recent ED shifts is the increasingly full ED, both in the patient care areas and in the waiting room. This has resulted in many patients 'roomed' in hallway stretchers, resulting in fairly intimate conversations overheard under bright lights in open hallways" (student 284). Another student described the struggle she felt when her patient was put in a hallway bed. She wrote, "I couldn't help but feel embarrassed for her as I asked the standard questions about GI happenings and sexual history. It was obvious this woman would find these topics sensitive in a private setting, but here we were in the hallway, across from a room with a very sick, intubated patient with providers and nurses gathered easily within hearing distance" (student 375).

**Clinical Challenges Related to the Opioid Epidemic.** Students also identified dilemmas related to the growing opioid epidemic including the challenges of pain and addiction management. They emphasized that "bioethical principles demand of us to not dismiss patients' complaints" (student 330). However, they also expressed how challenging it can be to make the decision to treat pain in this setting. One student stated this clearly when she wrote about a series of patients she cared for with opiate dependence. She wrote that, "often their pain is severe,

bringing them to tears during the interview ... The result is a 'dance' between the patient and physician in which the patient must say just the right things in order to leave with an opioid script. Physicians have a moral obligation to treat pain, but they are reluctant to contribute to the epidemic of opioid abuse. ED physicians are in a precarious position" (222). Students described feelings of frustration with patients, providers, and the health care system regarding opiates. A student who was frustrated with a patient wrote, "I had spent a significant portion of my night and an unnecessary amount of hospital resources on a pain-seeking patient" (student 422). While another student disagreed with the decision to withhold narcotic medications from a patient stating, "This seemed particularly unfair to me. From a harm reduction perspective, it seems that providing an addicted person a legal, safe form of pain control might prevent him from resorting to heroin use if the pain became intolerable" (student 293).

## DISCUSSION

Professional identity formation is the transformational process in which medical students establish the values and ethical principles that will guide them as physicians.<sup>1</sup> It is a process that is supported by encouraging reflection<sup>5,6</sup> and influenced by the clinical environment. Reflective narratives, in this case, gave us insight into how students experience ethical dilemmas in the ED. These dilemmas are not inherently harmful to the PIF of students; however, educators need to better understand how students internalize these experiences to intentionally facilitate positive professional growth within the clinical learning environment.

As medical students learn to navigate the clinical environment, they are faced with complex ethical dilemmas that may challenge them to act in ways that conflict with their ideals. The moral distress caused by this pressure to act against their moral judgment can result from various causes including conflicts with coworkers or hospital hierarchy, both of which have previously been shown to impact medical students.<sup>11</sup> However, in these narratives, moral distress seems to most often stem from the cognitive dissonance<sup>28</sup> students experienced when confronting these dilemmas. Cognitive dissonance is a psychological concept defined by the stress a person experiences when they are presented with conflicting values, beliefs, or actions.<sup>28,29</sup> Cognitive dissonance has been shown to promote reflection as it did in these

narratives but has consequences on professional development which are underrecognized by educators.<sup>28,30-32</sup>

The clinical and professional scenarios that underpinned the common themes in this study are like those identified in prior studies<sup>10,18</sup> including patient-provider conflict and systems issues. However, by situating these themes within the theoretical framework of cognitive dissonance, we were better able to understand the moral distress these challenges created for students. Students noted that patient-provider conflict challenged their ability to provide optimal clinical care while maintaining the value they place on autonomy, consent, and fairness. They understood that some compromise was necessary, but many appropriately worried about how to balance these conflicting values.

Students also struggled with uncertainty. Early in medical school, students are taught the aspirational value of using evidenced-based medicine to provide patients with a clear diagnosis and treatment plan.<sup>33</sup> When this value was challenged by the inherent uncertainty of the ED, they experienced cognitive dissonance around their treatment plan. Experiencing uncertainty is a normal part of practicing medicine; however, an inability to tolerate uncertainty in the clinical environment has been shown to predict psychological distress in medical students.<sup>34,35</sup> Although it has previously been suggested that tolerance of uncertainty increases throughout medical school,<sup>36</sup> these results indicate that senior medical students continue to struggle with this concept.

Finally, many students become physicians to help others, often with a particular desire to help the most vulnerable.<sup>37</sup> When they saw that the exact value they place on caring for vulnerable populations is not shared or supported by our society as a whole, they experienced cognitive dissonance, causing frustration. Encountering, often for the first time, the extreme challenges that so many ED patients experience when they do not have adequate follow-up, housing, psychiatric care, or addiction management many students felt disillusioned by the failings of our social structures and health care system.

The ED, possibly more than other clinical environments, exposes students to uncertainty, complex clinical scenarios, social injustices, and challenging patient interactions. This makes an EM clerkship both a high-risk environment, where students have the potential for significant moral distress, and a high-value one, where students face issues that challenge them to

reflect, which, when prepared properly, can lead to improved resilience. As students encounter difficult situations within the ED, positive experiences of reflection, especially when combined with other interventions such as small-group discussions and faculty mentoring, can lead to an increased ability to endure future instances of moral distress.<sup>38</sup>

Cognitive dissonance can lead to ethical distress and, subsequently, burnout and depression.<sup>3,39</sup> When confronted with cognitive dissonance, one will attempt to resolve the moral distress caused by these inconsistencies by changing their values, actions, or perception of actions.<sup>40</sup> These solutions may be appropriate or even beneficial to PIF in some scenarios but detrimental in other instances. For example, one would not want a student to change their actions or beliefs surrounding informed patient consent for procedures. However, it would be reasonable for a student to change their perception if informed consent was not obtained secondary to an emergent clinical condition.

In fact, this process of reconciling previously held values with the realities of the clinical environment plays a critical role in professional development by creating a distinction between principles and preferences. A principle is an ethical belief that is fixed. A preference, on the other hand, provides a student with a guideline that can be molded based on the scenario or changed over time.<sup>41</sup> Prior studies have looked at how learners navigate the threshold between principle and preference as they develop into professionals. The distinction is rarely discussed explicitly but develops as learners gain clinical experience.<sup>41</sup> Reflective narratives give students the ability to discuss their ethical principles and preferences, and both can be identified in this study. In fact, the vast majority of narratives describe the cognitive dissonance students feel as they realize that previously held principles might actually represent preferences that need to be molded to fit the realities in the ED. Educators may be able to guide students to reflect on this reality as a way for them to potentially ease the distress caused by cognitive dissonance without compromising their values entirely.

### Implications for Educators

While ethics is generally covered in the preclinical years, many students have little continued ethics education throughout the clinical years of training.<sup>42,43</sup> However, it is the ethically complex clinical experiences they encounter which will have one of the most significant impacts on their moral values and behavior.<sup>7</sup>

Educators can play an important role in providing students with the opportunity to reflect on the cognitive dissonance they may be feeling to help them find a solution that does not negatively impact their PIF. Reflective narratives may be a critical tool in this process, but it must also be accompanied by formative feedback and guidance.<sup>10,18</sup>

In addition to helping students reflect and identify instances of cognitive dissonance, educators may be able to provide students with more preparation prior to experiencing such encounters. The themes identified here and elsewhere including challenging communication, implicit bias, medical uncertainty, and system limitations—may provide a good starting point for this process.<sup>10,18</sup> However, the concepts discussed by students are complex and somewhat individualistic, underscoring the importance of providing students with the tools they need to identify their own triggers of moral distress.<sup>44</sup> The added benefit of supporting students before the start of their clinical experience would be to decrease the extraneous cognitive load from moral distress helping them focus on the intrinsic cognitive load of the environment and germane cognitive load of their learning.<sup>45</sup> Finally, we know that iterative reflection is necessary for transformative learning.<sup>46</sup> By asking probing questions and stimulating iterative reflection, educators can guide students as they process their experiences, provide context, and further support their identity formation.<sup>47,48</sup>

As result of this study, we permanently changed the submission format of our required reflective writing requirement. We now require students to post their reflective writing online, on a secure social media platform (Yammer, Microsoft Corp.) rather than as a private Microsoft Word (Microsoft Corp.) document to the clerkship directors alone. This change has made students' reflective writings visible to their student peers on the rotation, as well as the EM clerkship directors, and has allowed for real-time peer feedback, community discussion, and faculty guidance. Our hope is that this space for iterative discussion will prompt positive responses to ethical dilemmas and cultivate resilience.

### LIMITATIONS

Students were simply instructed to write about an ethical dilemma they encountered in the ED. Although we identified areas of cognitive dissonance within these essays that were most prevalent, these were not

the only topics discussed. Although some essays explored positive ways of responding to ethical dilemmas, there was a paucity of data to create specific codes related to lessons that the students learned. Some narratives involved little mention of personal internal struggle and as a result were not discussed here. The narratives also did not disclose how students resolved their struggles. Given that students were asked to discuss these dilemmas in narrative form, their actual lived experiences may not be fully expressed in their writing samples. While one of the authors (CS) was an EM resident, providing a near-peer perspective in the process of analysis, a medical student coauthor would have added a student voice to the analysis. Finally, while students did rotate at different sites, most students rotated at both our academic and our county hospital sites, which could limit transferability of our findings.

## CONCLUSIONS

Medical students' experiences, described in their reflective writing narratives, led to the themes of discord, indecision, and disillusion. Educators can use the theoretical framework of cognitive dissonance to help students prepare for morally challenging scenarios and to negotiate the differences between preferences and principles. Future work should explore how students positively manage their experiences of moral distress, develop specific interventions to promote educator understanding of learners' moral distress, and create novel models of support for learners.

The thank Alexis Rush and Kristen Seiler for their essential administrative and collaborative contributions to this study.

## References

1. Wald HS. Professional identity (trans)formation in medical education: reflection, relationship, resilience. *Acad Med* 2015;90:701–6.
2. Coulehan J, Williams PC. Vanquishing virtue: the impact of medical education. *Acad Med* 2001;76:598–605.
3. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ* 2016;50:132–49.
4. Mahoney S, Sladek RM, Neild T. A longitudinal study of empathy in pre-clinical and clinical medical students and clinical supervisors. *BMC Med Educ* 2016;16:270.
5. Lutz G, Scheffer C, Edelhauser F, Tauschel D, Neumann M. A reflective practice intervention for professional development, reduced stress and improved patient care—a qualitative developmental evaluation. *Patient Educ Couns* 2013;92:337–45.
6. Plack MM, Greenberg L. The reflective practitioner: reaching for excellence in practice. *Pediatrics* 2005;116:1546–52.
7. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. A schematic representation of the professional identity formation and socialization of medical students and residents: a guide for medical educators. *Acad Med* 2015;90:718–25.
8. Santen SA, Hemphill RR. A window on professionalism in the emergency department through medical student narratives. *Ann Emerg Med* 2011;58:288–94.
9. Karnieli-Miller O, Vu TR, Holtman MC, Clyman SG, Inui TS. Medical students' professionalism narratives: window on the informal and hidden curriculum. *Acad Med* 2010;85:124–33.
10. Peterson WJ, House JB, Sozener CB, Santen SA. Understanding the struggles to be a medical provider: view through medical student essays. *J Emerg Med* 2018;54:102–8.
11. Camp ME, Jeon-Slaughter H, Johnson AE, Sadler JZ. Moral distress in medical student reflective writing. *AJOB Empir Bioeth* 2019;10:70–8.
12. Zuzelo PR. Exploring the moral distress of registered nurses. *Nurs Ethics* 2007;14:344–59.
13. McAndrew NS, Leske JS, Garcia AJ. Influence of moral distress on the professional practice environment during prognostic conflict in critical care. *Trauma Nurs* 2011;18:221–30.
14. Dzung E, Colaizzi A, Roland M, Levine D, Kelly MP, Barclay S, Smith TJ. Moral distress amongst American physician trainees regarding futile treatments at the end of life: a qualitative study. *J Gen Intern Med* 2016;31:93–9.
15. Mullin J, Bogetz J. Moral distress can indicate inappropriate care at end-of-life. *Psychooncology* 2018;27:1490–92.
16. Camp ME, Jeon-Slaughter H, Johnson AE, Sadler JZ. Medical student reflections on geriatrics: moral distress, empathy, ethics and end of life. *Gerontol Geriatr Educ* 2018;39:235–48.
17. Thurn T, Anneser J. Medical students' experiences of moral distress in end-of-life care. *J Palliat Med* 2019. <https://doi.org/10.1089/jpm.2019.0049>. [Epub ahead of print].
18. House JB, Theyyanni N, Barnosky AR, et al. Understanding ethical dilemmas in the emergency department: views from medical students' essays. *J Emerg Med* 2015;48:492–8.
19. Bolander Lakso K, Dornan T, Teunissen PW. Making theory explicit - an analysis of how medical education research(ers) describe how they connect to theory. *BMC Med Educ* 2017;17:18.
20. Bordage G. Conceptual frameworks to illuminate and magnify. *Med Educ* 2009;43:312–9.



21. Ringsted C, Hodges B, Scherpbier A. "The research compass": an introduction to research in medical education: AMEE Guide no. 56. *Med Teach* 2011;33:695–709.
22. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005;15:1277–88.
23. Sandelowski M. Whatever happened to qualitative description? *Res Nurs Health* 2000;23:334–40.
24. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014;89:1245–51.
25. Ramani S, Könings KD, Mann K, van der Vleuten CP. A guide to reflexivity for qualitative researchers in education. *Acad Med* 2018;93:1257.
26. Strote J. When in doubt. *Ann Emerg Med* 2012;59:75.
27. Varpio L, Ajjawi R, Monrouxe LV, O'Brien BC, Rees CE. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Med Educ* 2017;51:40–50.
28. Timmermans S, Angell A. Evidence-based medicine, clinical uncertainty, and learning to doctor. *J Health Soc Behav* 2001;42:342–59.
29. Nevalainen MK, Mantyranta T, Pitkala KH. Facing uncertainty as a medical student—a qualitative study of their reflective learning diaries and writings on specific themes during the first clinical year. *Patient Educ Couns* 2010;78:218–23.
30. Lally J, Cantillon P. Uncertainty and ambiguity and their association with psychological distress in medical students. *Acad Psychiatry* 2014;38:339–44.
31. Han PK, Schupack D, Daggett S, Holt CT, Strout TD. Temporal changes in tolerance of uncertainty among medical students: insights from an exploratory study. *Med Educ Online* 2015;20:28285.
32. Goel S, Angeli F, Dhirar N, Singla N, Ruwaard D. What motivates medical students to select medical studies: a systematic literature review. *BMC Med Educ* 2018;18:16.
33. Berger J. Moral distress in medical education and training. *J Gen Intern Med* 2014;29:395–8.
34. Aronson E. Dissonance theory: progress and problems. In: Abelson R, Aronson E, McGuire WJ, Rosenberg MJ, Tanenbaum PH, editors. *Theories of Cognitive Consistency: A Sourcebook*. Chicago, IL: McNally, 1968. pp. 5–27.
35. Thompson BM, Teal CR, Rogers JC, Paterniti D, Haidet P. Ideals, activities, dissonance, and processing: a conceptual model to guide educators' efforts to stimulate student reflection. *Acad Med* 2010;85:902–8.
36. Lehmann LS, Kasoff WS, Koch P, Federman DD. A survey of medical ethics education at U.S. and Canadian medical schools. *Acad Med* 2004;79:682–9.
37. DuBois JM, Burkemper J. Ethics education in U.S. medical schools: a study of syllabi. *Acad Med* 2002;77:432–7.
38. Rogers D. Which educational interventions improve healthcare professionals' resilience? *Med Teach* 2016;38(12):1236–41.
39. Schei E, Johnsrud RE, Mildestvedt T, Pedersen R, Hjørleifsson S. Trustingly bewildered. How first-year medical students make sense of their learning experience in a traditional, preclinical curriculum. *Med Educ. Online* 2018;23:1500344.
40. Festinger L. *A Theory of Cognitive Dissonance*. Evanston, IL: Row & Peterson, 1957.
41. Apramian T, Cristancho S, Watling C, Ott M, Lingard L. Thresholds of principle and preference: exploring procedural variation in postgraduate surgical education. *Acad Med* 2015;90(11 Suppl):S70–76.
42. Campbell AV, Chin J, Voo TC. How can we know that ethics education produces ethical doctors? *Med Teach* 2007;29:431–6.
43. Fryer-Edwards K, Wilkins MD, Baernstein A, Braddock CH 3rd. Bringing ethics education to the clinical years: ward ethics sessions at the University of Washington. *Acad Med* 2006;81:626–31.
44. Wald HS, White J, Reis SP, Esquibel AY, Anthony D. Grappling with complexity: medical students' reflective writings about challenging patient encounters as a window into professional identity formation. *Med Teach* 2019;41:152–60.
45. Young JQ, Van Merriënboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Med Teach* 2014;36:371–84.
46. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract* 2009;14:595–621.
47. Reis SP, Wald HS, Monroe AD, Borkan JM. Begin the BEGAN (The Brown Educational Guide to the Analysis of Narrative) - a framework for enhancing educational impact of faculty feedback to students' reflective writing. *Patient Educ Couns* 2010;80:253–9.
48. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach* 2009;31:685–95.

# A Retrospective Cohort Study of the Effect of Home Institution on Emergency Medicine Standardized Letters of Evaluation

Megan Boysen-Osborn, MD, MHPE<sup>1</sup> , Jessica Andrusaitis, MD, MS<sup>2</sup>, Clelia Clark, MBA<sup>2</sup>, Soheil Saadat, MD, MPH, PhD<sup>1</sup>, John Billimek, PhD<sup>2</sup>, Sara Paradise, MD<sup>1</sup>, Alisa Wray, MD, MAEd<sup>1</sup>, Warren Wiechmann, MD, MBA<sup>1</sup>, and Shannon Toohey, MD, MAEd<sup>1</sup> 

## ABSTRACT

**Background:** The standardized letter of evaluation (SLOE) was developed to make letters of recommendation in emergency medicine (EM) more objective and discerning. Typically, students obtain one SLOE from a home EM rotation and at least one more SLOE from an away clerkship. It is unclear if students perform better on their home or away EM rotations.

**Objective:** The purpose of this study was to determine if students perform better on the group SLOE at their home institution compared to an away institution.

**Methods:** The authors performed a retrospective application review of all allopathic applicants to an urban, academic EM residency program. The authors calculated a composite score (CS) for each group SLOE, using the global assessment scores for comparative rank and rank list position. A lower CS indicates better performance. The authors compared mean CS for students' first home rotations with first away rotations. For students in the study who had a third (second away or second home site) SLOE available, the authors compared mean CS on the students' first SLOEs with mean CS on the students' third SLOEs.

**Results:** A total of 624 records were included in the primary analysis. There was a small, but significant difference between mean CS for students' home rotations when compared to away rotations (4.67 vs. 4.85,  $p = 0.024$ ). Students performed better on their home rotations. Students who had three SLOEs available performed worse on their third rotation (first = 4.40, second = 4.63, third = 4.77,  $p = 0.012$  for first vs. third). For all available SLOEs, more than 50% of students fell into the top 10% or top one-third categories.

**Conclusion:** Students perform slightly better on their home EM rotations. Students' mean SLOE CS is slightly worse for a third rotation when compared to a first rotation.

The standardized letter of evaluation (SLOE) was created by the Council of Residency Directors in Emergency Medicine (CORD-EM) to make letters in emergency medicine (EM) more standardized, concise, and discriminating.<sup>1,2</sup> Standardized letters are more efficient to review and offer better inter-rater reliability

From the <sup>1</sup>Department of Emergency Medicine, University of California at Irvine, Orange, CA; and the <sup>2</sup>University of California at Irvine, Irvine, CA; Dr. Andrusaitis is currently with the Kaiser Permanente Emergency Medicine Residency, San Diego, CA.

Received May 16, 2019; revision received June 11, 2019; accepted June 17, 2019.

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

Author contributions: MBO, JA, AW, WW, and ST conceived the study; JA and MBO designed the experiment and obtained IRB approval; MBO supervised the conduct of the trial and data collection; JA and CC performed the data collection; JB and SS provided statistical advice on study design; JB, JA, SS, ST, and MBO analyzed the data; JA, CC, MBO, SP, ST, AW, WW, and CC drafted the manuscript; and all authors contributed substantially to its revision. MBO takes responsibility for the paper as a whole.

Supervising Editor: Michael Gottlieb, MD, RDMS.

Address for correspondence and reprints: Megan Boysen-Osborn, MD, MHPE; e-mail: mbo@uci.edu.

AEM EDUCATION AND TRAINING 2019;3:340–346.

than narrative letters of recommendation (NLOR).<sup>3</sup> The SLOE stratifies students into quantiles based on their performance in various categories, including an overall assessment. The SLOE may provide a more objective assessment of student performance than a NLOR.<sup>2</sup>

Most advisors in EM recommend that students obtain at least one SLOE from an EM clerkship rotation at their home institution and at least one other SLOE from an “away” EM rotation.<sup>4</sup> These SLOEs typically replace at least two NLORs in a student’s electronic residency application service (ERAS) application to EM residency programs. Currently, the SLOE is the selection factor used most frequently by program directors (PDs) to make interview and ranking decisions.<sup>5,6</sup>

Despite efforts to maintain objectivity, SLOE authors may inflate scores.<sup>1,7-10</sup> Many SLOE authors do not adhere strictly to quantile categories because they are concerned about negatively impacting a student’s ability to match.<sup>11</sup> It is unclear whether students receive higher scores on the group SLOEs from home rotations versus away rotations. We hypothesized that faculty at a student’s home institution would be more concerned for their own student’s ability to match than for an external student. We sought to determine whether students perform better on the SLOE at their home institution compared to an away rotation.

## METHODS

### Study Design

We performed a retrospective application review of all U.S. MD applications submitted to the University of California at Irvine EM residency program through ERAS in the 2015 and 2016 application cycles. We obtained institutional review board (IRB) approval from the University of California at Irvine prior to commencing this study. The IRB did not require subject consent.

### Study Setting and Population

We included ERAS records for applicants from Liaison Committee for Medical Education accredited U.S. allopathic medical schools.

### Study Protocol

Two trained, nonblinded data abstractors collected the following data onto a standardized data abstraction

form: Association of American Medical Colleges identification number, year of application, sex of applicant, student’s home institution, step 1 score, step 2 clinical knowledge (CK) score, step 2 CS score, and SLOE data. For each SLOE, we collected date and location of rotation, author type, number of letters authored by author in the previous year, comparative rank score (CRS), and rank list position score (RLPS). The CRS is the response to question C1: “Compared to other EM residency candidates you have recommended in the last academic year the candidate is: top 10%, top third, middle third, lower third.” The RLPS is the response to the question C2b: “How highly would you estimate the candidate will reside on your rank list: top 10%, top third, middle third, lower third, unlikely to be on our rank list.” The senior investigator held periodic meetings with the abstractors. We resolved any questions via consensus. These data were stored in an online secure database, REDCap. The senior investigator reviewed all final data to ensure its accuracy, including reviewing whether each home and away institution had been coded correctly. The senior investigator sampled 5% of all records for review and calculated Cohen’s kappa for inter-rater reliability.

### Inclusion and Exclusion Criteria

We screened all SLOEs from traditional 4-week or 1-month EM rotations. We recorded data for all SLOEs from U.S. MD applicants. For the primary study analysis, we excluded SLOEs that were not authored by a faculty group, PD, clerkship director (CD), or any combination thereof. Thus, we did not include SLOEs authored by a single non-PD or CD faculty member, chair, or nonfaculty member, since we felt that they may not have the most accurate information regarding the student’s comparative clerkship performance and rank list position. We excluded SLOEs with incomplete data, such as a missing final global assessment or rank, and SLOEs from an author who had not written more than five SLOEs in the previous or current year. Finally, we excluded records from students whose first and second EM clerkships were not a home and away program (as determined by available SLOEs). Accordingly, we excluded applicants whose home SLOE occurred sequentially third, since not all applicants completed a third rotation. We considered a home program to be an EM training program affiliated with the student’s medical school, with which the faculty have academic appointments, within 30 miles of the medical school campus. We considered any

program to be an away if it was not considered a home program. When it was unclear whether a program should be considered a home or away, the primary and secondary authors e-mailed the PD for clarification.

### Key Outcome Measures

The primary outcome of this study was the home institution effect on SLOE outcome. Secondary outcomes included whether students did better on their first, second, or third SLOE rotation. We described the distribution of scores for all SLOEs. Although not in our original protocol, we compared mean SLOE scores for students who were excluded from the primary analysis for not having a home rotation with mean SLOE scores for students in primary study group.

### Data Analysis

To determine home institution effect on SLOE outcome, we compared performance on students' first home SLOE to performance on first away SLOE. We calculated a composite score (CS) for the SLOE by adding the students' CRS and RLPS. Since the CRS has four options (top 10%, top third, middle third, lower third) and the RLPS has five options (top 10%, top third, middle third, lower third, unlikely to be on our rank list), the best possible CS attainable was a "2" (top 10% in both categories). The worst possible score was a "9" (lower third and unlikely to be ranked). We calculated the CS of each student's first home SLOE and compared it to his or her first away SLOE using the related-samples Wilcoxon signed-rank test. We analyzed data using IBM SPSS Statistics for Windows, version 25.0. We wanted to detect at least a 15% difference in the scores of home and away SLOEs, which corresponds to approximately a one tertile difference in one category for a middle third, middle third SLOE. We assumed the least amount of reliability in the outcome at 50%. To have a power of 80% and a significance of  $p = 0.05$ , the required sample size was 167 participants in each group.

We also describe the distribution of student performance on home versus away rotations. We calculated the difference between the first home CS and the first away CS for each student and reported the percentage of students performing better, worse, or no difference on home SLOEs. For example, a score difference of two means that the student did one quantile better (or worse) in both categories (e.g., middle third, middle third to top third, top third) or did two quantiles

better in one category (e.g., lower third, middle third to top third, middle third).

For students who had a third SLOE available (either a second away or a second home site), we report the mean CS for the first, second, and third SLOEs for that cohort and compare performance between first and third SLOE, using paired-sample *t*-test.

Although not in our a priori protocol, we sought to describe performance for students without a home rotation. We calculated mean CS for students who were originally excluded from our primary analysis for not having a home rotation. We compared the CS from their second SLOE with the primary study mean away CS, using Mann-Whitney *U*-test.

## RESULTS

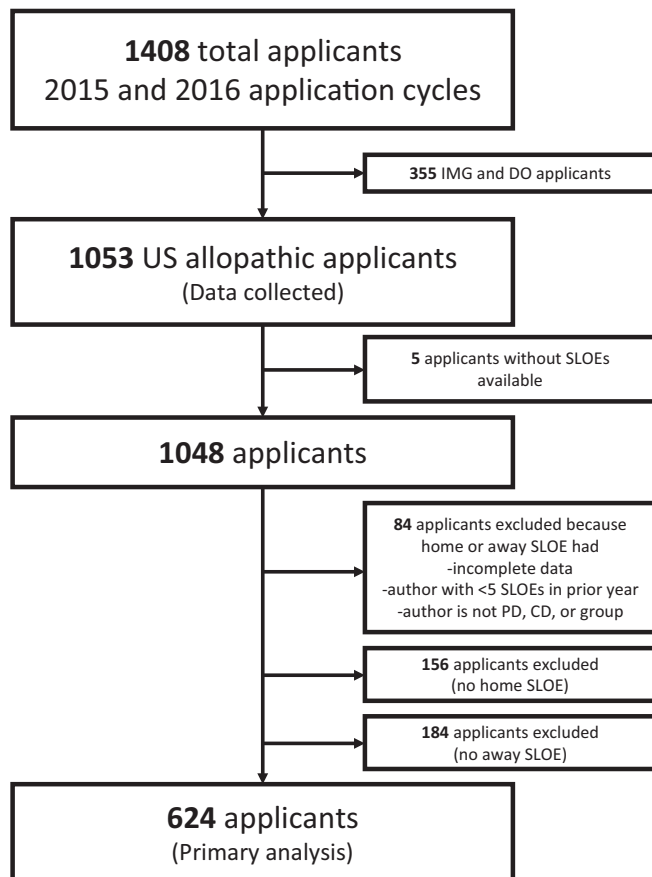
There were 2,078 US allopathic seniors applying to EM in 2015 and 1988 in 2016.<sup>12</sup> For the 2015 application cycle, the University of California at Irvine EM Residency Program received 640 applications. For 2016, the program received 768 applications, for a total of 1,408 subjects. Of these, 1,053 were applicants from U.S. allopathic medical schools. We recorded data for these applicants.

For the primary data analysis, we excluded five records because the student did not have any SLOEs in their application. We excluded the records of 59 applicants because their home or away SLOE was not a group, CD, or PD SLOE or the author had not written more than five SLOEs in the previous or current year. We excluded 25 records for not having complete data (e.g., the student was not given a CRS). We excluded 156 records because the student's first or second SLOE was not a home SLOE; we excluded 184 records for not having an away SLOE. We analyzed records from 624 applicants in our primary analysis (Figure 1). The measured agreement between reviewers was very good ( $\kappa = 0.919$ –1.000).

Of the 624 applicants, 213 (34%) were female. The mean USMLE Step 1 was 232 (standard deviation [SD] = 16, range 167–271); the mean USMLE Step 2 score was 244 (SD  $\pm$  14, range 192–283). Applicants were from medical schools in the West ( $n = 205$ , 33%), Midwest ( $n = 122$ , 19%), South ( $n = 145$ , 23%), and Northeast ( $n = 152$ , 24%). Twenty-five applicants were from our own medical school.

There was a small, but statistically significant difference between students' home rotation CS and away rotation CS, with students performing better (lower





**Figure 1.** Inclusion and exclusion of study subjects. CD = clerkship director; DO = doctor of osteopathic medicine; IMG = international medical graduate; PD = program director; SLOE = standardized letter of evaluation.

CS) on their home rotations (4.67 vs. 4.85,  $p = 0.024$ ). The majority of students completed their home rotation prior to their away rotation ( $n = 559$ , 90%). Students' home CRSs were top 10% ( $n = 117$ , 19%), top third ( $n = 242$ , 39%), middle third ( $n = 206$ , 33%), and lower third ( $n = 59$ , 9%). Their away CRSs were top 10% ( $n = 101$ , 16%), top third ( $n = 250$ , 40%), middle third ( $n = 211$ , 34%), and lower third ( $n = 62$ , 10%). Students' home RPLSs were top 10% ( $n = 121$ , 19%), top third ( $n = 249$ , 40%), middle third ( $n = 178$ , 29%), lower third ( $n = 73$ , 12%), and unlikely to be ranked ( $n = 3$ , <1%). Away RPLSs were top 10% ( $n = 95$ , 15%), top third ( $n = 216$ , 35%), middle third ( $n = 237$ , 38%), lower third ( $n = 71$ , 11%), and unlikely to be ranked ( $n = 5$ , 1%; Figure 2).

One-quarter of students had the same home and away CS ( $n = 158$ , 25%). More students performed better on their home rotation ( $n = 250$ , 40%). Of these, 185 performed two or more points better (30%). Thirty-five percent of students performed better on their away rotation

( $n = 216$ , 35%), with 142 (23%) performing two or more points better (Figure 3).

A total of 175 (28%) had a third SLOE available. Of these, we excluded two because they did not have complete data. We excluded seven because they were not written by a group, PD, or CD or author that wrote more than five SLOEs in the previous year, leaving 166 for use in our study. A student's third SLOE was usually an away rotation ( $n = 162$ , 98%). For this cohort, the mean CSs for students' first, second, and third rotations were 4.40, 4.63, and 4.77, respectively ( $p = 0.012$  for first vs. third).

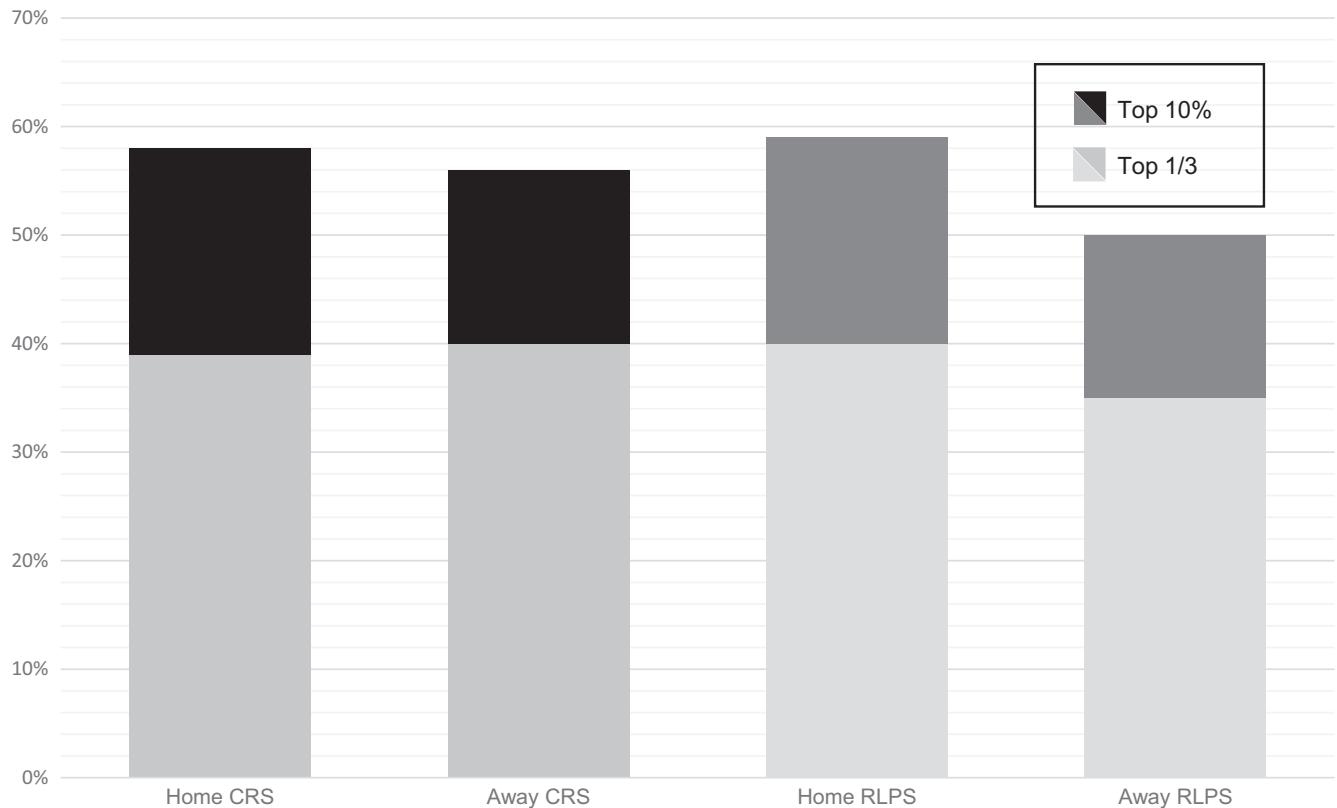
For all SLOEs, including the 624 home SLOEs, 624 away SLOEs, and 166 third SLOEs, 17% of CRSs were in the top 10% ( $n = 246$ ), 40% ( $n = 562$ ) were in the top third, 33% ( $n = 469$ ) were in the middle third, and 10% ( $n = 137$ ) were in the lower third. For the RPLS, 17% ( $n = 242$ ) of students fell in the top 10%, 38% ( $n = 531$ ) were in the top third, 33% ( $n = 465$ ) were in the middle third, 12% ( $n = 168$ ) were in the lower third, and less than 1% were "unlikely to be ranked" ( $n = 8$ ).

