EDITOR'S PICK

Increased Sensitivity of Focused Cardiac Ultrasound for Pulmonary Embolism in Emergency Department Patients With Abnormal Vital Signs
James I. Daley, Kristin H. Dwyer, Zachary Grunwald et al.

ORIGINAL CONTRIBUTIONS

Freestanding Emergency Department Entry and Market-level Spending on Emergency Care
Vivian Ho, Yingying Xu, Murtaza Akhter

Early Screening for Posttraumatic Stress Disorder and Depression Among Injured Emergency Department Patients: A Feasibility Study
Stephany Jaramillo, Brian Suffoletto, Clifton Callaway et al.

Emergency Medicine in the #MeToo Era
Catherine A. Marco, Joel M. Geiderman, Raquel M. Schears et al.

Integrated Use of Conventional Chest Radiography Cannot Rule Out Acute Aortic Syndromes in Emergency Department Patients at Low Clinical Probability
Peiman Nazerian, Emanuele Pivetta, Simona Veglia et al.

Exploring Gender Bias in Nursing Evaluations of Emergency Medicine Residents
Krista Brucker, Nash Whitaker, Zachary S. Morgan et al.
SYSTEMATIC REVIEWS (WITH OR WITHOUT META-ANALYSES)

Avoiding Misdiagnosis in Patients With Posterior Circulation Ischemia: A Narrative Review
Kiersten L. Gurley, Jonathan A. Edlow 1273

RESEARCH LETTER

Unexpected Deaths Among Patients With Normal Initial Vital Signs
Brandon W. Yan, Renee Y. Hsia 1285

HOT OFF THE PRESS

Hot Off the Press: Prospective Validation of a Checklist to Predict Short-term Death in Older Patients After ED Admission in Australia and Ireland
Justin Morgenstern, Corey Heitz, Chris Bond et al. 1289

THE BRASS TACKS: CONCISE REVIEWS OF PUBLISHED EVIDENCE

Tranexamic Acid for the Treatment of Epistaxis
Michael Gottlieb, Alex Koyfman, Brit Long 1292

Factors Predicting Difficult Endotracheal Intubation
Brit Long, Alex Koyfman, Michael Gottlieb. 1294

COMMENTARY—INVITED

Maybe It's Time to Rethink Freestanding Emergency Departments
Jesse M. Pines 1297

COMMENTARY—UNSOLICITED

Conflict of Interest Blind Spots in Emergency Medicine: Ideological and Financial Conflicts of Interest in the Gender Bias Literature
Chase M. Donaldson 1300

CORRESPONDENCE—RESPONSE TO LETTERS TO THE EDITOR

In Reply:
Christopher L. Bennett, Ali S. Raja, Dara Kass et al. 1303

REFLECTIONS

Providing Care or Caring
Jonathan T. Miller 1304
CME Information: Increased Sensitivity of Focused Cardiac Ultrasound for Pulmonary Embolism in Emergency Department Patients With Abnormal Vital Signs

CME Editor: Corey Heitz, MD

Authors: James I. Daley, MD, MS, MPH, Kristin H. Dwyer, MD, Zachary Grunwald, MD, Daniel L. Shaw, MD, Michael B. Stone, MD, Alexandra Schick, MD, Michael Vrablik, DO, M. Kennedy Hall, MD, MS, Jane Hall, PhD, Andrew S. Liteplo, MD, Rachel M. Haney, MD, Nancy Hun, MD, Rachel Liu, MD, and Chris L. Moore, MD

If you wish to receive credit for this activity, please refer to the website: www.wileyhealthlearning.com/aem

Educational Objectives
After reading the article, participants should be able to discuss the sensitivity of focused cardiac ultrasound (FOCUS) for PE in ED patients with abnormal vital signs.

Activity Disclosures
This activity received no commercial support.

CME Editor Corey Heitz discloses no relevant financial relationships.

This activity underwent peer review in line with standards of editorial integrity and publication ethics. Conflicts of interest have been identified and resolved in accordance with John Wiley and Sons, Inc.’s Policy on Activity Disclosure and Conflict of Interest.

Accreditation
John Wiley and Sons, Inc. designates this journal-based CME activity for a maximum of 1.0 AMA PRA Category 1 Credit™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

For information on applicability and acceptance of continuing medical education credit for this activity, please consult your professional licensing board.

This activity is designed to be completed within 1 hour. To successfully earn credit, participants must complete the activity during the valid credit period, which is up to two years from initial publication. Additionally, up to 3 attempts and a score of 70% or better is needed to pass the post test.
Increased Sensitivity of Focused Cardiac Ultrasound for Pulmonary Embolism in Emergency Department Patients With Abnormal Vital Signs

James I. Daley, MD, MS, MPH¹, Kristin H. Dwyer, MD², Zachary Grunwald, MD¹, Daniel L. Shaw, MD¹, Michael B. Stone, MD⁵, Alexandra Schick, MD², Michael Vrablik, DO³, M. Kennedy Hall, MD, MS³, Jane Hall, PhD⁶, Andrew S. Liteplo, MD⁴, Rachel M. Haney, MD⁴, Nancy Hun, MD¹, Rachel Liu, MD¹, and Chris L. Moore, MD¹

ABSTRACT

Background: Focused cardiac ultrasound (FOCUS) is insensitive for pulmonary embolism (PE). Theoretically, when a clot is large enough to cause vital sign abnormalities, it is more likely to show signs of right ventricular dysfunction on FOCUS, although this has not been well quantified. A rapid bedside test that could quickly and reliably exclude PE in patients with abnormal vital signs could be of high utility in emergency department (ED) patients. We hypothesized that in patients with tachycardia or hypotension, the sensitivity of FOCUS for PE would increase substantially.

Methods: We performed a prospective observational multicenter cohort study involving a convenience sample of patients from six urban academic EDs. Patients suspected to have PE with tachycardia (heart rate [HR] ≥ 100 beats/min) or hypotension (systolic blood pressure [sBP] < 90 mm Hg) underwent FOCUS before computed tomography angiography (CTA). FOCUS included assessment for right ventricular dilation, McConnell’s sign, septal flattening, tricuspid regurgitation, and tricuspid annular plane systolic excursion. If any of these were abnormal, FOCUS was considered positive, while if all were normal, FOCUS was considered negative. We a priori planned a subgroup analysis of all patients with a HR ≥ 110 beats/min (regardless of their sBP). We then determined the diagnostic test characteristics of FOCUS for PE in the entire patient population and in the predefined subgroup, based on CTA as the criterion standard. Inter-rater reliability of FOCUS was determined by blinded review of images by an emergency physician with fellowship training in ultrasound.

Results: A total of 143 subjects were assessed for enrollment and 136 were enrolled; four were excluded because they were non-English-speaking and three because of inability to obtain any FOCUS windows. The mean (±SD) age of enrolled subjects was 56 (±7) years, mean (±SD) HR was 114 (±12) beats/min, and 37 (27.2%) subjects were diagnosed with PE on CTA. In all subjects, FOCUS was 92% (95% confidence interval [CI] = 78% to 98%) sensitive and 64% specific (95% CI = 53% to 73%) for PE. In the subgroup of 98 subjects with a HR ≥ 110 beats/min, FOCUS was 100% sensitive (95% CI = 88% to 100%) and 63% specific (95% CI = 51% to 74%) for PE. There was substantial interobserver agreement for FOCUS (κ = 1.0, 95% CI = 0.31 to 1.0).

From the ¹Department of Emergency Medicine, Yale School of Medicine, New Haven, CT; the ²Department of Emergency Medicine, Warren Alpert Medical School of Brown University, Providence, RI; the ³Department of Emergency Medicine, University of Washington School of Medicine, Seattle, WA; the ⁴Department of Emergency Medicine, The Massachusetts General Hospital, and the ⁵Department of Emergency Medicine, The Brigham and Women’s Hospital, Harvard University School of Medicine, Boston, MA; and ⁶Data Science and Statistics, Jane Hall Biomed, Seattle, WA.

Received March 5, 2019; revision received April 15, 2019; accepted April 25, 2019.

This study was supported by the Society for Academic Emergency Medicine Academy of Emergency Ultrasound Research Grant. The authors have no relevant financial information or potential conflicts to disclose.

Supervising Editor: Michael S. Runyon, MD.

Address for correspondence and reprints: James Daleym, MD, MS, MPH; e-mail: james.i.daley@yale.edu.

ACADEMIC EMERGENCY MEDICINE 2019;26:1212–1220.

doi: 10.1111/acem.13774
Conclusions: A negative FOCUS examination may significantly lower the likelihood of the diagnosis of PE in most patients who are suspected of PE and have abnormal vital signs. This was especially true in those patients with a HR $\geq 110$ beats/min. Our results suggest that FOCUS can be an important tool in the initial evaluation of ED patients with suspected PE and abnormal vital signs.

The evaluation of a patient with chest pain or dyspnea in the emergency department (ED) often prompts the emergency physician (EP) to consider the diagnosis of pulmonary embolism (PE). A rapid bedside test that could quickly and reliably exclude PE in patients with tachycardia or hypotension could be of high utility in the management of these ED patients. Focused cardiac ultrasound (FOCUS) can be used to assess for signs of right ventricular dysfunction (RVD) due to PE. The routine application of FOCUS in all patients suspected of PE is not recommended as FOCUS is relatively insensitive for PE. A recent systematic review and meta-analysis concluded that the overall sensitivity of FOCUS for PE is 53% (95% confidence interval [CI] = 45% to 61%). However, ED patients with tachycardia who have a PE are more likely to show signs of RVD on FOCUS. RVD occurs in between 30% and 70% of patients with PE, and the absence of RVD in patients with hemodynamic compromise makes PE an unlikely etiology. EP-performed FOCUS has been shown to be effective in detecting findings of RVD. Typical findings of RVD on FOCUS include right ventricular dilation, the presence of McConnell’s sign, septal flattening, and tricuspid regurgitation (with regurgitant jet velocity on continuous-wave Doppler $> 2.6$ m/sec).

Recent literature has demonstrated that a less commonly utilized measure of RVD known as tricuspid annular plane systolic excursion (TAPSE) is more sensitive for the diagnosis of PE than other signs, is relatively easy to perform, and may be more reproducible as an objective measurement. TAPSE assesses for RVD by using M-mode to measure the dynamic movement of the tricuspid valve annulus over the course of a contraction (Figure 1). Any value over 1.7 cm is typically considered normal, while any value below 1.7 cm is considered indicative of RVD. TAPSE correlates well with other modalities that measure RVD and has been shown to have high degrees of inter-rater reliability among cardiologists and EPs with experience in FOCUS.

Our group recently described the diagnostic utility of TAPSE for PE by EPs. With a testing threshold of 2.0 cm (compared to the 1.7 cm threshold most commonly found in the literature), TAPSE was 72% (95% CI = 38% to 74%) sensitive for PE. However, we conducted a post hoc subgroup analysis of 17 patients that presented to the ED with tachycardia (heart rate [HR] $\geq 100$ beats/min) and/or hypotension (systolic blood pressure [sBP] $< 90$ mm Hg) and found that FOCUS was 100% (95% CI = 80% to 100%) sensitive for PE and the sensitivity of TAPSE for PE was 94% (95% CI = 71% to 99%) in this group. If FOCUS is shown to be sensitive in this population, it may allow for the exclusion of PE in patients who are too unstable to leave the ED for definitive imaging, who are in a resource-limited practice environment without access to computed tomography angiography (CTA), or who have a contraindication to CTA such as contrast allergy or acute kidney injury.

The objectives of the current study were to determine the diagnostic test characteristics of FOCUS and its components for PE in patients with tachycardia and/or hypotension. We hypothesized that in patients with a HR $\geq 100$ beats/min or sBP $< 90$ mm Hg, the sensitivity of FOCUS would be over 90%. Additionally, we planned a subgroup analysis of patients with a HR $\geq 110$ beats/min a priori and hypothesized...
that the sensitivity of FOCUS would be over 95% in this subgroup.

**METHODS**

**Study Design**
We performed a prospective observational multicenter cohort study involving a convenience sample of ED patients undergoing evaluation for suspected PE with CTA who underwent FOCUS. Enrollment took place from April 2016 to November 2018. This study was approved by the institutional review board (IRB) and conducted in accordance with the STARD guidelines for reporting diagnostic accuracy studies.20

**Study Setting and Population**
Subjects were recruited from six urban academic medical centers with annual ED visits ranging from 60,000 to 120,000 annually. Subjects were enrolled when an EP or study investigator trained in obtaining FOCUS was available. Subjects were eligible for enrollment if they were adults (18 years of age or older) with tachycardia and/or hypotension undergoing CTA for evaluation of possible PE in the ED. Investigators initially identified potential subjects by scanning the ED track board for patients undergoing CTA with abnormal vital signs to evaluate for PE. Investigators then confirmed the abnormal vital signs immediately prior to the FOCUS examination at the bedside. If a subject was enrolled due to hypotension alone, investigators measured their blood pressure twice and the subject was only enrolled if both measurements were believed to be reliable and were below 90 mm Hg. Prisoners, wards of the state, non–English-speaking patients, and those where investigators could not obtain any echocardiographic data due to technical challenges were excluded.

**Study Protocol**
Study personnel included seven ultrasound fellowship–trained attending EPs, three emergency medicine resident physicians, and three medical students. Subjects were primarily enrolled when study personnel were working clinically in the ED. At the primary study site, three medical students who were trained in the acquisition of FOCUS worked on a part-time basis in the ED to enroll subjects by screening the ED track board for patients who were undergoing CTA for suspected PE. When a potential subject was identified, subjects were then assessed for enrollment eligibility, and if deemed eligible, written consent was obtained. Vital signs were assessed by the investigator at the time of the FOCUS examination. Three patients were unable to provide consent at the time of enrollment due to the severity of their illness and provided consent later in their hospital stay, as permitted by the IRB.

All personnel received standardized training that consisted of a brief video and a 1-hour didactic meeting to ensure that standardized images were being obtained. Two of the resident physicians underwent an additional didactic session conducted by an ultrasound fellowship–trained EP. Residents later performed supervised practice examinations until the ultrasound fellowship–trained EP was satisfied that they could reliably perform all the components of FOCUS prior to enrolling patients in the study. These residents were PGY-3 in emergency medicine and had prior ultrasound experience consistent with their level of residency training. The third resident, the primary author, already had significant experience in FOCUS (including TAPSE) and did not undergo additional training for study purposes.2 The three medical students underwent a 1-hour didactic and 1-hour hands-on training session by the primary author. Medical students were in their third year of medical school and did not have significant experience in bedside ultrasound prior to becoming involved in this study. Each student then completed 20 FOCUS examinations with feedback under the supervision of the primary author prior to enrolling patients in the study.

Investigators performed and interpreted FOCUS at the bedside during the subject’s ED stay, if possible, prior to the patient undergoing CTA. If performed after CTA, echocardiographers were blinded to CTA results. Investigators conducted the FOCUS examination using four echocardiographic windows: the parasternal long, parasternal short, apical four chamber, and subxiphoid. Components of FOCUS included the measurement of TAPSE and evaluation for other measures of RVD, defined as right ventricular enlargement (visual appearance of the right ventricle being equal to or greater in size than the left ventricle), septal flattening (flattening of the interventricular septum typically seen on parasternal short axis, sometimes referred to as the “D-sign”), tricuspid regurgitation, and McConnell’s sign (hypokinesis of the right ventricle with apical sparing). While some degree of tricuspid regurgitation can be normal (if measured on continuous wave Doppler at less than 2.6 cm/sec), for the purposes of this study and to simplify FOCUS acquisition for novice users, any presence of tricuspid
regurgitation viewed on color doppler was considered abnormal. The normal testing threshold for TAPSE when used for the prognosis of disease was 1.7 cm. This study defined an abnormal TAPSE as <2.0 cm because prior research by the primary author that employed receiver operating characteristic curve analysis demonstrated that a higher testing threshold of 2.0 cm provided increased sensitivity for PE, while maintaining moderate specificity. TAPSE was obtained in the apical four-chamber view by placing the M-mode cursor along the lateral tricuspid valve annulus and measuring the change in height of the resultant tracing from trough to peak (Figure 1).  

If any of the components of FOCUS were abnormal, then the FOCUS examination was considered positive, while if all components of FOCUS were normal, the FOCUS examination was considered negative. The criterion standard for diagnosing PE was the presence of a filling defect on CTA consistent with PE as reported by radiology. Radiologists were not aware of FOCUS results at any time. All patients who were enrolled underwent CTA in the ED.

Inter-rater reliability of right ventricular enlargement, septal flattening, tricuspid regurgitation, and McConnell’s sign were determined by blinded review of images by the site principal investigator in 104 of 136 patients. If there was a disagreement between raters, the interpretation initially made at the bedside was used for statistical analysis to maintain study generalizability. Inter-rater reliability of TAPSE could not be conducted in this manner because TAPSE must be measured at the bedside. In a subset of eight patients, two investigators performed FOCUS on the same patient to determine inter-rater reliability of the entire FOCUS examination and TAPSE. They were blinded to each other’s results. These patients were selected based on convenience; if two study investigators were present at the time of enrollment, then each performed FOCUS. Selection was not related to image quality, the clinical condition of the patient, or any other criteria.

**Measures**

The primary outcome measure was the sensitivity of FOCUS for PE in both predescribed patient populations: 1) those with a HR ≥ 100 beats/min or sBP < 90 mm Hg (n = 136) and 2) those with a HR ≥ 110 beats/min (n = 98). Secondary outcomes include the specificity and likelihood ratios of FOCUS for PE in each population. Additional secondary outcome measures include the diagnostic test characteristics for PE of the individual component parts of FOCUS.

**Data Analysis**

All data analyses were performed using Stata 15.1 (StataCorp) by a statistician with a doctorate in statistics and significant work experience. Patient and demographic characteristics were tabulated for patients with and without PE. p-values were calculated using Student’s t-tests for continuous variables and chi-square tests for categorical variables. For inter-rater agreement, Fleiss kappa values were calculated for two raters (first operator and second operator), using the Stata “kap” package. Kappa values were interpreted based on recommendations by Landis and Koch. Diagnostic test characteristics of TAPSE and other measures of right heart strain were calculated using the Stata “diagnost” package, both for all patients and for a subset of patients with a HR ≥ 110 beats/min. To determine sample size, we chose to conduct a power calculation to minimize the 95% CI for the sensitivity of FOCUS in the diagnosis of PE. For the width of the 95% CI to be 20% or less, the authors calculated that it would require 120 subjects, assuming a FOCUS sensitivity greater than 90%.

Some patients had missing data from the FOCUS examination. In some instances, this was because the data were not recorded, while in others, this was because investigators were unable to obtain the required echocardiographic windows. Patients with incompletely documented or recorded FOCUS examinations were still included in the data analysis for the diagnostic test characteristics of the entire FOCUS examination, as the FOCUS examination could still be tabulated as positive or negative based on incomplete data. For example, a patient may have had data recorded for right ventricular enlargement, TAPSE, septal flattening, and McConnell’s sign, but nothing recorded for tricuspid regurgitation. In this example, it would still be possible to determine if the patient had a positive or negative FOCUS examination; however, presence or absence of tricuspid regurgitation would not contribute to that categorization, nor would this patient’s data have been used to determine the individual diagnostic test characteristics for tricuspid regurgitation.

**RESULTS**

There were 143 patients that underwent CTA during the study period. Four were not eligible for enrollment.
because they did not speak English, while three were excluded because personnel were unable to obtain any FOCUS windows (none of these subjects were diagnosed with a PE on CTA). There were 136 patients enrolled during the study period (Figure 2) and 37 (27.2%) were diagnosed as having a PE (Table 1). Of the 37 patients diagnosed with PE, six (16.2%) were in patients whose sBP was below 90 mm Hg, 28 (75.6%) were in normotensive patients who had evidence of RVD on FOCUS, and three (8.1%) were in normotensive patients without evidence of RVD on FOCUS. Patients diagnosed with a PE were more likely to be admitted to the hospital and had higher rates of admission to the step-down or intensive care unit (Table 1). Data depicting patient enrollment by site and sonographer level of training may be found in the Data Supplement S1 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13774/full).

In patients with a HR ≥ 100 beats/min or a sBP < 90 mm Hg (n = 136), the sensitivity of FOCUS for PE was 92% (95% CI = 78% to 98%; Table 2). The most sensitive component of FOCUS for PE was TAPSE with a testing threshold of 2.0 cm, which was 88% sensitive (95% CI = 72% to 97%). The sensitivity of TAPSE when using the traditional testing threshold of 1.7 cm was 67% (95% CI = 48% to 82%). The sensitivity for other components of the FOCUS examination ranged from 35% to 51% (Table 2). FOCUS had an overall specificity of 64% (95% CI = 53% to 73%).

The most specific component of FOCUS was McConnell’s sign with 99% (95% CI = 94% to 100%) specificity. The specificity of other components of FOCUS ranged from 64% to 93% (Table 2).

In the subgroup of patients with a HR ≥ 110 (n = 98), the sensitivity of FOCUS for PE was 100% (95% CI = 88% to 100%; Table 3). The most sensitive component of FOCUS was TAPSE when using a testing threshold of 2.0 cm, which was 93% (95% CI = 75% to 99%) sensitive for PE. The sensitivity of TAPSE with the traditional threshold of 1.7 cm was 77% (95% CI = 56% to 91%). The sensitivity for other components of the FOCUS examination ranged from 36% to 57% (Table 3). FOCUS was 63% (95% CI = 51% to 74%) specific for PE in this subgroup. The most specific component of FOCUS was McConnell’s sign with 100% (95% CI = 95% to 100%) specificity. The specificity of other components of FOCUS ranged from 63% to 93% (Table 3).

Inter-rater reliability for whether the FOCUS examination was found to be positive or negative by two separate sonographers was substantial with a kappa statistic of 1.0 (95% CI = 0.31 to 1.0). Inter-rater reliability of the components of the FOCUS examination was moderate to high with kappa statistics measuring as 0.61 (95% CI = 0.31 to 1.0) for TAPSE, 0.88 (95% CI = 0.69 to 1.0) for septal flattening, 0.89 (95% CI = 0.7 to 1.0) for right ventricular enlargement, 0.89 (95% CI = 0.7 to 1.0) for McConnell’s sign, and 0.81 (95% CI = 0.64 to 1.0) for tricuspid regurgitation. Nine patients had missing data for

**Figure 2.** Patient enrollment flow diagram. CTA = Computed tomographic angiogram; PE = pulmonary embolism.
TAPSE, two for right ventricular enlargement, two for septal flattening, two for McConnell’s sign, and 60 for tricuspid regurgitation.

**DISCUSSION**

Our results confirm that FOCUS is highly sensitive for PE in patient populations with abnormal vital signs who are suspected of having PE, especially in those with a HR ≥ 110 beats/min. A rapid bedside test that could reliably exclude or significantly lower the likelihood of PE at the time the history and physical examination is performed in ED patients with abnormal vital signs could be of significant utility. While CTA is the criterion standard for the diagnosis of PE in the ED, there are a variety of common.
clinical situations that may limit its accessibility, including the evaluation of patients with acute kidney injury, those with a contrast allergy, patients who are too obese to fit in a CT scanner, patients who are too hemodynamically unstable to leave the ED, and those who live in a resource-limited setting where CTA may not be available (e.g., global or rural health facilities).