We performed a secondary analysis on the cohort of applicants ( $n = 156$ ) who were excluded from our primary analysis for not having a home first or second SLOE. Of these 156 applicants, 29 had SLOEs from only one clerkship. Another three applicants had SLOEs without complete data or they were written by authors who had not written more than five SLOEs in the previous year. Two students had third (but not first or second) SLOEs that were home SLOEs, leaving 122 applicants to analyze. The mean SLOE CS for these 122 students' second SLOE was 5.43, when compared with our study mean for away SLOEs of 4.85 ( $p = 0.001$ ). This cohort's first SLOE CS was 5.12. Third SLOE, when available, averaged 4.85.

## DISCUSSION

It is well documented that SLOEs are the most highly rated and used selection factor by program leadership when selecting applicants for their EM program.<sup>5,6</sup> Most advisors recommend that a student obtain at least two SLOEs to be a competitive applicant in EM.<sup>4</sup> From our sample of U.S. allopathic medical students, most students (67%) did at least one home rotation and at least one away rotation. Most programs provided a group, PD, or CD SLOE, as very few applicants were excluded from the study for not having a group home or away SLOE.





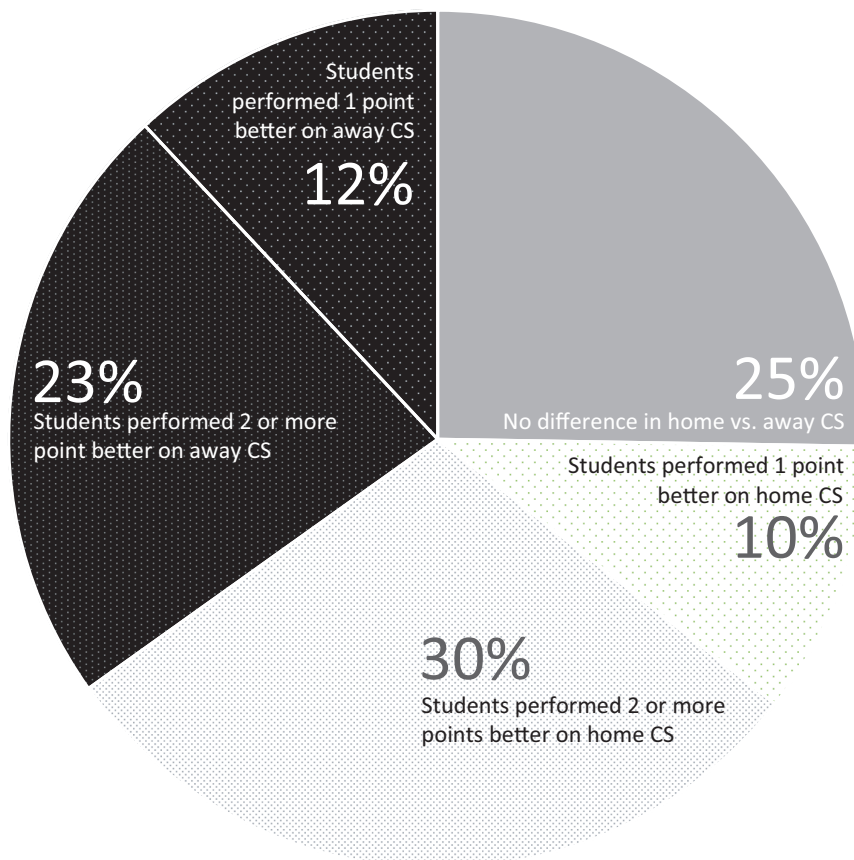
**Figure 2.** Percentage of students (total  $n = 624$ ) receiving a top 10% or top one-third rating CRS and RLPS for home versus away rotations. CRS = comparative rank score; RLPS = rank list position score.

Our study found a small, but statistically significant difference between students' scores on their home rotation versus their away rotation. We suspect that students tend to do better on their home rotation because program and clerkship leadership may be hesitant to rank a student in a lower category for fear they will not match. Furthermore, at a student's home site, students may be more familiar with the hospital system, staff, and consultants and thus perform better. Additionally, students may have been involved in activities at their home institution that gave them more exposure to the SLOE writers. Familiarity with the applicant has been associated with higher SLOE scores in a previous study.<sup>13</sup> We saw a trend of worsening SLOE scores with a student's first, second, and third rotations; it is possible our study difference could be explained by rotation sequence rather than home versus away.

The overall distribution of SLOE scores placed more applicants in the top 10% and top third categories than would be expected. In our study, 57% of candidates fell into either the top 10% or top one-third categories for CRS and 55% of candidates fell into either category for RLPS, exceeding the expected 33%. These findings are consistent with previous reports.<sup>1,10</sup>

The SLOE is intended to stratify applicants into quantile groups. Residency and clerkship leadership are in the roles of both advisor and evaluator. They may be faced with the struggle between objectively assessing students and wanting them to match. Program and clerkship directors likely acknowledge the impact that placing a candidate into the lower categories may have on their competitiveness for a residency position.<sup>11</sup> With only 4 weeks to observe a student, it may also be difficult to neatly place them into a quantile. Since it is common practice to place more students into higher categories, perhaps the SLOE should adopt a system where the lower categories are intended for only a smaller percentage of students; however, if such a system were adopted, lower quantile students would likely have even more difficulty matching and even fewer students would be placed in the lower quantile. It would be interesting to study whether students who receive a lower third on their SLOE have decreased likelihood of matching into EM.

Advisors should acknowledge that—for students with a home EM rotation—students' scores and rankings tend to decrease among their first, second, and third SLOEs. Students should be aware that they will



**Figure 3.** Percentage of students whose CS is better, worse, or the same for home versus away rotations. CS = composite score.

likely not do better on their third rotation. They may still have other reasons to do a third rotation, such as finding whether they “fit” with a program or expanding the regional reach of their application.

While not in our a priori study design, we examined the average SLOE score for students who did not have a home rotation. To make the comparison fair, we compared the scores from their second SLOEs with the scores of the study subjects’ away SLOEs. Students who did not have a home rotation performed significantly worse on their SLOE than students who did have a home rotation. Most of these students were from medical schools without EM residency training programs. These students may have been less prepared than students from schools with an EM program, since they may not have an EM advisor. Interestingly, their scores improved on their third rotation. It may be wise to advise students without a home rotation to do a third rotation.

## LIMITATIONS

There are several limitations to our study. First, we only reviewed SLOEs that were provided to our single

program through ERAS. Applicants may not have provided SLOEs from all the rotations they completed, as there are only four LOR spots available in ERAS. There were more students than expected in the top quantiles of the SLOEs. It is possible that students withheld SLOEs from programs at which they feel they did not perform well. It is also possible a stronger cohort of applicants applied to our program; however, our study mean USMLE was similar/slightly lower than the national mean (2016 = 233, Step 1 = 245, Step 2 = CKCKCK)<sup>14</sup> for matched U.S. senior EM applicants.

Universally, applicants who rotated at our program applied to our program through ERAS; thus, our program’s ranking practices could have been disproportionately represented in the applicant pool. While it is possible that there was selection bias in the study, we believe that our study represents the general population of EM applicants nationwide. While the study USMLE mean was slightly lower than the average for matched EM seniors, our study population includes both matched and unmatched U.S. MD EM applicants. Applicants were geographically distributed throughout the United States, but there were more

applicants from western medical schools. Applicants to our program represented approximately one-quarter of total U.S. MD EM applicants.

For our first, second, and third SLOE data, we relied on programs reporting accurate dates for the student's rotation. While we did see a small advantage at applicants' home institution, we did not adjust the study for quality of the home institution. It is possible that students rotated at more competitive away institutions, when compared to their home institution. It would be difficult to adjust for this variable, as there is not a reliable method for ranking residency training programs. Since most students did their away rotation second in the sequence of rotations, our study findings may be due to rotation sequence, rather than home versus away.

## CONCLUSION

The standardized letter of evaluation is one of the most important parts of a student's application to emergency medicine residency. Students perform slightly better on their home rotation, when compared to an away rotation. Advisors should make sure that students are aware that standardized letter of evaluation performance may decline on a third rotation, unless the student is from a medical school without an emergency medicine residency.

## References

1. Love JN, Deiorio NM, Ronan-Bentle S, et al. Characterization of the council of emergency medicine residency directors' standardized letter of recommendation in 2011-2012. *Acad Emerg Med* 2013;20:926-32.
2. Keim S, Rein J, Chisholm C, et al. A standardized letter of recommendation for residency application. *Acad Emerg Med* 1999;6:1141-6.
3. Girzadas DV, Harwood RC, Dearie J, Garrett S. A comparison of standardized and narrative letters of recommendation. *Acad Emerg Med* 1998;5:1101-4.
4. Lufty-Clayton L, Morrissey T, Pelletier-Bui A, et al. The Emergency Medicine Applying Guide. The Council of Emergency Medicine Residency Directors: Advising Students Committee in Emergency Medicine. Available at: <https://www.cordem.org/globalassets/files/committees/student-advising/2018-applying-guide-cord-cdem-emra-rsa.docx.pdf>. Accessed May 15, 2019.
5. Negaard M, Assimacopoulos E, Harland K, Van Heukelom J. Emergency medicine residency selection criteria: an update and comparison. *AEM Educ Train* 2018;2:146-53.
6. National Resident Matching Program, Data Release and Research Committee: Results of the 2018 NRMP Program Director Survey. Washington, DC: National Resident Matching Program, 2018. Available at: <https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf>. Accessed May 15, 2019.
7. Hegarty CB, Lane DR, Love JN, et al. Council of Emergency Medicine Residency Directors standardized letter of recommendation writers' questionnaire. *J Grad Med Educ* 2014;6:301-6.
8. Love JN, Smith J, Weizberg M, et al. Council of Emergency Medicine Residency Directors' standardized letter of recommendation: the program director's perspective. *Acad Emerg Med* 2014;21:680-7.
9. Grall KH, Hiller KM, Stoneking LR. Analysis of the evaluative components on the standard letter of recommendation (SLOR) in emergency medicine. *West J Emerg Med* 2014;15:419-23.
10. Jackson JS, Bond M, Love JN, Hegarty C. Emergency medicine standardized letter of evaluation (SLOE): findings from the new electronic SLOE format. *J Grad Med Educ* 2019;(April):182-6.
11. Pelletier-Bui A, Van Meter M, Pasirstein M, Jones C, Rimple D. Relationship between institutional standardized letter of evaluation global assessment ranking practices, interviewing practices, and medical student outcomes. *AEM Educ Train* 2017;2:73-6.
12. Association of American Medical Colleges and Electronic Residency Application Services, Emergency Medicine. Available at: <https://www.aamc.org/download/358770/data/emergencymed.pdf>. Accessed on June 7, 2019.
13. Girzadas DV, Harwood RC, Delis SN, et al. Emergency medicine standardized letter of recommendation: predictors of guaranteed match. *Acad Emerg Med* 2001;8:648-53.
14. National Resident Matching Program. Charting Outcomes in the Match for U.S. Allopathic Seniors. Washington, DC. 2016. Available at: <https://www.nrmp.org/wp-content/uploads/2016/09/Charting-Outcomes-US-Allopathic-Seniors-2016.pdf>. Accessed on May 15, 2019.

# Use of the “Step-back” Method for Education Research Consultation at the National Level: A Pilot Study

Jaime Jordan, MD<sup>1,2</sup>, Kaushal Shah, MD<sup>3</sup>, Andrew W Phillips, MD<sup>4</sup>, Nicholas Hartman, MD, MPH<sup>5</sup>, Jeffrey Love, MD, MHPE<sup>6</sup>, and Michael Gottlieb, MD<sup>7</sup>

## ABSTRACT

**Background:** There are a limited number of emergency medicine (EM) physicians with expertise in education research. The Harvard Macy “step-back” method is an emerging model utilized to gather group feedback. Despite its use in multiple educational settings, there are little published data demonstrating effectiveness.

**Objectives:** Our objective was to create and evaluate a national faculty development session providing consultation in education research utilizing the step-back method.

**Methods:** This was a pilot study. EM experts in education research from across the country served as facilitators for a faculty development session held at the 2018 Council of Emergency Medicine Residency Directors Academic Assembly. Small groups consisting of two or three facilitators and one or two participants were formed and each participant underwent a step-back consultation for their education research study. Participants wrote their study question before and after the session. After the session, facilitators and participants completed an evaluative survey consisting of multiple-choice, Likert-type, and free-response items. Descriptive statistics were reported. Qualitative analysis using a thematic approach was performed on free-response data. Participant study questions were assessed by the PICO (population, intervention, comparison, outcome) and FINER (feasible, interesting, novel, ethical, relevant) criteria. Both scales were evaluated using a two-way random-consistency intraclass correlation. Before and after scores were evaluated with a paired t-test.

**Results:** Twenty-four facilitators and 13 participants completed the step-back session. Evaluations from 20 facilitators and nine participants were submitted and analyzed. Sixteen of 20 facilitators felt that the step-back method “greatly facilitated” their ability to share their education research expertise. All facilitators and participants recommended that the session be provided at a future academic assembly. Regarding suggestions for improvement, qualitative analysis revealed three major themes: praise for the session, desire for additional time, and a room set up more conducive to small group work. Seven of nine responding participants felt that the session was “very valuable” for improving the strength of their study methods. Qualitative analysis regarding change in study as a result of the step-back session yielded four major themes: refinement of study question, more specific outcomes and measurements, improvement in study design, and greater understanding of study limitations. Both FINER and PICO scale comparisons showed improvement pre- and postintervention (PICO 60% relative increase; FINER 16% relative increase). Neither achieved statistical significance (PICO  $t(5) = -1.835$ ,  $p = 0.126$ ; and FINER  $t(5) = -1.305$ ,  $p = 0.249$ ).

From the <sup>1</sup>UCLA Department of Emergency Medicine, Ronald Reagan UCLA Medical Center, and the <sup>2</sup>David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA; the <sup>3</sup>Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai Medical Center, New York, NY; the <sup>4</sup>Department of Emergency Medicine, University of North Carolina, Chapel Hill, NC; the <sup>5</sup>Department of Emergency Medicine, Wake Forest School of Medicine, Winston-Salem, NC; the <sup>6</sup>Department of Emergency Medicine, George Washington University, Washington, DC; and the <sup>7</sup>Department of Emergency Medicine, Rush University Medical Center, Chicago, IL.

Received November 7, 2018; revision received March 21, 2019; accepted March 23, 2019.

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

Supervising Editor: Ting Dong, PhD.

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

AEM EDUCATION AND TRAINING 2019;3:347–352

**Conclusion:** A national-level education research consultation utilizing the step-back method was feasible to implement and highly valued by facilitators and participants. Potential positive outcomes include refinement of study question, more specific outcomes and measurements, improvement in study design, and greater understanding of limitations. These results may inform others who want to utilize this method.

The field of education research in emergency medicine (EM) is burgeoning. There has been a call for increased methodologic rigor in education research.<sup>1–6</sup> An estimated 43% of academic EM faculty are primarily involved in education, but a relatively small number possess formal training or expertise in performing and disseminating education research.<sup>7</sup> This problem is not unique to EM. Consensus groups of medical educators have concluded that lack of expertise and mentorship are among the most significant barriers to the production of high-quality education research.<sup>8</sup> Methods and venues that provide further training in education research techniques and spark cross-institutional mentorship are needed to address these gaps.

The “step-back” method is a technique for developing and critiquing project proposals in a collaborative as well as objective fashion. It was proposed and described by Dr. Robert Kegan of the Harvard Macy Institute for Physician Educators in 2002 and has been used in their courses.<sup>9</sup> During a step-back session, the project presenter provides a summary and then “steps back,” allowing the other members of the group to take on and develop the project as though it were their own, without input from the presenter. At the end, the presenter returns to the conversation. This technique allows new ideas and perspectives to be fully entertained without being inhibited by the presenter responding immediately. Given the format, active engagement and collaboration are encouraged, which has been shown to improve learning outcomes.<sup>10–12</sup> Despite potential benefits, little has been published about the efficacy of the step-back technique or the applicability to the EM education research community.

The Council of Emergency Medicine Residency Directors (CORD) Academy for Scholarship sought to provide faculty development in education research utilizing the step-back method to EM educators at the national level. The objective of this study was to evaluate the efficacy of this approach at a national meeting to allow for the sharing of education research expertise and the development of a community of practice in EM education research. Secondly, we sought to understand in what ways this technique may have aided the development of education research proposals.

## METHODS

### Study Design

This was a pilot survey study of EM educators from across the United States. This study was approved by the institutional review board at Rush Medical Center.

### Study Setting and Participants

This study was performed at the CORD Academic Assembly in April 2018. Attendees of the conference could sign up as participants for the special education research consult session utilizing the step-back method with their conference registration. The CORD Academy for Scholarship identified and recruited session facilitators with education research expertise from across the country based on personal knowledge of Academy leadership. All facilitators were education faculty with a successful track record for publishing education research. Nominations for faculty facilitators were solicited from Academy leadership, agreed upon by consensus, and then recruited by e-mail by session directors to participate. Participation in the session, both as a participant and as a facilitator, and completion of the evaluative survey were voluntary.

### Instrument Development

Two separate evaluation instruments were developed for program participants and facilitators by members of the study group, expert EM education researchers, after literature review to optimize content validity. Survey development followed established guidelines for survey research.<sup>13</sup> The facilitator instrument consisted of five items including multiple-choice, Likert-type, and free-response items. The participant instrument consisted of six items including multiple-choice, completion, and Likert-type items. Items were discussed amongst the study group to ensure response process validity and piloted with a small group of representative subjects. Revisions for clarity and readability were made. Final versions of the evaluations 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/aet2.10349/full>).



**Study Protocol**

Enrolled participants were instructed to prepare a brief education research project idea or proposal in advance of the session. They were also provided with an overview of the session including the goals and objectives as well as the process of the step-back. The objectives of the session were as follows:

- 1. To provide an opportunity for education researchers to have their projects and proposals reviewed by experts for methodologic issues especially focusing on research questions and study design.
- 2. To promote high-quality education research by refining the next wave of projects.
- 3. To connect more junior researchers with senior experts to create the opportunity for mentoring in the future.

Facilitators were also oriented to the goals and objectives of the session and how to perform the step-back consultation. During the session, participants were divided into small groups consisting of two or three facilitators and one or two participants. Each participant was asked to write down their research study question and subsequently underwent a step-back consultation for their education research study for approximately 20 minutes with the goals of receiving targeted feedback on strengths, weakness, potential barriers, solutions, and next steps for their study. The components of the step-back consultation are shown in Figure 1. At the end of the session participants rewrote their study question, and both facilitators and participants completed evaluative surveys.

**Data Analysis**

Descriptive statistics were reported for items with discrete answer choices. Qualitative analysis using a thematic approach was performed by two analysts (JJ and

MG) on data from free-response items. Data were reviewed line by line and assigned codes using the constant comparative technique.<sup>14</sup> The two analysts met to decide a final coding scheme after independent review. This coding scheme was then applied to all data by each of the analysts. Inter-rater agreement was 87.5%. Discrepancies were resolved by in-depth discussion and negotiated consensus. Participant study questions were evaluated by two reviewers (KS and MG), blinded to the time at which the question was written, utilizing the PICO (population, intervention, comparison, outcome) and FINER (feasible, interesting, novel, ethical, relevant) criteria. Each question was evaluated for the “presence” or “absence” of each item in the PICO scale. Each question was also rated as “yes” or “no” for each component of the FINER criteria. Both scales were evaluated using a two-way random-consistency intraclass correlation. Before and after scores were evaluated with a paired t-test. All calculations were performed using SPSS v25 (IBM Corp.).

**RESULTS**

**General Results**

Twenty-four facilitators and 13 participants completed the step-back session. Evaluations from 20 facilitators and nine participants were submitted and analyzed.

**Participants**

Seven of nine responding participants felt that the step-back discussion was “very valuable” for improving the strength of their study methods, one of nine felt that it was “moderately valuable,” and one of nine felt that it was “a little valuable.” No participant felt that it was “not valuable at all.” Five of nine participants had never published an education research manuscript.

Step Back Consultation Outline	
2 minutes:	Presentation of project by the participant presenter
3 minutes:	Group members may each ask clarifying questions of presenter
10 minutes:	Presenter should “Step Back” while the group deliberates and discusses the project as if it were their own. The presenter should be actively listening and taking notes on the groups deliberations.
5 minutes:	The presenter “steps back in” and an interactive discussion between group and presenter takes place.

**Figure 1.** Step-back consultation outline.

For the remaining four participants who had published education research previously, the mean number of publications was  $4.5 \pm 1.73$ . All responding participants recommended that the session be provided again at a future CORD academic assembly, and the majority (8/9) would be willing to serve as a facilitator in the future. Seven participants responded to a question regarding their plans to contact members of their group that were not previously known to them to discuss their project in the future; three responded “yes,” three responded “no,” and one responded “maybe.” When asked how their study changed as a result of the step-back session, qualitative analysis of responses yielded four major themes: refinement of study question, more specific outcomes and measurements, improvement in study design, and greater understanding of study limitations.

### Facilitators

Facilitators positively viewed the session with 16/20 facilitators reporting that the step-back method “greatly facilitated” their ability to share their education research expertise while the remaining 4/20 felt that it “somewhat facilitated” their ability. The majority of facilitators (19/20) considered the step-back session to be participation in a community of practice. All facilitators recommended that the session be provided at a future academic assembly and would participate as a facilitator again. Qualitative analysis of a question regarding suggestions for improvement revealed five major themes: praise for the session, desire for additional time, a room set up more conducive to small group work, greater number of participants, and more advanced preparation of participants. Exemplar quotes include:

“Awesome! I’m going to do this at my med-ed research group at home.”

“Need more time.”

“This was well designed, suggestions for improvement include round tables and more participants.”

“I think participants should submit their research questions in advance.”

### Participant Study Questions

Both FINER and PICO scales showed good interrater reliability, PICO ICC = 0.88,  $p = .001$  and FINER

ICC = 0.713,  $p = .025$ . Both scale comparisons showed improvement pre- and post-intervention [PICO mean(SD) 2.25(1.37) pre-intervention and 3.59(0.49) post-intervention, 60% relative increase; FINER 4.25(1.17) pre-intervention and 4.92(0.204) post-intervention, 16% relative increase]. Neither achieved statistical significance [PICO  $t(5) = -1.835$ ,  $p = .126$  and FINER  $t(5) = -1.305$ ,  $p = .249$ ].

## DISCUSSION

To the best of our knowledge, this is the first study to evaluate the feasibility and effectiveness of the Step Back method for education research consultation. Given that clinician educators often face multiple competing demands for their time and may have limited local education research expertise available, this type of faculty development holds great potential to advance the field by providing instruction in education research methodology, dedicated feedback specific to an individual’s project and needs, and an opportunity for collaboration and mentorship that might not have otherwise been accessible.<sup>15–17</sup>

Our pilot study found that this technique was feasible to implement and valued highly by both the facilitators and participants alike. Qualitative assessment identified multiple methodologic areas where improvements in research projects were made from study question to research design to potential limitations. This demonstrates the wide range of content that was addressed through this modality. Another benefit of this session was the ability to create new educational networks for future collaboration, which has been recommended by education researchers as a strategy for success.<sup>18</sup> The majority of facilitators perceived this activity to be participation in a community of practice. While a true community of practice requires multidimensional experiences over time between a group of likeminded educators, this may be the first step in creating a community of practice among the participants.<sup>19</sup> In fact, half of responding participants stated they had plans to follow up with their group and continue the collaboration after the session.

Based on PICO criteria, quality of initial questions designed by participants ranged widely. Not surprisingly, novice researchers often choose study topics based on interest and passion but don’t always apply the PICO criteria, resulting in less rigorous study questions, e.g. “How can we improve medical student

performance on a simulation scenario?” We expected that expert consultation would improve study question and design; however, while this study did show improvement in both scales, it was not statistically significant. Based on analysis by two reviewers, there was a 60% increase in conforming to PICO criteria after the study intervention. Interestingly, the assessment of FINER criteria did not demonstrate much improvement after the intervention which may simply speak to the fact that the questions were already worthwhile endeavors but not rigorously developed. This is also supported by the qualitative analysis identifying refinement of study question, more specific outcomes and measurements, improvement in study design and greater understanding of limitations as improvements in their project rather than change of topic or question content. A lack of power may also be contributory. It is important to note that these assessments only evaluate the study question and while an extremely important part of a research study, there are many other components, as highlighted through qualitative analysis, that are essential for a methodologically sound study that the Step-Back method could impact. It currently remains unclear whether a 20-min step-back consultation could achieve significant improvements, however, given that education research experts particularly in EM are not ubiquitous, the authors believe a forum at a national conference would add value to developing education research projects.

The main lesson learned was that for a “Step-Back” exercise to be successful, conditions should be conducive to a small group discussion. This pilot study occurred in a large hall with all the small group discussions occurring in one room; this was identified as a concern by the participants, facilitators and the program organizers. Future large-scale “Step-Back” exercises would benefit from separating into small meeting rooms with round tables after initial group instruction and adhering to optimal conditions for small group education.<sup>20,21</sup> Participants and facilitators both suggested allowing more time for the discussions. This requires further evaluation. In the studied session, time was allotted by the CORD Academic Assembly Program Committee and influenced by program leaders’ desire to accommodate the greatest number of participants given the available facilitators. The ideal amount of time remains unclear; additional time would probably be more satisfying for

all participants but whether the participants’ research studies would improve significantly requires further investigation.

This pilot study suggests that the “Step-Back” method can be utilized to provide much needed faculty development in education research content, methodology, as well as create an opportunity for collaboration and mentorship, thereby serving to address identified needs to improve the quality and quantity of education research.<sup>1–8,18</sup> Future studies should also assess what is the most effective time frame and number of group members for this technique. Additionally, studies evaluating objective outcomes in the short, intermediate and long term such as methodologic changes, future collaborations, research presentations, and successful publications will shed light on the true value of this program.

---

## LIMITATIONS

It is important to consider several limitations with respect to the current paper. First, this was performed at a national conference of residency educators and it is unclear whether similar benefits would be identified in a different educator population. Similarly, as no demographic data was collected, it is unclear in which populations this program may be beneficial, though we suspect that it would have value to anyone seeking to improve their current level of knowledge in education research. Additionally, the sample size was relatively small. However, we believe this was acceptable given that this was a pilot study of a new research evaluation strategy and the limitations of a conference setting. Additionally, the current study limited the session to 20 minutes. Based upon the feedback, it appears that more time would have been beneficial and it is unclear how this would influence the technique’s effectiveness.

---

## CONCLUSIONS

A national level education research consultation utilizing the step-back method was feasible to implement and was highly valued by both facilitators and participants. Potential positive outcomes include refinement of study question, more specific outcomes and measurements, improvement in study design, and greater understanding of limitations. These results may inform others who wish to utilize this method.

## References

1. Sullivan GM, Simpson D, Cook D, et al. Redefining quality in medical education research: a consumer's view. *J Grad Med Educ* 2014;6:424–9.
2. Cook DA, Beckman TJ, Bordage G. Quality of reporting of experimental studies in medical education: a systematic review. *Med Educ* 2007;41:737–45.
3. Cook DA, Levinson AJ, Garside S. Method and reporting quality in health professions education research: a systematic review. *Med Educ* 2011;45:227–38.
4. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and the quality of published medical education research. *JAMA* 2007;298:1002–9.
5. Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med* 2004;79:955–60.
6. Lurie SJ. Raising the passing grade for studies of medical education. *JAMA* 2003;290:1210–2.
7. Jordan J, Coates C, Clarke S, et al. Exploring scholarship and the emergency medicine educator: a workforce study. *West J Emerg Med* 2016;18:163–8.
8. Yarris LM, Juve AM, Artino AR, et al. Expertise, time, money, mentoring and reward: systemic barriers that limit education research productivity from the AAMC GEA workshop. *J Grad Med Educ* 2014;6:430–6.
9. Harvard Macy Institute. Twenty Years of Innovation: Educating to Innovate in Healthcare. Available at: [http://www.harvardmacy.org/images/documents/hmi\\_20\\_years\\_final.pdf](http://www.harvardmacy.org/images/documents/hmi_20_years_final.pdf). Accessed November 7, 2018.
10. DeClute J, Ladyshevsky R. Enhancing clinical competence using a collaborative clinical education model. *Phys Ther* 1993;73:683–9.
11. Hake RR. Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 1998;66:64–74.
12. Michael J. Where's the evidence that active learning works? *Adv Physiol Educ* 2006;30:159–67.
13. Rickards G, Magee C, Artino AR Jr. You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. *J Grad Med Educ* 2012;4:407–10.
14. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv Res* 2007;42:1758–72.
15. Farley H, Casaletto J, Ankel F, et al. An assessment of the faculty development needs of junior clinical faculty in emergency medicine. *Acad Emerg Med* 2008;15:664–8.
16. Welch J, Sawtelle S, Cheng D, et al. Faculty mentoring practices in academic emergency medicine. *Acad Emerg Med* 2017;24:362–70.
17. Gottlieb M, Dehon E, Jordan J, et al. Getting published in medical education: overcoming barriers to scholarly production. *West J Emerg Med* 2018;19:1–6.
18. Jordan J, Coates WC, Clarke S, et al. The uphill battle of performing education scholarship: barriers educators and education researchers face. *West J Emerg Med* 2018;19:619–29.
19. Lave J, Wenger E. *Situated Learning: Legitimate Peripheral Participation*. New York, NY: Cambridge University Press, 1991.
20. Tomcho TJ, Foels R. Meta-analysis of group learning activities: empirically based teaching recommendations. *Teach Psychol* 2012;39:159–69.
21. Walton H. Small group methods in medical teaching. *Med Educ* 1997;31:459–64.

## Supporting Information

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10349/full>

**Data Supplement S1.** Evaluative surveys.

# Clinical Improvement Interventions for Residents and Practicing Physicians: A Scoping Review of Coaching and Mentoring for Practice Improvement

Casey MacKenzie, MD Candidate,<sup>1</sup> Teresa M. Chan, MD, MHPE, FRCPC, DRCPC,<sup>2,3</sup> and Shawn Mondoux, MD, MSc, FRCPC<sup>4</sup>

## ABSTRACT

**Background:** Graduate medical education (GME) bodies are beginning to mandate coaching as an integral part of the learning process, in addition to current requirements for mentorship. Once an emergency medicine physician transitions beyond graduate training, there is no requirement and little focus on coaching as a method of improving or maintaining clinical practice. Our objective was to understand and describe the current state of the published literature with regard to the use of coaching and mentorship for both GME and practicing physicians.

**Methods:** We conducted a structured review of the literature through PubMed and Google Scholar and included all articles applying coaching or mentorship modalities to GME trainees or practicing physicians. A Google Form was used for standardized data abstraction. Data were collected pertaining to the settings of intervention, the nature of the intervention, its effect, and its resource requirements.

**Results:** A total of 3,546 papers were isolated during the literature review. After exclusion, 186 underwent full-text review by the authors of which 126 articles were included in the final data analysis. Eighty-two articles (65%) pertained to mentorship and 14 (11%) to coaching; the remainder of the articles discussed a combination or variation of these two concepts. Fifty-three (42%) articles were descriptive studies and 35 (28%) were narrative reviews or commentaries. Forty-seven (37%) articles originated from within surgical specialties and coaching was most commonly applied to procedural or manual skills with 22 (17%) instances among all studies.

**Conclusions:** Most literature on coaching and mentorship is descriptive or narrative, and few papers are in the specialty of emergency medicine. Most interventions are limited to single instances of coaching or mentorship without longitudinal application of the intervention. There is an important need to study and publish further evidence on coaching interventions.

Coaches have long been used to increase and augment performance. We are familiar with their place in professional sports and business, but it is a burgeoning area in the field of medical education.<sup>1–3</sup> Coaching in medicine has even made it to the pages of *The New Yorker*, where attending surgeon Atul Gawande recently

wrote about his experiment of employing a coach in the operating room. In his article, Gawande wondered, “Professional athletes use coaches to make sure they are as good as they can be . . . so why did I find it inconceivable to pay someone to come into my operating room and coach me on my surgical technique?”<sup>4</sup>

From the <sup>1</sup>Michael G. DeGroote School of Medicine and the <sup>2</sup>Division of Emergency Medicine, Department of Medicine, McMaster University, Hamilton, Ontario; the <sup>3</sup>McMaster program for Education Research, Innovation, and Theory (MERIT), Hamilton, Ontario; and the <sup>4</sup>Institute of Health Policy, Management and Evaluation (IHPE), University of Toronto, Toronto, Ontario, Canada.

Received January 15, 2019; revision received March 12, 2019; accepted March 18, 2019.

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

Supervising Editor: Daniel J. Egan, MD, FACEP.

Address for correspondence and reprints: Shawn Mondoux, MD, MSc, FRCPC; e-mail: shawn.mondoux@medportal.ca.

AEM EDUCATION AND TRAINING 2019;3:353–364



In our field and specialty of emergency medicine, there is little research examining the utility of coaches. In this paper, we sought to understand the state of literature regarding coaching in medical education and translate this to emergency medicine.

Defining Coaching and Mentoring

Many definitions for coaching are put forward in the literature,<sup>5–9</sup> and all have very similar elements. Common elements of various nonclinical coaching frameworks include interactions between coach and learner that occur over several observations and time and depend on a trusted respect-based bond; direct contemporaneous observation of practice in the clinical environment; clearly defined and specific areas of analysis and improvement; creation of a cycle of continuous skill/performance improvement in the learner (rather than reach a finite endpoint); encouragement of critical self-reflection and collaborative goal-setting; and feedback and reflections that are timely, specific, actionable, and task-directed.<sup>2,5,6</sup> There is some debate within the literature about the timeline for coaching, with some advocating for a longitudinal relationship and others noting that it can be time-limited.<sup>10</sup> Table 1 shows a summary of various definitions of coaching and mentorship in the literature.