Our results suggest that CTA may not provide additional diagnostic information regarding PE in patients with abnormal vital signs and a normal FOCUS examination, although given the width of the reported 95% CIs (Tables 3 and 4), further study in a larger cohort of patients is required before this can be said with certainty. In this scenario, a clinician may choose to further resuscitate a patient before attempting a potentially hazardous trip to radiology to obtain a CTA. Additionally, in patients with abnormal vital signs where confirmatory imaging is delayed due to a contrast allergy or acute kidney injury, an abnormal FOCUS examination may help the EP in their decision to begin anticoagulation earlier on a potentially unstable patient, since the early initiation of anticoagulation is associated with reduced mortality in patients with PE.22

There have been many studies in the literature that delineate the diagnostic test characteristics of the more typical components of the FOCUS examination for PE,1,2,3,24 although many of these have not included TAPSE.2 Our prior work demonstrated the high reliability of EP measured TAPSE in patients with suspected or confirmed PE and suggested a higher testing threshold (2.0 cm vs. 1.7 cm) to increase the examinations sensitivity.2 Consistent with our prior results, this study demonstrated that TAPSE is the most sensitive measure of PE when compared to other components of the FOCUS exam (Tables 2 and 3). Additionally, using a higher testing threshold yielded increased sensitivity for PE, although the difference was not statistically significant (Tables 2 and 3). While TAPSE is the most sensitive measure of the FOCUS examination, it is important to retain the other components of the FOCUS examination when assessing a patient with abnormal vitals for PE. Four of 37 patients diagnosed with PE in this study had a TAPSE greater than 2.0 cm (normal) but demonstrated other signs of PE on the FOCUS examination.

Of 37 patients diagnosed with PE in this study, three had false-negative FOCUS examinations without any signs of RVD. In all three of these patients, investigators were unable to perform a complete FOCUS examination due to difficult cardiac windows. One patient did not have a TAPSE calculated while in the ED examination due to difficult cardiac windows. One patient did not have a TAPSE calculated while in the ED, and the other two patients, investigators were unable to assess for tricuspid regurgitation. Additionally, these patients had a high pretest probability of PE (all three were being treated for cancer and had a prior history of PE). Furthermore, these patients were younger (aged 28, 46, and 62 years) and in good cardiovascular health, which makes it less likely that their heart would manifest signs of RVD on FOCUS. For these reasons, we caution the use of FOCUS to exclude PE in patients with an incomplete FOCUS examination, younger patients, those with a high pretest probability of PE, or those with a HR < 110 beats/min.

While FOCUS demonstrates high sensitivity for PE in this patient population, it is not surprising that the specificity of FOCUS for PE is only moderate (63%). This is in part explained by the fact that only one

### Table 3

<table>
<thead>
<tr>
<th>FOCUS Component</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Positive Likelihood Ratio (95% CI)</th>
<th>Negative Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAPSE threshold (cm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>93 (75–99)</td>
<td>73 (60–83)</td>
<td>3.4 (2.3–5.1)</td>
<td>0.11 (0.03–0.40)</td>
</tr>
<tr>
<td>1.7</td>
<td>77 (56–91)</td>
<td>88 (78–95)</td>
<td>6.4 (3.2–12.6)</td>
<td>0.26 (0.13–0.53)</td>
</tr>
<tr>
<td>RVE</td>
<td>57 (37–76)</td>
<td>84 (73–92)</td>
<td>3.6 (1.9–6.7)</td>
<td>0.51 (0.32–0.79)</td>
</tr>
<tr>
<td>Septal flattening*</td>
<td>47 (28–66)</td>
<td>93 (84–98)</td>
<td>6.3 (2.5–16.0)</td>
<td>0.58 (0.41–0.82)</td>
</tr>
<tr>
<td>TR</td>
<td>47 (21–73)</td>
<td>75 (59–87)</td>
<td>1.9 (0.9–4.0)</td>
<td>0.71 (0.43–0.89)</td>
</tr>
<tr>
<td>McConnell’s sign†</td>
<td>36 (19–56)</td>
<td>100 (95–100)</td>
<td>Undefined</td>
<td>0.64 (0.49–0.85)</td>
</tr>
</tbody>
</table>

FOCUS = focused cardiac ultrasound; HR = heart rate; PE = pulmonary embolism; RVE = right ventricular enlargement (appearance of right ventricle as being equal to or larger than the left ventricle); TAPSE = tricuspid annular plane systolic excursion; TR = tricuspid regurgitation (any regurgitant jet visualized on color Doppler).

*Abnormal flattening of the interventricular septum during systole.
†Visualization of hypokinesis of the right ventricle with apical sparing.
component of the FOCUS examination had to be abnormal for the examination to be deemed positive for PE, which maximizes the examination’s sensitivity at the expense of its specificity. Additionally, higher false-positive rates were noted in patients with heart failure or obstructive lung disease (e.g., chronic obstructive pulmonary disease, asthma), likely due to underlying mild right heart failure secondary to these disease processes. However, conditions that lead to chronic right heart strain will lead to global thickening of the RV which may help determine if the noted RVD is acute or chronic.25

While the overall specificity of FOCUS is moderate, our results demonstrate that certain components of FOCUS have a very high specificity for PE. McConnell’s sign was 99% specific for PE and septal flattening was 93% specific for PE (Tables 2 and 3). When present, especially in the absence of a prior history of pulmonary hypertension, McConnell’s sign, and septal flattening suggest an extremely high likelihood that the patient has a PE. Patients with these findings may merit empiric anti-coagulation or thrombolysis if the EP is unable to obtain a CTA within a reasonable period of time.

LIMITATIONS

This study was limited by derivation of subjects from a convenience sample of patients, which may introduce selection bias. Additionally, the observational nature of the study introduces potential for bias. In two subjects, investigators were unintentionally unblinded to results and had a high suspicion that the subjects had been diagnosed with PE, due to the fact that they were receiving a heparin infusion during the FOCUS examination (however, independent blinded review of these particular subjects agreed with the initial FOCUS results).

Some patients had missing data concerning various components of the FOCUS examination (e.g., missing tricuspid regurgitation data). Because the FOCUS examination only required one abnormal component to be categorized as a positive examination, the missing data may have contributed to a reported sensitivity that is lower than the actual sensitivity and a reported specificity that is higher than the actual specificity of the FOCUS examination.

Investigators chose to power the study so the width of the 95% CI for the sensitivity of the FOCUS examination was no greater than 20%. Investigators felt that a 20% width would most appropriately balance the need for diagnostic data and the team’s ability to meet the required enrollment numbers given limited resources.

Investigators had significant experience in bedside echocardiography and received dedicated training in measuring TAPSE, which limits the generalizability of these results to EP populations with more variable ultrasound experience. However, three medical students with limited experience in bedside echocardiography were able to learn the technique, suggesting that EPs with less experience in bedside ultrasound could also acquire the necessary skill set. Prior work by this author has demonstrated high rates of interobserver reliability between medical students who were taught the FOCUS examination and EPs with significant experience in bedside echocardiography.2

CONCLUSIONS

Focused cardiac ultrasound performed by emergency physicians with advanced training in emergency ultrasound may significantly lower the likelihood of the diagnosis of pulmonary embolism in most patients who are suspected of pulmonary embolism and have abnormal vital signs. This was especially true in those patients with a heart rate > 110 beats/min. Further study in a larger cohort of patients (which would yield narrower 95% confidence intervals) is required before focused cardiac ultrasound can be used to reliably exclude pulmonary embolism in this patient population. Our results suggest that focused cardiac ultrasound can be an important tool in the initial evaluation of ED patients with suspected pulmonary embolism and abnormal vital signs.

References

4. Jackson RE, Rudoni RR, Hauser AM, Pascual RG, Hussey ME. Prospective evaluation of two-dimensional

Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13774/full

Data Supplement S1. Patient enrollment by site and level of experience.
ABSTRACT

Background: Freestanding emergency departments (FrEDs) could reduce wait times in overcrowded emergency departments (EDs), but they might also increase usage and overall spending for emergency care. We investigate the relationship between the number of FrEDs entering a local market and overall spending on emergency care.

Methods: We accessed data from Arizona, Florida, North Carolina, and Texas in Blue Cross Blue Shield Axis; a limited data set of deidentified insurance data claims that we linked to Public Use Microdata Area (PUMA) data from the American Community Survey; and lists of licensed FrEDs from state agencies. Regression analysis was used to estimate the association between changes in the number of FrEDs in 495 PUMAs and total spending on emergency care, out-of-pocket spending, utilization, and price per visit from January 2013 to December 2017. Final estimates came from a PUMA-level fixed-effects model, with controls for state, quarter, and PUMA-level demographics.

Results: Entry of an additional FrED in a PUMA was associated with a 3.6 percentage point (pp; CI = 2.4 to 4.9) increase in emergency provider reimbursement per insured beneficiary in Texas, Florida, and North Carolina. There was no change in spending (2.5 pp; CI = −8.2 to 3.1) associated with a FrED’s entry in Arizona. Entry of an additional FrED was associated with a 0.18 (CI = 0.12 to 0.23) increase in the number of emergency care visits per 100 enrollees in Texas, Florida, and Arizona. In contrast, entry of another FrED was not associated with a change in utilization (−0.03; CI = −0.09 to 0.02) in North Carolina. Estimated out-of-pocket payments for emergency care increased 3.6 pp (CI = 2.5 to 4.8) with the entry of a FrED in Texas, Florida, and Arizona, but declined by 15.3 pp (CI = −26.8 to −3.7) in North Carolina.

Conclusions: Rather than functioning as substitutes for hospital-based EDs, FrEDs have increased local market spending on emergency care in three of four states’ markets where they have entered. State policy makers and researchers should carefully track spending and utilization of emergency care as FrEDs disseminate to better understand their potential health benefits and cost implications for patients.

The annual number of emergency department (ED) visits rose by 18.4% between 2006 and 2014, accompanied by an increase in the average age and number of comorbidities among ED patients.1 Correspondingly, the share of civilian, noninstitutionalized persons’ health expenditures devoted to ED care rose from 3.6% to 4.4% during this time period.2 Providers in some states have met this increased demand by opening freestanding emergency departments (FrEDs).
FrEDs deliver emergency care in a facility that is physically separate from an acute care hospital. Some FrEDs are owned by a parent hospital and are referred to as a “satellite” to that hospital, while other FrEDs have no such hospital affiliation.\(^3\) Texas, Ohio, and Colorado were documented as having the most FrEDs in 2015 (181, 34, and 24, respectively), but 360 were located across 30 states.\(^4\) Proponents of FrEDs claim that these facilities can relieve the burden of overcrowded waiting rooms in hospital-based EDs, while promptly caring for patients in more convenient locations.\(^5,6\)

However, critics of FrEDs argue that the facilities increase spending, because they serve as supplements to traditional EDs rather than substitutes, delivering care that could be provided in alternative lower cost settings.\(^7\) Furthermore, policy researchers are concerned that existing private insurer and Medicare payment policies are encouraging providers to shift services from lower paying settings such as urgent care centers and physicians’ offices to higher paying settings such as FrEDs.\(^8,9\)

Similar to hospital EDs, FrEDs charge a facility fee for each visit. Retail and urgent care clinics or doctors’ offices do not charge this fee.\(^10,11\) The facility fee originated under the Medicare program and was intended to compensate hospitals for the operational expenses of maintaining an outpatient facility, but facility fees are also charged to patients with private insurance coverage.\(^12,13\) A recent study found that prices for patients with similar diagnoses were 10 times higher at FrEDs in Texas compared to urgent care clinics, with the majority of the price difference attributable to the facility fee charged by FrEDs. The average price for a visit to a FrED in 2015 was $2,199, which was comparable to a hospital-based ED visit, with 82% of that price being the facility fee.\(^14\)

Although past studies have documented the rapid growth of FrEDs in multiple states, there is only limited information about how entry of FrEDs has influenced overall spending on EDs and how much changes in spending are attributed to shifts in price versus utilization. In this study, we analyze ED claims data from four U.S. states to examine the association between the entry of FrEDs in local markets and total ED spending, utilization, price, and out-of-pocket spending. The analysis provides useful insights on how much FrEDs serve as a substitute for care provided at existing hospital-based EDs versus contributing to a rise in overall use of EDs. This information is valuable to state regulators and public and private insurers who must consider regulations and reimbursement policies for FrEDs.