When reading on the topic of coaching, however, the historical confusion between mentorship and coaching is worth considering and understanding. Coaching differs from mentorship in that the latter

requires no direct observation of skills, is based largely on general advice and guidance, may not frame discussion within the goals of the mentee, and largely uses personal or lived experiences as a method of motivating or guiding learners.<sup>5</sup> In the period of data collection and analysis for this article, the definition of coaching in the medical context has changed. While experts such as Watling had previously posted a definition of medical coaching on his academic website, subsequent publications<sup>9</sup> have identified that coaching suffers from a lack of clear definition in the medical literature. As a result, there is some confusion in the field as to what the terms *coaching* and *mentorship* mean to us in medicine.<sup>9</sup> While we must acknowledge that mentorship has had a substantive impact on a number of academic outcomes (e.g., career satisfaction, academic productivity, funding),<sup>12–17</sup> the use of mentorship in the clinical sphere blurs substantively with the more recent literature on coaching—and hence it is important to bear in mind how this concept might overlap in the clinical environment.

Interestingly, although confusion exists, there has been a large push from major educational bodies to adopt a coaching-type framework. As graduate medical education (GME) accrediting bodies expect a shift toward coaching relationships, understanding how these differ from current training modalities and learner relationships will be important. For instance, the Royal College of Physicians and Surgeons of Canada has recently adopted the new “Competency By Design”

Table 1  
Various Definitions of Mentorship and Coaching From the Literature

Mentorship (Launer) <sup>5</sup>	Coaching (Watling) <sup>9</sup>	Coaching (Deiorio) <sup>6</sup>	Coaching (Ives) <sup>11</sup>
<ul style="list-style-type: none"><li>• Generally guidance and support offered by a more experienced colleague</li><li>• Can arise both formally and informally</li><li>• Can cover a broad range of topics which include professional and clinical spheres but also often include topics of personal concern</li><li>• Generally a private encounter</li><li>• Confidentiality is expected on both sides</li><li>• Arrangement is entirely voluntary</li></ul>	<ul style="list-style-type: none"><li>• Mutual engagement</li><li>• Clear bilateral orientation toward learner growth</li><li>• Embeds bidirectional reflection</li><li>• Acknowledges that failure as inevitable and invaluable</li><li>• Implied longitudinal relationship (which may not be the case in the medical education setting)</li></ul>	<ul style="list-style-type: none"><li>• Longitudinal learner and coach relationship based in trust</li><li>• Built on respect rather than friendship</li><li>— Should include both negative and positive reactions</li><li>— Includes the review of objective assessments</li><li>— Coaches help develop increased self-monitoring</li><li>• Explicit creation of goals</li><li>• Identification of strategies to manage existing and potential challenges</li><li>• Improvement of learner academic performance</li><li>• Development of professional identity</li></ul>	<ul style="list-style-type: none"><li>• A systematic process designed to facilitate development (change), whether cognitive, emotional, or behavioral</li><li>• Intended for a nonclinical population</li><li>• An individualized, tailor-made approach</li><li>• Aims to encourage coachees to assume charge of their life</li><li>• Based on the twin growth areas of awareness and responsibility</li><li>• Reliant of the twin skills of listening and questioning</li><li>• Involve a collaborative and egalitarian relationship, rather than one based on authority</li><li>• Creates a relationship within which the client agrees to be held accountable for the choices she makes</li><li>• Designed to access the inner resourcefulness of the client and built on her wealth of knowledge, experience</li><li>• Focused on the achievement of a clear stated goal, rather than problem analysis</li></ul>

(CBD) framework,<sup>1</sup> which involves a heavy emphasis on providing all CBD residents with individualized educational planning via assistance from academic coaches.<sup>1</sup> This push likely comes from the observation that coaching is seen to be learner-centered and has proven its effectiveness in other domains.<sup>9</sup>

### Coaching and the Practicing Physician

Compared to GME, there is little emphasis on coaching in the realm of continuous professional development. To identify areas of improvement and design interventions, we rely heavily on physician self-assessment,<sup>18,19</sup> but it is increasingly clear that physicians do not effectively identify opportunities for personal improvement.<sup>20</sup> Specifically, while physicians are often privy to various forms of practice reports (e.g., peer or multi-source feedback, practice analysis reports, and patient reviews), internalizing this information for change and improvement is not ubiquitous. It is possible this is why many have turned to the constructs of coaching or mentorship as methods to overcome this challenge. However, it is unclear if coaching fits the bill in overcoming these challenges to self-improvement—and, despite the lack of evidence observed in the GME landscape, there is high interest in the concept of coaching.<sup>4</sup>

### Objectives

Our primary outcome was to perform a scoping review to understand the current state of the literature of coaching as applied to residents and active practicing physicians. Specifically, we sought to understand what currently exists in the field of medical education as it pertains to effectively improving and/or maintaining competence in clinical practice. Our secondary outcomes were to understand the trends in the literature, as well as the contexts in which the interventions were applied (e.g. specialty, level of training, and clinical environment), their structure of these coaching interventions, the types of skill improvement sought, and the outcomes of the interventions.

## METHODS

### Study Design

We performed this study using the scoping review approach similar to the one described by Arksey and O'Malley.<sup>21,22</sup> A scoping review was the preferred approach to this concept as our review focused primarily on the assessment of the state of a field of educational work and approaches we called coaching and

mentorship rather than a focused question within these fields. This approach offers researchers and scholars a method to map the existing literature, allowing us to find gaps in the literature.<sup>21,22</sup> Furthermore, we expected the types and quality of these articles to be varied, a scenario best suited to a scoping review.<sup>21,22</sup>

### Data Sources and Search Strategy

With the assistance of a trained health sciences librarian, we conducted an initial search using PubMed in January 2018. We included literature from 1960 to January 15, 2018. The search was limited to English-language papers using “and/or” combinations of the following keywords/phrases: “mentor,” “coach,” “professional development,” “continuing education,” “leader coaching,” “peer coach,” “leader skill development,” “continuing professional development,” “reciprocal peer coach,” “staff development,” “career success,” “facilitation,” “personal development,” “surgery,” and “medicine.” Data Supplement S1, Appendix S1 (available as supporting information in the online version of this paper, which is available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10345/full>), shows our full search strategy in Boolean and plain-text terminology.

To supplement the results from PubMed, an adjunctive search was performed using Google Scholar as per previously described methods.<sup>23–28</sup> Google Scholar has been previously shown as a stand-alone alternative search method, which can replace the use of other databases for review papers.<sup>29</sup> For our Google Scholar search, the keywords “surgeons,” “mentor,” “coach,” “professional development,” “continuing education,” “leadership,” “peer teaching,” “faculty,” and “skill” were used. We did not specify a time period for publications and accessed all articles resulting from the two searches. Duplicates from the two searches were excluded.

### Article Selection and Retrieval

For the scoping analysis, we included all papers that provided guidance to readers wishing to conduct, report, or write studies on GME coaching in medicine. Our inclusion criterion was that articles had to include participants that were in medical education at a GME level (not only about medical students) or actively practicing. We did not limit the inclusion of articles to EM literature and instead looked more broadly into the general medical and medical education literature. We included commentaries because we wished to be inclusive in our understanding of the scope of the field—many times commentaries on a

field will include important thought-advancing ideas that can shed light on the ways that the field is moving. We elected to include these, since we know that the concept of coaching is relatively new to medical education and, hence, wanted to capture some of these thoughts and opinions. Exclusion criteria were articles that were concerning only undergraduate medical education or articles that concerned only nonmedical fields.

**Title Review.** Two investigators (TMC, CKM) performed the literature search and three investigators (TMC, CKM, SM) reviewed titles for relevancy. We ensured inter-rater with our title search, and we independently reviewed titles and then met to calibrate. We set a threshold of a Cronbach's alpha of greater than 0.80 to proceed with our reviewing. Our initial baseline scoring was 0.70 for 75 titles. This triggered a calibration exercise, which was conducted with all three reviewers actively discussing the include and excluded studies from the first group of titles. After calibration, Cronbach's alpha was 0.83. We then distributed the remainder of the titles evenly among the three reviewers.

**Abstract Review.** We then reviewed the abstracts of papers that remained after assessing titles. The same inclusion and exclusion criteria were used. Again, inter-rater consistency was done using the same method described above (Cronbach's alpha of 0.82 for 33 citations after a calibration exercise with all three reviewers).

### Data Extraction and Analysis

Following this, we reviewed the full text of the remaining papers that had selected abstracts. These papers were reviewed for inclusion and exclusion criteria as described. We used a Google Form to collect, categorize, and document each paper that met inclusion criteria and passed exclusion criteria. In the design of this paper, three investigators agreed upon what information would be extracted from each study. To ensure consistency between the investigators, two samples were selected to review as a group. The remaining papers were divided equally among the investigators for a final review of the text. For any papers for which a reviewer was unsure, a second reviewer was asked to engage in a review of that same paper. We excluded any papers where we could not obtain the full text via our library or via personal correspondence with the authors.

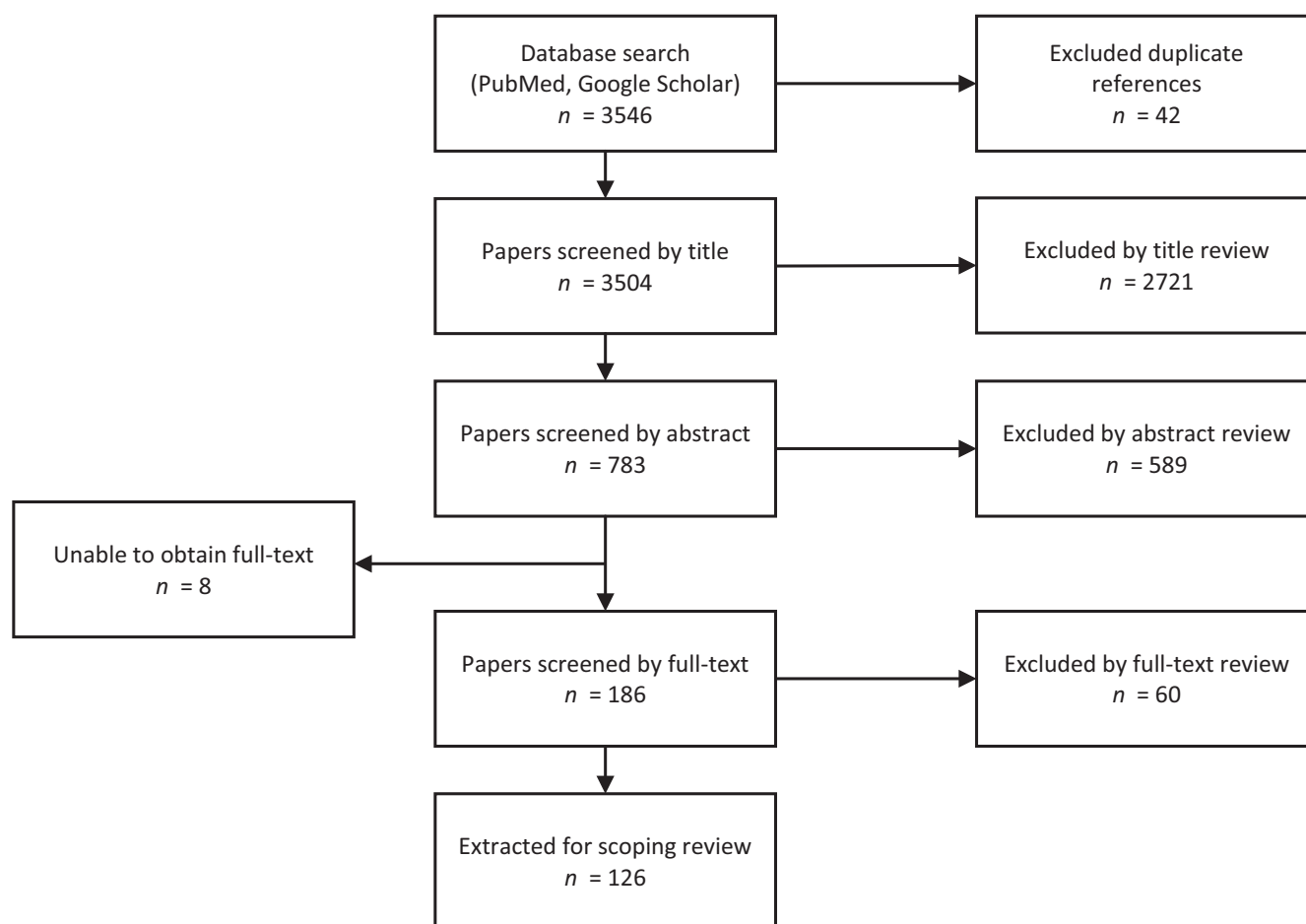
We extracted the following data on our papers: population of study, country of origin, type of study, number of individuals to which the intervention was applied, reason for the study, specialty and setting, description of the intervention, other interventions simultaneously applied, quantitative and qualitative outcomes described, type of change sought by the authors (cognitive skills, emotional skill, manual skill, etc.), level of implementation, Kirkpatrick level outcomes (level 1—acceptability/satisfaction; level 2—knowledge acquisition; level 3—behavioral change; level 4—organizational/systemwide/patient care level outcomes), results of interventional package, and cost of the intervention.<sup>30</sup> Data extraction elements for each paper are enumerated in Data Supplement S1, Appendix S2.

**Thematic Analysis.** After data extraction, the papers left for full-text review were further thematically analyzed to examine for key themes and trends. The primary investigator (SM) conducted a content analysis of these remaining papers and presented these themes for triangulation and discussion with the other two members of the research team (CM, TC).

## RESULTS

Our initial search yielded a total of 3,456 papers, 42 of which were immediately excluded as duplicates. Our title and abstract reviews excluded another 2,721 and 589 articles, respectively. We were unable to obtain the full text of eight articles. A total of 186 articles underwent full-text review, of which 60 were further excluded upon review of the entire article. This resulted in a total of 126 articles that were included in our final scoping review and data analysis. Figure 1 shows the flow diagram of our search, selection, and extraction.

Within these final 126 articles, 82 (65%) papers included content in the area of mentorship, 14 (11%) on coaching, 15 (12%) included some variation on mentorship or coaching, nine (7%) manuscripts fit into the maybe category, and 6 (5%) papers were on both coaching and mentorship. Sixty-three (50.4%) articles studied populations in the United States and 24 (19.2%) studied populations in Canada. The breakdown of the article types can be seen in Table 2. An example of those that included a variation of mentorship or coaching include performance feedback interventions<sup>31</sup> and effectiveness of enhanced peer assessment programs.<sup>32</sup>



**Figure 1.** Flow diagram showing the review of literature.

**Table 2**  
Types of Articles Identified in the Scoping Review on Coaching or Mentorship

Type of Article	Number of Articles	% of Total
Descriptive studies (minimal evaluation and data)	52	34.2
Commentaries or narrative reviews	35	28.0
Knowledge synthesis (systematic review, narrative review, scoping, meta-analysis)	12	9.6
Survey results	12	9.6
Clarification study (why does this work?)	5	4.0
Justification studies	3	2.4
Observational or qualitative studies	3	2.4
Working group or consensus statements	2	1.6
Cognitive engagement as measured by electroencephalogram in OR during observation	1	0.8

A total of 30 articles (24%) were targeted to residents exclusively; 52 (41%) applied only to practicing physicians. Of those papers that discuss practicing physicians, 12 were commentary articles (23%) and 28 (54%) were descriptive studies. Four of the studies about practicing physician education were surveys, one study was an observational study, and three were review articles (one white paper on mentorship relationships, one on techniques for surgical coaching,

and one systematic review on models for mentorship). Of the descriptive studies, 18 (64%) originated exclusively from within surgical specialties (no inclusion of any other specialties in the population) and seven of these papers (39%) were on manual skill acquisition for a specific surgical task.

A variety of specialties were represented in the review, some of which were applied to highly specialized groups such as clinician educators, academic

**Table 3**  
Coaching Articles and Associated Interventions

Classification	Article	Interventions Mentioned	Article Subtype
Coaching	Palamara K, Kauffman C, Stone VE. Promoting success: a professional development coaching program for interns in medicine. <i>J Grad Med Educ</i> 2015;7:630–7.	Program development, coach selection and training, and coachee perceptions. <sup>33</sup>	Descriptive study
	Greenberg CC, Ghousseini HN, Pavuluri Quamme SR, et al. A statewide surgical coaching program provides opportunity for continuous professional development. <i>Ann Surg</i> 2018;267:868–73.	Statewide implementation of evidence-based coaching curriculum. <sup>34</sup>	Descriptive study
	Iyasere CA, Baggett M, Romano J, Jena A, Mills G, Hunt DP. Beyond continuing medical education: clinical coaching as a tool for ongoing professional development. <i>Acad Med</i> 2016;91:1647–50.	Description of the implementation of a coaching program. <sup>23</sup>	Descriptive study
	Mutabdzic D, Mylopoulos M, Murnaghan ML, et al. Coaching surgeons: is culture limiting our ability to improve? <i>Ann Surg</i> 2015;262:213–6.	Study of physician responses to ideas of surgical coaching. <sup>35</sup>	Descriptive study
	Hu Y, Peyre SE, Arriaga AF, et al. Postgame analysis: using video-based coaching for continuous professional development. <i>J Am Coll Surg</i> 2012;214:115–24.	Video-based coaching for improvement of specific surgical skill. <sup>36</sup>	Clarification study
	Sekerka LE. Peer coaching as a technique to foster professional development in clinical ambulatory settings. <i>J Contin Educ Health Prof</i> 2003;23:30–37.	Intervention not described. Study of the perceptions of coaches and coachees to educational program. <sup>37</sup>	Clarification study and survey
	van de Wiel MW, Van Den Bossche P, Janssen S, Jossberger H. Exploring deliberate practice in medicine: how do physicians learn in the workplace? <i>Adv Heal Sci Educ</i> 2011;16:81–95.	Intervention not described. A survey of how physicians best learn in the workplace. <sup>38</sup>	Survey
	Sarkiss CA, Philemond S, Lee J, et al. Neurosurgical skills assessment: measuring technical proficiency in neurosurgery residents through intraoperative video evaluations. <i>World Neurosurg</i> 2016;89:1–8.	Intervention not described. Development of a “skill grading score” from peer coaches. <sup>39</sup>	Survey
	Min H, Morales DR, Orgill D, Smink DS, Yule S. Systematic review of coaching to enhance surgeons’ operative performance. <i>Surgery</i> 2015;158:1168–91.	Intervention not described. Summarize current evidence for surgical coaches. <sup>40</sup>	Knowledge synthesis—systematic review
	Sargeant J, Bruce D, Campbell CM. Practicing physicians’ needs for assessment and feedback. <i>J Contin Educ Health Prof</i> 2013;33 Suppl 1:S54–62.	Intervention not described as the focus. Commentary on how assessment and feedback can guide CPD. <sup>41</sup>	Commentary
	Deiorio NM, Carney PA, Kahl LE, Bonura EM, Juve AM. Coaching: a new model for academic and career achievement. <i>Med Educ Online</i> 2016;21:1–4.	Intervention not described. Development of definitions and academic constructs for coaching. <sup>6</sup>	Commentary
	Linney B. Coach your physicians to care, listen and connect with patients. <i>Physician Exec</i> 2001;27:36–9.	Intervention not described. Commentary on coaching to improvement physician/patient relationship. <sup>42</sup>	Commentary
	Zahid A. Coaching experts: applications to surgeons and continuing professional development. <i>Surg Innov</i> 2018;25:77–80.	Intervention not described. Narrative on coaching to identify practice gaps. <sup>43</sup>	Commentary
	Peyre SE, Frankl SE, Thorndike M, Breen EM. Observation of clinical teaching: interest in a faculty development program for surgeons. <i>J Surg Educ</i> 2011;68:372–6.	Intervention not described. Survey on faculty perceptions of an observation of teaching program. <sup>44</sup>	Electronic survey

(Continued)



Table 3 (continued)

Classification	Article	Interventions Mentioned	Article Subtype
Coaching and mentorship	Beasley HL, Ghousseini HN, Wiegmann DA, Brys NA, Quamme SRP, Greenberg CC. Strategies for building peer surgical coaching relationships. <i>JAMA Surg</i> 2017;152:e165540.	Strategies used by peer coaches to develop effective relationships. <sup>45</sup>	Descriptive study
	Challacombe B, Kandaswamy R, Dasgupta P, Mamode N. Telementoring facilitates independent hand-assisted laparoscopic living donor nephrectomy. <i>Transplant Proc</i> 2005;37:613–6.	Telementoring in laparoscopic kidney transplant <sup>46</sup>	Descriptive study
	Sargeant J, Lockyer J, Mann K, et al. Facilitated reflective performance feedback: developing an evidence- and theory-based model that builds relationship, explores reactions and content, and coaches for performance change (R2C2). <i>Acad Med</i> 2015;90:1698–706. Sargeant J, Mann K, Manos S, et al. R2C2 in action: testing an evidence-based model to facilitate feedback and coaching in residency. <i>J Grad Med Educ</i> 2017;9:165–70.	Development of evidence-based method for performance feedback. <sup>2,3</sup>	Descriptive study and then followed by an empirical study
	Yardley S, Westerman M, Bartlett M, Walton JM, Smith J, Peile E. The do's, don't and don't knows of supporting transition to more independent practice. <i>Perspect Med Educ</i> 2018;7:8–22.	Intervention not described. Synthesis of evidence in supporting physician transition to practice. <sup>47</sup>	Knowledge synthesis—scoping review
	Ahmed K, Patel S, Aydin A, et al. European Association of Urology Section of Urolithiasis (EULIS) consensus statement on simulation, training, and assessment in urolithiasis. <i>Eur Urol Focus</i> 2018;4:614–20.	Intervention not described. Use of simulation to teach ureteroscopy. <sup>48</sup>	Consensus statement
	Sachdeva AK, Blair PG, Lupi LK. Education and training to address specific needs during the career progression of surgeons. <i>Surg Clin North Am</i> 2018;96:115–28.	Intervention not described. Review of new methods to increase effectiveness of CME. <sup>49</sup>	Commentary

Nomenclature using the Cook et al. classification<sup>50</sup> for medical education studies and other commonly found types of scholarship.  
CPD = continuous professional development.

writers, or administrators. Overall, surgical specialties were best represented, exclusively featured in 47 articles (37%); medicine was exclusively featured in nine (7%), and 11 (9%) papers exclusively featured internal medicine or medicine subspecialties. Please note that some articles mentioned multiple specialties. A total of 18 (14%) articles discussed mentorship generally or applied interventions across many specialties.

## Coaching Interventions

To better understand the variety of coaching interventions described in the literature, papers directly pertaining to coaching or combinations of coaching and mentorship are presented in Table 3. A total of 15 papers fulfill these criteria and include a variety of interventions such as telementoring in surgical settings and statewide implementations of coaching curricula. These range broadly from the application of technology, the use of simulation, the assessment of physician perspectives and the knowledge synthesis studies.

## Outcomes in the Coaching and Mentorship Literature

Table 4 describes that the outcomes sought by the papers we found in our review. Data Supplement S1, Appendix S3, contains a listing all the citations we extracted. Of note, we did not find any empirically based coaching articles that originated within the field of emergency medicine.

Thirty studies (24%) achieved Kirkpatrick level 1 outcomes, seven (5%) achieved level 2 outcomes, 12 (10%) achieved level 3 outcomes, and two (2%) achieved Kirkpatrick level 4 outcomes. Eighty-five (68%) were not studies that looked to achieve an outcome and therefore the Kirkpatrick framework could not be applied.

Of the 14 articles that focused on coaching exclusively, five of the articles were commentaries, four were descriptive studies, three were surveys, two were clarification studies, and one article was a knowledge synthesis study. None of these 14 papers on coaching

**Table 4**  
Changes Sought as Part of the Mentorship or Coaching Literature in the Scoping Review

Outcomes Sought	Number of Papers	Sample Article
Procedural/manual skill acquisition	20	Challacombe B, Kandaswamy R, Dasgupta P, Mamode N. Telementoring facilitates independent hand-assisted laparoscopic living donor nephrectomy. <i>Transplant Proc</i> 2005;37:613–6. <sup>46</sup>
Cognitive skill acquisition	19	Greenberg CC, Ghouseini HN, Pavuluri Quamme SR, et al. A statewide surgical coaching program provides opportunity for continuous professional development. <i>Ann Surg</i> 2018;267:868–73. <sup>34</sup>
Emotional skill acquisition	16	Sekerka LE. Peer coaching as a technique to foster professional development in clinical ambulatory settings. <i>J Contin Educ Health Prof</i> 2003;23:30–37. <sup>37</sup>
Improvement of knowledge	15	Berian JR, Thomas JM, Minami CA, et al. Evaluation of a novel mentor program to improve surgical care for US hospitals. <i>Int J Qual Health Care</i> 2017;29:234–42. <sup>51</sup>
Attitudinal change	11	Beasley HL, Ghouseini HN, Wiegmann DA, Brys NA, Quamme SR, Greenberg CC. Strategies for building peer surgical coaching relationships. <i>JAMA Surg</i> 2017;152:e165540. <sup>45</sup>
Understanding elements of current curriculum/programming	5	Healy NA, Glynn RW, Malone C, Cantillon P, Kerin MJ. Surgical mentors and role models: prevalence, importance and associated traits. <i>J Surg Educ</i> 2012;69:633–7. <sup>52</sup>
Synthesis of information or vision documents	5	Patel VM, Warren O, Ahmed K, et al. How can we build mentorship in surgeons of the future? <i>ANZ J Surg</i> 2011;81:418–24. <sup>53</sup>
Aggregation of best practices	1	Pellegrini VD Jr. Mentoring during residency education: a unique challenge for the surgeon? <i>Clin Orthop Relat Res</i> 2006;449:143–8. <sup>54</sup>
Analysis of predictors of choice for a career	1	Buddeberg-Fischer B, Stamm M, Buddeberg C. Academic career in medicine—requirements and conditions for successful advancement in Switzerland. <i>BMC Health Serv Res</i> 2009;9:70. <sup>55</sup>
Not described	43	NA

analyzed participants' ability to retain skills or habits over the long term. One paper, by Palamara et al.,<sup>33</sup> contains the largest cohort of coaches and learners with a total of 98 participants. The other descriptive studies were small and contained 35, 31, and 20 participants, respectively.

### Thematic Analysis

**Surgeons Are Early Adopters of Coaching.** Most coaching articles are attributed to surgical specialties and generally explore the use of coaches to ensure the acquisition of a manual skill. Examples included surgical coaching from specialties such as: urology,<sup>46,48</sup> neurosurgery,<sup>39</sup> general surgeons,<sup>35,36</sup> and multiple other surgical specialties.<sup>34,40,44</sup> This is perhaps the most logical use of coaches in medicine. Greenberg and colleagues<sup>34</sup> proposed a statewide surgical coaching model which centered on the review of previously recorded video and direct surgical coaching. Information technology and video conferencing solutions were also broadly applied to allow the coaching surgeon to attend remotely. Specifically, there were very few papers in emergency medicine, with those papers mainly mentioning career and academic mentoring.<sup>12,56,57</sup>

**Temporality and Design of Interventions.** Within our literature search, there are no papers that reported the results of a longitudinal long-term coaching program within a group of practicing physicians. Furthermore, no paper commented on the design or structure of a sustainable coaching program. Most papers described short-term interventions and skill acquisition and some tested the retention of this skill at a fixed time frame. Additionally, no coaching intervention used an iterative design approach (e.g., plan-do-study-act) to increase satisfaction with and adherence to a coaching framework.

### DISCUSSION

The vast majority of the identified content had mentorship as its main subject area. Descriptive studies and commentaries accounted for most articles. One-third of all studies originated with surgical specialties describing the acquisition of a procedural or manual skill. A large proportion did not describe what outcome was sought and very few achieved observed and sustained behavior change (i.e., Kirkpatrick<sup>30</sup> level 3) or any level of organizational impact (i.e., Kirkpatrick 4).<sup>30</sup> The vast majority did not assess acquisition of skills or abilities in a longitudinal way, at several

proximal and distal time points, after the intervention was performed.

### Confounding Mentorship and Coaching

To be as inclusive as possible, we attempted to include all papers that referenced both mentorship and coaching. We hypothesized that some authors, citing interventions of mentorship, might also be using coaching modalities. In doing so, we found important and impactful systematic reviews on mentorship in the literature. No strong systematic reviews were found explicit to coaching. In fact, a significant amount of the literature identified in both mentorship and coaching realms were commentaries or narrative reviews. These provide a significant amount of opinion surrounding the optimal use of coaching in medicine but provide little empirical evidence on this topic.

As the literature on coaching evolves, so does the definition of coaching itself.<sup>7,9</sup> It has been raised by some authors that the traditional definitions of coaching as applied to sports and music are unlikely to be as easily translated to medicine.<sup>9</sup> This work was published after our literature review and supports the assertion that there is much confusion between notions of feedback, mentorship, and coaching in medical education. A search strategy including mentorship and coaching will be needed until a clear definition is established. The definition used in this paper was put forward by Dr. Watling on his academic webpage, which was no longer available during the peer review process and revisions for this paper. Surprisingly, despite the absence of a clear definition, coaching remains an explicit educational strategy advocated by organizations such as the Royal College of Physicians and Surgeons of Canada.<sup>1</sup> Future publications on coaching interventions will require an explicit definition as part of their manuscripts until a broadly-agreed-upon definition is established.

When designing and publishing interventions on the topics of coaching or mentorship, authors must insist on more clarity and accuracy in the use of the terms coaching and mentorship. As described by Watling and colleagues, coaching implies an ongoing relationship with the goal of increasing skill.<sup>9</sup> Their article clearly describes that the desire for longitudinal coaching in medicine seems less than those in other fields, such as music. This is supported by further work in comparing learning expectations between athletes and physicians.<sup>7,9</sup> To date our field has wrestled with how best to encapsulate the teaching relationships

we have in our medical education environments and this has resulted in a rapid evolution of terminology in our literature. However, we feel that adherence to Watling's call for better definitions when discussing coaching interventions might lead to less confusion and overlap in the medical education literature.

When the current state of the literature on coaching and mentoring is considered, most published studies report satisfaction-based outcomes. There is a great opportunity for emergency medicine to enter into this field, since our field is one which allows for direct observation and assessment quite readily (at least in the GME setting).<sup>58</sup> Future work must go beyond descriptive studies and move to designs that evaluate the acquisition of a skill set. Furthermore, future work must consider longitudinal evaluations that measure both the acquisition and the retention of skill sets when coaching is used as the educational modality. Findings from such studies can then be compared to traditional didactic learning and mentorship with regards to patient-centered outcomes.

### LIMITATIONS

Our scoping review did not explicitly include conference abstracts or the gray literature on this topic. It also excluded textbook chapters and other forms of publication that were not featured in peer-reviewed journals. Some of the knowledge in such texts may be applicable to the design of a longitudinal intervention, which required the application of coaching theory in the clinical environment and should be included when designing such an intervention. Data abstraction was also performed by a single author (SM) with periodic data checks by the remaining authors. Finally, readers are to be reminded that our present synthesis is a scoping review, which is meant to help researchers and scholars map a broad field and look for gaps within the literature base on a topic. The studies included within our review pertain to a broad range of subjects, methods, and outcomes; our summary of the results is not meant for educators to draw any significant conclusions about the efficacy of coaching methods. Finally, due to the lack of clarity in terminology around coaching and mentorship, certain papers, which may have used coaching- or mentorship-like strategies (e.g., group peer review and discussion of practice patterns during audit and feedback<sup>59</sup>), may have been missed in our present search strategy.

## FUTURE DIRECTIONS

This paper clearly demonstrates that there is a gap in the medical education literature; to date few studies have properly examined the needs of clinicians, educational modalities, clinical data needs, and sustainability plans in clinical coaching. While coaching is a term that should not merely supplant mentorship or other concepts in medical education, we must work as a field to define the limitations and best usages of this particular instructional strategy.

### Needs of Practicing Physicians Versus Trainees

Many articles have included trainees, faculty members, and other health care workers as part of mentorship or coaching structures. The assumption in many papers is that trainees require competence in specific skills which benefit from training. Some papers surveyed or tested for competence before the intervention. No studies assessed competence throughout the spectrum of their clinicians for targeted recruitment and coaching, nor did any studies assess whether the coaching needs of trainees versus faculty members differed in any meaningful way or described whether the approach to coaching these individuals may differ. Furthermore, there were no papers that described coaching senior members of the faculty in skills that required improvement.

### The Relationship Between Coaching and Mentorship

There is some clear overlap between coaching and mentorship, and this intersection should be further explored. What tensions might lie, for instance, as trainees transition them from needing direct coaching for their day-to-day tasks (e.g., when they are junior residents) to more mentorship in later periods of training (e.g., about supervisory styles of their teaching)? High-quality coaching interventions should adhere to clear, previously outlined definitions of coaching and not blur the lines with mentorship or other teaching strategies. Perhaps more importantly, we must begin establishing evidence that coaching in medicine leads to improved educational outcomes, patient outcomes, and system-level outcomes while simultaneously being able to replicate these interventions in our own jurisdictions. These will provide a strong foundation for current recommendations for increased coaching in graduate-level and faculty education.

## CONCLUSION

There is a paucity of literature on the effectiveness of coaching in medical education and more specifically in emergency medicine. Moving forward the field should focus on proving the effectiveness of coaching interventions on practice change for practicing physicians and residents or organizational improvement for systems or patients. In addition to no clear definition of coaching in medicine, current evidence to support coaching is limited to very few outcomes-oriented studies. Instead, it mainly consists of many descriptive studies and commentaries. Researchers must begin to design interventions and explicitly measure the ability of coaching to address variability in clinical practice (e.g., rates of ordering imaging, adherence to guidelines), attitudes/professionalism, and other patient- or systems-oriented outcomes.

## References

1. CBD and Coaching. Ottawa: The Royal College of Physicians and Surgeons of Canada, 2019. <http://www.royalcollege.ca/rcsite/cbd/implementation/wbas/coaching-and-cbd-e>
2. Sargeant J, Lockyer J, Mann K, et al. Facilitated reflective performance feedback: developing an evidence- and theory-based model that builds relationship, explores reactions and content, and coaches for performance change (R2C2). *Acad Med* 2015;90:1698–706.
3. Sargeant J, Mann K, Manos S, et al. R2C2 in action: testing an evidence-based model to facilitate feedback and coaching in residency. *J Grad Med Educ* 2017;9:165–70.
4. Gawande A. Personal best: top athletes and singers have coaches. Should you? *New Yorker* 2011:44–53.
5. Launer J. Supervision, mentorship and coaching. In: *Understanding Medical Education: Evidence, Theory and Practice*, 2nd ed. Hoboken (NJ): Wiley-Blackwell, 2013. p. 111–22.
6. Deiorio NM, Carney PA, Kahl LE, Bonura EM, Juve AM. Coaching: a new model for academic and career achievement. *Med Educ Online* 2016;21:1–4.
7. Watling C, Driessen E, van der Vleuten CP, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Med Educ* 2014;48:713–23.
8. Watling C, Driessen E, van der Vleuten CP, Vanstone M, Lingard L. Music lessons: revealing medicine's learning culture through a comparison with that of music. *Med Educ* 2013;47:842–50.
9. Watling CJ, LaDonna KA. Where philosophy meets culture: exploring how coaches conceptualise their roles. *Med Educ* 2019 [Epub ahead of print].



10. Simpson D, Holmboe E, Marcdante K, Anderson A, Souza KH. Job roles of the 2025 medical educator. *J Grad Med Educ* 2018;10:243–6.
11. Ives Y. What is ‘coaching’? An exploration of conflicting paradigms. *Int J Evid Based Coach Mentor* 2008;6:100–13.
12. Yeung M, Nuth J, Stiell IG. Mentoring in emergency medicine: the art and the evidence. *Can J Emerg Med* 2010;12:143–9.
13. Straus SE, Chatur F, Taylor M. Issues in the mentor-mentee relationship in academic medicine: a qualitative study. *Acad Med* 2009;84:135–9.
14. Sambunjak D, Straus SE, Marusic A. A systematic review of qualitative research on the meaning and characteristics of mentoring in academic medicine. *J Gen Intern Med* 2010;25:72–8.
15. Luckhaupt SE, Chin MH, Mangione CM, et al. Mentorship in academic general internal medicine: results of a survey of mentors. *J Gen Intern Med* 2005;20:1014–8.
16. Donovan A, Donovan A. Views of radiology program directors on the role of mentorship in the training of radiology residents. *AJR Am J Roentgenol* 2010;194:704–8.
17. Sambunjak D, Straus SE, Marušić A. Mentoring in academic medicine. *JAMA* 2006;296:1103.
18. Eva K, Regehr G. “I’ll never play professional football” and other fallacies of self-assessment. *J Contin Educ Health Prof* 2008;28:14–9.
19. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med* 2005;80(10 Suppl):S46–54.
20. Eva KW, Regehr G. Effective feedback for maintenance of competence: from data delivery to trusting dialogues. *CMAJ* 2013;185:463–4.
21. Arksey H, O’Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol Theory Pract* 2005;8:19–32.
22. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Res Synth Methods* 2014;5:371–85.
23. Iyasere CA, Baggett M, Romano J, Jena A, Mills G, Hunt DP. Beyond continuing medical education: clinical coaching as a tool for ongoing professional development. *Acad Med* 2016;91:1647–50.
24. Chan TM, Wallner C, Swoboda TK, Leone KA, Kessler C. Assessing interpersonal and communication skills in emergency medicine. *Acad Emerg Med* 2012;19:390–402.
25. Thoma B, Camorlinga P, Chan TM, Hall AK, Murnaghan A, Sherbino J. A writer’s guide to education scholarship: quantitative methodologies for medical education research (part 1). *CJEM* 2018;20:125–31.
26. Chan TM, Ting DK, Hall AK, et al. A writer’s guide to education scholarship: qualitative education scholarship (part 2). *Can J Emerg Med* 2018;20:284–92.
27. Hall AK, Hagel C, Chan TM, Thoma B, Murnaghan A, Bhanji F. The writer’s guide to education scholarship in emergency medicine: education innovations (part 3). *CJEM* 2018;20:563–70.
28. Murnaghan A, Weersink K, Thoma B, Hall AK, Chan T. The writer’s guide to education scholarship in emergency medicine: systematic reviews and the scholarship of integration (part 4). *Can J Emerg Med* 2018;20:626–33.
29. Gehanno J, Rollin L, Darmoni S. Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Med Inform Decis Mak* 2013;13:7.
30. Yardley S, Dornan T. Kirkpatrick’s levels and education ‘evidence’. *Med Educ* 2012;46:97–106.
31. Kaye AD, Okanlawon OJ, Urman RD. Clinical performance feedback and quality improvement opportunities for perioperative physicians. *Adv Med Educ Pract* 2014;5:115–23.
32. Wenghofer EF, Way D, Moxam RS, Wu H, Faulkner D, Klass DJ. Effectiveness of an enhanced peer assessment program: introducing education into regulatory assessment. *J Contin Educ Health Prof* 2006;26:199–208.
33. Palamara K, Kauffman C, Stone VE. Promoting success: a professional development coaching program for interns in medicine. *J Grad Med Educ* 2015;7:630–7.
34. Greenberg CC, Ghousseini HN, Pavuluri Quamme SR, et al. A statewide surgical coaching program provides opportunity for continuous professional development. *Ann Surg* 2018;267:868–73.
35. Mutabdzic D, Mylopoulos M, Murnaghan ML, et al. Coaching surgeons: is culture limiting our ability to improve? *Ann Surg* 2015;262:213–6.
36. Hu Y, Peyre SE, Arriaga AF, et al. Postgame analysis: using video-based coaching for continuous professional development. *J Am Coll Surg* 2012;214:115–24.
37. Sekerka LE. Peer coaching as a technique to foster professional development in clinical ambulatory settings. *J Contin Educ Health Prof* 2003;23:30–37.
38. van de Wiel MW, Van Den Bossche P, Janssen S, Jossberger H. Exploring deliberate practice in medicine: how do physicians learn in the workplace? *Adv Heal Sci Educ* 2011;16:81–95.
39. Sarkiss CA, Philemond S, Lee J, et al. Neurosurgical skills assessment: measuring technical proficiency in neurosurgery residents through intraoperative video evaluations. *World Neurosurg* 2016;89:1–8.
40. Min H, Morales DR, Orgill D, Smink DS, Yule S. Systematic review of coaching to enhance surgeons’ operative performance. *Surgery* 2015;158:1168–91.
41. Sargeant J, Bruce D, Campbell CM. Practicing physicians’ needs for assessment and feedback. *J Contin Educ Health Prof* 2013;33(Suppl 1):S54–62.
42. Linney B. Coach your physicians to care, listen and connect with patients. *Physician Exec* 2001;27:36–9.



43. Zahid A. Coaching experts: applications to surgeons and continuing professional development. *Surg Innov* 2018;25:77–80.
44. Peyre SE, Frankl SE, Thorndike M, Breen EM. Observation of clinical teaching: interest in a faculty development program for surgeons. *J Surg Educ* 2011;68:372–6.
45. Beasley HL, Ghousseini HN, Wiegmann DA, Brys NA, Quamme SR, Greenberg CC. Strategies for building peer surgical coaching relationships. *JAMA Surg* 2017;152:e165540.
46. Challacombe B, Kandaswamy R, Dasgupta P, Mamode N. Telementoring facilitates independent hand-assisted laparoscopic living donor nephrectomy. *Transplant Proc* 2005;37:613–6.
47. Yardley S, Westerman M, Bartlett M, Walton JM, Smith J, Peile E. The do's, don't and don't knows of supporting transition to more independent practice. *Perspect Med Educ* 2018;7:8–22.
48. Ahmed K, Patel S, Aydin A, et al. European Association of Urology Section of Urolithiasis (EULIS) consensus statement on simulation, training, and assessment in urolithiasis. *Eur Urol Focus* 2018;4:614–20.
49. Sachdeva AK, Blair PG, Lupi LK. Education and training to address specific needs during the career progression of surgeons. *Surg Clin North Am* 2018;96:115–28.
50. Cook DA, Bordage G, Schmidt HG. Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Med Educ* 2008;42:128–33.
51. Berian JR, Thomas JM, Minami CA, et al. Evaluation of a novel mentor program to improve surgical care for US hospitals. *Int J Qual Health Care* 2017;29:234–42.
52. Healy NA, Glynn RW, Malone C, Cantillon P, Kerin MJ. Surgical mentors and role models: prevalence, importance and associated traits. *J Surg Educ* 2012;69:633–7.
53. Patel VM, Warren O, Ahmed K, et al. How can we build mentorship in surgeons of the future? *ANZ J Surg* 2011;81:418–24.
54. Pellegrini VD Jr. Mentoring during residency education: a unique challenge for the surgeon? *Clin Orthop Relat Res* 2006;449:143–8.
55. Buddeberg-Fischer B, Stamm M, Buddeberg C. Academic career in medicine—requirements and conditions for successful advancement in Switzerland. *BMC Health Serv Res* 2009;9:70.
56. Regehr G, Bandiera G, Leblanc C, et al. Education scholarship in emergency medicine part 2: supporting and developing scholars. *CJEM* 2014;16(Suppl 1):S6–12.
57. Love JN, Coates WC, Santen SA, Hobgood CD, Mavis BE, Farrell SE. The MERC at CORD scholars program in medical education research: a novel faculty development opportunity for emergency physicians. *Acad Emerg Med* 2009;16 Suppl 2:S37–41.
58. Colmers-Gray IN, Walsh K, Chan TM. Assessment of emergency medicine residents: a systematic review. *Can J Med Educ* 2017;8:e106–22.
59. Cooke LJ, Duncan D, Rivera L, Dowling SK, Symonds C, Armson H. How do physicians behave when they participate in audit and feedback activities in a group with their peers? *Implement Sci* 2018;13:104.

## Supporting Information

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10345/full>

**Data Supplement S1.** Supplemental material.

# Training in the Management of Psychobehavioral Conditions: A Needs Assessment Survey of Emergency Medicine Residents

Jason Pickett, MD<sup>1</sup>, Mary Rose Calderone Haas, MD<sup>2</sup>, Megan L. Fix, MD<sup>3</sup>, Ramin R. Tabatabai, MD<sup>4</sup>, Angela Carrick, DO<sup>5</sup>, Jennifer Robertson, MD<sup>6</sup>, Angelica Veronica Barnes, MD<sup>6</sup>, Amy Ondeyka, MD<sup>7</sup>, Mary Jane Brown, MD<sup>8</sup>, Andrew R. Edwards, MD<sup>9</sup>, and Erin Dehon, PhD<sup>1</sup> 

## ABSTRACT

**Objective:** Mental health–related ED visits are increasing. Despite this trend, most emergency medicine (EM) residency programs devote little time to psychiatry education. This study aimed to identify EM residents' perceptions of training needs in emergency psychiatry and self-confidence in managing patients with psychobehavioral conditions.

**Methods:** A needs assessment survey was distributed to residents at 15 Accreditation Council for Graduate Medical Education–accredited EM programs spanning the U.S. Survey items addressed amount and type of training in psychiatry during residency, perceived training needs in psychiatry, and self-confidence performing various clinical skills related to emergency psychiatric care. Residents used a 5-point scale (1 = nothing; 5 = very large amount) to rate their learning needs in a variety of topic areas related to behavioral emergencies (e.g., medically clearing patients, substance use disorders). Using a scale from 0 to 100, residents rated their confidence in their ability to independently perform various clinical skills related to emergency psychiatric care (e.g., differentiating a psychiatric presentation from delirium).

**Results:** Of the 632 residents invited to participate, 396 (63%) responded. Twelve percent of respondents reported completing a psychiatry rotation during EM residency. One of the 15 participating programs had a required psychiatry rotation. Residents reported that their program used lectures (56%) and/or supervised training in the ED (35%) to teach residents about psychiatric emergencies. Most residents reported minimal involvement in the treatment of patients with psychiatric concerns. The majority of residents (59%) believed that their program should offer more education on managing psychiatric emergencies. Only 14% of residents felt “quite” or “extremely” prepared to treat psychiatric patients. Overall, residents reported the lowest levels of confidence and highest need for more training related to counseling suicidal patients and treating psychiatric issues in special populations (e.g., pregnant women, elderly, and children).

**Conclusions:** Most EM residents desire more training in managing psychiatric emergencies than is currently provided.

From the <sup>1</sup>Department of Emergency Medicine, University of Mississippi Medical Center, Jackson, MS; the <sup>2</sup>Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI; the <sup>3</sup>Department of Emergency Medicine, University of Utah, Salt Lake City, UT; the <sup>4</sup>Department of Emergency Medicine, School of Medicine of the University of Southern California, Los Angeles County + USC Medical Center, Los Angeles, CA; the <sup>5</sup>Department of Emergency Medicine, Norman Regional Health System, Norman, OK; the <sup>6</sup>Department of Emergency Medicine, Emory University School of Medicine, Atlanta, GA; the <sup>7</sup>Department of Emergency Medicine, Inspira Health Network, Vineland, NJ; the <sup>8</sup>Department of Emergency Medicine, University of Tennessee Health Science Center, Nashville, TN; and the <sup>9</sup>Department of Emergency Medicine, University of Alabama at Birmingham, Birmingham, AL.

Received April 5, 2019; revision received June 19, 2019; accepted July 1, 2019.

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

Supervising Editor: Jason Wagner, MD.

Address for correspondence and reprints: Erin Dehon, PhD; e-mail: edehon@umc.edu.

AEM EDUCATION AND TRAINING 2019;3:365–374.

Mental health (MH) and substance use disorders (SUD) affect an estimated 43.6 million and 21.5 million adults, respectively, in the United States each year.<sup>1</sup> Emergency departments (EDs) nationwide are increasingly providing care for individuals with MH and SUD. In 2007, approximately one in eight ED visits were related to a MH and/or SUD.<sup>2</sup> More recent data show a 44.1% increase in MH and SUD visits to the ED between 2006 and 2014.<sup>3</sup> The overall rate of ED visits related to behavioral/MH concerns is increasing at a significantly faster rate than ED visits related to injuries or medical conditions.<sup>3</sup> The increasing number of MH-related visits combined with a national shortage of inpatient psychiatric beds have led to widespread boarding of psychiatric patients. Patients with primary psychobehavioral complaints have been found to wait 3.2 times longer for inpatient placement than patients with nonpsychiatric chief complaints.<sup>4</sup> Given that EDs have become a main source of care for patients with MH complaints, it is important to identify whether EM residents are well prepared to meet this demand and treat these patients.

According to the American College of Emergency Physicians, “the practice of emergency medicine includes the initial evaluation, diagnosis, treatment coordination of care among multiple providers, and disposition of any patient requiring expeditious medical, surgical, or psychiatric care.”<sup>5</sup> Currently, the 2018 Accreditation Council for Graduate Medical Education (ACGME) emergency medicine (EM) residency program requirements do not specify that programs ensure residents have ample experiences treating psychiatric patients. Instead, the ACGME states that “residents must demonstrate knowledge of established and evolving biomedical, clinical, epidemiological and social-behavioral sciences, as well as the application of this knowledge to patient care.”<sup>6</sup> The most recent version of the EM Model, the specialty’s guiding document for curricula, includes the following conditions under psychobehavioral disorders as core content areas: SUD, mood disorders, thought disorders, factitious disorders, neurotic disorders, organic psychoses, patterns of violence/abuse/neglect, personality disorders, psychosomatic disorders, and feeding and eating disorders.<sup>7</sup> The EM Model’s section on procedures and skills integral to the practice of EM includes two components within the psychobehavioral category: psychiatric screening examination and violent patient management/restraint.<sup>7</sup>

Based on these guidelines, it appears that EM residents are expected to develop skills in treating psychobehavioral conditions through on-the-job training in the ED. This leaves EM programs with a wide array of variation and emphasis regarding the appropriate amount of education to EM residents in the area of psychobehavioral conditions. Although somewhat outdated, due to the paucity of literature in this area, one study from 2003 on the scope of psychiatric education provided by EM training programs showed that only 14% of the surveyed programs included a 1-month psychiatry rotation and 67% of these programs did not provide or require any formal training in the acute management of psychobehavioral conditions.<sup>8</sup> Additionally, EM residents have little incentive to study material related to MH/SUD since only 4% of the questions on the American Board of Emergency Medicine Certification Exam pertain to psychobehavioral disorders.<sup>9</sup> The importance of psychobehavioral disorders is also deemphasized in core EM textbooks (i.e., Tintinalli and Rosen) that allot only 2% to 3% of content to psychobehavioral disorders.<sup>10,11</sup>

There is limited research on EM residents perceived educational needs in managing psychobehavioral conditions. A survey conducted in 1988 among program directors of nonpsychiatric residencies (i.e., EM, family medicine, pediatrics, obstetrics-gynecology, and surgery) found a lack of training in emergency psychiatric interventions.<sup>12</sup> However, this study is outdated and included only the program directors’ perspectives. To begin to address these gaps in the literature, we conducted an educational needs assessment survey that aims to identify EM residents’ self-perceived training needs and self-confidence in evaluating and treating patients with psychobehavioral conditions.

## METHODS

### Study Design and Population

We designed and distributed an anonymous and voluntary needs assessment survey to EM resident physicians at 15 ACGME-accredited EM residency programs across the United States from July 2018 to October 2018. We identified a geographically diverse sample of residency programs to participate through directed emails to program directors who serve on the Council of Emergency Medicine Residency Directors (CORD) Resilience Committee. The e-mail did not provide any incentive to participate. EM residents at these select residencies were surveyed using a self-

administered internet or paper questionnaire. The results were then mailed or e-mailed back to our home institution. The study received institutional review board approval from the University of Mississippi Medical Center. Informed consent was waived to preserve participant anonymity.

### Survey Content and Administration

The authors collaborated to create a survey instrument aimed to determine the breadth and depth of psychiatric education, both optional and required, provided by EM residency programs. The survey was modeled using similar formats to existing needs assessments surveys (i.e., Cook et al.<sup>13</sup>). Initial items were identified based on literature about the emergency physician's role in psychobehavioral emergencies.<sup>7,14</sup> A list of potential items were reviewed, edited, and finalized by several of the authors, many of whom are members of the CORD Resilience Committee. Survey items asked EM residents to report their own levels of formal training in psychiatry, comfort treating psychiatric patients, and self-perceived training needs and self-confidence in 15 specific clinical skills related to psychiatric care. Items about demographic information were also included. Most items assessing residents self-perceived training needs and comfort treating psychiatric patients consisted of a 5-point unipolar response scale (1 = nothing or not at all comfortable; 5 = very large amount or extremely comfortable). Residents also rated their confidence in their ability to independently perform various clinical skills related to psychobehavioral conditions (e.g., differentiating a psychiatric presentation from delirium, developing a safety plan with a suicidal patient) using a 100-point scale, ranging from 0 (cannot do at all) to 100 (highly certain can do). The use of a 0 to 100 scale was based on existing recommendations for developing confidence scales.<sup>15</sup> Items can be seen in Tables 2 and 3.

Finally, residents' attitudes toward individuals with mental illness was assessed using the Attitude Subscale of the Opening Minds Scale for Health Care Providers (i.e., Kassam et al.<sup>16</sup>) an instrument that measures providers' attitudes toward people with mental illness. Higher scores suggest a more stigmatizing attitude. Cronbach's alpha was 0.75. Items are listed in Table 4.

### Data Analysis

Results were analyzed with IBM SPSS Statistics for Windows, Version 24.0. Descriptive statistics were

used to analyze resident characteristics and survey responses. Analysis of variance was used to examine differences in self-perceived educational needs and confidence by postgraduate year (PGY) level.

## RESULTS

### Characteristics of Participating Residents

A total of 632 EM residents representing 15 programs and 12 different states were invited to complete the survey. Of those invited to participate, 396 residents completed the survey for a response rate of 62.7%. The respondents were 59.8% male and 40.2% female and included a relatively equal breakdown of PGYs (Table 1).

### Characteristics of Participating Programs

Response rate by institution ranged from 35% to 50% (3/15), 50% to 65% (3/15), 65% to 80% (5/15), and >80% (3/15; 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/aet2.10377/full>, for a list of participating institutions and response rate by institution and PGY level). Eight of the programs used a 3-year training format, five programs used a 4-year format, and two programs were in the midst of transitioning from a 4-year to a 3-year program. Programs spanned all five regions of the U.S. One of the 15 programs had a required psychiatry rotation. Twelve of the 15 participating programs had a "psychiatrist available in real time to consult on ED patients 24 hours a day."

### Training in Psychiatry and Current Involvement in Psychiatric Emergencies

Almost all of the resident participants (>99%) reported that they completed a formal rotation in psychiatry in medical school. Overall, 12% of resident respondents reported completing a psychiatry rotation during their EM residency. The vast majority of these residents were attending a residency program that required a psychiatry rotation. Three residents from other programs reported that they opted to complete an elective psychiatry rotation. All residents reported seeing at least one to two patients with a psychiatric or behavioral complaint during a typical 8-hour ED shift. Although residents frequently evaluate patients with psychobehavioral conditions, over half (55%) identified their level of involvement in managing these patients (beyond medical clearance) as "minimal" or "none."

**Table 1**  
Participant Demographics

Participant Characteristics	<i>n</i> (%) <sup>†</sup>
Sex	
Male	232 (59.8)
Female	156 (40.2)
Postgraduate year	
PGY-1	137 (34.8)
PGY-2	119 (30.2)
PGY-≥3	138 (35.2)
Prior residency training	
No	376 (95.4)
Yes	3 (0.8)
Partially	15 (3.8)
Region	
Northeast	121 (30.6)
Southeast	94 (23.7)
Southwest	23 (5.8)
Midwest	61 (15.4)
West	97 (24.5)
Completed a rotation in psychiatry	
During residency	36 (11.8)
During medical school	314 (99.7)
Average number of patients with psychiatric chief complaints seen in a typical 8-hour shift	
0	0
1–2	143 (46.4)
2–3	98 (31.8)
3–4	39 (12.7)
>4	28 (9.1)

<sup>†</sup>The number of respondents (*n*) varies for categories due to missing data

Residents from institutions who did not have a psychiatrist available for consultation 24 hours a day reported similar levels of involvement as those who did have a psychiatry consultant available.

### Perceptions About Training in Psychiatry

Sixty-nine percent of residents felt it is quite or extremely important to develop competency in the assessment and treatment of patients with psychobehavioral conditions. However, only 13% felt well prepared to assess and treat such patients, and a majority of residents (59%) felt that their program should offer more training in managing psychobehavioral conditions. Residents who felt that their program should offer more training in managing psychobehavioral emergencies thought it should be delivered through a variety of formats including lectures (29%), simulation exercises (26%), supervised training in the ED (21%), a core rotation (9%), and an elective (13%). Residents

indicated that their programs currently provided education about psychobehavioral conditions through use of lectures (56%), supervised training in the ED (35%), and simulation (6%). Six percent of residents reported having no formal education dedicated to management of patients with psychobehavioral conditions.

The majority of residents (mean = 60%) felt a need for at least “moderate improvement” in their clinical practice across 15 areas relevant to emergency psychiatric care. Overall, residents reported needing the least amount of training pertaining to medically clearing patients and the most amount of training pertaining to caring for special patient populations (e.g., children, elderly, and pregnant women). Verbatim wording of the survey items and response data about resident attitudes about their training in emergency psychiatry are presented in Table 2.

### Comfort and Confidence Managing Psychiatric Patients

Only 6% of residents reported feeling quite or extremely comfortable independently performing the assessment and prescribing initial treatment for patients with psychobehavioral conditions. A minority of residents (36%) believed that their attendings are quite or extremely comfortable treating patients with psychobehavioral conditions. Residents’ self-reported confidence performing various clinical skills related to emergency psychiatric care ranged from 0 (cannot do at all) to 100 (highly certain can do). On average, most residents rated their level of confidence performing various skills related to emergency psychiatric care in the moderate range (mean = 60). As expected, confidence levels generally increased by PGY; however, this was not always the case among PGY-4s whose confidence ratings were lower than the mean across several skill areas. A one-way between-groups analysis of variance was performed to investigate the impact of PGY level on levels of confidence. There was a significant difference in overall (average) confidence scores by PGY level ( $F(3,364) = 12.4$ ,  $p < 0.001$ ). Post hoc analyses using the Tukey HSD test indicated that the mean score for PGY-1s (mean  $\pm$  SD =  $50.64 \pm 17.4$ ) was significantly lower than PGY-2s (mean  $\pm$  SD =  $61.42 \pm 17.8$ ), and PGY-3s (mean  $\pm$  SD =  $63.94 \pm 16.40$ ). Supplemental analyses were also performed to examine differences in confidence levels among residents who had completed a psychiatry rotation during residency versus those who had not completed a psychiatry rotation. Surprisingly, residents



**Table 2**  
Resident Attitudes About Training in Emergency Psychiatry

Item	Total N	Mean ( $\pm$ SD), Median	1 <sup>†</sup>	2 <sup>†</sup>	3 <sup>†</sup>	4 <sup>†</sup>	5 <sup>†</sup>
How much do you think you need to learn or improve in each of the following topic areas?			Nothing	A small amount	A moderate amount	A large amount	A very large amount
Medically clearing psychiatric patients	388	2.46 ( $\pm$ 1.1), 3	10.3 (40)	37.6 (146)	29.4 (114)	15.5 (60)	7.2 (28)
Performing a mental status examination	387	2.61 ( $\pm$ 1.0), 3	5.4 (21)	34.4 (133)	37.5 (145)	16.5 (64)	6.2 (24)
Conducting a psychiatric interview	389	2.77 ( $\pm$ 1.0), 3	6.2 (24)	30.8 (120)	38.3 (149)	16.7 (65)	8.0 (31)
Risk assessment of harm to self or others	388	2.69 ( $\pm$ 1.0), 3	5.4 (21)	34.5 (134)	33.0 (128)	19.8 (77)	7.2 (28)
Lethal means counseling	387	3.23 ( $\pm$ 1.0), 3	2.8 (11)	21.4 (83)	33.9 (131)	26.6 (103)	15.2 (59)
Safety planning with suicidal patients	388	3.23 ( $\pm$ 1.0), 3	2.3 (9)	14.7 (57)	41.0 (159)	27.8 (108)	14.2 (55)
Etiologies of altered mental status	388	3.15 ( $\pm$ 1.1), 3	7.7 (30)	37.9 (147)	27.6 (107)	16.5 (64)	10.3 (40)
SUD	387	3.08 ( $\pm$ 1.0), 3	5.2 (20)	37.2 (144)	33.3 (129)	15.5 (60)	8.8 (34)
Mood disorders	388	2.77 ( $\pm$ 0.9), 3	2.6 (10)	29.1 (113)	43.6 (169)	19.3 (75)	5.4 (21)
Use of physical restraints	388	3.08 ( $\pm$ 1.1), 3	6.4 (25)	37.1 (144)	26.5 (103)	20.1 (78)	9.8 (38)
Use of chemical restraints	388	3.15 ( $\pm$ 1.1), 3	5.2 (20)	37.9 (147)	24.7 (96)	21.4 (83)	10.8 (42)
Acutely agitated or psychotic patient	388	3.00 ( $\pm$ 1.1), 3	4.6 (18)	32.2 (125)	33.0 (128)	19.6 (76)	10.6 (41)
Psychiatric emergencies in pregnant women	388	3.62 ( $\pm$ 0.9), 4	0.5 (2)	8.8 (34)	37.1 (144)	32.5 (126)	21.1 (82)
Psychiatric emergencies in the elderly	388	3.38 ( $\pm$ 0.9), 3	0.8 (3)	17.0 (66)	42.5 (165)	26.3 (102)	13.4 (52)
Psychiatric emergencies in children	388	3.38 ( $\pm$ 0.9), 3	0.5 (2)	15.7 (61)	37.1 (144)	30.2 (117)	16.5 (64)
Comfort assessing and treating psychiatric patients			Not at all comfortable	Mildly comfortable	Somewhat comfortable	Quite comfortable	Extremely comfortable
If there were no psychiatrist in your institution, how comfortable would you feel performing the assessment and prescribing initial treatment for psychiatric patients by yourself?	373	2.00 ( $\pm$ 0.9), 2	38.9 (145)	35.7 (133)	19.3 (72)	5.9 (22)	0.3 (1)
Based on the amount of supervision and consultant support at your institution, how comfortable are you performing the assessment and prescribing initial treatment for psychiatric patients? <sup>‡</sup>	312	2.71 ( $\pm$ 1.0), 2	24.4 (76)	35.9 (112)	26.9 (84)	11.9 (37)	1.0 (3)
How comfortable do the majority of your ED attendings appear when assessing and treating psychiatric patients? <sup>‡</sup>	306	3.71 ( $\pm$ 0.9), 3	4.9 (15)	24.8 (76)	34.6 (106)	30.1 (92)	5.6 (17)
Importance of developing competence to assess and treat psychiatric patients			Not at all important	Mildly important	Somewhat important	Quite important	Extremely important
How important is it for you to develop competency in the assessment and treatment of patients with psychiatric complaints?	385	3.69 ( $\pm$ 0.9), 4	1.6 (6)	7.8 (30)	21.6 (83)	43.9 (169)	25.2 (97)
Preparedness assessing and treating psychiatric patients			Not at all prepared	Mildly prepared	Somewhat prepared	Quite prepared	Extremely prepared
How prepared are you to assess and treat patients presenting with psychiatric chief complaints?	385	2.61 ( $\pm$ 0.9), 3	9.6 (37)	32.7 (126)	43.9 (169)	12.7 (49)	1.0 (4)

Numbers may not sum to 396 because of missing data. Percentages are calculated using all available data.

SUD = substance use disorders.

<sup>†</sup>Data are reported as % (No.).

<sup>‡</sup>PGY-1s did not answer these items

who had completed a psychiatry rotation during residency reported, on average, significantly lower levels of confidence (mean  $\pm$  SD =  $41.40 \pm 18.87$ ) compared to residents who had completed a psychiatry rotation (mean  $\pm$  SD =  $63.53 \pm 15.63$ ;  $t(291) = -7.83$ ,  $p < 0.001$ ). Specific survey items and response data related to resident attitudes about their training in emergency psychiatry are presented in Table 3.

### Attitudes Toward Patients With Mental Illness

Emergency medicine residents in this study had significantly higher (more stigmatizing) scores on the Attitude Subscale of the Opening Minds Scale for Health Care Providers (mean  $\pm$  SD =  $13.42 \pm 3.96$ ) compared to the sample of physicians included in the development of the measure (mean  $\pm$  SD =  $12.7 \pm 3.4$ ;  $t(1050) = 3.11$ ,  $p = 0.002$ ); however, the effect size was small ( $d = 0.195$ ). As expected, the majority of residents (84%) reported that they are more comfortable treating patients with a physical illness than a mental illness. Nonetheless, most respondents reported feeling compassionate toward patients with mental illness and believed that they could help a patient with mental illness (see Table 4).

## DISCUSSION

We conducted a needs assessment using a sample of 15 EM residency programs across the U.S. to gather information about EM resident exposure to education regarding psychobehavioral conditions, perceived confidence treating patients with such conditions, and perceived educational needs with regard to psychobehavioral conditions. Overall, we found that the focus on psychobehavioral conditions in EM residency education is relatively minimal compared to the frequency with which these conditions are encountered clinically. Consequently, most EM residents feel uncomfortable and lack confidence independently managing patients with psychobehavioral conditions and would like more training in this area.

Our survey results also demonstrated that opportunities for ongoing dedicated educational experiences managing patients with psychobehavioral conditions are rather limited, possibly in part due to a greater emphasis placed on teaching the management of medical and surgical conditions. The ACGME EM Program Requirements do not specifically mention that residents should be able to competently evaluate and

provide initial treatment for patients with psychobehavioral conditions.<sup>6</sup> Instead, these requirements place a heavy emphasis on the importance of residents developing medical and surgical procedural competencies.<sup>6</sup> To meet these requirements, most programs require residents to rotate through settings designed to develop their medical and surgical procedural skills including obstetrics and gynecology, medical, surgical and pediatric intensive care units, trauma and acute care surgery, and anesthesiology. Nearly all (>99%) of respondents completed a psychiatry rotation during medical school, but only 12% reported rotating through a psychiatry rotation in residency. Not surprisingly, the vast majority (84%) of respondents agreed or strongly agreed that they feel more comfortable managing a patient with a physical illness than a MH illness. The relatively limited opportunity for ongoing dedicated educational experiences managing patients with psychobehavioral conditions may be one reason why EM residents have less confidence in this area when compared to managing patients with medical or surgical conditions.