**METHODS**

**Data**

**Blue Cross Blue Shield Axis Claims Data.** We accessed the Blue Cross Blue Shield (BCBS) Axis limited data set, which is estimated to contain claims for 175 million active and inactive (previously enrolled) commercially insured members between 2012 and 2018.\(^15\) The BCBS Axis data are a limited data set under HIPPA privacy rules, because it excludes 16 categories of direct identifiers and is used for research purposes without obtaining prior authorization from patients.\(^16\) The BCBS companies in BCBS Axis are licensees of BCBS Association, an association of independent, locally operated Blue Cross and Blue Shield companies.

We restricted the analysis to claims from Arizona, Florida, North Carolina, and Texas. We chose these states, because they all experienced entry of FrEDs during the sample period and because BCBS companies have the largest market share in each of these states, which increased the number of claims available to analyze.\(^17\) These states also vary in population size, presence of Certificate of Need (CON) regulations, whether they impose policies specific to FrEDs, and other legal requirements (see 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/10.1111/acem.13848/full).\(^18\) We analyzed claims from 2013 to 2017, because these were the only years available in BCBS Axis when we began our analysis. This secondary analysis of facility and professional claims for EDs was accessed through a secure data portal. The institutional review board of Rice University considered this study exempt from review.

Claims data for ED visits were identified using the National Committee for Quality Assurance’s methodology. The NCQA counts any claim as emergency related if it contains a CPT code of 99281-5, which are procedure codes for ED visits for evaluation and management of a patient, or if the claim contains a UB revenue code of 0450-2, 0459, or 0981—codes for hospital services delivered in the ED. The NCQA also counts claims with a place of service code of 23 (emergency room–hospital) and one of more than
5,000 ED-related procedure codes as emergency claims. Any claims data satisfying these criteria for the years 2013 through 2017 were drawn for the sample. No 2013 claims data were available for Arizona; no claims data for Florida were available between 2015 Q3 and 2017 Q2.

For each claim within the sample we recorded the “total allowed amount,” which is the combined amount the provider should receive from the insurer and out of pocket from the patient. We subtracted from this number the “paid amount,” which is the amount the insurer paid to the provider, to estimate the amount billed to the patient to be paid out of pocket.

**FrED Entry and Markets.** Our approach for identifying FrEDs in Texas was described previously.20 We relied on licensing data from the Texas Department of State Health Services, as well as Internet searches, e-mails, and phone calls. In Arizona, Florida, and North Carolina, only hospitals are allowed to open FrEDs. The names, addresses, and effective dates of operation for FrEDs were obtained from the state health department in each of these states.

Using the address of each FrED, we determined which Public Use Microdata Area (PUMA) it was located in. Developed by the Census Bureau, PUMAs are geographic units constructed by combining census tracts (or counties in sparsely populated areas) with geographical contiguity. Each PUMA must contain over 100,000 residents, and the majority of PUMAs contain 100,000 to 200,000 residents.20 For example, Texas has 212 PUMAs. A total of 38 of those are located in Harris County, which covers 1,778 square miles and includes Houston, the fourth largest city in the U.S. In contrast, the “Rio Grande COG & Permian Basin Regional Planning Commission” PUMA in west Texas spans 36,606 square miles and contains 14 counties.

Defining local markets using PUMAs is consistent with the economics literature, which finds that population density plays a central role in firm entry decisions.21,22 PUMAs are much smaller in geographic size than the more well-known Hospital Service Areas defined by the Dartmouth Atlas.23 For example, the HSA for Houston contains 38 hospitals and is divided into roughly 50 PUMAs. Market sizes at the PUMA level are more likely to reflect the decision facing consumers considering the use of EDs—visiting a nearby FrED with little or no wait time for care or traveling slightly farther to a hospital ED where roughly one-third of patients waited an hour or more for care.24

**Constructing PUMA-level Variables.** Emergency claims data were attributed to PUMAs based on the zip code of residence of the patient on the date of treatment. Claims data were then aggregated to obtain the total amount of expenditures on EDs by PUMA, year, and quarter. ED spending per enrollee was calculated by summing the allowed amount in the facility and professional claims in each PUMA and dividing by the number of enrollees (whether or not they had any claims) in the PUMA. Out-of-pocket ED spending per enrollee was calculated in a similar manner.

We subdivided spending on EDs into utilization versus price. The number of emergency visits per 100 enrollees was calculated by dividing the number of unique facility claims on each date for each enrollee by the number of enrollees/100 in each PUMA. An estimate for the “price” of each visit was obtained by dividing ED spending for each PUMA by the number of facility claims.

PUMA-level data on the percentage of the population with any insurance along with the percentage covered by Medicare and Medicaid; median household income; the percentage of the population Hispanic, black, or with a high school diploma; and population count by year were obtained from the American Community Survey. Based on previous studies, these variables were hypothesized to influence demand for ED services among insured persons.25,26

Although we are only measuring factors associated with ED use, previous studies have found that patients with different coverage types use EDs at different rates. Higher or lower propensity to utilize EDs by insurance type can influence the patient load and wait time at hospital EDs, which may affect decisions by consumers on whether and where to seek care in an ED when new facilities become available.

**Data Analyses**

Descriptive statistics were used to compare the number of FrEDs by PUMA at the beginning of the sample (2013 Q1) and the end (2017 Q4) for Arizona, Florida, North Carolina, and Texas. Insured status and other sociodemographic characteristics included in the American Community Survey, the mean volume of enrollees by PUMA, total PUMA population, and mean values of the dependent variables in the regressions by PUMA for these time periods are listed by state. We graphed mean ED spending per enrollee by quarter and state.
We then applied regression analysis to test whether the entry of one or more FrEDs to a PUMA was associated with a change in ED spending per enrollee by year and quarter, adjusting for other factors that might influence spending. We also tested for an association between FrED entry and the number of ED visits per enrollee and the price per visit. The unit of analysis for the regressions is a PUMA during the quarter of a given year. The explanatory variable of interest was a continuous measure of the number of FrEDs in a PUMA in a given year and quarter. We included an interaction of state indicator variables with FrED counts to test whether the association between facility entry and ED spending differed across states. In cases where the state interaction term was statistically significant, the change in the dependent variable associated with FrED entry was the linear combination (sum) of the coefficients of the number of FrEDs and the state interaction term.

The multiple measures of insured status and socioeconomic status that were considered as explanatory variables were highly correlated. To avoid potential problems of multicollinearity, variables with a variance inflation factor greater than 2.5 were excluded from the regressions. The excluded variables were the percentage covered by Medicaid, median household income, the percentage of the population Hispanic, and population count in each PUMA.

The regressions included fixed effects for each of the 20 quarters in the sample, PUMA fixed effects, and interaction effects of a linear time trend with each PUMA fixed effect. Inclusion of these fixed effects and interactions controls for potential systematic trends in ED spending across PUMAs that may have coincided with the entry of FrEDs. With the inclusion of PUMA fixed effects, the FrED explanatory variables measure the association between within-PUMA changes in the number of FrEDs and ED spending.

The regressions involving per-capita spending and price were estimated using a generalized linear model (GLM) with a log link. Expenditures and price data commonly follow a skewed rather than normal distribution, which suggests that the relation between the explanatory variables and these monetary values is better modeled with a log distribution. The GLM allows one to estimate this relation while avoiding biases that can result from estimates derived from ordinary least squares with the log of spending or price as the dependent variable. The regressions were estimated using Stata 15.1. The glm command in STATA allows one to adjust the standard errors to account for correlation in the error terms within PUMAs, and the regressions were weighted by the number of enrollees in each PUMA.

**Sensitivity Analysis**

Census experts form PUMAs based on population counts rather than land area. Therefore, one might be concerned that PUMAs may not be the correct unit of analysis in rural areas. As a sensitivity analysis, we limited the sample to PUMAs where 100% of the land area was located in a Metropolitan Statistical Area as defined by the census. Major insurers are trying to control rising ED spending by denying reimbursement for ED visits that they deem unnecessary. We estimate an additional regression with the insurer’s amount paid to the provider per enrollee as the dependent variable to determine whether the results are consistent with those obtained with spending per enrollee.

**RESULTS**

Table 1 presents descriptive statistics on the 495 PUMAs by state at the beginning and end of the sample period. Despite similar population numbers across PUMAs, FrED entry was more widespread in Texas than in the other three states. By 2017 Q4, 74% of PUMAs in Texas had at least one FrED, while only 28% of PUMAs in Arizona, 22% of PUMAs in Florida and 14% of PUMAs in North Carolina had one or more FrEDs.

Figure 1 illustrates trends in ED spending per enrollee by state during the sample period. All states displayed a general upward trend in ED spending per enrollee, with Texas showing the steepest increase. Florida and North Carolina had the lowest spending per enrollee, while spending per enrollee was substantially higher in Texas.

Figure 1 illustrates trends in ED spending per enrollee by state during the sample period. All states displayed a general upward trend in ED spending per enrollee, with Texas showing the steepest increase. Florida and North Carolina had the lowest spending per enrollee, while spending per enrollee was substantially higher in Texas.

A graph of the number of FrEDs by state and year is in Data Supplement S1, Figure S1. The coefficients on the quarter fixed effects in Column 1 indicate that spending on EDs steadily increased throughout the sample period. Figure 2 reports adjusted estimates of the association between FrED entry into a PUMA and ED spending per enrollee. The glm regression specification with a log link implies that coefficients are interpreted as the percentage change in spending per enrollee with a one-unit change in the explanatory variable. Texas is coded as
the base state for comparison, because it contains the most FrEDs overall and the most FrEDs per PUMA. Entry of each additional FrED was associated with a 3.6 (CI = 2.4 to 4.9) percentage point increase in ED spending per enrollee in a PUMA. The state interaction terms for Florida and North Carolina were not significantly different from 0, so the association between FrED entry and spending for these two states was assumed to be similar to Texas. However, there was no change in ED spending per enrollee (CI = 0.03, CI = 0.09 to 0.02) associated with entry of an additional FrED in Arizona.

Entry of an additional FrED was associated with a 0.18 (CI = 0.12 to 0.23) increase in the number of ED visits per 100 enrollees in Texas, Florida, and Arizona. In contrast, entry of another FrED was not associated with a reduction (CI = 0.03, CI = 0.09 to 0.02) in utilization of EDs per 100 enrollees in North Carolina. The regression results suggested no relationship between the number of FrEDs entering a PUMA and price for Texas, Florida, and North Carolina. However, entry of an additional FrED in Arizona was associated with a 10.1 percentage point decrease (CI = 7.0 to 3.3) in the price per ED visit in Arizona. Entry of a FrED was associated with a 3.6 percentage point increase (CI = 2.5 to 4.8) in estimated out-of-pocket spending per beneficiary in Texas, and the magnitude was not significantly different from Texas for Florida and Arizona. However, FrED entry was associated with a 15.3 percentage point (CI = -26.8 to -3.7) reduction in out-of-pocket spending per enrollee in North Carolina.

In a sensitivity analysis, we limited the sample to PUMAs where 100% of the land area was located in a Metropolitan Statistical Area. The regression results remained virtually unchanged. An additional sensitivity analysis revealed that entry of an additional FrED was also associated with a 3.6 percentage point (CI = 2.3 to 5.0) increase in the insurer’s amount paid per enrollee in Florida, North Carolina, and Texas. Similar to the spending estimates, FrED entry was not associated with a significant change in paid amounts by the insurer in Arizona (CI = -5.1, CI = -11.3 to 1.0).

**DISCUSSION**

FrEDs have been touted as a means to reduce wait times and excess demand at overburdened hospital EDs. They have also been associated with lower inpatient hospital admission rates relative to hospital-based EDs. However, the same sources that suggest that FrEDs could function as a substitute for hospital EDs also remark that these facilities could lead to an increase of ED visits for nonemergency conditions.
One study of a hospital ED at a tertiary care center found that patient volume fell 7.5% during a period of 3 years when two FrEDs opened nearby, while total patient volume for all three facilities combined rose 45%. Another study found that Medicare expenditures per beneficiary were $55 higher for each FrED that entered a county between 2003 and 2009. However, this study could not isolate the relationship between FrED entry and ED spending.

This study is the first we know of that measures the relationship between entry of FrEDs and overall spending on emergency services in a large sample of local markets. Our sample contains data from Arizona, Florida, North Carolina, and Texas, which include 20.4% of the U.S. population. The observed differences in spending between states were generally consistent with price differentials for an ED visit reported by the Health Care Cost Institute (HCCI). The HCCI found the price of an ED visit to be between 110 and 133% of the national average in Texas, between 100 and 110% of the national average in Arizona and North Carolina, and between 90 and 100% of the national average in Florida.

In three of four states, we found a positive association between entry of a FrEDs in a PUMA and average ED spending per enrollee. Entry of an additional FrED was associated with a 3.6% increase in spending per enrollee in Texas PUMAs, and the results for Florida and North Carolina were not statistically significantly different from 3.6%. In Arizona, entry of a FrED in a PUMA was not associated with a change in spending per enrollee.

When we separated the changes in spending into parts attributable to changes in utilization versus price, there was a significant increase in utilization for Texas and Florida, but no significant change in price. For North Carolina, there was no significant change in utilization associated with FrED entry. We also found no significant increase in price associated with entry of FrEDs in North Carolina, although the magnitude of the estimate was relatively large (1.8 percentage points, CI = -4.7 to 8.4).

For Arizona, the increase in utilization was not significantly different from Texas, but there was also a statistically significant 10.1% fall in price. It is possible that FrED entry was associated with more aggressive

Figure 1 Mean ED spending per BCBS enrollee by state and quarter. BCBS = Blue Cross Blue Shield.
price competition in Arizona than in other states. More FrEDs in Arizona may have become in-network for insurers than in other states, which would have lowered prices. Unfortunately, we could not distinguish between in-network and out-of-network claims in the BCBS Axis data. More light could be shed on...
these hypotheses with access to an All-Payers Claims Database. 

Entry of a FrED in a PUMA was associated with an increase in out-of-pocket spending per enrollee for Texas, Florida, and Arizona, but out-of-pocket spending per capita dropped a remarkable 15.3 percentage points for North Carolina. Of the four states we examined, Texas is the only state that allows independent FrEDs. Previous studies have noted potentially different incentives between independent and satellite FrEDs. Hospitals may open satellite FrEDs in the hopes of attracting more inpatient admissions, but in the three states with only satellite FrEDs, we observe different relationships between entry and out-of-pocket spending. We were able to confirm that all of the visits by BCBS enrollees to EDs in North Carolina were to in-network facilities. However, some of the hospital EDs had out-of-network physicians providing emergency care. North Carolina patients may have shifted their ED utilization away from hospital EDs with out-of-network physicians and toward FrEDs with in-network physicians, which may have reduced their payment burden.

Health insurers and policy researchers have been critical of FrEDs because of their potential for raising overall ED expenditures. The Medicare Payment Advisory Commission recommended in June 2018 to reduce Type A ED payment rates by 30% for satellite EDs that are within 6 miles of a hospital-based ED. MedPAC’s rationale was that the current Medicare payment system creates incentives for providers to treat lower-intensity patients in the ED setting rather than in urgent care centers, which are paid less than half the Type A payment rates for ED services. If insurers and FrED owners could reach an agreement to reduce facility fees for patients with nonemergent conditions at FrEDs, patients could benefit by receiving timely and affordable access to care, while FrED operators could still earn additional revenues for filling an unmet need in the market.

One might be concerned that the increase in prevalence of high-deductible plans and insurers’ attempts to deny claims for allegedly unnecessary visits to EDs have placed a greater burden of FrED spending on patients versus insurers. However, a sensitivity analysis indicated that both total spending per enrollee and payments solely by the insurer rose by the same amount (3.6 percentage points) with entry of a new FrED. This result suggests that the proportion of the total bill that patients are responsible for does not increase with FrED entry.

LIMITATIONS

We mention several caveats to our analysis. Our data do not allow us to distinguish between claims filed by FrEDs versus those that came from hospital EDs. For this reason, we can only measure the association between FrED entry and overall ED spending per enrollee in a local market. Nevertheless, this analysis is much more precise than a recent study comparing the number of FrEDs in a county and total Medicare expenditures.

Our analyses did not adjust for patient case-mix severity, which has been found to be lower at FrEDs compared to hospital-based EDs. However, our dependent variable is spending per BCBS enrollee not spending per visit. Spending per enrollee can rise even if entering FrEDs treat lower-severity patients, as long as utilization per enrollee and the price per visit do not fall.

We aggregated ED claims based on the zip code of each enrollee because the claims lack accurate information on the provider’s location. Hospitals tend to submit claims for their satellite FrEDs using the location of their hospital-based ED. Some enrollees may visit an ED outside of their PUMA of residence. Assuming that the tendency to obtain emergency care outside an enrollee’s PUMA is random across PUMAs, the association between FrED entry and the dependent variables of interest would be biased toward zero. Therefore, we may have underestimated the association between FrED entry and spending, utilization, and price.

We lack data on the number of primary care or urgent care practices by PUMA. FrEDs may have entered in areas where such facilities existed, which would legitimately increase ED utilization. However, other studies suggest that FrEDs enter in high-income, well-insured neighborhoods, which are the same factors that attract both primary care physicians and urgent care clinics.

Previous research found that state policies regarding FrEDs vary widely. Florida and North Carolina maintain CON regulations, while Arizona and Texas do not. North Carolina has no FrED-specific policies, while the other states do. Other differences in rules across the states we examined are reproduced in Data Supplement S1. With multiple differences in
regulations, we are unable to determine what role any specific policy played in the spending, utilization, and price changes we observed.

Patients may not have paid their portion of the allowed amount reported in the insurance claims data, which would have led to overestimated spending. However, we have no reason to believe that hospitals and FrEDs differ in their ability to collect payments from patients. If the rate of underpayment is the same for hospital EDs and FrEDs, then the percentage changes in spending per enrollee associated with FrED entry that we estimated remain the same.

The BCBS Axis limited data set had no claims data for Arizona in 2013 and no claims information for eight quarters in Florida. However, the trends in Figure 1 are consistent with an upward trend in spending for all four states throughout the sample period. These missing claims data do not bias the regression estimates, because the regressions include quarter fixed effects that account for the quarter and year in which each observation occurs.

CONCLUSION

In conclusion, entry of freestanding EDs has increased local market spending on EDs in three of four states’ markets where they have entered. The increase in spending for Texas and Florida was accompanied by a rise in utilization of EDs. The increase in spending observed for North Carolina was accompanied by a relatively large but statistically insignificant price increase. Entry of freestanding EDs were associated with a decline in the price of ED visits in Arizona, an increase in utilization, and no change in spending. State policy makers and researchers should carefully track spending and utilization of emergency care as freestanding EDs disseminate to better understand their potential health benefits and cost implications for patients.

Rice University participates in the Blue Cross Blue Shield Alliance for Health Research. The Blue Cross Blue Shield Association established the Blue Cross Blue Shield Alliance for Health Research to engage leading U.S. health care researchers in collaborative efforts to use a limited data set drawn from Blue Cross Blue Shield companies to explore critical health care issues to improve the health of Americans. The Blue Cross Blue Shield Alliance for Health Research provides researchers with use of a secure data portal to access a limited data set from Blue Cross Blue Shield Axis, the largest collection of commercial insurance claims, medical professional, and cost-of-care information. Blue Cross Blue Shield Association is an association of independent Blue Cross Blue Shield companies.

References

15. BCBS Axis Overview and Details. Chicago, IL: Blue Cross Blue Shield Association, 2019.


**Supporting Information**

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13848/full

Data Supplement S1. Supplemental material.
Early Screening for Posttraumatic Stress Disorder and Depression Among Injured Emergency Department Patients: A Feasibility Study

Stephany Jaramillo, MD1, Brian Suffoletto, MD1, Clifton Callaway, MD, PhD1, and Maria Pacella-LaBarbara, PhD1

ABSTRACT

Background: Despite the risk of developing posttraumatic stress disorder (PTSD) and associated comorbidities after physical injury, few emergency departments (EDs) in the United States screen for the presence of psychological symptoms and conditions. Barriers to systematic screening could be overcome by using a tool that is both comprehensive and brief. This study aimed to determine 1) the feasibility of screening for posttraumatic sequelae among adults with minor injury in the ED and 2) the relationship between ED screening and later psychological symptoms and poor quality of life (QOL) at 6 weeks postinjury.

Methods: In the EDs of two Level I trauma centers, we enrolled injured patients (n = 149) who reported serious injury and/or life threat in the past 24 hours. Subjects completed the Posttraumatic Adjustment Scale (PAS) to screen for PTSD and depression in the ED, and 6 weeks later they completed assessments for symptoms of PTSD, depression, and trauma-specific QOL (T-QoL).

Results: Our retained sample at 6 weeks was 84 adults (51.2% male; mean ± SD age = 33 ± 11.88 years); 38% screened positive for PTSD, and 76% screened positive for depression in the ED. Controlling for age, hospital admission, and ED pain score, regression analyses revealed that a positive ED screen for both PTSD and depression was significantly associated with 6 weeks PTSD (p = 0.027, 95% confidence interval [CI] = 0.92 to 15.14) and depressive symptoms (p = 0.001, 95% CI = 2.20 to 7.74), respectively. Further, a positive ED screen for depression (p = 0.043, 95% CI = –16.66 to –0.27) and PTSD (p = 0.015, 95% CI = –20.35 to –2.24) was significantly associated with lower T-QoL.

Conclusions: These results suggest that it is feasible to identify patients at risk for postinjury sequelae in the ED; screening for mental health risk may identify patients in need of early intervention and further monitoring.

Traumatic injury is a leading cause of mortality and morbidity in the United States,1,2 and approximately 29% of annual emergency department (ED) visits are attributable to physical injury.3 The majority of these injuries are mild, with only 7% requiring hospital admission.4 However, the stress of
injury can lead to alterations in the psychological processing of the event and in the perception of pain, independent of tissue damage.\textsuperscript{5} Psychological sequelae including depression and anxiety are common after trauma exposure.\textsuperscript{6–8} Research conducted among hospitalized injury populations suggests that up to 31% report a psychiatric disorder 1 year postinjury, with 22% of these being new diagnoses.\textsuperscript{9} Posttraumatic stress disorder (PTSD) and depression are particularly common, with rates of PTSD at 12 months postinjury ranging from 6% to 29% and rates of depression at 12 months postinjury ranging from 9% to 28%.\textsuperscript{9–11} Further contributing to a complicated recovery process, early psychiatric symptoms after trauma are associated with lower health-related quality of life (QOL) and long-term disability.\textsuperscript{12,13} Notably, postinjury psychiatric symptoms (including anxiety, depressive, and posttraumatic stress symptoms) are a major contributor to disability over time, above and beyond physical factors and pain severity.\textsuperscript{14}

Regarding recovery after minor injury, Pacella and colleagues\textsuperscript{15} previously found that psychological symptoms are associated with physical symptoms among ED patients discharged home immediately from the ED; specifically, hyperarousal symptoms of PTSD are associated with daily pain during the first 14 days after injury. Similarly, patients with mild–moderate motor vehicle collision (MVC)-related injuries report reduced health-related QOL at 2 years post-MVC\textsuperscript{12} and at least one psychiatric diagnosis (50%) at 6-, 12-, and 24-months post-MVC.\textsuperscript{16} Despite these associations, patients with minor physical injuries who do not require hospitalization receive little or no psychological support services, and few studies have enrolled and monitored ED subjects to determine their health risk after minor injury. Thus, there may be an underused opportunity for interventions to prevent chronic pain and PTSD after trauma, but instruments are needed to identify patients most at risk.