We found that most residents would prefer more education on managing psychobehavioral conditions through various methods including lectures, simulation, and hands on training in the ED. Review of the EM literature demonstrated a paucity of data surrounding the best methods for teaching psychobehavioral emergencies. The use of high-fidelity simulation followed by structured group debriefs has previously been well received by psychiatry residents and nurses learning how to manage psychobehavioral conditions.<sup>17-19</sup> Teaching strategies using standardized patients and role-playing have also been found to increase learner comfort with assessing and managing psychobehavioral conditions.<sup>20,21</sup> In medical students, case-based independent study was found to be an effective method to improve exposure to emergency psychiatry cases and could be adapted for resident use.<sup>20,22</sup> Dedicated off-service emergency psychiatry rotations could also be developed to address educational needs.<sup>23</sup> MacLean et al.<sup>24</sup> previously described learning objectives for an off-service psychiatry rotation for EM residents, and the American Association for Emergency Psychiatry has published guidelines and a model curriculum with specific training objectives for psychiatry residents to learn emergency psychiatry that could be adapted for the EM resident.<sup>24,25</sup> Zun<sup>26</sup> previously advocated for developing a fellowship in emergency psychiatry as well as a course for psychiatric

**Table 3**  
Resident Confidence in Psychiatric Skills

Item	Total (n = 381)	PGY-1 (n = 131)	PGY-2 (n = 117)	PGY-3 (n = 89)	PGY-≥4 (n = 44)	Psych Rotation (n = 38)	No Psych Rotation (n = 261)
Conduct an H&P to determine whether psychiatric complaints are due to an organic or psychological cause	65.57 (±23.0)	56.10 (±25.3)	67.57 (±21.4)	73.88 (±16.6)	71.61 (±22.0)	51.45 (±25.96)	71.04 (±18.22)
Perform a thorough mental status exam (behavioral, cognitive, and emotional functioning)	52.27 (±24.6)	49.62 (±23.7)	54.88 (±21.9)	55.88 (±26.8)	45.53 (±27.8)	26.49 (±26.68)	57.77 (±21.48)
Differentiate a psychiatric presentation from delirium	63.17 (±20.8)	53.40 (±23.3)	65.47 (±17.5)	70.93 (±16.0)	70.20 (±19.2)	53.34 (±18.97)	68.00 (±17.51)
Conduct a psychiatric interview to diagnose common psychiatric disorders (anxiety, depression, substance abuse)	60.41 (±25.6)	57.82 (±24.5)	62.55 (±24.6)	64.64 (±25.2)	53.68 (±30.5)	29.11 (±28.64)	65.83 (±21.20)
Determine whether a patient is at risk of harm to self or others	69.97 (±21.1)	62.70 (±23.6)	72.86 (±19.2)	75.61 (±17.4)	72.61 (±19.3)	60.05 (±19.90)	74.48 (±17.49)
Develop a safety plan with a suicidal patient	46.21 (±25.2)	44.32 (±22.9)	47.56 (±26.3)	48.34 (±26.1)	43.63 (±27.0)	23.22 (±26.68)	51.48 (±23.40)
Counsel a suicidal patient about reducing access to firearms and other lethal means	51.11 (±26.8)	47.06 (±24.7)	54.75 (±27.5)	53.84 (±27.1)	47.19 (±28.9)	23.24 (±24.89)	56.76 (±23.60)
Initiate treatment for patients presenting with drug overdose	67.24 (±25.0)	55.45 (±24.4)	71.57 (±22.7)	75.31 (±22.9)	74.52 (±24.6)	56.61 (±23.60)	73.62 (±21.32)
Initiate treatment for patients presenting with acute alcohol or drug withdrawal	72.59 (±22.1)	60.31 (±23.4)	77.47 (±18.1)	79.62 (±20.0)	81.7 (±16.2)	74.76 (±18.10)	76.75 (±19.03)
Use physical or chemical restraints to facilitate work-up in a trauma patient who has a psychiatric condition	66.90 (±24.8)	49.27 (±23.5)	74.12 (±19.2)	78.28 (±20.4)	74.80 (±23.1)	52.47 (±29.05)	73.93 (±20.38)
Perform the assessment and prescribe initial treatment of the acutely agitated or psychotic patient	63.02 (±24.2)	47.88 (±23.1)	68.50 (±21.1)	72.18 (±19.7)	74.59 (±21.7)	66.97 (±20.57)	67.73 (±22.46)
Perform the assessment and prescribe initial treatment of pregnant women with psychiatric complaints	40.78 (±24.5)	32.77 (±20.9)	44.79 (±25.3)	44.96 (±24.6)	44.41 (±26.7)	21.45 (±20.14)	46.41 (±24.02)
Perform the assessment and prescribe initial treatment of elderly patients with psychiatric complaints	49.69 (±24.7)	42.27 (±20.5)	52.98 (±25.1)	54.27 (±25.0)	52.61 (±29.3)	26.55 (±25.54)	55.90 (±22.65)
Perform the assessment and prescribe initial treatment of children with psychiatric complaints	42.55 (±24.7)	33.96 (±20.3)	45.49 (±25.0)	47.48 (±24.9)	49.33 (±28.9)	23.89 (±22.47)	48.29 (±24.13)

Data are reported as mean (±SD). *Psych Rotation* refers to residents who completed a psychiatry rotation during and EM or non-EM residency. *No psych rotation* refers to those residents who did not complete a psychiatry rotation during residency. H&P = history and physical examination.

**Table 4**  
Resident Attitudes Toward People With Mental Illness

Item	Total N	Mean ( $\pm$ SD), Median	1 = Strongly Disagree	2 = Disagree	3 = Neutral	4 = Agree	5 = Strongly Agree
I am more comfortable helping a person who has a physical illness than I am helping a person who has a mental illness	386	4.13 ( $\pm$ 0.12), 4.00	0.5 (2)	4.4 (17)	11.1 (43)	49.0 (189)	35.0 (135)
Despite my professional beliefs, I have negative reactions toward people who have mental illness	387	2.21 ( $\pm$ 0.97), 2.00	25.6 (99)	39.8 (154)	23.8 (92)	9.8 (38)	1.0 (4)
There is little I can do to help people with mental illness	387	2.20 ( $\pm$ 0.92), 2.00	21.2 (82)	49.6 (192)	19.4 (75)	8.0 (31)	1.8 (7)
More than half of people with mental illness don't try hard enough to get better	386	1.88 ( $\pm$ 0.83), 2.00	35.8 (138)	44.3 (171)	16.6 (64)	2.6 (10)	0.8 (3)
Healthcare providers do not need to be advocates for people with mental illness	387	1.52 ( $\pm$ 0.79), 1.00	59.9 (232)	33.1 (128)	3.9 (15)	1.3 (5)	1.8 (7)
I struggle to feel compassion for a person with mental illness	387	1.93 ( $\pm$ 0.92), 2.00	36.2 (140)	43.2 (167)	12.7 (49)	7.2 (28)	0.8 (3)
There is little to nothing I can do for a patient with suicidal ideation <sup>†</sup>	375	1.75 ( $\pm$ 0.85), 2.00	45.3 (170)	40.5 (152)	8.8 (33)	4.8 (18)	0.5 (2)
Total score	386	13.42 ( $\pm$ 3.96), 14.00					

The items were taken from the Attitudes of Health Care Providers Towards People with Mental Illness Subscale of the Opening Minds Scale for Health Care Providers Scale OMS-HC.

<sup>†</sup>Data are reported as % (No.).

<sup>‡</sup>This item is not include in the original scale. It was added for the purpose of this study.

emergencies similar to advanced trauma life support, advanced cardiovascular life support, and pediatric advanced life support. Increased bedside teaching of concepts related to evaluation and management of psychobehavioral conditions by experienced EM attendings may also promote hands-on resident learning on shift.

Opportunities to learn through hands-on training in the ED are available given that all residents reported typically seeing at least one or more patients with a psychobehavioral chief complaint during a typical 8-hour shift. Nonetheless, over half of respondents described minimal involvement with psychiatric patients. Residents reported that only about one-third of their attendings are quite or extremely comfortable treating these patients, which may explain why educational discussions on these patients are not occurring. Furthermore, EM residents may be taking a less active role in the care of patients with psychobehavioral conditions due to the availability of consultants, social workers, or other support services. While this may not pose a problem in an academic medical setting with access to psychiatric consultants and other support services, many residents will go on to practice in rural or lower-resource settings that offer

limited additional support. Thus, ensuring that residents feel confident independently managing these patients through increased autonomy and involvement in their care is critical to successful practice in diverse environments upon completion of training.

The results of our study showed that residents who completed a psychiatry rotation actually reported lower levels of confidence performing various clinical skills related to psychobehavioral conditions compared to residents who had not completed a psychiatry rotation. These results were initially surprising; however, it may be that after completing a psychiatry rotation residents had developed a greater understanding of the skills and amount of training needed to effectively treat patients with psychobehavioral conditions. It is also important to note that physician self-reported levels of confidence are often poor predictors of observational measures of performance. In fact, several studies have found that that physicians who are the least skilled are often the most confident and least likely to recognize their learning needs.<sup>27</sup> Given these implications, future studies examining residents' training needs would benefit from including objective measures of performance in addition to self-assessments.

Our study identified several specific content areas within the broader category of psychobehavioral conditions in which residents wish to improve their knowledge and skills: 1) management in special populations such as pregnant women, children, and elderly patients; 2) lethal means counseling; and 3) safety planning with suicidal patients. Correspondingly, residents expressed the lowest confidence levels with performing the assessment and prescribing initial treatment of special populations. Simulation exercises and didactic content could target these identified areas of need. The current lack of a clear “best practices” guide for training methods may account for the survey results demonstrating that the amount of time spent and the method of training in psychobehavioral conditions are widely variable among residency programs. Residency programs would benefit from an improved and more standardized curriculum.

## LIMITATIONS

There are several limitations of this study. First, participants were a convenience sample of EM residents from 15 ACGME-accredited EM programs that were identified based on existing professional relationships. We did not systematically include residents from programs with varying levels of emergency psychiatric services (e.g., availability of consultants, dedicated psychiatric emergency unit) and resident educational offerings (e.g., required psychiatry rotation). However, by limiting participating programs to those with known colleagues available to help with data collection, we were able to maximize our response rate, which was rather high especially for a survey study. Additionally, we were able to include residents completing programs in all areas of the United States. A second concern involved the timing of data collection (July and August). Early in the academic year, the PGY-1s would have little experience and contact with patients. While patients presenting with psychobehavioral complaints and subsequent boarding of these patients are increasing, patient presentations can vary seasonally. Changing the survey timing may alter responses for both junior and senior residents.

## CONCLUSIONS

In summary, residents in emergency medicine desire more training in the management of psychobehavioral conditions. There are wide variations among residency programs with regard to the time spent in training

and the methods of training for psychobehavioral conditions. Residency programs would benefit from more education in psychobehavioral conditions and a clear set of best practices for improved curriculum standardization.

We thank Alicia Pilarski, MD, Nicholas Schwartz, MD, Robert Cooney, MD, Andy Grock, MD, Kyle Ragins, MD, Steven Hochman, MD, and Loice Swisher, MD for their help with data collection.

## References

1. Center for Behavioral Health Statistics and Quality. Behavioral health trends in the United States: results from the 2014 national survey on drug use and health. HHS Publication No. SMA 15-4927, NSDUH Series H-50. 2015. at <https://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf>. Accessed October 12, 2018.
2. Owens PL, Mutter R, Stocks C. Mental Health and Substance Abuse-related Emergency Department Visits Among Adults, 2007. Healthcare Cost and Utilization Project (HCUP) Statistical Brief #92. 2010. Available at: <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb92.pdf>. Accessed November 1, 2018.
3. Moore BJ, Stocks C, Owens PL. Trends in Emergency Department Visits, 2006–2014. Healthcare Cost and Utilization Project (HCUP) Statistical Brief #227. 2017. Available at: [www.hcup-us.ahrq.gov/reports/statbriefs/sb227-Emergency-Department-Visit-Trends.pdf](http://www.hcup-us.ahrq.gov/reports/statbriefs/sb227-Emergency-Department-Visit-Trends.pdf). Accessed November 2, 2018.
4. Nicks BA, Manthey DM. The impact of psychiatric patient boarding in emergency departments. *Emerg Med Int* 2012;2012:1–5.
5. American College of Emergency Physicians. Definition of Emergency Medicine. Policy Compendium. 2015. Available at: <http://www.acep.org/workarea/DownloadAsset.aspx?xml:id=101698>. Accessed November 10, 2018.
6. Accreditation Council for Graduate Medical Education (ACGME). ACGME Common Program Requirements (Residency). 2018. Available at: <https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRResidency2019.pdf>. Accessed November 6, 2018.
7. Counsellman FL, Babu K, Edens MA, et al. The 2016 Model of the Clinical Practice of Emergency Medicine. *J Emerg Med* 2017;52:846–9.
8. Santucci KA, Santher J, Baker DM. Emergency medicine training programs’ educational requirements in the management of psychiatric emergencies. *Pediatr Emerg Care* 2003;19:154–6.
9. American Board of Emergency Medicine (ABEM). Qualifying Exam Content Specifications. Available at: <https://www.abem.org/Qualifying-Exam-Content-Specifications>.






- www.abem.org/public/become-certified/qualifying-exam/exam-content. Accessed November 8, 2018.
10. Tintinalli JE, Stapczynski JS, Ma OJ, Yealy DM, Meckler GD, Cline DM. *Tintinalli's Emergency Medicine: A Comprehensive Study Guide*, 8th ed. New York: McGraw-Hill, 2016.
  11. Walls R, Hockberger R, Gausche-Hill M. *Rosen's Emergency Medicine: Concepts and Clinical Practice*, 8th ed. Philadelphia: Elsevier/Saunders, 2014.
  12. Weissberg M. The meagerness of physicians' training in emergency psychiatric intervention. *Acad Med* 1990;65:747–50.
  13. Cook DA, Blachman MJ, Price DW, West CP, Berger RA, Wittich CM. Professional development perceptions and practices among U.S. physicians. *Acad Med* 2017;92:1335–45.
  14. American College of Emergency Physicians. The emergency physician's role in behavioral emergencies. *Ann Emerg Med* 1984;13:972–3.
  15. Bandura A. Guide for constructing self-efficacy scales. in: Pajares F, Urdan T, eds. *Self-Efficacy Beliefs of Adolescents*. Greenwich, CT: Information Age Publishing, 2006:307–38.
  16. Kassam A, Papish A, Modgill G, Patten S. The development and psychometric properties of a new scale to measure mental illness related stigma by health care providers: the Opening Minds Scale for Health Care Providers (OMS-HC). *BMC Psychiatry* 2012;12:1–12.
  17. Masters J, Kane M, Pike M. The suitcase simulation: an effective and inexpensive psychiatric nursing teaching activity. *J Psychosoc Nurs Ment Health Serv* 2014;52:39–44.
  18. Thomson AB, Cross S, Key S, Jaye P, Iversen AC. How we developed an emergency psychiatry training course for new residents using principles of high-fidelity simulation. *Med Teach* 2013;35:797–800.
  19. Murray BA. The use of high-fidelity simulation in psychiatric and mental health nursing clinical education. *Int J Health. Sci Educ* 2014;2;:Art.3.
  20. Lofchy J, Boyles P, Delwo J. Emergency psychiatry: clinical and training approaches. *Can J Psychiatry* 2015;60:1–7.
  21. Lofchy J. The use of standardized patients in the teaching of emergency psychiatry. *Emerg Psychiatry* 1997;3: 78–79.
  22. Hirshbein LD, Gay T. Case-based independent study for medical students in emergency psychiatry. *Acad Psychiatry* 2005;29:96–9.
  23. Bode A, Jackson JS. The current emergency medicine residency curriculum: missing psychiatry. *Am J Emerg Med* 2017;35:1771–2.
  24. MacLean TA, Bourgeois JA, Hamilton GC, Kay J. Objectives to direct the training of emergency medicine residents on off-service rotations: psychiatry. *J Emerg Med* 1995;13:545–51.
  25. Brasch J, Glick RL, Cobb TG, Richmond J. Residency training in emergency psychiatry: a model curriculum developed by the education committee of the American Association of Emergency Psychiatry. *Acad Psychiatry* 2004;28:95–103.
  26. Zun L. Care of psychiatric patients: the challenge to emergency physicians. *West J Emerg Med* 2016;17:174–6.
  27. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;296:1094–102.

### Supporting Information

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10377/full>

**Data Supplement S1.** Resident program characteristics.

# Systematic Online Academic Resource (SOAR) Review: Renal and Genitourinary

Andrew Grock, MD<sup>1</sup>, Anuja Bhalerao, MD (Candidate)<sup>2</sup>,  
Teresa M. Chan, MD, MHPE<sup>3</sup> , Brent Thoma, MD, MA, MSc<sup>4</sup> ,  
Annie B. Wescott, MLIS<sup>5</sup>, and N. Seth Trueger, MD, MPH<sup>6</sup> 

## ABSTRACT

**Background:** Online resources for emergency medicine (EM) trainees and physicians have variable quality and inconsistent coverage of core topics. In this first entry of the Society for Academic Emergency Medicine Systematic Online Academic Resource (SOAR) series, we describe the application of a systematic methodology to comprehensively identify, collate, and curate online content for topic-specific modules.

**Methods:** A list of module topics and related terms was generated from the American Board of Emergency Medicine's Model of the Clinical Practice of Emergency Medicine. The authors selected "renal and genitourinary" for the first module, which contained 35 terms; all MeSH headers and colloquial synonyms related to the topic and related terms were searched both within the 100 most impactful online educational websites per the Social Media Index and the FOAMsearch.net search engine. Duplicate entries, journal articles, images, and archives were excluded. The quality of each article was rated using the revised METRIQ (rMETRIQ) score.

**Results:** The search yielded 13,058 online resources. After 12,717 items were excluded, 341 underwent quality assessment. All renal/genitourinary topics were covered by at least one resource. The median rMETRIQ score was 11 of 21 (interquartile range = 8–14). Calculus of urinary tract was most prominently featured with 60 posts. Thirty-four posts (10% of full-text screened FOAM articles) covering 12 core topics were identified as high quality (rMETRIQ ≥ 16).

**Conclusions:** We demonstrated the feasibility of systematically identifying and curating FOAM resources for a specific EM topic and identified an overrepresentation of some subtopics. This curated list of resources may guide trainees, teacher recommendations, and resource producers. Further entries in the series will address other topics relevant to EM.

The well-documented growth and use, both formally (in residency curricula<sup>1,2</sup>) and informally,<sup>3–5</sup> of free online medical education resources (FOAM) in emergency medicine (EM) have been remarkable.<sup>3,4,6,7</sup> However, resources are scattered across an enormous number of sites, of variable and difficult to evaluate

quality,<sup>8–10</sup> and may not be at the appropriate level for all learners.<sup>11</sup> Additionally, various topics are covered extensively while others receive no or scant coverage<sup>12</sup> with no synthesis describing specific topics coverage, extent, level of intended learner, and quality. As individuals, educators, and program leaders look to

From the <sup>1</sup>Division of Emergency Medicine, Greater Los Angeles VA Healthcare System, and the Department of Emergency Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA; the <sup>2</sup>School of Medicine, University of Toronto, Toronto, Ontario, Canada; the <sup>3</sup>Department of Medicine, Division of Emergency Medicine, McMaster University, and the McMaster Education Research, Innovation, and Theory Program, Hamilton, Ontario, Canada; the <sup>4</sup>Department of Emergency Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; and <sup>5</sup>Galter Health Sciences Library & Learning Center, Feinberg School of Medicine, and the <sup>6</sup>Department of Emergency Medicine, Northwestern University, Chicago, IL.

Received January 3, 2019; revision received March 1, 2019; accepted March 6, 2019.

The authors have no relevant financial information.

AB reports salary support from the Society for Academic Emergency Medicine (SAEM) for her work on this project. TMC holds grants from the Physician Services Incorporated Foundation for work on social media knowledge translation. NST receives salary support from the American Medical Association for his role as Digital Media Editor, *JAMA Network Open*.

Related articles appear on pages 396 and 398.

Supervising Editor: Esther H. Chen, MD.

Address for correspondence and reprints: Teresa M. Chan, MD, MHPE; e-mail: teresa.chan@medportal.ca.

AEM EDUCATION AND TRAINING 2019;3:375–386

incorporate FOAM into both extra- and co-curricular activities, we see an opportunity for a centralized resource for systematically identified and quality-assessed content. This paper aims to create a new type of synthesis scholarship that borrows from the systematic review format to aggregate and filter high-quality, online academic resources for the purposes of assisting educators in finding useful content within certain topics. Please see Box 1 for details about this initiative.

## METHODS

### Study Design

The design of this review of educational content was conducted similar to that of a traditional systematic literature review. We attempted to adhere to the PRISMA guidelines as closely as possible.<sup>13</sup>

### Topic Identification

Appropriate subtopic searches was based on the 2016 American Board of Emergency Medicine's Model of the Clinical Practice of Emergency Medicine (MCPEM) document as it "represents essential information and skills necessary for the clinical practice of EM by board-certified emergency physicians."<sup>13</sup> An

important component of the document includes a listing of common conditions, symptoms, and disease presentations often seen in the emergency department under each system heading.<sup>14</sup>

The "renal and genitourinary" system search used the MCPEM renal and genitourinary topic's six headings and over 20 subheadings to create a list of subtopics for the database search. In addition, the authors also reviewed signs, symptoms, and presentations of the MCPEM for other relevant subtopics. Two medical librarians assisted our teams in translating the subtopic list into MeSH headers and colloquial synonyms.

### Database Search

We searched using FOAMsearch.net<sup>15</sup> and top 100 FOAM websites per the Social Media Index (SMI)<sup>16,17</sup> as the two search repositories. We selected these two methods to identify online educational resources targeted to health professionals; we were concerned that a nonrestricted search of a generic search engine (e.g., Google) would result in an excess of irrelevant and noneducational content aimed at the lay public. As this was the first attempt at this type of review, we restricted our search to the top 100 sites on the SMI because of resource limitations and the availability of data from the originators of the tool. Additionally, the sites beyond 100 in the SMI have too minimal a presence online to adequately correlate with quality. The results of the initial search were extracted with their heading, source type (journal, blog, podcast, archive), and URL.

**Inclusion Criteria.** All open educational resources on renal and genitourinary topics were included, as determined by matching with a topic listed on the MCPEM list. Of note, items related to female genitoreproductive topics were not included, since they are classified by MCPEM under gynecology. Of note, podcast show notes were specifically included in this review.

**Exclusion Criteria.** Items were excluded if they were deemed not relevant to the renal genitourinary topic (e.g., a toxicology post that mentions renal injury as a sequela), irrelevant source types (including posts without text to review, such as pure audio or video links without associated written notes or "show notes"), articles in peer-reviewed journals, or reblogging repositories of posts published initially elsewhere.

#### Box 1

##### The Objective of the SAEM SOAR Topic Review Series

This is the first entry in the SAEM Systematic Online Academic Resource (SOAR) Topic Review series. We aim to systematically identify online resources by topic to determine FOAM's current coverage of EM topics and subtopics, assess the quality of these resources with a validated tool, and collate links to identified high-quality online resources in a user-friendly framework. Thus, the SAEM SOAR Topic Reviews will enable all users (individual learners, educators, and program leaders) to easily find relevant, high-quality online resources. Additionally, this ongoing series will help create a durable scientific evaluation of FOAM in EM.

This initial entry in the SAEM SOAR Topic Review series will identify and evaluate online education resources in the Renal and Genitourinary systems. Future editions will review other topics, and this SAEM SOAR Topic Review is featured as an article-type in *Academic Emergency Medicine: Education & Training*, open to any and all authors. Additionally, this article and its methodology can serve as a guide to future authors for future topics (and, eventually, updated reviews of prior topics) to provide a comprehensive description of the FOAM landscape.

If you are interested in submitting a future SAEM SOAR Topic Review, please contact SAEM at [sroseen@saem.org](mailto:sroseen@saem.org)

## Data Extraction and Quality and Usage Assessment

Google Forms (Mountainview, CA) was used to organize author abstraction tool for the final list of resources and included publication date, word count, accompanying media, described and inferred audience level, subtopic reviewed, author information, and quality assessment via the revised METRIQ (rMETRIQ)

score.<sup>18</sup> The rMETRIQ score contains seven quality-related scales, each on a scale of 0 to 3, for a maximal score of 21 (Table 1). Each source was extracted and quality assessed by one study author, who also determined the appropriate audience level (as indicated by the post or as determined by the rater if not indicated) and appropriate usage (e.g., journal club, postshift reading, or appropriate for on-shift ‘just-in-time’

**Table 1**  
rMETRIQ Score Questions

Questions	Options
Q1: Does the resource provide enough background information to situate the user?	3—Yes, the resource provides sufficient background information to situate the user and also directs users to other valuable resources related to the topic. 2—Yes, the resource provides sufficient background information to situate the user. 1—No, the information presented within the resource cannot be situated within its broader context, but users are directed to resources with this information. 0—No, the information presented within the resource cannot be situated within its broader context without looking up information independently.
Q2: Does the resource contain an appropriate amount of information for its length?	3—No unnecessary, redundant or missing content, all content was essential. 2—Some unnecessary, redundant, or missing content, but most content was essential. 1—Lots of unnecessary redundant or missing content. 0—Insufficient content.
Q3: Is the resource well written and formatted?	3—The resource is very well written and formatted in a way that optimized and benefits learning. 2—The resource is reasonably well written and formatted, but aspects of the organization or presentation are distracting or otherwise detrimental to learning. 1—The resource is somewhat well written and formatted, but could benefit from substantive editing (e.g., grammatical errors are seen or better organized). 0—The resource is poorly written and/or formatted and should not be a resource for learning.
Q4: Does the resource cite its references?	3—Yes, the references are cited, clearly map to specific statements within the resource, and all statements of fact that are not common knowledge are supported with a reference. 2—Yes, the references are cited and clearly map to specific statements within the resource, but statements of fact that are not common knowledge are made without the support of a reference. 1—Yes, there are references listed but they do not map to specific statements within the resource. 0—No, no references are cited.
Q5: Is it clear who created the resource and do they have any conflicts of interest?	3—Yes, the identity and qualifications of the author are clear and they specify that they have no relevant conflicts of interest. 2—Yes, the identity and qualifications of the author are clear, but they do not disclose whether they have any conflicts of interest. 1—Yes, the identity of the author is clear, but they do not list their qualifications or disclose whether they have any conflicts of interest. 0—No, the author of the resource has significant conflicts of interest or is not clearly identified (e.g., no name or a pseudonym is used).
Q6: Are the editorial and prepublication peer review processes that were used to create the resource clearly outlined?	3—Yes, a clear review process is described on the website and it was clearly applied to the resource. 2—Yes, a clear review process is described on the website, but it was not clear whether it was applied to the resource. 1—Yes, a review process is mentioned on the website, but it was not clearly described. 0—No, it is unclear whether or not the website has a review process or there is no process.
Q7: Is there evidence of postpublication commentary on the resource's content by its users?	3—Yes, a robust discussion of the resource's content has occurred that expands upon the content of the resource. 2—Yes, some comments have been made on the resource, but a robust discussion about the resource's content has not occurred. 1—There was a mechanism to leave comments but none had been made. 0—No, there was no mechanism to leave comments or comments that were present were either unrelated to the post or unprofessional.

reference). We initially used a small set of papers to calibrate our team for judging the audience and the best usage for individual posts. Disagreements were resolved via consensus. We used a modified Angoff method to determine a threshold for high quality.

RESULTS

The FOAMsearch and SMI search yielded 13,058 results. Figure 1 displays the search and review results. After excluding for duplicate entries (1,215), nonrelevant titles (3,373), non-English sources, journals, and non-FOAM (i.e., not free) sites, nonfunctioning links, tweets, or republished resources (8,097), 330 resources met the inclusion criteria, with the remaining 43 flagged for secondary review. Of these 43 resources, 11 met inclusion criteria on secondary review for a total of 341 resources assessed for quality. The high-quality threshold was determined by author consensus via modified Angoff method<sup>19</sup> at 16 yielding 34 posts with a score ≥ 16. The high-quality posts can be found at the SAEM SOAR website at <https://www.saem.org/education/saem-online-academic-resources>.

Topic Coverage

Of the 35 renal/genitourinary topics, all were covered at least once; high-quality posts covered 12 of the topics (plus one additional post on Foley catheter problems and one additional “top nephrology topics,” which covered a range of topics). There is an uneven distribution of FOAM posts within the renal topic,

with 17.6% of posts focused on renal calculi. Among the 34 high-quality posts, 41% are on renal calculi followed by urinary retention (8.8%) and cystitis (8.8%). Although certain topics are covered extensively online, only a small number of these resources met the high-quality threshold. Renal calculi was most extensively covered online (60 of 341, 18%) but only 23.3% of these posts met the high-quality threshold. Resources that were relevant but not on the topic list made up 13% of the posts but only 4.4% of these posts were high quality. A total of 7.6% of the posts were on acute and chronic renal failure but only 3.85% were high quality (Table 2).

Types of Posts

For each of the 341 articles reviewed, each reviewer determined appropriate uses yielding a total of 771 assigned uses. Topics were determined to be appropriate for personalized reading (87.0%), postshift reading assignment (61.5%), flipped classroom (42.4%), individualized interactive instruction credit (27.9%), adjunct to journal club (13.9%), and on shift review of best evidence on a topic or just-in-time resources (0.3%; Figure 2).

Quality Assessment and Trainee Level of Recommendation.

**Quality Assessment.** The rMETRIQ scores ranged from 0 to 20 (mean ± SD = 10.8 ± 3.9). It approximated a normal distribution with a mode of 14. Table 3 lists the rank order of the top posts within our review. Overall, 34 (10%) posts

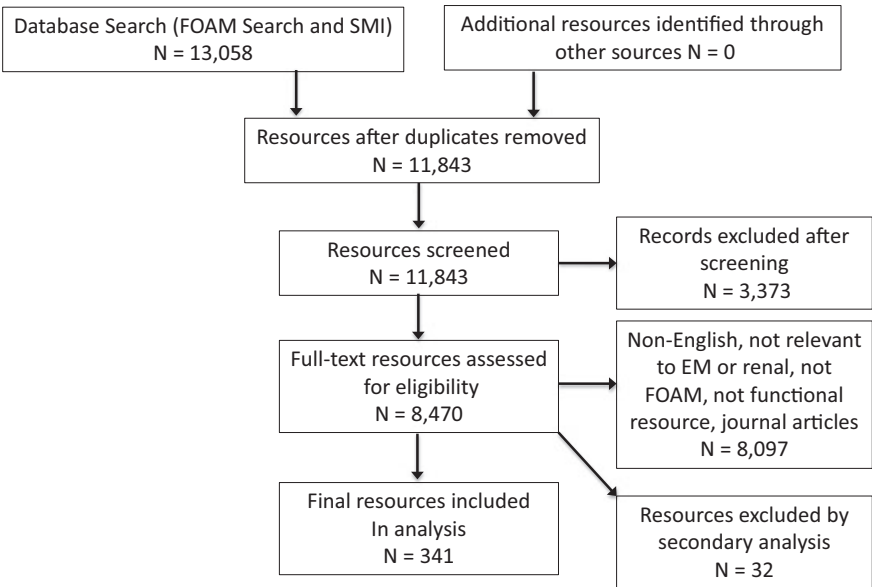


Figure 1. Search and review results. FOAM = free online medical education resources; SMI = Social Media Index.



**Table 2**  
Renal and Genitourinary SOAR Review Subtopic Analysis

	All	For High-quality Posts
Calculus of urinary tract	60	14
Relevant to topic, but not on list	45	2
Acute and chronic renal failure	26	1
Cystitis	21	3
Torsion	21	1
Hematuria	16	1
Renal tumors	13	0
Urinary retention	13	3
Epididymitis/orchitis	12	1
Gangrene of the scrotum (Fournier's gangrene)	13	2
Paraphimosis/phimosis	12	0
Polycystic kidney disease	8	0
Pyelonephritis	8	2
Hemolytic uremic syndrome	9	3
Priapism	8	0
Dysuria	6	1
Prostatitis	6	0
Nephrotic syndrome	5	0
Prostatic hypertrophy (BPH)	4	0
Glomerulonephritis	4	0
Hernias	4	0
Genital lesion	4	1
Complications of renal dialysis	3	0
Balanitis/balanoposthitis	1	0
Obstructive uropathy	2	0
Urinary incontinence	2	0
Urogenital tumors	2	0
Anuria	1	0
Asymptomatic bacteriuria	1	0
Glomerular disorders	1	0
Testicular masses	1	0
Urethritis	1	0
Not relevant to the topic (e.g., a toxicology post that mentions the kidney)	8	0
<b>Total</b>	<b>341</b>	<b>35</b>

SOAR = Systematic Online Academic Resource.

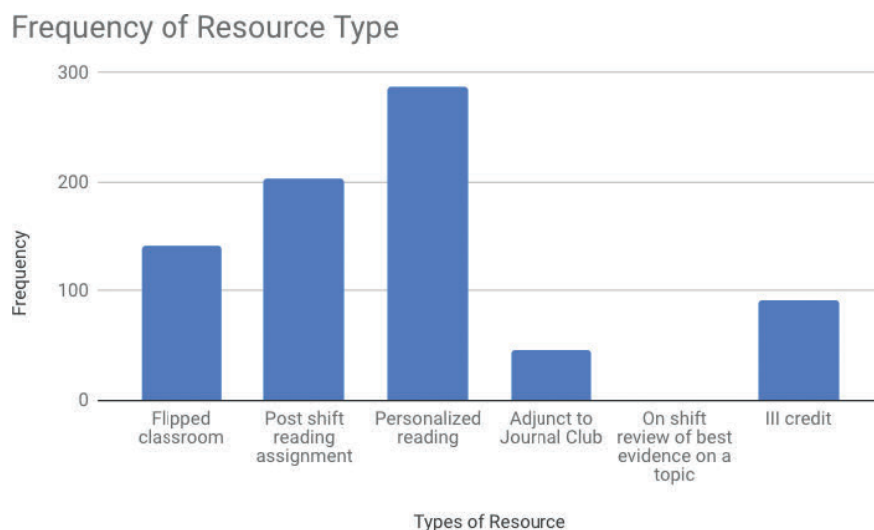
met our high-quality cutoff score of  $\geq 16$ . In evaluating rates of subtopic meeting our quality criteria, the topics with the highest numbers and rates of high-quality posts were renal colic (60 total; 14 [23%] high quality), hemolytic uremic syndrome (seven total; one [14%] high quality), genital lesions (four total; one [25%] high quality), and urinary retention (13 total; three [23%] high quality). Renal colic was by far the most frequently covered subtopic overall and among resources that met our quality criteria.

**Recommended Target Audience.** Of the 341 posts reviewed, each was assigned to be appropriate for at least one type of learner, ranging from preclerk to attending physician. Authors of the posts specified target audience levels for only 12 posts; the remaining were determined by reviewers. Each of these were deemed appropriate for multiple levels (mean = 3.2 levels). In order of frequency, the most common audience was junior resident (76.3%), intern (64.2%), senior resident (56.4%), clerk (i.e., third- or fourth-year medical student, 52.1%), attending (40%), and preclerk (i.e., first- or second-year medical student, 29.0%). The distribution among high-quality resources was similar: junior resident, 100%; senior resident, 94%; attending, 74%; intern, 56%, clerk, 29%; and preclerk, 12%. Among the high-quality posts, topics that had the highest number of posts relevant to at least one learner were calculus of urinary tract (renal colic; 41.2%), cystitis (8.8%), and urinary retention (8.8%). Figure 3 graphically depicts the distribution of covered topics, appropriate level, and rMETRIQ score for high-quality resources.