Few studies examine risk factors for developing postinjury PTSD and/or depression in ED patients. Although multiple validated ED screening tools are available for assessing PTSD and depression in pediatrics,\textsuperscript{17,18} no tools exist to identify adults at risk for posttraumatic symptoms during the initial hours after trauma. Barriers to systematic screening could be overcome by using a tool that is both comprehensive and brief. In hospitalized adults, the Posttraumatic Adjustment Scale (PAS) is a brief (10-item) screening instrument that predicts subsequent PTSD and depression, with sensitivities of 0.82 and 0.72 for PTSD and depression, respectively.\textsuperscript{19} An advantage of the PAS is that it assesses three general categories of risk: 1) pre-trauma items reflecting psychiatric and trauma history and social support; 2) peritrauma items reflecting the response to and severity of the event; and 3) post-trauma items reflecting acute pain, cognitive response, coping self-efficacy, and social support.

**Current Study**

Consistent with prior research determining feasibility and effectiveness for ED-based screening programs,\textsuperscript{20} our study objectives are twofold: we aim to determine 1) the feasibility of screening for posttraumatic sequelae among adults with minor injury in the ED and 2) the relationship between psychosocial risk screening in the ED and symptoms of PTSD, depression, and QOL at 6 weeks postinjury. We included patients who perceived their trauma to be serious irrespective of the actual physical injuries identified. Our primary outcomes were PTSD symptoms (PTSS) and depressive symptoms at 6 weeks postinjury. Given the association between posttraumatic symptoms and QOL, we included the secondary outcome of Trauma-Specific Quality of Life (T-QoL), assessed using an instrument specifically designed to evaluate QOL in patients with physical injury.\textsuperscript{21} We hypothesized that 1) ≥75% of patients approached in the ED will agree to psychological screening and 2) a positive ED screening will be associated with elevated symptoms of PTSD and depression and with reduced QOL at 6 weeks postinjury.

**METHODS**

**Participants**

We conducted an observational, prospective, cohort study of adults admitted within 24 hours of injury to the EDs of two Level I trauma centers between January 2016 and May 2017. Eligible patients met the following criteria: 1) suffered a physical injury (e.g., MVC, general trauma, fall, assault) within the past 24 hours; 2) were between the ages of 18 and 60; 3) were medically and emotionally stable to understand and provide medical consent; 4) were not being treated primarily for a mental health or substance use issue directly related to the injury (e.g., MVC related to alcohol abuse); 5) did not present with self-inflicted injury or suicidal thoughts; 6) did not have a neurologic disease (e.g., seizure disorder, stroke, multiple sclerosis); and 7) met criterion A of the PTSD diagnosis (self-
Participants self-reported their psychological history at baseline by answering yes/no to the question “Has there ever been a point in your life when you received help for emotional or mental health problems?”

**Medical Record Review.** We abstracted the chief complaint/mechanism of injury (later recoded as MVC-related, fall, work, or general accident [e.g., sports, minor trips, interpersonal physical assault]), type of injury (later coded into six categories reflecting contusions, sprains and strains, fractures and dislocations, open wounds, crushing injury, and burns), injury location (e.g., head and neck, extremities, spine and back, and torso), and the initial ED pain severity score (ranging from 0 to 10) from the medical chart. The ED pain score was not available for 13 participants for varied reasons (e.g., no pain assessment conducted or only after medication administration). We used the discharge pain score \( (n=6) \) and the admission pain score \( (n=2) \) when available for eight of these participants, resulting in 79 participants with complete ED pain score data and five participants with incomplete/missing data for ED pain score.

**Baseline Assessment in ED: Predictors. PTSD and Depression.** Participants completed the PAS, a 10-item survey including pretrauma, peritrauma, and posttrauma risk factors used to screen for the risk of developing PTSD and depression after acute trauma\(^1\) (Figure 1). Each of the 10 items are scored on a Likert scale ranging from 0 (“not at all”) to 4 (“totally”). The PAS-PTSD score is calculated by summing all 10 items and applying a cutoff score of 16. The PAS-depression score is calculated by summing five of the 10 items and applying a cutoff score of 4 (see O’Donnell et al.\(^1\) for detailed scoring procedures). Cronbach’s alpha in our sample was acceptable for the PAS-depression (\( \alpha = 0.70 \)) and PAS-PTSD subscales (\( \alpha = 0.82 \)).

**Six Weeks After Injury Assessment: Outcomes. Psychiatric Symptoms.** Participants completed the DSM-5 PCL and PHQ-8 to assess the symptoms and presence of possible PTSD and major depressive disorder, respectively. The PCL is a 20-item self-report measure that assesses DSM-5 PTSS.\(^2\) The items are summed together to provide a continuous measure of PTSS severity with a total score ranging from 0 to 80 and a cut-point of \( \geq 30 \) (internal
consistency was high in our sample, $\alpha = 0.96$). The PHQ-8 is a valid severity scale for depressive symptoms composed of eight items that are summed to obtain a total score from 0 to 24, with a cutoff value of $\geq 10$ (internal consistency was high in our sample, $\alpha = 0.91$).25

QOL. Participants completed the 43-item T-QoL survey,27 specifically created for a traumatized population. The T-QoL has a five-component structure with subscales that assess emotional well-being (16 items), functional engagement (eight items), recovery/resilience (six items), peritraumatic experience (five items), and physical well-being (eight items). Response options for each item of the T-QoL are presented as a four-choices (for 41 questions) and five-choices (for two questions in the peritraumatic experiences subscale) Likert scale; however, to maintain consistency and allow for ease of administration to participants, we used a modified version that excluded the fifth response option for both items of the peritraumatic scale. This resulted in removing the “neutral” response option in the scale ranging from 0 (strongly disagree) to 4 (strongly agree); in this way, the four choices reflected the same response options for the remaining items in the entire scale. We summed the responses to create a single numerical measurement of total QOL, with a higher score indicating better T-QoL (internal consistency was high, $\alpha = 0.91$). Example items include: “I currently have physical limitations”; “I have trouble sleeping at night”; and “I felt fear when I was injured.”

Statistical Analysis
We used SPSS version 24.026 for statistical analyses; significance was defined using an alpha level of 0.05. We described the frequency of positive PTSD and depression screenings in the ED and probable rates of PTSD and depression at 6 weeks postinjury. We used

---

### T-QoL Survey Example Items

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Response Options</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>I currently have physical limitations</td>
<td>Strongly Disagree</td>
<td>0</td>
</tr>
<tr>
<td>I have trouble sleeping at night</td>
<td>Strongly Agree</td>
<td>4</td>
</tr>
<tr>
<td>I felt fear when I was injured</td>
<td>Strongly Agree</td>
<td>4</td>
</tr>
</tbody>
</table>

**Note:** Add all items to calculate the posttraumatic stress disorder score on the Posttraumatic Adjustment Scale (PAS), cutoff score $\geq 16$. Add items marked with an * to calculate the depression score on the PAS, cutoff score $\geq 4$.

---

analysis of variance (ANOVA) to determine whether there were any group differences based on sex in the outcomes of PTSS, depression symptoms, and QoL at 6 weeks. We also measured the strength and direction of linear relationships between continuous variables (e.g., ED pain score, age, PCL, PHQ-8, and T-QoL) using Pearson correlations.

We used Pearson chi-square to test for associations between participant retention at 6 weeks and categorical variables of sex, race, hospital admission, and positive ED screen. ANOVA tested whether retention was associated with age and ED pain score.

For our primary analyses, we created separate hierarchical linear regression models to examine the PAS-PTSD score as a predictor of 1) 6-week PTSS and 2) T-QoL. Similarly, two additional models were conducted to examine the PAS-depression score as a predictor of 1) 6-week depression symptoms and 2) T-QoL. Variables with a significant association with the predictors or outcomes were used as covariates.

The basic assumptions were tested before conducting each of the ANOVA, chi-square and hierarchical regression analyses.27-29 With one exception for regression analyses (see Results), all assumptions were satisfied.

Sample Size and Power. Using GPower version 3.1,30 we calculated the sample size required for a linear multiple regression test of the increase in variance explained in the outcome due to the inclusion of the predictor variable (PAS screen), above and beyond the effects of three covariates, given \( \alpha = 0.05 \) and a medium effect size (\( \eta^2 = 0.15 \)). For a power of 0.90, total sample size required for this test is \( N = 73 \); given that our retained sample size (\( N = 84 \)) is higher than the target value, power was adequate for these models.

RESULTS

We approached 534 patients in the ED, of whom 148 refused to be screened (28%). Of those who refused, 38 (26%) did not agree to talk to the RA or wish to hear about the study, 61 (41%) were not interested in participating after a brief description by the RA, 32 (22%) stated that they were in too much pain to participate, seven (5%) had no time to commit to the study, eight (5%) were not willing to complete the surveys, and two (1%) stated that the compensation was too low. A total of 386 patients (72%) agreed to be screened for the study, and of those screened, 213 (55%) were not eligible. Of those ineligible, 201 (94%) did not report exposure to threatened death or actual or serious injury (criterion A of the PTSD diagnosis), and 12 (6%) subjects were injured \( \geq 24 \) hours prior to ED admission. Of the 173 eligible patients, 154 completed baseline data prior to discharge, and 91 (59%) completed the follow-up survey at 6 weeks postinjury. Of the latter subjects, five patients were excluded for the following reasons: physical injury secondary to a medical condition (\( n = 4 \)), illiteracy (\( n = 1 \)), and hospital admission \( > 1 \) day (\( n = 4 \)).

Our final retained sample (\( N = 84 \)) included nearly equal proportions of males and females of predominantly white race (Table 1). Participants reported a high ED pain score, and the most common mechanism of injury was general accident. Eight patients (9%) were admitted to the hospital for their injuries for 1 day only (primarily for observation or minor procedures). Half of the sample (\( n = 42 \)) reported receiving help for emotional or mental health problems in the past. Contusion was the most common type of injury (32%) and it was frequently located on head and neck (55.5%) and extremities (37%). Sprains and strains (29.8%) and fractures and dislocations (23.8%) were also common; other type of injuries were open wounds, crushing injuries, and burns. All participants reported threat of serious injury (eligibility criterion) and approximately one-third reported threat of death. Males (52% retained) were less likely than females (71% retained) to complete the 6-week assessment (\( \chi^2(1) = 5.05, p = 0.025 \)). Subjects who screened positive for depression in the ED were more likely to complete the 6-week assessment (66% compared to 46%; \( \chi^2(1) = 5.29, p = 0.021 \)). Retention rates did not differ by race, hospital admission, or positive PTSD ED screening.

ED Screening. We screened a total of 155 patients in the ED, of whom 37% (\( n = 58 \)) screened positive for PTSD and 67.7% (\( n = 105 \)) for depression. Of the final sample retained at 6 weeks postinjury, 38% (\( n = 32 \)) screened positive for PTSD and 76% (\( n = 64 \)) for depression (Figure 2).

Outcomes. At 6 weeks, 18% (\( n = 15 \)) of participants scored above the cutoff for probable PTSD via the PCL, and 26% (\( n = 22 \)) for probable depression via the PHQ-8. Hospital admission (\( n = 8 \)) was associated with T-QoL, such that patients admitted to the
hospital reported lower QOL (mean ± SD = 115.37 ± 21.22) than those not admitted to the hospital (mean ± SD = 133.21 ± 20.48, F = 5.45, p = 0.02).