## DISCUSSION

This study marks the first detailed description of a comprehensive, systematic review of FOAM resources. We performed a topic-based review and curation of FOAM resources to identify high-quality resources for educators to integrate into their curricula. Our results demonstrate that it can be quite difficult to navigate and curate resources on a particular topic. The challenges of searching and reviewing online literature are not new.<sup>20</sup> Prior studies looking at online patient information resources have demonstrated high variability in quality.<sup>21–26</sup> Since both educators and trainees find it difficult to reliably rate online resources,<sup>9,10</sup> it is likely that trainees and clinicians have difficulty identifying high-quality resources for their own learning, which frequently occurs at the point of care.<sup>5</sup>

Other initiatives, such as the Academic Life in Emergency Medicine (ALiEM) Approved Instructional Resources (AIR) series,<sup>27,28</sup> collect and curate topic specific resources, but utilize a much less rigorous and comprehensive search strategy than in this review. Further, the goal of the AIR series is to build modules sufficient for individual interactive instruction, not to comprehensively describe and assess the coverage of FOAM. This review extends the AIR series' pioneering search technique to the top 100 of the SMI instead of the top 50, as well as performing a broader



**Figure 2.** Frequency of resource types found in the renal/genitourinary SOAR review. SOAR = Systematic Online Academic Resource.

and more comprehensive FOAMsearch. Additionally, this review applies a newly refined scoring system (the rMETRIQ score). Finally, this review attempts to classify resources based on trainee level to identify appropriate resources for different levels of learner.

### Topic Coverage

While we found a plethora of resources within our selected topic, many were not of high quality. In fact, 90% of the identified resources did not reach the high-quality threshold set by our study group. The reasons for heterogeneity of quality as measured by the rMETRIQ score were quite variable. Some sites (e.g., Wikis) were limited in their ability to attribute the work to a single, well-documented author, raising the potential for author conflict of interest or insufficient knowledge or education skill. On other sites, it was difficult to discern whether an author had academic credentials and/or training, or conflicts of interest, since they used informal biographies. Although we performed this review primarily to identify resources for EM educators and learners, a high level of contribution exists from other specialties that are avidly contributing to the FOAM space. Radiopaedia,<sup>29,30</sup> notably, contributed over 18% of the search results that we reviewed.

The coverage of the topics that we identified for the search was quite heterogeneous. Every topic was covered by at least one post; some (particularly renal colic) were discussed extensively, while others were only covered once. This is consistent with the findings of the less granular review of all FOAM topics by Stuntz and Clontz,<sup>12</sup> which found markedly uneven coverage of topics, with those thought to be more interesting

discussed more frequently. While all topic areas were covered by at least one resource, it is notable that resources did not have to provide comprehensive coverage of a topic to be classified under it and many subtopics are incompletely covered. This suggests that many renal and genitourinary topics have yet to be covered in depth anywhere. Numerous resources were also created on topics that did not make our topic list, suggesting that, while superficial, the coverage of this area within FOAM is vast.

### Best Use of Resources

Very few resources were deemed by our team to be worthy of on-shift review, mostly because the resources were either too in depth (high-quality posts) or incomplete (low-quality posts), although online content on renal/genitourinary topics may have less on-shift utility than other topics. The literature describes trainees reporting that they do use the content from FOAM to review during shifts at the point of care.<sup>5</sup> Perhaps, in other topic areas (e.g., cardiology or critical care) there may be other FOAM resources that are of high quality and appropriate for such usage; our review largely yielded only a very small fraction of posts (0.3%) that we found would be appropriate for referencing on shift, as they were generally broader reviews of topics as opposed to references for clinical utility (e.g., clinical decision aides or medication dosage recommendations).

### LIMITATIONS

There were a number of limitations to our present protocol and study. First and foremost, we excluded

**Table 3**  
Rank Order of Posts in the Renal and Genitourinary SOAR Review by rMETRIQ Score

Subtopic	Name of First Author	Name of Blog Post	URL	Level of Trainee (Starred = Author; Unstarred = Team's Recommendation)	rMETRIQ score
Calculus of urinary tract (renal colic)	Helman, Anton	Journal Jam 3 – Ultrasound vs CT for Renal Colic View Larger Image	<a href="https://emergencymedicinencases.com/ultrasound-vs-ct-for-renal-colic/">https://emergencymedicinencases.com/ultrasound-vs-ct-for-renal-colic/</a>	Junior, senior, attending	20
Calculus of urinary tract (renal colic)	Chartier, Lucas	Episode 5: Renal Colic, Toxicology Update & Body Packers	<a href="https://emergencymedicinencases.com/episode-5-renal-colic-toxicology-update-body-packers/">https://emergencymedicinencases.com/episode-5-renal-colic-toxicology-update-body-packers/</a>	Clerk, intern, junior, senior, attending	19
Calculus of urinary tract (renal colic)	Swaminathan, Anand	Optimal First Line Analgesia in Ureteric Colic	<a href="https://coreem.net/journal-reviews/analgesia-ureteric-colic/">https://coreem.net/journal-reviews/analgesia-ureteric-colic/</a>	Intern, junior, senior, attending	19
Topic not on list (Foley catheter problems)	Schneider, Aaron	Foley Catheter Patients: Common ED Presentations / Management / Pearls & Pitfalls	<a href="http://www.emdocs.net/foley-catheter-patients-common-ed-presentations-management-pearls-pitfalls/">http://www.emdocs.net/foley-catheter-patients-common-ed-presentations-management-pearls-pitfalls/</a>	Intern, junior, senior	19
Calculus of urinary tract (renal colic)	Firestone, Daniel	Top 10 reasons NOT to order a CT scan for suspected renal colic	<a href="https://www.aliem.com/2014/04/top-10-reasons-order-ct-scan-suspected-renal-colic/">https://www.aliem.com/2014/04/top-10-reasons-order-ct-scan-suspected-renal-colic/</a>	Intern, junior, senior, attending	19
Cystitis	Sahsi, Rupinder	Diagnosing Pediatric Urinary Tract Infections	<a href="http://epmonthly.com/article/diagnosing-pediatric-urinary-tract-infections/">http://epmonthly.com/article/diagnosing-pediatric-urinary-tract-infections/</a>	Clerk, intern, junior, senior, attending	18
Hemolytic uremic syndrome	Swartz, Jordan	Hemolytic uremic syndrome	<a href="https://wikem.org/w/index.php?title=Hemolytic_uremic_syndrome&amp;redirect=no">https://wikem.org/w/index.php?title=Hemolytic_uremic_syndrome&amp;redirect=no</a>	Intern, junior, senior	18
Urinary retention	Benitez, Javier	First ALIEM journal article: Trial of void for acute urinary retention	<a href="https://www.aliem.com/2013/03/trial-of-void-acute-urinary-retention/">https://www.aliem.com/2013/03/trial-of-void-acute-urinary-retention/</a>	Clerk, intern, junior, senior, attending	18
Calculus of urinary tract (renal colic)	Shih, Jeffrey	Ultrasound For The Win: 46F with Right Abdominal and Flank Pain #US4TW	<a href="https://www.aliem.com/2015/02/ultrasound-win-46f-right-abdominal-flank-pain-us4tw/">https://www.aliem.com/2015/02/ultrasound-win-46f-right-abdominal-flank-pain-us4tw/</a>	Clerk, intern, junior, senior	18
Calculus of urinary tract (renal colic)	Swaminathan, Anand	Medical Expulsion Therapy with Tamsulosin in Ureteral Colic	<a href="http://www.emdocs.net/medical-expulsion-therapy-with-tamsulosin-in-ureteral-colic/">http://www.emdocs.net/medical-expulsion-therapy-with-tamsulosin-in-ureteral-colic/</a>	Junior, senior, attending	17
Calculus of urinary tract (renal colic)	Swaminathan, Anand	Medical Expulsion Therapy in Ureteral Colic: An Update	<a href="http://rebelem.com/medical-expulsion-therapy-in-ureteral-colic-an-update/">http://rebelem.com/medical-expulsion-therapy-in-ureteral-colic-an-update/</a>	Junior, senior, attending	17
Calculus of urinary tract (renal colic)	Swaminathan, Anand	Medical Expulsive Therapy (MET) in Renal Colic	<a href="https://coreem.net/journal-reviews/medical-expulsive-therapy-met-in-renal-colic/">https://coreem.net/journal-reviews/medical-expulsive-therapy-met-in-renal-colic/</a>	Junior, senior, attending	17
Calculus of urinary tract (renal colic)	Chitnis, Subhanir	Nephrolithiasis: Diagnosis and Management in the ED	<a href="http://www.emdocs.net/nephrolithiasis-diagnosis-management-ed/">http://www.emdocs.net/nephrolithiasis-diagnosis-management-ed/</a>	Junior, senior, attending	17
Pyelonephritis	Nickson, Chris	Urosepsis	<a href="https://lifeinthefastlane.com/ccu/urosepsis/">https://lifeinthefastlane.com/ccu/urosepsis/</a>	Clerk, intern, junior	17
Dysuria	Lin, Michelle	Paucis Verbis card: Urinary tract infection	<a href="https://www.aliem.com/2010/02/paucis-verbis-card-urinary-tract-infection/">https://www.aliem.com/2010/02/paucis-verbis-card-urinary-tract-infection/</a>	Intern, junior, senior	17

(Continued)

Table 3 (continued)

Subtopic	Name of First Author	Name of Blog Post	URL	Level of Trainee (Starred = Author; Unstarred = Team's Recommendation)	rMETRIQ score
Cystitis	Shenvi, Christina	Uncomplicated Urinary Tract Infection in Older Adults: Diagnosis and Treatment (Part 1)	<a href="https://www.alleim.com/2014/03/unc-omplicated-urinary-tract-infection-older-adults-diagnosis-treatment-1/">https://www.alleim.com/2014/03/unc-omplicated-urinary-tract-infection-older-adults-diagnosis-treatment-1/</a>	Preclerk, clerk, intern, junior, senior	17
Cystitis	Shenvi, Christina	Uncomplicated Urinary Tract Infections in Older Adults: Diagnosis and Treatment (Part 2)	<a href="https://www.alleim.com/2014/04/unc-omplicated-urinary-tract-infections-older-adults-diagnosis-treatment-part-2/">https://www.alleim.com/2014/04/unc-omplicated-urinary-tract-infections-older-adults-diagnosis-treatment-part-2/</a>	Preclerk, clerk, intern, junior, senior, attending	17
Acute and chronic renal failure	LaPietra, Alexis	Piperacillin/Tazobactam and Risk of Acute Kidney Injury with Vancomycin	<a href="https://www.alleim.com/2014/05/piperacillin-tazobactam-acute-kidney-injury/">https://www.alleim.com/2014/05/piperacillin-tazobactam-acute-kidney-injury/</a>	Junior, senior	17
Calculus of urinary tract (renal colic)		Intravenous lidocaine for renal colic	<a href="https://www.alleim.com/2018/02/intravenous-lidocaine-for-renal-colic/">https://www.alleim.com/2018/02/intravenous-lidocaine-for-renal-colic/</a>	Intern, junior, senior, attending	17
Hematuria	Saunders, Fiona	Investigating microscopic haematuria in blunt abdominal trauma	<a href="https://bestbets.org/bets/bet.php?id=41">https://bestbets.org/bets/bet.php?id=41</a>	Junior, senior, attending	16
Urinary retention	Rezaie, Salim	Urinary Retention: Rapid Drainage or Gradual Drainage to Avoid Complications?	<a href="http://rebelem.com/urinary-retention-rapid-drainage-gradual-drainage-avoid-complications/">http://rebelem.com/urinary-retention-rapid-drainage-gradual-drainage-avoid-complications/</a>	Junior, senior, attending	16
Calculus of urinary tract (renal colic)	Milne, Ken	SGEM#202: Lidocaine for Renal Colic?	<a href="http://thesgem.com/2018/01/sgem202-lidocaine-for-renal-colic/">http://thesgem.com/2018/01/sgem202-lidocaine-for-renal-colic/</a>	Intern, junior, senior, attending	16
Calculus of urinary tract (renal colic)	Wightman, Rachel	Ultrasonography vs CT in Renal Colic	<a href="https://coreem.net/journal-reviews/ultrasonography-vs-ct-in-renal-colic/">https://coreem.net/journal-reviews/ultrasonography-vs-ct-in-renal-colic/</a>	Intern, junior, senior, attending	16
Hemolytic uremic syndrome	Long, Brit	EM@3AM – Hemolytic Uremic Syndrome	<a href="http://www.emdocs.net/em3am-hemolytic-uremic-syndrome/">http://www.emdocs.net/em3am-hemolytic-uremic-syndrome/</a>	Junior, senior, attending	16
Gangrene of the scrotum (Fournier's gangrene)	Cohen, Paul	How do we misdiagnose and mismanage necrotizing fasciitis?	<a href="http://www.emdocs.net/misdiagnose-mismanage-necrotizing-fasciitis/">http://www.emdocs.net/misdiagnose-mismanage-necrotizing-fasciitis/</a>	Junior, senior, attending	16
Gangrene of the scrotum (Fournier's gangrene)	Santistevan, Amie	Necrotizing Fasciitis: Pearls & Pitfalls	<a href="http://www.emdocs.net/necrotizing-fasciitis-pearls-pitfalls/">http://www.emdocs.net/necrotizing-fasciitis-pearls-pitfalls/</a>	Preclerk, clerk, intern, junior, senior, attending	16
Urinary retention	Fox, Sean	Urinary Retention in Kids	<a href="http://pedemorsels.com/urinary-retention-in-kids/">http://pedemorsels.com/urinary-retention-in-kids/</a>	Junior	16
Torsion	Shih, Jeffrey	Ultrasound For The Win: 22M with Scrotal Pain #US4TW	<a href="https://www.alleim.com/2015/02/ultrasound-win-22m-scrotal-pain-us4tw/">https://www.alleim.com/2015/02/ultrasound-win-22m-scrotal-pain-us4tw/</a>	Junior, senior, attending	16
Genital lesions	Lin, Michelle	Paucis Verbis: Genital Ulcers	<a href="https://www.alleim.com/2012/05/paucis-verbis-genital-ulcers/">https://www.alleim.com/2012/05/paucis-verbis-genital-ulcers/</a>	Clerk, intern, junior, senior	16
Pyelonephritis	Long, Brit	Pyelonephritis: It's not always so straightforward...	<a href="http://www.emdocs.net/pyelonephritis-not-always-so-straightforward/">http://www.emdocs.net/pyelonephritis-not-always-so-straightforward/</a>	Preclerk, clerk, intern, junior, senior, attending	16
Topic not on list	Sparks, Matt	Renal Fellow Network: Top nephrology-related stories of 2016	<a href="http://renalfellow.blogspot.com/2016/12/top-nephrology-related-stories-of-2016_25.html">http://renalfellow.blogspot.com/2016/12/top-nephrology-related-stories-of-2016_25.html</a>	Junior, senior, attending	16

(Continued)

Table 3 (continued)

Subtopic	Name of First Author	Name of Blog Post	URL	Level of Trainee (Starred = Author, Unstarred = Team's Recommendation)	rMETRIQ score
Epididmitis/Orchitis	Shih, Jeffrey	Ultrasound For The Win! Case – 64M with Fever and Scrotal Pain #US4TW	<a href="https://www.aliem.com/2018/02/ultrasound-for-the-win-case-fever-scrotal-pain-us4tw/">https://www.aliem.com/2018/02/ultrasound-for-the-win-case-fever-scrotal-pain-us4tw/</a>	Intern, junior, senior, attending	16
Calculus of urinary tract (renal colic)	Lin, Michelle	Author Insight: Ultrasonography versus CT for suspected nephrolithiasis   NEJM	<a href="https://www.aliem.com/2015/03/author-insight-ultrasonography-versus-ct-for-suspected-nephrolithiasis-nejm/">https://www.aliem.com/2015/03/author-insight-ultrasonography-versus-ct-for-suspected-nephrolithiasis-nejm/</a>	Junior, senior, attending	16
Calculus of urinary tract (renal colic)	Swaminathan, Anand	Alpha Blockers in Renal Colic: A Systematic Review	<a href="http://rebelem.com/alpha-blockers-in-renal-colic-a-systematic-review/">http://rebelem.com/alpha-blockers-in-renal-colic-a-systematic-review/</a>	Junior, senior, attending	16

SOAR = Systematic Online Academic Resource.

audio and video entries from our review. Previous literature has shown that trainees have some level of preference for podcasts,<sup>3,4</sup> and so this may have resulted in the exclusion of some popular and leading FOAM content. However, podcast show notes were included in our search strategy and included for review.

Also, for pragmatic reasons (primarily resource constraints), only a single reviewer extracted information and rated the items (rMETRIQ score, appropriate target audience, usage). To mitigate the problems with this single extractor, we met multiple times as a team to discuss the interpretation of each aspect of the rMETRIQ score. However, despite the presumed improvements in reliability of the rMETRIQ score,<sup>18</sup> it is likely that additional raters would be needed to reach reliable ratings for each resource.<sup>31</sup>

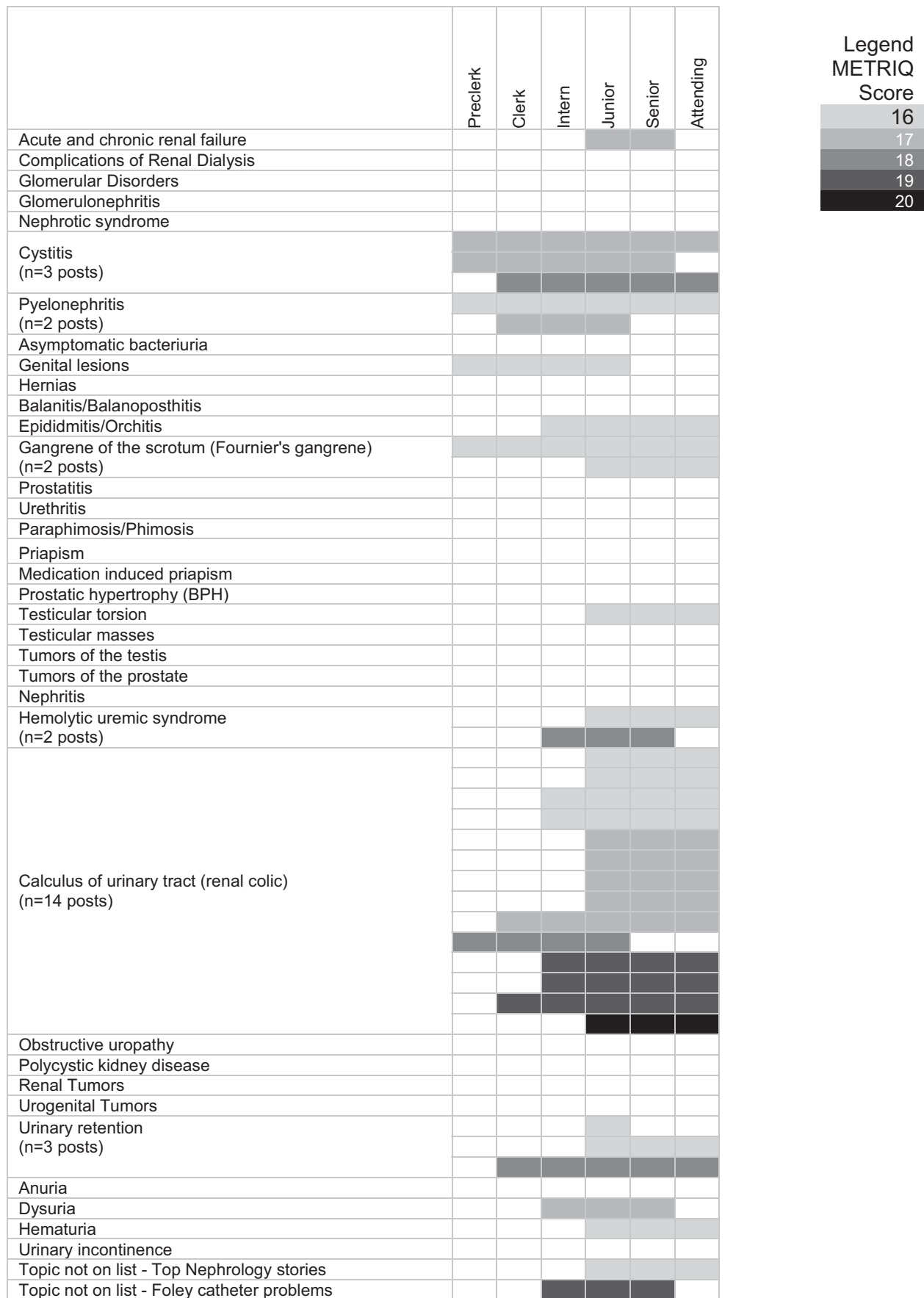
Moreover, we did not directly assess the accuracy of content included in posts. Assessing accuracy of online educational resources is difficult and no criterion standard exists. The rMETRIQ score does not directly ask for a quality appraisal of the blog post; it was designed and derived to be used by junior trainees who may be unable to accurately judge quality.<sup>32</sup> Further, the rMETRIQ score includes elements that are likely to correlate with and/or help ensure accuracy (e.g., peer review processes, postpublication commentary, references, COI). In the future, applying another scoring tool which incorporates an educator's judgment on accuracy (such as the ALiEM AIR score<sup>27,28</sup>) may potentially improve the quality assessment of online content (n.b., prior work has shown correlation between the ALiEM AIR and METRIQ scores<sup>31</sup>). Next, while we did discuss and calibrate our raters with regards to the target audience and usage suitability, we did not compute any inter-rater reliability metric on this part of our extraction.

Finally, our analysis likely overestimates topic coverage, as posts are included if they cover any part of a topic. Finally, we utilized a new and as-of-yet untested methodology to identify both the resources for inclusion and the highest quality resources.

## CONCLUSION

We present a systematic approach to reviewing online educational resources, showing that only 10% of posts that met our inclusion criteria were of high quality according to the revised METRIQ scoring tool. In our first review, we found that free online medical





**Figure 3.** FOAM nephrology coverage at a glance. FOAM = free online medical education resources.

education resources covered renal/genitourinary topics broadly but with variable quality and high-quality posts only covered approximately one-third of topics. We hope that educators and learners can use this information to find educational resources and that free online medical education resources content producers can use this information to choose topics to produce new content.

We thank the SAEM for their support of this project.

## References

1. Reiter DA, Lakoff DJ, Trueger NS, Shah KH. Individual interactive instruction: an innovative enhancement to resident education. *Ann Emerg Med* 2013;61:110–3.
2. Scott KR, Hsu CH, Johnson NJ, Mamtani M, Conlon LW, DeRoos FJ. Integration of social media in emergency medicine residency curriculum. *Ann Emerg Med* 2014;64:396–404.
3. Purdy E, Thoma B, Bednarczyk J, Migneault D, Sherbino J. The use of free online educational resources by Canadian emergency medicine residents and program directors. *CJEM* 2015;1717:101–6.
4. Mallin M, Schlein S, Doctor S, Stroud S, Dawson M, Fix M. A survey of the current utilization of asynchronous education among emergency medicine residents in the United States. *Acad Med J Assoc Am Med Coll* 2014;89:598–601.
5. Patocka C, Lin M, Voros J, Chan T. Point-of-care resource use in the emergency department: a developmental model. *AEM Educ Train* 2018;2:221–8.
6. Cadogan M, Nickson C. Free Open Access Medical Education. *Life in the Fast Lane*. 2014. Available at: <http://lifeinthefastlane.com/foam/>. Accessed January 9, 2016.
7. Cadogan M, Thoma B, Chan TM, Lin M. Free open access medical education (FOAM): the rise of emergency medicine and critical care blogs and podcasts (2002-2013). *Emerg Med J* 2014;31:e76–7.
8. Schrager DL. Does everything need to be “scientific”? *Ann Emerg Med* 2016;68:738–9.
9. Krishnan K, Thoma B, Trueger NS, Lin M, Chan TM. Gestalt assessment of online educational resources may not be sufficiently reliable and consistent. *Perspect Med Educ* 2017;6:91–8.
10. Thoma B, Sebok-Syer SS, Krishnan K, et al. Individual gestalt is unreliable for the evaluation of quality in medical education blogs: a METRIQ study. *Ann Emerg Med* 2017;70:394–401.
11. Trueger NS. Yelp, but for blogs? *Ann Emerg Med* 2016;68:736–7.
12. Stuntz R, Clontz R. An evaluation of emergency medicine core content covered by free open access medical education resources. *Ann Emerg Med* 2016;67:649–53.e2.
13. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
14. Counselman FL, Babu K, Edens MA, et al. The 2016 Model of the Clinical Practice of Emergency Medicine. *J Emerg Med* 2017;52:846–9.
15. Raine T, Thoma B, Chan TM, Lin M. FOAMSearch.net: a custom search engine for emergency medicine and critical care. *Emerg Med Australas* 2015;27:363–5.
16. Thoma B, Sanders J, Lin M, Paterson Q, Steeg J, Chan T. The Social Media Index: measuring the impact of emergency medicine and critical care websites. *West J Emerg Med* 2015;16:242–9.
17. Thoma B, Chan TM, Kapur P, et al. The Social Media Index as an indicator of quality for emergency medicine blogs: a METRIQ study. *Ann Emerg Med* 2018;72:696–702.
18. Colmers-Gray IN, Krishnan K, Chan TM, et al. P022: the revised METRIQ score: an international, social-media based usability analysis of a quality evaluation instrument for medical education blogs. *Can J Emerg Med* 2018;20 (Suppl S1):S64.
19. George S, Haque MS, Oyebo F. Standard setting: comparison of two methods. *BMC Med Educ* 2006;6:46.
20. Lo A, Shappell E, Rosenberg H, et al. Four strategies to find, evaluate, and engage with online resources in emergency medicine. *Can J Emerg Med* 2018;20:293–9.
21. Raikos A, Waidyasekara P. How useful is YouTube in learning heart anatomy? *Anat Sci Educ* 2014;7:12–8.
22. Azer SA. Mechanisms in cardiovascular diseases: how useful are medical textbooks, eMedicine, and YouTube? *AJP Adv Physiol Educ* 2014;38:124–34.
23. Rees CE, Ford JE, Sheard CE. Evaluating the reliability of DISCERN: a tool for assessing the quality of written patient information on treatment choices. *Patient Educ Couns* 2002;47:273–5.
24. Khazaal Y, Chatton A, Cochand S, et al. Brief DISCERN, six questions for the evaluation of evidence-based content of health-related websites. *Patient Educ Couns* 2009;77:33–7.
25. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999;53:105–11.
26. Shepperd S, Charnock D, Cook A. A 5-star system for rating the quality of information based on DISCERN. *Health Info Libr J* 2002;19:201–5.
27. Lin M, Joshi N, Grock A. Approved Instructional Resources (AIR) series: a national initiative to identify quality emergency medicine blog and podcast content for resident education. *J Grad Med Educ* 2016;8:219–25.
28. Chan TM, Grock A, Paddock M, Kulasegaram K, Yarris LM, Lin M. Examining reliability and validity of an online

- score (ALiEM AIR) for rating free open access medical education resources. 2016;68:729–35.
29. Hoang JK, McCall J, Dixon AF, Fitzgerald RT, Gaillard F. Using social media to share your radiology research: how effective is a blog post? *J Am Coll Radiol* 2015;12:760–5.
  30. Dixon A, Fitzgerald RT, Gaillard F. Letter by Dixon et al regarding article, “A randomized trial of social media from circulation”. *Circulation* 2015;131:e393.
  31. Thoma B, Sebok-Syer SS, Colmers-Gray I, et al. Quality evaluation scores are no more reliable than gestalt in evaluating the quality of emergency medicine blogs: a METRIQ study. *Teach Learn Med* 2018;30:294–302.
  32. Chan TM, Thoma B, Krishnan K, et al. Derivation of two critical appraisal scores for trainees to evaluate online educational resources: a METRIQ study. *West J Emerg Med* 2016;17:574–84.

# The Revised METRIQ Score: A Quality Evaluation Tool for Online Educational Resources

Isabelle N. Colmers-Gray, MD, MSc<sup>1</sup>, Keeth Krishnan, MD<sup>2</sup>, Teresa M. Chan, MD, MHPE<sup>3</sup> , N. Seth Trueger, MD, MPH<sup>4</sup>, Michael Paddock, DO, MS<sup>5</sup>, Andrew Grock, MD<sup>6</sup>, Fareen Zaver, MD<sup>7</sup>, and Brent Thoma, MD, MA, MSc<sup>8</sup> 

## ABSTRACT

**Background:** With the rapid proliferation of online medical education resources, quality evaluation is increasingly critical. The Medical Education Translational Resources: Impact and Quality (METRIQ) study evaluated the METRIQ-8 quality assessment instrument for blogs and collected feedback to improve it.

**Methods:** As part of the larger METRIQ study, participants rated the quality of five blog posts on clinical emergency medicine topics using the eight-item METRIQ-8 score. Next, participants used a 7-point Likert scale and free-text comments to evaluate the METRIQ-8 score on ease of use, clarity of items, and likelihood of recommending it to others. Descriptive statistics were calculated and comments were thematically analyzed to guide the development of a revised METRIQ (rMETRIQ) score.

**Results:** A total of 309 emergency medicine attendings, residents, and medical students completed the survey. The majority of participants felt the METRIQ-8 score was easy to use (mean  $\pm$  SD =  $2.7 \pm 1.1$  out of 7, with 1 indicating strong agreement) and would recommend it to others ( $2.7 \pm 1.3$  out of 7, with 1 indicating strong agreement). The thematic analysis suggested clarifying ambiguous questions, shortening the 7-point scale, specifying scoring anchors for the questions, eliminating the “unsure” option, and grouping-related questions. This analysis guided changes that resulted in the rMETRIQ score.

**Conclusion:** Feedback on the METRIQ-8 score contributed to the development of the rMETRIQ score, which has improved clarity and usability. Further validity evidence on the rMETRIQ score is required.

With increasing expansion of emergency medicine (EM) blogs and podcasts, residents frequently use these open educational resources to supplement and potentially replace traditional tools.<sup>1-4</sup> Unlike textbooks and journals, these online resources are rarely peer-reviewed<sup>5-7</sup> and critics raise concerns that learners

From the <sup>1</sup>Department of Emergency Medicine, University of Alberta, Edmonton, Alberta, Canada; the <sup>2</sup>Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada; the <sup>3</sup>Division of Emergency Medicine, Department of Medicine, McMaster University, Hamilton, Ontario, Canada; the <sup>4</sup>Department of Emergency Medicine, Northwestern University, Chicago, IL; the <sup>5</sup>Department of Emergency Medicine, University of Minnesota, Minneapolis, MN; the <sup>6</sup>USC and LAC Departments of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; the <sup>7</sup>Department of Emergency Medicine, University of Calgary, Calgary, Alberta, Canada; and the <sup>8</sup>Department of Emergency Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Received January 26, 2019; revision received June 8, 2019; accepted June 17, 2019.

Presented at the Canadian Association of Emergency Physicians Annual Conference, Calgary, Alberta, Canada, May 2018.

Funding for this research was provided by the Canadian Association of Emergency Physicians (Junior Investigator Grant) and the Royal College of Physicians and Surgeons of Canada (Robert Maudsley Fellowship for Studies in Medical Education).

BT (ALiEM.com, Debrief2Learn.org, EMSimCases.com, and CanadiEM.org), FZ (ALiEM.com, and CanadiEM.org), NST (Mdaware.org), AG (ALiEM.com), and TMC (ALiEM.com, CanadiEM.org, FeminEM.org, ICENet.royalcollege.ca) edit or operate medical education blogs. NST receives salary support from the American Medical Association for his role as Digital Media Editor, *JAMA Network Open*, and a stipend for his role as Social Media Editor for *Emergency Physicians Monthly*; he previously received a stipend for his role as Social Media Editor for *Annals of Emergency Medicine* during portions of the METRIQ Study. The remaining authors declare no conflicts of interest.

Supervising Editor: Lalena M. Yarris, MD.

Address for correspondence and reprints: Brent Thoma, MD, MA, MSc; e-mail: brent.thoma@usask.ca.

AEM EDUCATION AND TRAINING 2019;3:387–392.

are being misled.<sup>8-10</sup> Supporting these concerns, the Medical Education Translational Resources: Impact and Quality (METRIQ) study found that gestalt evaluations of these resources were unreliable.<sup>11-14</sup> This suggests that a systematized appraisal of these resources may be more appropriate.<sup>11,15</sup>

The METRIQ-8 score is a structured rating tool resulted from a rigorous derivation process, which included a systematic review and qualitative analysis designed to identify appropriate quality indicators for blogs,<sup>16</sup> a modified Delphi process with expert bloggers and podcasters,<sup>17</sup> a modified Delphi process with medical educators,<sup>18</sup> and a derivation study.<sup>19</sup> However, along with another structured assessment tool (the ALiEM AIR score<sup>19,20</sup>), the METRIQ study found that METRIQ-8 was no more reliable than staff physician gestalt in a general population of raters.<sup>12</sup> As part of a planned secondary analysis of data collected within the METRIQ study, we analyzed feedback on the METRIQ-8 score with the goal of improving its usability and reliability.

## METHODS

This was a planned secondary analysis of data from the METRIQ study (<http://metriqstudy.org>), which recruited students, EM trainees, and EM attendings to rate the quality of 20 clinically oriented EM blog posts via an online survey between March 1, 2016, and June 1, 2016.<sup>11,13,14</sup> After rating five blog posts with the METRIQ-8 score (outlined in Data Supplement S1, Table S1, available as supporting information in the online version of this paper, which is available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10376/full>), participants rated the METRIQ-8 score on usability and whether they would recommend it using 7-point Likert scales (1 = strongly agree). Participants also reviewed the eight METRIQ-8 items to identify unclear elements. Each question was followed by an open-ended question asking participants to explain their rationale. Only data from participants who completed the full METRIQ study survey were used. The METRIQ study protocol was reviewed by the University of Saskatchewan's Research Ethics Board and met the requirements for exemption (BEH 16-09).

Quantitative analysis was conducted using parametric descriptive statistics and tests of significance. Analysis of variance with a two-tailed significance of  $\alpha = 0.05$  was used to determine whether the ease of

use or likelihood of recommendation differed significantly based upon level of training, frequency of blog reading, or region of origin.

Anonymized narrative data were analyzed using a content analysis to identify areas for improvement and common points of feedback.<sup>21</sup> The researchers included six staff emergency physicians, one emergency medicine resident, and a senior medical student. Two authors had formal training in qualitative methods (TMC, BT). All authors were familiar with open-access medical education resources. Two raters (INCG, KK) independently coded the data, compared their analyses, and resolved discrepancies through consensus. The final codebook was organized into themes and subthemes with quotes from participants demonstrating each subtheme.