Emergency department pain score (Table 2) was positively correlated with poorer outcomes, and older age was associated with greater depressive symptoms and poorer QOL. There were no sex differences in T-QoL, PCL, and PHQ-8 scores at 6 weeks postinjury (p-values ≥ 0.2), but hospital admission was significantly associated with poorer QOL.

### Primary Regression Analyses.
Controlling for age, hospital admission, and ED pain score, a positive ED screening for PAS-PTSD was associated with PTSS at 6 weeks (B = 0.24, p = 0.027) (see Table 3), and a positive ED screening for PAS-depression was associated with PHQ-8 score at 6 weeks (B = 0.36, p = 0.001) (see Table 4). A positive ED screening for PAS-PTSD (B = −0.19, p = 0.043) and PAS-depression (B = −0.23, p = 0.015) was associated with lower T-QoL score at 6 weeks postinjury (see Table 5). The assumption of homoscedasticity was violated by the presence of two outliers with residual values > 2 for the outcome of PTSS; however, upon excluding these subjects and repeating the analysis, the results remained unchanged. Therefore, we retained these individuals to maintain integrity of the data.

### DISCUSSION
Consistent with our hypothesis, our data suggest that it is feasible to screen and identify ED patients at risk for postinjury psychological sequelae within 24 hours of minor physical injury. Specifically, >70% of patients approached about this research study agreed to be screened for eligibility, and 88% of eligible patients agreed to participate. The primary reason for ineligibility was no self-reported life threat or actual or threatened serious injury, consistent with the notion that we attempted to approach patients with minor injury who were likely to be discharged directly home from the ED. Further, high rates of positive ED screens for PTSD and depression emerged from the PAS: 38% of our sample screened positive for PTSD, and 76% screened positive for depression in the ED.

### Table 1
Descriptive Statistics of the Final Sample (N = 84)

<table>
<thead>
<tr>
<th>Demographics</th>
<th>N (%) or Mean (±SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>33 (±11.88); range 18–60</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>43 (51%)</td>
</tr>
<tr>
<td>Female</td>
<td>41 (49%)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>46 (55%)</td>
</tr>
<tr>
<td>African American</td>
<td>27 (32%)</td>
</tr>
<tr>
<td>Other</td>
<td>11 (13%)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>High school/GED or less</td>
<td>35 (42%)</td>
</tr>
<tr>
<td>Some college/technical or vocational school</td>
<td>33 (39%)</td>
</tr>
<tr>
<td>College degree or higher education</td>
<td>13 (15%)</td>
</tr>
<tr>
<td>Declined to answer</td>
<td>3 (4%)</td>
</tr>
<tr>
<td>ED pain score</td>
<td>6.90 (2.64); Range 1–10</td>
</tr>
<tr>
<td>Hospital admission</td>
<td></td>
</tr>
<tr>
<td>Admitted to hospital</td>
<td>8 (9%)</td>
</tr>
<tr>
<td>Discharged from ED</td>
<td>76 (91%)</td>
</tr>
<tr>
<td>Mechanism of injury/chief complaint variables</td>
<td></td>
</tr>
<tr>
<td>Motor vehicle crash</td>
<td>20 (24%)</td>
</tr>
<tr>
<td>Falls</td>
<td>20 (24%)</td>
</tr>
<tr>
<td>Work-related accident</td>
<td>14 (17%)</td>
</tr>
<tr>
<td>General accidents</td>
<td>30 (35%)</td>
</tr>
<tr>
<td>Type of injury</td>
<td></td>
</tr>
<tr>
<td>Contusion</td>
<td>27 (32.1%)</td>
</tr>
<tr>
<td>Sprain/strain</td>
<td>25 (29.8%)</td>
</tr>
<tr>
<td>Fracture/dislocation</td>
<td>20 (23.8%)</td>
</tr>
<tr>
<td>Open wound</td>
<td>9 (10.7%)</td>
</tr>
<tr>
<td>Crushing injury</td>
<td>2 (2.4%)</td>
</tr>
<tr>
<td>Burn</td>
<td>1 (1.2%)</td>
</tr>
<tr>
<td>Received preinjury help for emotional or mental health problems</td>
<td>42 (50%)</td>
</tr>
</tbody>
</table>

Continuous variables are presented as mean (±SD), dichotomous variables are presented as n (%). GED = general education diploma; PTSD = posttraumatic stress disorder.

Figure 2 Rates of PTSD and depression in the ED and at 6 weeks postinjury (N = 84). PTSD = posttraumatic stress disorder; PTSD + = PTSD positive; Depression + = depression positive.
Similar findings regarding positive PTSD screens in ED patients have been reported by Downey and colleagues\(^{17}\) (35% of their sample of children and adults) and by Richmond and colleagues\(^ {31}\) (36% of adults with minor injuries); higher rates (48.9%) were reported by Hunt and colleagues\(^ {32}\) in hospitalized trauma patients. Characteristic symptoms of PTSS include intrusions (e.g., nightmares, flashbacks), avoidance of injury-related reminders, negative alterations in cognitions and emotions, and altered physiological arousal and reactivity.\(^ {22}\) The presence of these symptoms may negatively impact functioning and recovery, even in the absence of meeting criteria for full PTSD.\(^ {33,34}\) Two recent studies also highlight the critical impact of early postinjury PTSD on long-term psychological recovery. First, among an Australian

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Pearson Correlations ((N = 79))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>Age</td>
</tr>
<tr>
<td>Age</td>
<td>—</td>
</tr>
<tr>
<td>Pain score</td>
<td>0.048</td>
</tr>
<tr>
<td>6-week PTSS</td>
<td>0.197</td>
</tr>
<tr>
<td>6-week depressive symptoms</td>
<td>0.232*</td>
</tr>
<tr>
<td>6-week QOL</td>
<td>0.313**</td>
</tr>
</tbody>
</table>

6-week PTSS assessed with the PCL; 6-week Depressive symptoms assessed with the PHQ-8. PTSS = posttraumatic stress symptoms; QOL = quality of life. *\(p < 0.05\); **\(p < 0.01\).

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Summary of Hierarchal Regression Analysis Demonstrating the Association of Positive Screening With Development of PTSD Symptoms at 6 Weeks ((N = 79))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>Outcome: 6-Week PTSD Symptoms</td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.20</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>0.17</td>
</tr>
<tr>
<td>ED pain score</td>
<td>0.26</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.16</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>0.13</td>
</tr>
<tr>
<td>ED pain score</td>
<td>0.21</td>
</tr>
<tr>
<td>PTSD + screen in ED</td>
<td>0.24</td>
</tr>
</tbody>
</table>

\(B\) = standardized beta coefficient; \(SE\ B\) = standard error of \(B\). PTSD = posttraumatic stress disorder; PTSS = PTSD symptoms; PTSD + screen = PTSD-positive screen.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Summary of Hierarchal Regression Analysis Demonstrating the Association of Positive Screening With Development of Depression Symptoms at 6 Weeks ((N = 79))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>Outcome: 6-week Depression Symptoms</td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.23</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>0.18</td>
</tr>
<tr>
<td>ED pain score</td>
<td>0.31</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.21</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>0.18</td>
</tr>
<tr>
<td>ED pain score</td>
<td>0.23</td>
</tr>
<tr>
<td>Depression + screen in ED</td>
<td>0.36</td>
</tr>
</tbody>
</table>

\(B\) = standardized beta coefficient; \(SE\ B\) = standard error of \(B\). Depression + screen = depression-positive screen.
cohort of patients injured in road traffic crashes, those with an early diagnosis of PTSD (rather than depression or anxiety) were significantly more likely to report a psychological diagnosis at 2 years postinjury.35 Next, among hospitalized injury patients, those who reported at least subsyndromal PTSD at 3 months postinjury were significantly more likely to report poor QOL at 12 months compared to those who never developed PTSD, even after controlling for preinjury QOL, pain, and depression.36

Regarding depression risk, this sample screened positive for depression at a much higher rate (76%) than reported in previous studies conducted among ED patients with minor injuries31 and hospitalized trauma adults32 (37%–44%). The presence of depressive symptoms after injury further complicates recovery and may contribute to poor QOL and long-term disability.14,37

Our ED screening rates may be higher given the brief period of time that had passed between the injury and symptom assessment; symptoms are naturally high immediately postinjury and tend to diminish over time. A probable explanation also resides in the nature of the PAS screening tool; whereas all 10 items are used to create the PTSD risk score (with a cutoff of 16), five of these same items are used to calculate the depression risk score with a low cutoff of 4. Consequently, the PAS items may be more specific to PTSD than to depression risk. To this end, in the development of the PAS, O’Donnell and colleagues19 reported a low positive predictive value of 24% (i.e., the probability that a subject who screens positive for the disorder actually has the disorder), a limitation that likely contributes to our high positive screen rate. Further, the low cutoff score for depression compared to PTSD increases the sensitivity to screen positive for depression risk; namely, a patient must only endorse one item at the highest level (each item is scored 0–4) on the PAS to screen positive for depression. Given that half of the sample endorsed receiving help for mental health problems in the past (an item which overlaps with the first question on the PAS), it is likely that many patients met criteria for depression risk by endorsing this item alone. It is important to keep in mind that the PAS was designed to serve as a screener and identify those who may benefit from monitoring and reassessment; outcomes should be assessed via interview-administered criterion standard diagnostic tools to accurately estimate the utility of the screening instrument.

Our data also revealed that QOL is lower in individuals at risk for postinjury depression and PTSS, which is consistent with prior research that found a negative correlation between QOL scores and the presence of PTSS.21 These findings along with previous data that display the association between mental health and QOL with disability over time14 suggest that

<table>
<thead>
<tr>
<th>Variables</th>
<th>Outcome: 6-week Quality of Life</th>
<th>B</th>
<th>SE B</th>
<th>95% CI</th>
<th>p-value</th>
<th>ΔR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Model #1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.30</td>
<td>0.16</td>
<td>-0.85 to -0.19</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital admission</td>
<td>-0.33</td>
<td>6.93</td>
<td>-38.38 to -10.76</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED pain score</td>
<td>-0.42</td>
<td>0.75</td>
<td>-4.86 to -1.88</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.26</td>
<td>0.16</td>
<td>-0.79 to -0.14</td>
<td>0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital admission</td>
<td>-0.30</td>
<td>6.87</td>
<td>-36.07 to -8.70</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED pain score</td>
<td>-0.38</td>
<td>0.74</td>
<td>-4.57 to -1.59</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD + screen in ED</td>
<td>-0.19</td>
<td>4.11</td>
<td>-16.66 to -0.27</td>
<td>0.043</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2: Model #2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.28</td>
<td>0.15</td>
<td>-0.81 to -0.18</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital admission</td>
<td>-0.33</td>
<td>6.70</td>
<td>-38.10 to -11.38</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED pain score</td>
<td>-0.37</td>
<td>0.74</td>
<td>-4.45 to -1.50</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression + screen in ED</td>
<td>-0.23</td>
<td>4.54</td>
<td>-20.35 to -2.24</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model #1 tests a positive PTSD screen as a predictor of QOL; Model #2 tests a positive depression screen as a predictor of QOL. Step 1 including the covariates is the same in both models. B = standardized beta coefficient; SE B = standard error of B.

PTSD = posttraumatic stress disorder; PTSD + screen = PTSD-positive screen; depression + screen = depression-positive screen.
disability is not solely a consequence of physical insult but also of early psychological symptoms. This novel implementation of the T-QoL in ED patients showed utility in short-term assessments; further research is needed to examine its usefulness in long-term follow-ups.

We found that the covariate of hospitalization was related to QOL, but not with PTSS or depressive symptoms. Patients admitted to the hospital may have longer recovery periods and physical limitations that can lead to reduced functioning across both psychological and physical domains that are later reflected on lower QOL scores. This study also confirmed that ED pain score is an important factor to consider when assessing risk for mental health sequelae and QOL, even among patients with minor physical injury. Among hospitalized patients after a traumatic orthopedic injury, Archer et al.38 similarly found that increased pain at hospital discharge was associated with depression and PTSD.