The revised METRIQ (rMETRIQ) score was developed through an iterative process. A subgroup of the authors (INCG, KK, BT) modified each item of the METRIQ-8 score and developed specific scoring criteria for each revised item. The remaining authors provided feedback, and consensus on each item was reached via group discussion. This version was then piloted by the five authors not involved in revising the score (TMC, NST, MP, AG, FZ) on a new set of blog posts. Consistent with methods used in previous work,<sup>12</sup> one new clinically relevant blog post was selected for review from each of the 10 websites used in the METRIQ study. Average intraclass correlation coefficients (ICCs) were calculated for each item and the total score (the sum of each item's scores). Minor additional edits were made to the final version to clarify items with a lower ICC (indicating lower reliability).

## RESULTS

Participant demographics are described in Table 1. A total of 309 of the 330 (93.6%) individuals enrolled in the METRIQ study completed the survey. As outlined in Figure 1, the majority of participants agreed that "the METRIQ-8 score was easy to use" (mean  $\pm$  standard deviation [SD] =  $2.7 \pm 1.1$  on a 7-point scale, with 1 indicating "strongly agree") and "would recommend the METRIQ-8 score for the evaluation of blog posts" (mean  $\pm$  SD =  $2.7 \pm 1.3$  on a 7-point scale, with 1 indicating "strongly agree"). Neither ease of use nor recommendation of the METRIQ-8 score varied significantly by level of training, frequency of blog reading, or global region of origin.



**Table 1**  
The METRIQ Study Participant Demographics

Variable	Category	<i>n</i> ( <i>N</i> = 309)	%
Age	Years	31.1 (mean)	7.3 (SD)
Sex	Female	123	39.8
	Male	184	59.5
	Other	2	0.6
Level of training	Medical student	121	39.2
	EM resident	88	28.5
	Emergency attending physician	100	32.4
Frequency of reading medical education blogs	Daily	48	15.5
	Several times weekly	141	45.6
	Once weekly	43	13.9
	Several times monthly	38	12.3
	Once monthly	21	6.8
	Less than once monthly	15	4.9
	Never	2	0.6
Manage, edit, own, or operate a medical education blog(s)	Yes	45	14.5
	No	261	84.5
	No response	3	1.0

Qualitative analysis of the comments and feedback on the METRIQ-8 score revealed nine main themes and 51 subthemes (detailed in Data Supplement S1, Table S2). Main themes included usability, interpretation, length, application, structure, validity and reliability, scale, completeness, and comparison to the ALiEM AIR score.

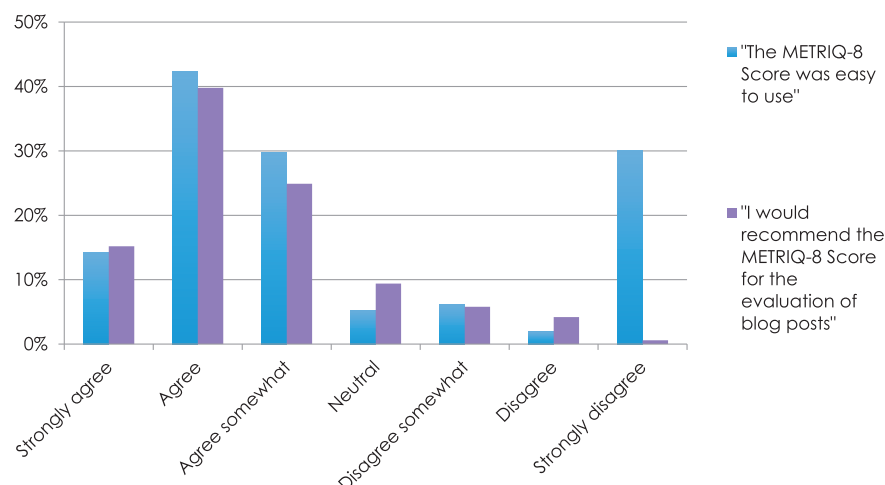
The feedback summarized in the thematic analysis (Figure 2) informed the creation of the rMETRIQ score from the METRIQ-8 score. Significant changes

from the METRIQ-8 score are summarized in Data Supplement S1, Table S1, and included clarifying ambiguous questions, shortening the 7-point scale to a 4-point scale, specifying scoring anchors for each question, eliminating the “unsure” option, and grouping-related questions. The choice of a 4-point scale was consistent with multiple participant recommendations to reduce the number of options. The scoring criteria and question refinement clarified terminology previously identified as ambiguous. Finally, we changed the order of the questions to group them into three broad categories: content, credibility, and review. The qualitative analysis of feedback provided on each item is outlined in Data Supplement S1, Table S3.

The results of pilot testing are shown in Table 2. Reliability of the aggregate score was high (ICC = 0.94, 95% confidence interval [CI] = 0.84–0.98). ICCs for individual items were also high ( $\geq 0.80$ ) with the exception of rQ3 (“Is the resource well written and formatted?”; 0.72) and rQ5 (“Is it clear who created the resource and do they have any conflicts of interest?”; 0.59). Further changes to these items suggested by the pilot testers were made as outlined in Data Supplement S1, Table S2. The final rMETRIQ Score is presented in Figure 2.

## DISCUSSION

The rMETRIQ score was developed from the METRIQ-8 score by leveraging quantitative and qualitative feedback provided by a large population of users at various stages in training, geographic location, and levels of involvement with online medical education. This diverse group of participants mirrors the range of



**Figure 1.** The METRIQ study participants' perspectives on the ease of use and recommendation of the METRIQ-8 score.

The rMETRIQ Score					
		0	1	2	3
CONTENT	1 Does the resource provide enough background information to situate the user?	No, the information presented within the resource cannot be situated within its broader context without looking up information independently.	No, the information presented within the resource cannot be situated within its broader context, but users are directed to other valuable resources with this information.	Yes, the resource provides sufficient background information to situate the user.	Yes, the resource provides sufficient background information to situate the user and also directs users to other valuable resources related to the topic.
	2 Does the resource contain an appropriate amount of information for its length?	Insufficient content.	Lots of unnecessary, redundant, or missing content.	"Any initial reactions?" Some unnecessary, redundant, or missing content, but most content was essential.	No unnecessary, redundant, or missing content; all content was essential.
	3 Is the resource well written and formatted?	The resource is poorly written and/or formatted, and should not be a resource for learning.	The resource is somewhat well-written and formatted, but could benefit from substantive editing (e.g. grammatical errors are seen, or better organized).	The resource is reasonably well-written and formatted, but aspects of the organization or presentation are distracting or otherwise detrimental to learning.	The resource is very well-written and formatted in a way that optimizes and benefits learning.
CREDIBILITY	4 Does the resource cite its references?	No, no references are cited.	Yes, there are references listed but they do not map to specific statements within the resource.	Yes, the references are cited and clearly map to specific statements within the resource, but statements of fact that are not common knowledge are made without the support of a reference.	Yes, the references are cited, clearly map to specific statements within the resource, and all statements of fact that are not common knowledge are supported with a reference.
	5 Is it clear who created the resource and do they have any conflicts of interest?	No, the author of the resource has significant conflicts of interest; or the author is not clearly identified (e.g. no name or a pseudonym is used).	Yes, the identity of the author is clear, but they do not list their qualifications or disclose whether they have any conflicts of interest.	Yes, the identity and qualifications of the author are clear, but they do not disclose whether they have any conflicts of interest.	Yes, the identity and qualifications of the author are clear and they specify that they have no relevant conflicts of interest.
REVIEW	6 Are the editorial and pre-publication peer review processes that were used to create the resource clearly outlined?	No, it is unclear whether or not the website has a review process; or, there is no process.	Yes, a review process is mentioned on the website, but it was not clearly described.	Yes, a clear review process is described on the website, but it was not clear whether it was applied to the resource.	Yes, a clear review process is described on the website and it was clearly applied to the resource.
	7 Is there evidence of post-publication commentary on the resource's content by its users?	No, there was no mechanism to leave comments; or comments that were present were either unrelated to the post or unprofessional.	There was a mechanism to leave comments but none had been made.	Yes, some comments have been made on the resource, but a robust discussion about the resource's content has not occurred.	Yes, a robust discussion of the resource's content has occurred that expands upon the content of the resource.

**Figure 2.** The revised METRIQ (rMETRIQ) score.

**Table 2**  
ICCs for the Items of the Revised METRIQ (rMETRIQ) Score

rMETRIQ Score Item	Average-measures ICC (95% CI)
Aggregate score (sum of rQ1–rQ7)	0.94 (0.84–0.98)
Revised Question 1 (rQ1). Does the resource provide enough background information to situate the user?	0.89 (0.70–0.97)
Revised Question 2 (rQ2). Does the resource contain an appropriate amount of information for its length?	0.80 (0.51–0.94)
Revised Question 3 (rQ3). Is the resource well written and formatted?	0.72 (0.30–0.92)
Revised Question 4 (rQ4). Does the resource cite its references?	0.96 (0.90–0.99)
Revised Question 5 (rQ5). Is it clear who created the resource and do they have any conflicts of interest?	0.59 (0–0.88)
Revised Question 6 (rQ6). Are the editorial and prepublication peer review processes that were used to create the resource clearly outlined?	0.82 (0.56–0.95)
Revised Question 7 (rQ7). Is there evidence of postpublication commentary on the resource's content by its users?	0.95 (0.87–0.99)

ICC = intraclass correlation coefficient.

typical end-users that our instrument targets and identified correctable aspects for improvement.

The development of the rMETRIQ score is important given recent studies demonstrating that both gestalt<sup>11,15</sup> and the current structured evaluation tools (METRIQ-8 and ALiEM AIR)<sup>12</sup> lack reliability in general populations of raters. Reliability is an important component of modern validity theory and is generally felt to be a necessary (but not sufficient) aspect of validity.<sup>22</sup> Reliability improves with the number of raters, but the gestalt rating of blog posts requires a prohibitive number to achieve adequate reliability.<sup>11</sup> The reliability of instruments (i.e., higher value in an average-measures ICC) can be increased through rater training and instrument improvement.<sup>23</sup> In light of the disappointing results of reliability testing, we felt that revising our instrument would be the next reasonable approach to improving evaluation of these resources. Pilot testing of the rMETRIQ score suggests that its reliability has improved. However, it will require further evaluation in a larger validation study with general readers of emergency medicine blogs. We anticipate that the rMETRIQ score will impact three separate areas within EM: first, by guiding quality assessment of online resource among readers; second, by improving quality of online content by providing a framework of quality metrics for content producers to incorporate into future online content; and finally, by supporting the development of more robust methods of reviewing and assessing the online emergency medicine.

The rMETRIQ score was recently used to appraise the quality of blog posts in the new SAEM Systematic Online Academic Resource (SOAR) review series of online educational content on EM renal and genitourinary conditions.<sup>24</sup> Although our work and the METRIQ study are centered around EM content, the quality principles of the rMETRIQ score can easily be applied to other domains within medicine. Additionally, we anticipate that with minor modification of the wording of the instrument, it will be possible to apply it to other types of online resources such as podcasts, videos, and other open educational resources that are not vetted through traditional peer review processes. Further studies will also be required to classify the numeric scores (i.e., what score constitutes high vs. medium vs. low quality).

## LIMITATIONS

First, the data used in this study was collected in 2016 and it is possible that the feedback received on the

METRIQ-8 score may have differed with a sample of blog posts published more recently. Second, given the significant modifications made to develop the rMETRIQ score, new validity evidence will need to be collected before its use can be recommended broadly. Finally, the rMETRIQ score was developed specifically using blogs and will need to be modified for application to other popular online educational resources.

## CONCLUSIONS

Direct feedback on the METRIQ-8 score spurred the development of the revised METRIQ score with improved usability and reliability. We anticipate that it will be used widely to assess the quality of blog posts and, potentially, other online resources. Further validity evidence for use of the revised METRIQ score will be required before it can be broadly recommended.

The authors thank the medical students, residents, and staff physicians who participated in the METRIQ study.

## References

1. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. *J Am Med Inform Assoc* 2012;19:777–81.
2. Matava CT, Rosen D, Siu E, et al. eLearning among Canadian anesthesia residents: a survey of podcast use and content needs. *BMC Med Educ* 2013;13:59.
3. Purdy E, Thoma B, Bednarczyk J, et al. The use of free online educational resources by Canadian emergency medicine residents and program directors. *Can J Med Educ* 2015;17:101–6.
4. Mallin M, Schlein S, Doctor S, et al. A survey of the current utilization of asynchronous education among emergency medicine residents in the United States. *Acad Med* 2014;89:598–601.
5. Azim A, Beck-Esmay J, Chan TM. Editorial processes in free open access medical educational (FOAM) resources. *AEM Educ Train* 2018;2:204–12.
6. Thoma B, Chan T, Desouza N, et al. Implementing peer review at an emergency medicine blog: bridging the gap between educators and clinical experts. *Can J Emerg Med* 2015;17:188–91.
7. Sidalak D, Purdy E, Lockett-Gatopoulos S, et al. Coached peer review: developing the next generation of authors. *Acad Med* 2017;92:201–4.
8. Genes N. Pro/con: why #FOAMed Is Not Essential to EM Education[Internet]. *Emergency Physicians Monthly*. 2014. Available at: <http://epmonthly.com/article/pro-con-is-foam-essential-to-em-education-no/>. Accessed June 6, 2019.

9. Brabazon T. The Google effect: googling, blogging, wikis and the flattening of expertise. *Libri* 2007;56:157–67.
10. Schriger DL. Does everything need to be “scientific?”. *Ann Emerg Med* 2016;68:738–9.
11. Thoma B, Sebok-Syer SS, Krishnan K, et al. Individual gestalt is unreliable for the evaluation of quality in medical education blogs: a METRIQ study. *Ann Emerg Med* 2017;70:394–401.
12. Thoma B, Sebok-Syer SS, Colmers-Gray I, et al. Quality Evaluation Scores are no more reliable than gestalt in evaluating the quality of emergency medicine blogs: a METRIQ study. *Teach Learn Med* 2018;30:294–302.
13. Thoma B, Paddock M, Purdy E, et al. Leveraging a virtual community of practice to participate in a survey-based study: a description of the METRIQ study methodology. *AEM Educ Train* 2017;1:110–3.
14. Thoma B, Chan TM, Kapur P, et al. The social media index as an indicator of quality for emergency medicine blogs: a METRIQ study. *Ann Emerg Med* 2018;72:696–702.
15. Krishnan K, Thoma B, Trueger NS, et al. Gestalt assessment of online educational resources may not be sufficiently reliable and consistent. *Perspect Med Educ* 2017;6:91–8.
16. Paterson QS, Thoma B, Milne WK, et al. A systematic review and qualitative analysis to determine quality indicators for health professions education blogs and podcasts. *J Grad Med Educ* 2015;7:549–54.
17. Thoma B, Chan TM, Paterson QS, et al. Emergency medicine and critical care blogs and podcasts: establishing an international consensus on quality. *Ann Emerg Med* 2015;66(396–402):e4.
18. Lin M, Thoma B, Trueger NS, et al. Quality indicators for blogs and podcasts used in medical education: modified Delphi consensus recommendations by an international cohort of health professions educators. *Postgrad Med J* 2015;91:546–50.
19. Chan TM, Thoma B, Krishnan K, et al. Derivation of two critical appraisal scores for trainees to evaluate online educational resources: a METRIQ study. *West J Emerg Med* 2016;17:574–84.
20. Chan TM, Grock A, Paddock M, et al. Examining reliability and validity of an online score (ALiEM AIR) for rating free open access medical education resources. *Ann Emerg Med* 2016;68:729–35.
21. Cooper S, Endacott R. Generic qualitative research: a design for qualitative research in emergency care? *Emerg Med J* 2007;24:816–9.
22. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006;166:e7–16.
23. Norman GR, Streiner DL. *Measurement Scales: A Practice Guide to Their Development and Use*, 3rd ed. Oxford: Oxford University Press, 2008.
24. Grock A, Bhalerao A, Chan TM, et al. Systematic Online Academic Resource (SOAR) review: renal and genitourinary. *AEM Educ Train* 2019;19:000–00.

### Supporting Information

The following supporting information is available in the online version of this paper available at <http://onlinelibrary.wiley.com/doi/10.1002/aet2.10376/full>

**Data Supplement S1.** Supplemental material

# Spokes for Our Folks: Public Health Bike Tour

Christian Rose, MD<sup>1,2</sup> , Brian Chang, MD<sup>3</sup>, and John Brown, MD<sup>3,4</sup>

## ABSTRACT

Nearly half of medical care in the United States is managed through the emergency department, a large portion of which could be managed by “lateral” health services provided by public health facilities like human immunodeficiency virus (HIV) prophylaxis, alcohol and drug treatment programs, emergency psychiatric resources, and medical respite or rehabilitation centers. These options may be underutilized due to lack of knowledge of their services and demographics by patients and health care workers alike. We aimed to educate all levels of emergency medicine trainees and staff to citywide services via bike tour. Participants reported an improved understanding of health services as well as a sense of “camaraderie” toward lateral health services and other providers on the rides.

## BACKGROUND

Nearly half of the medical care provided in hospitals across the United States comes from the emergency department (ED).<sup>1</sup> However, up to 27.1% of all ED visits could take place at alternative sites.<sup>2</sup> Appropriate use of “lateral” health services available from outpatient clinics and community health facilities allows for more appropriate allocation of services and has been regarded as a potential solution for ED overcrowding.<sup>3</sup> What is more, emergency physicians report keeping patients in the ED longer and admit to the hospital more due to inability to arrange outpatient follow-up.<sup>3</sup> We believe that this may be mitigated by physicians learning about the landscape of outpatient opportunities available for their patients. Furthermore, patients report higher satisfaction with their care when given options outside of the hospital.

Emergency medicine (EM) residents at the University of California at San Francisco often report unfamiliarity with the cornucopia of services available in their city and that they are less likely to utilize them due to not being sure what services are offered or who qualifies for them—despite summaries from ED social

workers as well as department of public health website information about available services.

The World Health Organization reports that inter-professional education and collaborative practice can improve health outcomes and is a necessary step in preparing a health workforce that is better prepared to respond to local health needs. It occurs when students from two or more professions learn about, from, and with each other to enable effective collaboration.<sup>4</sup> Thus, now heading into its fourth year, we set out to use the public bike lanes and an multidisciplinary team to visit a subset of public health facilities from which our patients arrive, depart to, or get care from regularly. We believe that by familiarizing residents, faculty, nursing, and emergency personnel with the patient demographic and services offered by local health care facilities, we might improve the use of lateral health services and comfort recommending them.

## EXPLANATION

Location selection was determined by feedback from residents and staff regarding which services they would like to learn more about in conjunction with

From the <sup>1</sup>VA Palo Alto Health Care System, Palo Alto, CA; <sup>2</sup>Stanford University, Stanford, CA; the <sup>3</sup>University of California, San Francisco, CA; and the <sup>4</sup>San Francisco Department of Public Health, San Francisco, CA.

Received March 20, 2019; revision received May 24, 2019; accepted June 6, 2019.

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

Supervising Editor: Anne Messman, MD.

Address for correspondence and reprints: Christian Rose, MD; e-mail: christianrosemd@gmail.com.

AEM EDUCATION AND TRAINING 2019;3:393–395



recommendations from social workers and department of public health administrators as to which they observe to be regular destinations for ED patients. These included a medically assisted sobering center for alcohol intoxication, a primary care wound clinic, a medical respite for homeless patients with ongoing acute medication needs, and a psychiatric urgent care and long-term rehabilitation center (further details and demographics in Table 1). All staff (physicians, nurses, social workers, EMTs, and students) from each training program ED were invited to participate by e-mail.

Mode of transit was determined by program and city-specific features like location, sprawl, or accessibility to roads or public transit. San Francisco is a cycling city—with a high “bike score” (from [www.walkscore.com](http://www.walkscore.com)) based on bike infrastructure, number of bike commuters, and over 82,000 bike trips per day.<sup>5</sup> Many UCSF residents report bicycle use as a primary means of transit or own a bicycle. Participants who did not own a bicycle borrowed one from a colleague or utilized city bike shares like Jump or CitiBike. To ensure optimal safety, all participants were asked to wear helmets and reflective attire, and the ride was scheduled at a nadir between the highest commuter traffic times in the Bay Area at approximately 12:00PM. This time also coincided with the completion of our weekly departmental conference.

A unidirectional route was chosen using Google Maps (Figure 1), which maximized bike lane usage.

The average distance of each ride was approximately 7 miles. Thirty minutes were allotted at each location with a representative there awaiting our arrival. Sites were contacted at least 1 month in advance of the ride and final arrival times coordinated according to the route choice. The site representative at each location gave a brief description of the services they offer, walked the team through the facility and performed a question/answer session upon completion. Finally, trivia based on historical reports of the health care system and found in the EMS rotation background reading materials was performed to further knowledge of San Francisco medical services and history.

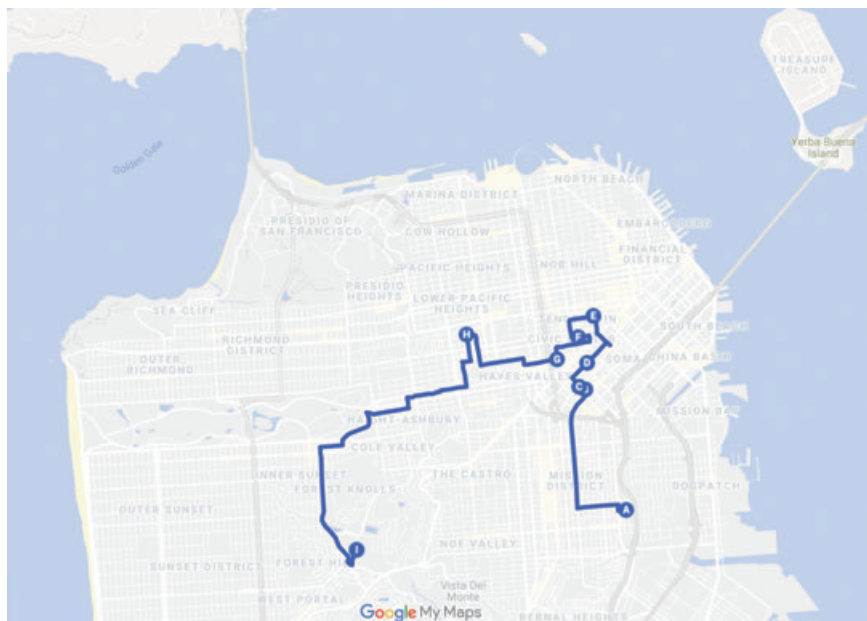
## DESCRIPTION

Participants reported an increased understanding of the capabilities of the facilities visited as well as greater understanding of the context of medical practice within the city structure. We believe that by exposing the health care team to this wider breadth of health care delivery we may improve utilization of these services. This, in turn, may impact ED and hospital crowding but also allow for greater impact on subsequent quality improvement or EMS initiatives by residents and staff.

Furthermore, the ride engendered a sense of camaraderie and community service from the multidisciplinary team of physicians, nurses, social workers, and

**Table 1**  
Public Health Resources Visited Over 3 Years of Bike Tours

Site	Services Offered	Patient Demographic	Setting
Dore Urgent Care Center	Psychiatric urgent care open 24 hours a day providing evaluation and short term crisis stay	All	Outpatient
Integrated Soft Tissue Infection Service (ISIS) Clinic	Acute and chronic wound care services	All	Outpatient
SF Sobering Center	12-bed facility for patients with alcohol intoxication, also providing resources for detoxification	Primarily homeless	Outpatient
Tom Waddell Urban Health Center	Primary care and drop-in services utilizing harm reduction models including office-based opiate treatment, HIV treatment, and dental and podiatry care	Homeless and marginally housed	Outpatient
SF City Clinic	Free to low-cost clinic for sexual health including diagnosis and treatment, HIV prophylaxis, and contraception	All	Outpatient
Joe Healy Medical Detoxification Center	San Francisco's only residential alcohol and opiate detoxification program that can provide medical management	All	Inpatient
Harbor Light Center	A residential 6- to 12-month program for drug and alcohol rehabilitation run by The Salvation Army	All	Inpatient
Laguna Honda Hospital	Publicly funded skilled nursing facility and rehabilitation center that provides specialties including palliative, geriatric, and HIV/AIDS care	All	Inpatient
Medical respite	60-bed facility for postacute patients with low-level medical supervision	Homeless	Inpatient



**Figure 1.** Example bike tour map through San Francisco—beginning at San Francisco General Hospital and ending at Laguna Honda Hospital.

EMTs. Some reported that they can now “call (the facility) directly for help in managing a patient.” Others reported exploring new quality improvement projects based on their exposure to these lateral health services. Several staff members at sites visited voiced appreciation to put “faces with names” of doctors they refer patients to and were happy to orient providers at the start of their EM careers to the services at their site.

Finally, with a portion of the EM department commuting by bicycle, the ride itself allowed for an opportunity to learn safe bikeways and practices, which can be a barrier to entry for new riders. While not all cities are as bikeable or have as moderate a climate as San Francisco, we believe that the use of approximately 4 hours (or the average EM weekly conference time) would be sufficient to arrange visits to core public health facilities by bicycle, public transportation, chartered bus, or even walking tour and might be equally efficacious at improving understanding and utilization of these lateral health services.

No injuries were reported from the six iterations of the event to date and several participants expressed their enthusiasm for an academic activity of the residency program that incorporated a wellness activity. The ride attracted participation from members of the EM faculty and fellowship programs, and has

produced at least two “spin-off” rides for those participating in ED academic retreats. In our experience, multifaceted activities embracing educational, social, and wellness benefits become easy to sustain and potential touchstones to improving organizational culture toward community awareness/service and work satisfaction.

## References

1. Marcozzi D, Carr B, Liferidge A, Baehr N, Browne B. Trends in the contribution of emergency departments to the provision of hospital-associated health care in the USA. *Int J Health Serv* 2018;48:267–88.
2. Weinick RM, Burns RM, Mehrotra A. Many emergency department visits could be managed at urgent care centers and retail clinics. *Health Aff* 2010;29:1630–6.
3. Pope I, Burn H, Ismail SA, Harris T, McCoy D. A qualitative study exploring the factors influencing admission to hospital from the emergency department. *BMJ Open* 2017;7:e011543.
4. Gilbert JH, Yan J, Hoffman SJ. A WHO report: framework for action on interprofessional education and collaborative practice. *J Allied Health* 2010;39(Suppl 1):196–7.
5. Gorsi T. Bicycle Ridership Data. SFMTA. 2017. Available from: <https://www.sfmta.com/bicycle-ridership-data>. Accessed July 27, 2018.

# Embracing FOAM as an Academic Community

Nicole Munz, DO and Esther H. Chen, MD 

We read with particular interest and enthusiasm the article written by Dr. Grock et al., describing his team's strategy to search for, collate, categorize, and assess the quality and content of written free online open-access medical education (FOAM).<sup>1</sup> As educators and consumers of FOAM, we have waited too long for someone to undertake this effort and applaud their collaboration with a national organization dedicated to scientific inquiry. This article highlights some important points about the appropriate use of FOAM and the shared responsibility between the consumer and contributor.

We were most surprised by the authors' assessment that only 0.3% of the written posts were appropriate as a point-of-care resource. In our experience, trainees frequently reference FOAM while on shift, and widespread use has been reported in the literature.<sup>2</sup> As the authors point out, this 0.3% statistic may be due to the nature of the renal and genitourinary topic selected and be different for other topics such as critical care. Moreover, some resources were considered to be too detailed to be useful for on-shift review. However, we can imagine a scenario in which an attending or senior resident, who may have already read a resource in its entirety, uses a specific section to highlight some important points during bedside teaching or is able to search quickly for an infographic to help guide patient management. Our clinical environment has become so technologically prevalent that it is often easier to look something up on the computer than to find a textbook.

Another important but less surprising finding is that only 10% of the posts were determined to be high quality based on the rMETRIQ score. As there is no criterion standard for assessing the quality and accuracy of FOAM, this tool provides a set of objective criteria that expands on prior work to determine quality, such as the quality checklists for blogs and podcasts and the ALiEM-approved institutional resources score.<sup>3,4</sup> One benefit of FOAM is rapid dissemination of and reference to primary literature. While we believe that creators of FOAM aim to present accurate and high-yield information, Dr. Grock's analyses suggest that there is significant room for improvement. This article reminds creators of FOAM that they need to be extra careful and precise in what they write, because consumers may view their content without critically appraising the primary sources themselves.<sup>5</sup> Just as researchers must follow guidelines when writing their manuscripts, we encourage future FOAM authors to use objective criteria, like the rMETRIQ score, to guide them in creating high-quality FOAM content. Furthermore, this article reminds educators, who may also be using FOAM for asynchronous learning, to use objective criteria as a guide for evaluating online content.

Free online open-access medical education can be an incredibly useful tool in our pursuit of lifelong learning as physicians. With a quick online search, we can find easily digestible summaries on almost any topic or recent article relevant to our clinical practice. There will likely be generations of learners who

---

Department of Emergency Medicine, University of California at San Francisco, San Francisco, CA

Received April 1, 2019; accepted April 16, 2019.

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

Related articles appear on pages 375 and 398.

Supervising Editor: John H. Burton, MD.

Address for correspondence and reprints: Esther H. Chen, MD; e-mail: Esther.Chen@ucsf.edu

EDUCATION AND TRAINING 2019;3:396–397

primarily use FOAM for their core learning and have gaps in their medical knowledge.<sup>6</sup> Dr. Grock and colleagues found that the relevant FOAM posts did not sufficiently cover the breadth or depth of most renal and genitourinary topics in the Model of the Clinical Practice of Emergency Medicine but was heavily weighted toward common conditions (e.g., renal calculi, urinary retention, and cystitis).<sup>7</sup> Therefore, FOAM should not be a stand-alone tool for learning core content and cannot replace textbooks or other structured curricula for junior learners, even if it is not their preferred learning resource.<sup>6,8</sup> We encourage learners to read the primary literature referenced in FOAM posts before changing their clinical practice.

Finally, as part of an academic community, we have a responsibility to encourage the use of a formal evaluation process like the SOAR Review to vet online content that will be featured on a national organization's website. SAEM's Online Academic Resources (SOAR) project, which serves as SAEM's home for online education, is a repository of FOAM resources. In this role, how much responsibility should the academy have to ensure this content is accurate and of high quality, like its journals? Is it enough to use the rMETRIQ score to evaluate these written posts? As academic physicians let's use this project as a launching pad for advancing FOAM, by approaching it with the same critical eye as we do scientific literature.

## References

1. Grock A, Bhalerao A, Chan T, et al. Systematic online academic resource (SOAR) review: renal and genitourinary. *AEM Educ Train* 2019;3:000–00.
2. Patocka C, Lin M, Voros J, Chan T. Point-of-care resource use in the emergency department: a developmental model. *AEM Educ Train* 2018;2:221–8.
3. Thoma B, Joshi N, Trueger NS, et al. Five strategies to effectively use online resources in emergency medicine. *Ann Emerg Med* 2014;64:392–5.
4. Lin M, Joshi N, Grock A, et al. Approved Instructional Resources (AIR) Series: a national initiative to identify quality emergency medicine blog and podcast content for resident education. *J Grad Med Educ* 2016;8:219–25.
5. Thoma B, Murray H, Huang SY, et al. The impact of social media promotion with infographics and podcasts in research dissemination and readership. *CJEM* 2018;20:300–6.
6. Mallin M, Schlein S, Doctor S, et al. A Survey of the current utilization of asynchronous education among emergency medicine residents in the US. *Acad Med* 2014;89:598–601.
7. Counselman FL, Babu K, Edens MA, et al. The 2016 Model of the Clinical Practice of Emergency Medicine. *J Emerg Med* 2017;52(6):846–9.
8. Purdy E, Thoma B, Bednarczyk J, et al. The use of free online educational resources by Canadian emergency medicine residents and program directors. *CJEM* 2015;1717:101–6.

# Thinking Critically About Appraising FOAM

Teresa M. Chan, MD, MHPE,<sup>1</sup>  Anuja Bhalerao, MD Candidate,<sup>2</sup> Brent Thoma, MD, MA, MSc,<sup>3</sup>  N. Seth Trueger, MD, MPH,<sup>4</sup>  and Andrew Grock, MD<sup>5</sup>

The dawn of the 21st century has certainly left us with new issues and problems that we have seldom encountered before. “Fake news,” misinformation are rife on social media, and the need for increased criticality in reading anything in our daily lives has never been so important. Critical appraisal is now being taught by elementary school teachers and encouraged all along the educational spectrum; online medical education should not be exempt from these discussions. As such, we thank Drs. Munz and Chen for their commentary<sup>1</sup> on our paper,<sup>2</sup> which highlights many key findings in our study and agrees with us that it is imperative for our discipline to bring critical appraisal to online educational resources.

Indeed, as we continue to develop digital scholarship, we must evaluate the nature and usefulness of the innovations that we create. Most movements begin with an engaged initial set of innovators and early adopters,<sup>3</sup> who seek to create new spaces for themselves by breaking apart from the pack to discover bluer waters.<sup>4</sup>

FOAM has evolved over time in a similar way. Initially created as a reaction to paywalls, a desire to democratize education through the power of Web 2.0 and provide informal, focused educational resources,<sup>5</sup> FOAM has undergone successive waves of evolution to adopt “structural” elements and organizational/

systems thinking. This enhanced collaboration, but at the cost of incorporating a formerly disruptive field into more mainstream influences. As trainees continue to use FOAM resources,<sup>6–8</sup> many educators began to engage with students in that space. Not surprisingly, many skeptics emerged, doubting the veracity of a field where textbooks had formerly dominated with historical precedent but no evidence.<sup>9</sup> Blogs then evolved to these response calls for increasing rigor by adapting traditional structures including peer review,<sup>10,11</sup> editorial boards,<sup>12</sup> and online journal clubs.<sup>13–15</sup>

This is where the METRIQ study collaboration began.<sup>16</sup> After describing the expansion and use of FOAM within emergency medicine,<sup>8,17</sup> it has innovated to quantify the impact<sup>18,19</sup> and quality<sup>19–22</sup> of FOAM. Impact was investigated first by developing an impact metric, the Social Media Index (SMi), which consolidated metrics of online readership and followership. When applied to journals, the correlated with traditional measures of journal impact. Initially, this effort was criticized for focusing on impact and/or popularity rather than resource quality,<sup>23,24</sup> but these objections have faded away with the evolution of this program of research.

Research on quality assessment of FOAM resources was spurred by this and the finding that gestalt assessment of online resources was quite unreliable.<sup>25,26</sup>

From the <sup>1</sup>Department of Medicine, Division of Emergency Medicine, McMaster University, and the McMaster Education Research, Innovation, and Theory Program, Hamilton, Ontario, Canada; the <sup>2</sup>School of Medicine, University of Toronto, Toronto, Ontario, Canada; the <sup>3</sup>Department of Emergency Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; the <sup>4</sup>Department of Emergency Medicine, Northwestern University, Chicago, IL; the <sup>5</sup>Division of Emergency Medicine, Greater Los Angeles VA Healthcare System, and the Department of Emergency Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA.

Received April 1, 2019; accepted April 16, 2019.

AB reports salary support from the Society for Academic Emergency Medicine (SAEM) for her work on this project. TMC hold grants from the Physician Services Incorporated Foundation for work on social media knowledge translation. NST receives salary support from the American Medical Association for his role as Digital Media Editor, *JAMA Network Open*.

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

Related articles appear on pages 375 and 396.

Supervising Editor: Susan B. Promes, MD.

Address for correspondence and reprints: Teresa M. Chan, MD, MHPE; e-mail: teresa.chan@medportal.ca.

AEM EDUCATION AND TRAINING 2019;3:398–402



Efforts were made to develop standardized approaches to improving the quality of FOAM resources. In response to the criticism that these resources were not peer reviewed, novel ways to incorporate expert review into online resources were developed and described.<sup>10,11</sup> The evaluation of FOAM was supported through systematic reviews<sup>27</sup> and Delphi studies<sup>28,29</sup> of quality metrics that were incorporated into quality assessment checklists<sup>22</sup> and scores.<sup>21</sup> Review processes to identify high-quality resources were developed using these tools<sup>20,30</sup> and assessed through the online recruitment of a large number of FOAM users.<sup>31</sup>

As Munz and Chen point out, it is crucial for us be critical of our educational interventions—and FOAM is not exempt from this process. While much of the recent literature around social media has focused on selling scientists and knowledge producers a method to engage with the lay public and front-line clinician colleagues,<sup>32–36</sup> we have also encouraged a healthy skepticism when consuming online education resources. In the age of “fake news,” we must both role model and encourage our trainees to think critically of what they read and what they adopt into their practice.<sup>37,38</sup>

Unfortunately, the decentralized explosion of FOAM resources is a double-edged sword, as consumers of FOAM are faced with a potentially unworkable abundance of options. Some very popular sites have proven themselves by consistently producing high-quality content, but there is no guarantee that popularity or word of mouth is a consistent marker of quality.<sup>39</sup> Similarly, trainees may not have the skills to identify quality resources, nor the ability to recognize their peers’ or own educational needs.<sup>40</sup> While many consumers of FOAM currently rely on (and will likely to continue to rely on) word of mouth and ad hoc recommendations, our prior work has found that these gestalt recommendations may be unreliable.<sup>25,26</sup> The goal of critical appraisal tools is to help square this circle—to provide rating schemes that learners can use to identify quality online educational resources, either by applying the tools themselves or by using others’ ratings. Unfortunately, to date, critical appraisal tools have not been shown to be more reliable than gestalt recommendations.<sup>41</sup> While this is disappointing, substantial work in this area is ongoing and this body of research has led to an increase discussion around how to evaluate online, secondary educational resources.<sup>38</sup> It is imperative that we remain vigilant and encouraging our trainees and teachers to remain skeptical of resources regardless of their origin and triangulate the

findings reported in such resources with the primary literature. For ease of reference, Table 1 summarizes a list of available critical appraisal tools for guiding readers toward high-quality, secondary online educational content.

We appreciate Drs. Munz and Chen and colleagues questions about the implications of SAEM “endorsing” specific FOAM resources. To clarify, SOAR topic reviews are not intended as an endorsement of any specific resource or a definitive declaration of any resource’s accuracy. Rather, the SOAR topic reviews are intended only as a starting point. Though the enormous amount of decentralized FOAM content seems daunting for users or even educators to collate and curate, in fact a finite amount of FOAM resources exist. We found nearly 400 in our first review of renal/genitourinary content, which was initially selected partly due to its fairly small FOAM footprint. We hope that trainees and educators (particularly those who are less immersed in the FOAM community) find the SOAR topic reviews useful as a centralized resource in a potentially overwhelming world of online resources. Further, as more trainees and educators continue to produce FOAM content, we hope that by identifying which areas have been covered, particularly with high-quality resources, producers will tend toward uncovered topics in creating new FOAM content. Structured critical appraisal tools (including the rMETRIQ score<sup>42</sup>) are relatively new and, like clinical decision instruments, should not replace good judgment. Given the abundance of online resources, and the variation in quality among them, we used the rMETRIQ score not as a guarantee of quality but as a filter to help educators and trainees eliminate the “noise” so they can focus their efforts to applying their critical eyes to higher quality posts.

For now, FOAM alone does not provide a comprehensive curriculum. However, we would argue that neither is a comprehensive and rigorously edited textbook.<sup>43</sup> While primary literature currently remains the best method of advancing our field, we must also acknowledge that simply reading the primary literature is a fool’s errand as vast amounts are produced daily. Trainees need to be guided by curated content bringing them toward the leading edges of where they can then read primarily. One potential resource that may help shorten the knowledge translation interval is FOAM. While recent efforts have been made to boil down the essential papers for our specialty,<sup>44</sup> these processes take a daunting amount of resources.

**Table 1**  
Critical Appraisal Tools for FOAM

Appraisal Tool (Citation for Paper)	Intended Audience for Tool	Warnings/Tips for Use
ALiEM AIR score <sup>20</sup>	<ul style="list-style-type: none"> <li>Developed by and for a team of expert evaluators</li> </ul>	<ul style="list-style-type: none"> <li>Evidence supporting its validity and reliability among an expert team who received instruction and feedback on using the score.</li> <li>Little data to support its use among learners or other evaluators.</li> <li>An average score of &gt;27 believed to reflect quality by the AIR team.</li> <li>Use of BEEM score favors posts focusing on newer content as opposed to core content.</li> <li>Unpublished data indicate that users describe difficulties with:               <ul style="list-style-type: none"> <li>Applying the BEEM score</li> <li>Scoring without a description for every anchor.</li> </ul> </li> </ul>
METRIQ <sup>5,8</sup> and rMETRIQ <sup>21,42</sup>	<ul style="list-style-type: none"> <li>Intended to be used by FOAM users</li> <li>Developed through multiple modified Delphi processes</li> </ul>	<ul style="list-style-type: none"> <li>Checklist for specific qualities consistent with quality.</li> <li>Users describe difficulties with:               <ul style="list-style-type: none"> <li>Lack of full anchors for each numerical score</li> <li>Lack of a fully anchored rubric.</li> </ul> </li> </ul>
Social Media index <sup>18,19</sup>	<ul style="list-style-type: none"> <li>FOAM users</li> </ul>	<ul style="list-style-type: none"> <li>Initially a list of the most impactful FOAM sites.</li> <li>Data support a correlation between the sites listed on SMi and quality scoring.</li> </ul>
Colmers et al. checklists <sup>22</sup>	<ul style="list-style-type: none"> <li>Checklist of 19 quality indicators developed from educators, users, a systematic review and multiple modified Delphi processes</li> </ul>	<ul style="list-style-type: none"> <li>Simple to use.</li> <li>Little data exist on what evaluation would differentiate quality from not quality.</li> </ul>

SMi = Social Media Index.

Based on our review of one tiny area of #FOAMed, we found that it lacked a robust curricular approach—but this too is changing. Several groups have begun using modified versions of classic curriculum development techniques (e.g., needs assessments) and adapting them for the online sphere.<sup>45–47</sup> Some groups, like the CRACKcast podcast affiliated with the CanadiEM group,<sup>12</sup> are even mapping their entire podcast to major emergency medicine textbooks. Naturally, a rhizomatic and naturally evolving movement like #FOAMed was never going to start out being programmatic. Instead, just as Wikipedia started with humble beginnings, #FOAMed similarly has emerged as a crowdsourced method by which an egalitarian effort by a worldwide network of educators has spontaneously sprung up to serve the online community. We are generally optimistic that efforts such as ours, which map out the areas of unwritten blog posts or podcasts, can help to guide new authors and eager contributors to expand the offerings that are available online. The incorporation of digital scholarship<sup>48,49</sup> by some institutions like Mayo Clinic<sup>50</sup> is paving the way for FOAM to become a legitimate method by which future generations of clinician educators might carve out their academic niches. When

Wikipedia first started it too was incomplete ... and now, it is a crowdsourced haven for information that can rival any other institutionally support source of knowledge aggregation. After all, when is the last time you cracked open a physical encyclopedia? Digital is here to stay, but it can only fulfill its potential if our academic community members contribute, collaborate, critique, and comment.

We thank SAEM for their support of the original SOAR review project.

## References

1. Munz N, Chen E. Embracing FOAM as an academic community. *AEM Educ Train* 2019;3:000–00.
2. Grock A, Bhalerao A, Chan TM, Thoma B, Trueger NS. Systematic Online Academic Resource (SOAR) review: renal and genitourinary. *AEM Educ Train* 2019;3:000–00.
3. Rogers EM. *Diffusion of Innovations*. 4th ed. New York, NY: The Free Press, A Division of Simon & Schuster Inc., 1995.
4. Kim WC, Mauborgne R. *Blue Ocean Shift: Beyond Competing-Proven Steps to Inspire Confidence and Seize New Growth*. New York: Hachette Books, 2017.

5. Nickson CP, Cadogan MD. Free open access medical education (FOAM) for the emergency physician. *Emerg Med Australas* 2014;26:76–83.
6. Mallin M, Schlein S, Doctor S, Stroud S, Dawson M, Fix M. A survey of the current utilization of asynchronous education among emergency medicine residents in the United States. *Acad Med J Assoc Am Med Coll* 2014;89:598–601.
7. Thurtle N, Banks C, Cox M, Pain T, Furyk J. Free open access medical education resource knowledge and utilisation amongst emergency medicine trainees: a survey in four countries. *African J Emerg Med* 2016;6:12–7.
8. Purdy E, Thoma B, Bednarczyk J, Migneault D, Sherbino J. The use of free online educational resources by Canadian emergency medicine residents and program directors. *CJEM* 2015;1717:101–6.
9. Thoma B, Chan T, Benitez J, Lin M. Educational scholarship in the digital age: a scoping review and analysis of scholarly products. *The Winnower* 2014;1(e141827.77297):1–13.
10. Thoma B, Chan T, Desouza N, Lin M. Implementing peer review at an emergency medicine blog: bridging the gap between educators and clinical experts. *CJEM* 2015;17:188–91.
11. Sidalak D, Purdy E, Luckett-Gatopoulos S, Murray H, Thoma B, Chan TM. Coached peer review: developing the next generation of authors. *Acad Med* 2017;92:201–4.
12. Ting DK, Thoma B, Luckett-Gatopoulos S, et al. CanadiEM: accessing a virtual community of practice to create a Canadian national medical education institution. *AEM Educ Train* 2019;3:86–91.
13. Chan TM, Thoma B, Radecki R, et al. Ten steps for setting up an online journal club. *J Contin Educ Health Prof* 2015;35:1484–54.
14. Lin M, Joshi N, Hayes BD, Chan TM. Accelerating knowledge translation: reflections from the online ALiEM-Annals Global Emergency Medicine Journal Club Experience. *Ann Emerg Med* 2017;69:469–74.
15. Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D. Globalization of continuing professional development by journal clubs via microblogging: a systematic review. *J Med Internet Res* 2015;17:e103.
16. The METRIQ Study Research Agenda. Available at: <https://metriqstudy.org/research-agenda/>. Accessed May 7, 2019.
17. Cadogan M, Thoma B, Chan TM, Lin M. Free Open Access Meducation (FOAM): the rise of emergency medicine and critical care blogs and podcasts (2002-2013). *Emerg Med J* 2014;31:e76–7.
18. Thoma B, Sanders JL, Lin M, Paterson QS, Steeg J, Chan TM. The Social Media Index: measuring the impact of emergency medicine and critical care websites. *West J Emerg Med* 2015;16:242–9.
19. Thoma B, Chan TM, Kapur P, et al. The Social Media Index as an indicator of quality for emergency medicine blogs: a METRIQ study. *Ann Emerg Med* 2018;72:696–702.
20. Chan TM, Grock A, Paddock M, Kulasegaram K, Yarris LM, Lin M. Examining Reliability and Validity of an Online Score (ALiEM AIR) for rating free open access medical education resources. *Ann Emerg Med* 2016;68:729–35.
21. Chan TM, Thoma B, Krishnan K, et al. Derivation of two critical appraisal scores for trainees to evaluate online educational resources: a METRIQ study. *West J Emerg Med* 2016;17:574–84.
22. Colmers IN, Paterson QS, Lin M, Thoma B, Chan TM. The quality checklists for medical education blogs and podcasts. *The Winnower*. 2015. <https://doi.org/10.15200/winn.144720.08769>
23. Weingart S. Why EMCrit is opting out of the social media index. Soundcloud. Available at: <https://soundcloud.com/emcrit/why-emcrit-is-opting-out-of>. Accessed May 7, 2019.
24. Carley S. The Social Media Index (SMi): Can & should we measure #FOAMed? Available at: <http://www.stemlynblog.org/the-social-media-index-smi-is-it-flawed/>. Accessed May 7, 2019.
25. Thoma B, Sebok-Syer SS, Krishnan K, et al. Individual gestalt is unreliable for the evaluation of quality in medical education blogs: a METRIQ study. *Ann Emerg Med* 2017;70:394–401.
26. Krishnan K, Thoma B, Trueger NS, Lin M, Chan TM. Gestalt assessment of online educational resources may not be sufficiently reliable and consistent. *Perspect Med Educ* 2017;6:91–8.
27. Paterson QS, Thoma B, Milne WK, Lin M, Chan TM. A systematic review and qualitative analysis to determine quality indicators for health professions education blogs and podcasts. *J Grad Med Educ* 2015;7:549–54.
28. Thoma B, Chan TM, Paterson QS, Milne WK, Sanders JL, Lin M. Emergency medicine and critical care blogs and podcasts: establishing an international consensus on quality. *Ann Emerg Med* 2015;66:396–402.
29. Lin M, Thoma B, Trueger NS, Ankel F, Sherbino J, Chan T. Quality indicators for blogs and podcasts used in medical education: modified Delphi consensus recommendations by an international cohort of health professions educators. *Postgrad Med J* 2015;91:546–50.
30. Grock A. New AIR Series: ALiEM Approved Instructional Resources. Academic Life in Emergency Medicine. Available at: <https://www.aliem.com/aliem-approved-instructional-resources-air-series/>. Accessed May 7, 2019.
31. Thoma B, Paddock M, Purdy E, et al. Leveraging a virtual community of practice to participate in a survey-based

- study: a description of the METRIQ study methodology. *AEM Educ Train* 2017;2:110–3.
32. Chan T, Trueger NS, Roland D, Thoma B. Evidence-based medicine in the era of social media: scholarly engagement through participation and online interaction. *Can J Emerg Med* 2018;20:3–8.
  33. Chan TM, Stukus D, Leppink J, et al. Social media and the 21st-century scholar: how you can harness social media to amplify your career. *J Am Coll Radiol* 2017;15:142–8.
  34. Bigham BL, Chan TM. Going viral and interacting with the press. *J Am Coll Radiol* 2018;15:153–4.
  35. Thoma B, Murray H, Huang SY, et al. The impact of social media promotion with infographics and podcasts on research dissemination and readership. *Can J Emerg Med* 2018;20:300–6.
  36. Huang S, Martin LJ, Yeh CH, et al. The effect of an infographic promotion on research dissemination and readership: a randomized controlled trial. *Can J Emerg Med* 2018;20:826–33.
  37. Thoma B, Joshi N, Trueger NS, Chan TM, Lin M. Five strategies to effectively use online resources in emergency medicine. *Ann Emerg Med* 2014;64:392–5.
  38. Lo A, Shappell E, Rosenberg H, et al. Four strategies to find, evaluate, and engage with online resources in emergency medicine. *CJEM* 2018;20:293–9.
  39. Tilburt JC, Allyse M, Hafferty FW. The case of Dr. Oz: ethics, evidence, and does professional self-regulation work? *AMA J Ethics* 2017;19:199–206.
  40. Trueger NS. Yelp, but for blogs? *Ann Emerg Med* 2016;68:736–7.
  41. Thoma B, Sebok-Syer SS, Colmers-Gray I, et al. Quality evaluation scores are no more reliable than gestalt in evaluating the quality of emergency medicine blogs: a METRIQ study. *Teach Learn Med* 2018c;30:294–302.
  42. Colmers-Gray I, Krishnan K, Chan TM, Trueger NM, Paddock M, Grock A, Zaver FT. P022: the revised METRIQ Score: an international, social-media based usability analysis of a quality evaluation instrument for medical education blogs. *Can J Emerg Med* 2018;20 (Suppl 1):S64.
  43. Luckett-Gatopoulos S. “Rosen’s on the causes of metabolic alkalosis. . . or did I just have a stroke? #PGWHY5” [Tweet]. Twitter. Available at: <https://twitter.com/SLuckettG/status/1102728960490852353?s=20>. Accessed March 4, 2019.
  44. Bazak SJ, Sherbino J, Upadhye S, Chan T. Ascertaining top evidence in emergency medicine : a modified Delphi study. *CJEM* 2019;21:291–5.
  45. Chan TM, Jo D, Shih AW, et al. The Massive Online Needs Assessment (MONA) to inform the development of an emergency haematology educational blog series. *Perspect Med Educ* 2018;7:219–23.
  46. Tseng EK, Jo D, Shih AW, de Wit K, Chan TM. Window to the unknown : using storytelling to identify learning needs for the intrinsic competencies within an online needs assessment. *AEM Educ Train* 2019;3:179–87.
  47. Khan T, Chan TM. Intrinsic CanMEDS competencies expected of medical students during emergency medicine core rotation: a needs assessment. *Cureus* 2018;10:e3316.
  48. Cabrera D, Vartabedian BS, Spinner RJ, Jordan BL, Aase LA, Timimi FK. More than likes and tweets: creating social media portfolios for academic promotion and tenure. *J Grad Med Educ* 2017;9:421–5.
  49. Sherbino J, Arora VM, Van Melle E, Rogers R, Frank JR, Holmboe ES. Criteria for social media-based scholarship in health professions education. *Postgrad Med J* 2015;91:551–5.
  50. Cabrera D. Mayo Clinic includes Social Media Scholarship Activities in Academic Advancement. *MCSMN Blog*. 2016. Available at: <https://socialmedia.mayoclinic.org/discussion/mayo-clinic-includes-social-media-scholarship-activities-in-academic-advancement/>. Accessed June 5, 2016.

# Pupil Prose Appraisal: Four Practical Solutions to Medical Student Documentation and Feedback in the Emergency Department

Mark F. Olaf, DO, FACEP 

## ABSTRACT

Documentation is part of a critical foundation of skills in the undergraduate medical education curriculum. New compliance rules from the Centers for Medicare and Medicaid Services will impact student documentation practices. Common barriers to student documentation include limited access to the electronic medical record, variable clerkship documentation expectations, variable advice regarding utilizing the electronic medical record, and limited time for feedback delivery. Potential solutions to these barriers are suggested to foster documentation skill development. Recommendations are also given to mitigate compliance and legal risk.

Documentation in the patient chart is a critical foundational skill that must be developed in undergraduate medical education. However, despite encouragement from the Association of American Medical Colleges (AAMC), the Liaison Committee for Medical Education, and other specialty groups,<sup>1-6</sup> a significant proportion of medical students do not receive adequate instruction and feedback to develop written communication skills.<sup>7,8</sup> This important skill development is highlighted by recent Centers for Medicare and Medicaid Services (CMS) standards, which allow teaching physicians to use student documentation for billable services. Emergency department directors and AAMC representatives agree that student documentation trains students to be concise and efficient, helps students learn detailed patient evaluation, improves decision-making skills, and helps students feel involved in patient care. The AAMC and ED directors endorse use of the EMR by students to enter and retrieve information.<sup>1,9</sup>

Several barriers limit the development of documentation skills. Institutional barriers include medical legal liability and limiting student access to the EMR.<sup>10,11</sup> Student challenges include variable instruction in navigating the EMR and lack of feedback specific to documentation.<sup>5,11</sup> Preceptors have limited time to review lengthy student notes and may not fully understand the expectations for student documentation to give effective feedback.<sup>9-11</sup>

## PRACTICAL SOLUTIONS TO CHALLENGES

### Challenge 1: Institutional Barriers to Student Access

Institutional policy limiting student access to the EMR is the greatest barrier to student documentation.<sup>12</sup> The AAMC advocates for the ability of students to fully interact with the EMR<sup>1</sup> including updating problems lists, documenting patient care into the note, and the appropriate use of templates.<sup>4</sup> Advocates for change

From the Geisinger Commonwealth School of Medicine, Geisinger Health, Danville, PA.

Received April 24, 2019; revision received July 22, 2019; accepted July 31, 2019.

Presented at Council of Emergency Medicine Residency Directors Academic Assembly, Seattle, WA, March 2019.

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

Supervising Editor: Esther H. Chen, MD.

Address for correspondence and reprints: Mark F. Olaf, DO, FACEP; e-mail: mfolaf@geisinger.edu.

AEM EDUCATION AND TRAINING 2019;3:403-407.



have offered specific steps to assist institutions in integrating the EMR into education.<sup>12</sup> In a study of medical school deans, over ninety percent believed student notes belong in medical records and indicated that without student notes, student education would be negatively affected. Most of the deans indicated that limiting students' notes would negatively affect several other issues, including preparation for internship and students' sense of involvement.<sup>13</sup> Additional barriers included concern for medical liability, an inability for student notes to support medical billing, lack of computer workspace, and inability for the clerkship to review notes.<sup>11</sup>

Legal liability is cited as a barrier to student note writing,<sup>11</sup> although Gliatto et al.<sup>14</sup> noted a paucity of information in the available literature, including no report of a medical student note being used in, or leading to, a legal case. In addition, medical students are trainees and are therefore not considered experts in the legal arena and their notes should not routinely be admissible or impactful in court.<sup>12,15</sup> With the new CMS allowance of student documentation for billable services,<sup>8</sup> the legal liability of student notes becomes less clear and may necessitate a novel approach to mitigating this new frontier.

### Challenge 2: Documentation Goals and Objectives for Clerkships

Emerging data seem to show that EMR use is a valuable learning tool and exercise.<sup>16</sup> Student note writing and feedback are associated with students' perceptions of high-quality teaching<sup>17</sup> and having students participate in the documentation of the patient care that they witness and engage in is a sound educational activity.<sup>18–20</sup> Students value the use of the EMR, which seems to generate reflection and insight into history taking and data synthesis development.<sup>21</sup> Examples of insight include enhanced understanding of recommended preventive practices and anticipatory guidance.<sup>22</sup> In some respects, EMR use may be viewed to inhibit education, but steps can be taken to mitigate these effects.<sup>23</sup>

Given the educational benefits, clearly defined documentation expectations should be set and should include components of quality, accuracy and completeness.<sup>5</sup> These goals and objectives should be based on the assessment of the development of documentation competency<sup>24</sup> and be guided by the medical student milestones in emergency medicine<sup>24</sup> and the Emergency Medicine Milestones.<sup>25</sup> Expectations will differ

for learners in different levels of training with consistent attention to appropriate content and communication components and variable emphasis on efficiency and billing compliance.<sup>5</sup>

The specific objectives for note writing should be particular, but allow for some interpretation.<sup>5,9,24</sup> A reasonable checklist for the evaluation of notes may include assessment of:<sup>5,9,24</sup>

1. Appropriate note components.
2. Documentation of pertinent positives and negatives and sufficient context.
3. Clarity of communication, including treatment rationale.
4. Appropriate use of automated EMR functions, including avoiding inappropriate abbreviations, copy-and-paste functions, and incorrect or inappropriate fields.

The Reporter–Interpreter–Manager–Educator (RIME) model has been proposed as a tool to contextualize, teach, and evaluate EMR documentation and may assist preceptors in evaluating learners.<sup>5,9,24,26,27</sup>

### Challenge 3: Managing the Pitfalls of Electronic Support

Students are at risk of using the EMR inappropriately and benefit from a graduated use of the EMR.<sup>5</sup> While note templates ease the burden of documentation on the experienced clinician, they can become a crutch for students who are not familiar with documentation practices. Even so, students may reap patient care skills improvement from use of EMR templates<sup>28</sup> including improvements in medical student objective structured clinical examination scores with template use.<sup>29</sup> Without the use of an EMR, most medical student notes lack complete or appropriate details with regard to a clinical encounter of a patient with chest pain<sup>30</sup> and would be down-coded with respect to reimbursement guidelines.<sup>31</sup> Experience with EMRs is necessary to develop appropriate efficiency and skills,<sup>32</sup> and limiting students' ability to document has been shown to have consequences on medical training.<sup>33</sup> Students should be allowed to use a template to develop familiarity, but not be confined by it, and should be encouraged to provide thoughtful, detailed, contextualized information, regardless of the template prompts.<sup>5,9</sup> Students should attempt to navigate the template and correct those areas of the note which are inappropriately populated by the EMR.<sup>5,34,35</sup> An evaluation of the student's ability to interact with the

EMR, including avoidance of inappropriate EMR uses (like copy and paste functions), are necessary for students to learn to overcome the pitfalls of EMR use.<sup>9,24</sup>

#### Challenge 4: Expectations for Assessment and Feedback Delivery

The clerkship director should develop an appropriate mechanism for evaluation and feedback that suits students' needs.<sup>5</sup> Verbal, written, or electronic delivery of the feedback are all viable mechanisms.<sup>9</sup>

To facilitate feedback, a reasonable and manageable number of notes should be written during the clerkship.<sup>5</sup> In one study, half or fewer of student notes were reviewed in 70% of cases,<sup>11</sup> which is concerning if poor behaviors go unaddressed. Another study showed improved quality of feedback and note quality when an electronic medical record was used to as a means to evaluate notes every three to four shifts.<sup>36</sup>

Competencies for each objective should be developed, with appropriate anchors so that an evaluator may maintain consistency.<sup>4,9</sup> The RIME model can thus become the basis of the rubric to assess a student note and can be useful in developing anchors for the rubric.<sup>5,33</sup> Electronic documentation can demonstrate insight into medical student clinical skills and may even augment clinical reasoning, leading to more robust feedback.<sup>5,20,24</sup> Feedback should be delivered in an iterative manner that allows the preceptor to assess a student's improvement over the course of time.<sup>37</sup> The audit and feedback methodology<sup>38</sup> and a Web-based tool to perform self and peer evaluation of notes have both been efficacious.<sup>39</sup>

#### Medical Student Scribing and Compliance: Words of Caution

The AAMC importantly acknowledges that a scribe may be a medical student, but a student may not be both a learner and a scribe at the same time.<sup>1</sup> Medical students are primarily learners and are not granted independent licenses and therefore their notes should not be routine parts of the billable medical record, although this may change with evolving insurance regulations. Each individual health care system must assess the capabilities of the EMR, the role of the student, and the associated compliance risks, especially in light of new billing regulations. Appropriate attention to these components will help to reduce and mitigate risks.<sup>1</sup>

## CONCLUSIONS

Student documentation is recognized as a critical clinical skill to teach and develop. Recommendations may help encourage systems and preceptors to allow students opportunities document and receive feedback. Aligning clear expectations with evaluation processes will help the student and the preceptor to efficiently and effectively develop this crucial skill.

## References

1. AAMC Compliance Advisory: Electronic Health Records (EHRs) in Academic Health Centers. 2014. Available at: [https://www.aamc.org/download/316610/data/advisory3ac\\_hallengefortheelectronichealthrecordsofacademicinsti.pdf](https://www.aamc.org/download/316610/data/advisory3ac_hallengefortheelectronichealthrecordsofacademicinsti.pdf). Accessed June 27, 2019.
2. Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the MD Degree. Publication of the Liaison Committee for Medical Education. Available at: <http://lcme.org/publications/>. Accessed June 27, 2019.
3. Association of American Medical Colleges. Learning objectives for medical student education—guidelines for medical schools: Report I of the Medical School Objectives Project. *Acad Med* 1999;74:461–2.
4. Carter TJ, Drusin R, Moeller J, Obeso V, Brown D, Philippi C, editors, for Core EPAs for Entering Residency Pilot Program. Adapted from the Association of American Medical Colleges (AAMC). Core entrustable professional activities for entering residency. 2014. Available at: <https://www.aamc.org/download/484778/data/epa13toolkit.pdf>. Accessed on June 27, 2017.
5. Hammound MM, Dalymple JL, Christner JG, et al. Medical student documentation in electronic health records: a collaborative statement from the Alliance for Clinical Education. *Teach Learn Med* 2012;24:257–66.
6. Manthey DE, Ander DS, Gordon DC, et al. Emergency medicine clerkship curriculum: an update and revision. *Acad Emerg Med* 2010;17:638–43.
7. United States 111th Congress. Health Information Technology for Economic and Clinical Health Act, an Act within the American Recovery and Reinvestment Act of 2009. Public Law 111-5. Page 123 STAT. 115. Available at: <https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf>. Accessed June 27, 2019.
8. The American Hospital Association. E/M Service Documentation Provided by Students (Manual Update). CR 10412, MM Article # 10412. Transmittal #R4068CP. Available at: <https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNMattersArticles/Downloads/MM10412.pdf>. Accessed June 27, 2019.

9. Hammound MM, Margo K, Christner JG, Fisher J, Fischer SH, Pangaro LN. Opportunities and challenges in integrating electronic health records into undergraduate medical education: a national survey of clerkship directors. *Teach Learn Med* 2012;24:19–224.
10. Wittels K, Wallenstein J, Patwari R, Patel S. Medical student documentation in the electronic medical record: patterns of use and barriers. *West J Emerg Med* 2016;18:133–6.
11. Virden RA, Sonnett FM, Khan AN. Medical student documentation in the emergency department in the electronic health record era: a national survey. *Pediatr Emerg Care* 2019;35:220–5.
12. Pageler NM, Friedman CP, Longhurst CA. Refocusing medical education in the EMR era. *JAMA* 2013;310:2249–50.
13. Friedman E, Sainte M, Fallar R. Taking note of the perceived value and impact of medical student chart documentation on education and patient care. *Acad Med* 2010;85:1440–4.
14. Gliatto P, Masters P, Karani R. Medical student documentation in the medical record: a liability? *Mt Sinai J Med* 2009;76:357–64.
15. Committee on the Development of the Third Edition of the Reference Manual on Scientific Evidence. Committee on Science, Technology, and Law Policy and Global Affairs. Federal Judicial Center. Reference Manual on Scientific Evidence, Third Edition. National Research Council of the National Academies. Washington, DC: The National Academies Press, 2011.
16. Tierney MJ1, Pageler NM, Kahana M, Pantaleoni JL, Longhurst CA. Medical education in the electronic medical record (EMR) era: benefits, challenges, and future directions. *Acad Med* 2013;88:748–52.
17. Torre DM, Simpson D, Sebastian JL, Elnicki DM. Learning/feedback activities and high-quality teaching perceptions of third-year medical students during an inpatient rotation. *Acad Med* 2005;80:950–4.
18. Mann KV. The role of educational theory in continuing medical education: has it helped us? *J Contin Educ Health Prof* 2004;24 Suppl 1:S22–30.
19. Chi J, Kugler J, Chu IM, et al. Medical students and the electronic health record: ‘an epic use of time’. *Am J Med* 2014;127:891–5.
20. Keenan CR, Nguyen HH, Srinivasan M. Electronic medical records and their impact on resident and medical student education. *Acad Psychiatry* 2006;30:522–7.
21. Rouf E, Chumley HS, Dobbie AE. Electronic health records in outpatient clinics: perspectives of third year medical students. *BMC Med Educ* 2008;8:1–7.
22. O’Connell RT, Cho C, Shah N. Take note(s): differential EHR satisfaction with two implementations under one roof. *J Am Med Inform* 2004;11:43–9.
23. Schenarts P, Schenarts K. Educational impact of the electronic medical record. *J Surg Educ* 2012;69:105–12.
24. Santen S, Peterson W, Khandelwal S, House J, Manthey D, Sozener C. Medical student milestones in emergency medicine. *Acad Emerg Med* 2014;21:905–11.
25. Beeson MS, Carter WA, Christopher TA, et al. Emergency medicine milestones. *J Grad Med Educ* 2013;5(1 Suppl 1):5–13.
26. Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med* 1999;74:1203–7.
27. Stephens M, Gimbel R, Pangaro L. The RIME/EHR scheme: an education approach to clinical documentation in electronic medical records. *Acad Med* 2011;86:11–4.
28. Frenzel J. Using electronic medical records to teach patient-centered care. *Am J Pharm Educ* 2010;74:71.
29. Gaeta T, Pyle M. Documentation templates for clinical encounters improves medical student performance of USMLE Step 2 CS essential actions on an objective structured clinical exam. *Ann Emerg Med* 2013;62:S120.
30. Wald D. Written documentation of the chest pain history by fourth-year medical students using a simulated emergency department patient encounter. *Acad Emerg Med* 2018;11:500.
31. Hoonpongsimanont W, Velarde I, Gilani C, Louthan M, Lotfipour S. Assessing medical student documentation using simulated charts in emergency medicine. *BMC Med Educ* 2018;18:203.
32. Crawford S, Kushner I, Wells R, Monks S. Electronic health record documentation times among emergency medicine trainees. *Perspect Health Inf Manag* 2019;16:1f.
33. Hammound MM, Margo K, Christner JG, Fisher J, Fischer SH, Pangaro LN. Opportunities and challenges in integrating electronic health records into undergraduate medical education: a national survey of clerkship directors. *Teach Learn Med* 2012;24:19–224.
34. Dye D. Enhancing critical reflection of students during a clinical internship using the self-S.O.A.P. note. *Internet J Allied Health Sci Pract* 2005;3:4.
35. Mintz M, Narvarte HJ, O’Brien KE, Papp KK, Thomas M, Durning SJ. Use of electronic medical records by physicians and students in academic internal medicine settings. *Acad Med* 2009;84:1698–704.
36. Spickard A, Gigante J, Stein G, Denny JC. Automatic capture of student notes to augment mentor feedback and student performance on patient write-ups. *J Gen Intern Med* 2008;23:979–84.
37. DeLeo S, Mothner B, Middleman A. Improving student documentation using a feedback tool. *Clin Teach* 2018;15:48–51.

38. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006;(2):CD000259.
39. McCarty R, Parkes MV, Anderson TT, Mines J, Skipper BJ, Grebosky J. Improved patient notes from medical students during web-based teaching using faculty-calibrated peer review and self-assessment. *Acad Med* 2005;80:567–70.