Taken together, ED screening allows for the identification of those at risk for PTSD close to the point of injury and provides an opportunity to monitor symptoms and administer interventions to prevent both psychological and physical health complications among a high-risk group. Although recent efforts on identifying risk factors for postinjury psychopathology have led to the development of predictive screeners to detect early psychological symptoms, research among acutely injured ED patients is limited. This lack of specificity among ED patients may be problematic given the many distinctions between hospitalized injured patients and those not requiring hospital admission, including the types of potentially traumatic events that contribute to hospitalization, the procedures likely involved in treatment for hospitalized patients, and the time between the event and assessment.32

Regarding screeners for hospitalized patients, Hunt et al.32 developed the Injured Trauma Survivor Screen, a nine-item binary (yes/no) response questionnaire ideal for a brief inpatient bedside evaluation of posttrauma PTSD and depression risk among hospitalized trauma patients. Although Richmond et al.31 developed and evaluated a screening tool for PTSD and depression designed for ED patients with minor injuries, the authors designed the scale to be administered to patients within 2 weeks posttrauma. It is not practical to administer this scale in the ED setting because some items would be difficult for ED patients to answer due to the proximal timing of the event to the assessment (e.g., “Has someone responded badly when you told them about what happened?” “Have you wanted to (or tried hard to) stay away from reminders of the event?” “Have you been staying away from people, even people you are usually close to?”). As such, the content of this survey limits its use in the acute setting given that the answers to these questions require additional processing time between the injury and the assessment of symptoms. Finally, although Mason et al.39 developed a brief screening tool for PTSD, depression, and anxiety that is easy to administer in the acute setting, the high false-positive rate led the authors to conclude that the tool was not cost effective or acceptable.

The majority of extant research in acute settings has tested established screeners for PTSD, as opposed to creating a new instrument for acute populations. Specifically, Walters and colleagues40 assessed the utility of the Trauma Screening Questionnaire (TSQ) in ED patients, but their sample was limited to assault-related injuries, and the TSQ should only be used 3- to 4-weeks posttrauma to allow for normal recovery processes to take place. Further, the TSQ only includes arousal and experiencing items; the avoidance cluster of DSM-IV is not represented.40 Similarly, the brief 4-item Primary Care-Posttraumatic Stress Disorder screener has been used among patients hospitalized after acute injury;41,42 however, this screener is anchored solely to posttrauma symptoms, provides a binary screen (yes/no) rather than assessing the severity of symptoms, and requires additional time for processing of the event (e.g., there is an item that asks about nightmares from the event).

Although ED patients not requiring hospitalization represent a significant proportion of injured patients and may benefit from further monitoring and follow-up care, these patients are often overlooked for psychological screening.8 Consistent with prior research suggesting a weak and inconsistent relationship between the severity of injury and psychological consequences, a recent study among 460 patients admitted to a Level I trauma center revealed that no significant relationships emerged between injury severity score and symptoms of PTSD, depression, pain, and physical and mental health throughout 6 months postinjury.43 These results support screening for PTSD in those treated in the ED, but whose injuries may not be sufficient to warrant hospitalization. However, the results of a 2014 trauma center survey conducted in Level I and II trauma centers throughout the United States
n = 391) revealed that early psychological screening procedures are lacking in general, as only 7% of trauma centers routinely screen for PTSD. Moreover, psychological screening is often limited to: 1) the inclusion of hospitalized patients with severe trauma and 2) a focus on long-term outcomes (rather than acute outcomes) with follow-ups that occur months and years postinjury.

In summary, existing screening tools do not assess pre- and peritrauma risk factors and would not be practical to administer in ED patients to assess symptoms from the index injury. Our study built on prior research by testing the utility of using the PAS among injured ED patients; the advantages of the PAS include its assessment of symptom severity to calculate risk and its inclusion of pre-, peri-, and posttrauma factors that can be used in close proximity to the index injury. The PAS also includes an item that reflects a history of psychological problems (factored into the risk score for both PTSD and depression), an important predictor variable of recovery after subsequent injury. Further, our study extended prior research by including an assessment of risk for PTSD and depression within 24 hours postinjury among adults with any trauma-related chief complaint; we also reevaluated the presence of psychological symptoms at a short follow-up time frame after discharge (6 weeks), allowing for the early identification of individuals at risk who may benefit from acute intervention. Although the PAS was sufficient for identifying at-risk patients, future research is warranted to develop a screen specifically to meet the needs of ED patients.

**LIMITATIONS**

Several limitations should be considered: As this was a feasibility study with limited personnel and compensation resources, our retention rate was low (59%) and therefore introduces the possibility of bias. Additionally, females and subjects with positive ED screening for depression were more likely to be retained, and this may weaken the external validity of our results. Low retention has also been reported in prior behavioral studies performed in acutely injured patients, with dropout rates ranging from 41% to 50% among acute injury survivors with follow-up 4 to 12 weeks postinjury. Given the myriad challenges that ED subjects are faced with after acute injury, this population requires a higher level of resources devoted specifically to retention. In the context of low retention rate, small sample size, and wide confidence intervals, we caution that these results are preliminary and must be replicated with larger samples with resources available to increase retention.

In addition, the PAS was designed to reflect the DSM-IV criteria and does not include the negative cognition and mood symptom cluster in DSM-5; an updated screener may be beneficial for more accurate assessments. Another important limitation is the short follow-up period to assess QOL. The construct of QOL may take longer than 6 weeks to evolve and its relationship with depression and PTSD may change over time; longer follow-up assessments would allow recognition of a better relationship between PTSS and QOL. Further, our primary outcomes of PCL and PHQ-8 are self-report and reflect the potential presence and severity of psychological symptoms; they cannot indicate a diagnosis. Finally, injury severity scores are not available for these patients; although the scores are likely low, future research among injured ED patients should consider adjusting for injury severity in models.

Despite these limitations, this is the first study to assess risk for depression and PTSD using the PAS in the acute setting of the ED after general injury and to include mostly minor injuries, which may be a good representation of the patients commonly seen in the EDs throughout the country. For example, 90% of MVC patients are not admitted to the hospital for their injuries. Given the high rates of risk and posttraumatic psychological symptoms found in this study, it is worth further investigating the utility of routine screening during ED trauma care.

Early cognitive behavioral interventions such as education, exposure, cognitive restructuring, anxiety management, and even Internet–virtual reality interventions show promise in the management of postinjury psychological symptoms; further research is needed to determine efficacy, particularly among injured ED patients.

**CONCLUSION**

It is feasible to identify ED patients with minor injury at risk for future psychological effects of trauma using brief screens like the Posttraumatic Adjustment Scale in the ED. The early identification of individuals at risk may allow further monitoring of psychological symptoms and/or intervention, both of which may promote a better recovery and QOL. These results add to a growing literature demonstrating the
importance of understanding the psychological and emotional responses that contribute to the processing of stress and pain after minor physical injury; our findings support and highlight the need for future research among larger samples of ED patients.

The authors acknowledge R. Kurtz, J. T. Kristen, J. M. Rupp, A. Keller, S. Diulus, D. Wang, and M. Kelly recruited and interviewed subjects for this study. R. Kurtz also prepared the data for analysis.

References


43. Lewis GC, Platts-Mills TF, Liberzon I, et al. Incidence and predictors of acute psychological distress and
Emergency Medicine in the #MeToo Era

Catherine A. Marco, MD1, Joel M. Geiderman, MD2, Raquel M. Schears, MD, MPH, MBA3, and Arthur R. Derse, MD, JD4

ABSTRACT
Sexual harassment is a serious threat to a safe and productive workplace. The emergency department (ED) environment poses unique threats, including stress, time constraints, working in close physical proximity, and frequent personal contacts with staff, colleagues, consultants, and difficult patients. Sexual harassment must be recognized and addressed in individual cases, in policy and in law, to protect staff members and patients.

This article addresses the scope of the problem of sexual harassment known to date. It describes the ED environment and culture and why they may be conducive to harassment or abusive behavior. The authors examine relationships among staff, legal and regulatory issues, and strategies for prevention and remediation of inappropriate behavior. The article ends with a call for future research.

ENSURING a professional and respectful working environment in medicine is crucial to patient care; the learning; the environment; and the success, health, and well-being of health care providers. Sexual harassment and related issues have recently received increased attention as serious threats to the professional workplace. Since medicine, like most professions, strives to regulate and govern itself, professional organizations, publications, health care administrators, and individual practices have a duty to address the issue. This article will focus on the definition, history, and law of sexual harassment; the emergency department (ED) environment and culture; sexual harassing and other inappropriate behaviors in the ED; proper, improper, and questionable relationships among staff; prevention and remediation of inappropriate behavior; and future directions in emergency medicine (EM), including a call for future research.

HISTORY

The term “sexual harassment” is often attributed to Mary Rowe, PhD, Chancellor for Women at Work at the Massachusetts of Technology (MIT), who used the term in 1973 in her work entitled, “Saturn’s Rings.” Others claim that credit is due to author and journalist, Susan Brownmiller, and other activists who brought this issue to the forefront more than 45 years ago. Another seminal work is that of Catherine MacKinnon in her book, Sexual Harassment of Working Women (1979). The concept advanced and regulations and case law further defined this behavior. The problem of inappropriate sexual behavior by physicians has been recognized as a deplorable practice in medicine for ages, as noted in the Oath of Hippocrates: “Whatever houses I may visit, I will come for the benefit of the sick, remaining free of all intentional injustice, of all mischief and in particular of sexual relations with both female and male persons, be they free or slaves.”

As early as 1998, sexual harassment was called out for its prevalence in medicine but received little attention. Increased public awareness of the societal scope of the problem has occurred recently. In October 2017, reports of the alleged abuse by media magnate

From the 1Department of Emergency Medicine, Wright State University, Dayton, OH; 2Ruth and Harry Roman Emergency Department, Department of Emergency Medicine, Cedars-Sinai Medical Center, Los Angeles, CA; the 3Department of Emergency Medicine, University of Central Florida College of Medicine, Orlando, FL; and the 4Center for Bioethics and Medical Humanities and Department of Emergency Medicine, Medical College of Wisconsin, Milwaukee, WI.

Received March 31, 2019; revision received June 2, 2019; accepted June 3, 2019.

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

Author contributions: drafting of the manuscript—CAM, JMG, RMS, and ARD; and critical revision of the manuscript for important intellectual content—CAM, JMG, RMS, and ARD.

Supervising Editor: Tammie E. Quest, MD.

Address for correspondence and reprints: Catherine A. Marco, MD; e-mail: cmarco2@aol.com.

ACADEMIC EMERGENCY MEDICINE 2019;26:1245–1254.

© 2019 by the Society for Academic Emergency Medicine
doi: 10.1111/acem.13814

ISSN 1553-2712

1245
Harvey Weinstein called public attention to the problem of harassment that is all too common and often the result of abuse of power. Following Weinstein’s rapid fall from grace, many others in the fields of film, television, business, politics, sports, academia, and medicine have been accused of sexual misconduct and many have been removed from their jobs or even gone to jail. The majority, but not all, of these cases involved alleged male-on-female abuse.

Recently, Pope Francis admitted for the first time that “priests and even bishops” had sexually abused nuns and subsequently held a 4-day conference at the Vatican to address sexual abuse. Whether or not these actions were a direct result of the #MeToo movement is unknown, but temporally it follows a pattern of alleged victims speaking out and institutions and organizations forced by the disclosures to examine their environments and practices.

Adding to this milieu is the 2018 report by the National Academies of Sciences, Engineering, and Medicine (NASEM) on the harassment of women in academia. The report was a pivotal account of the pervasiveness of this behavior, describing it as common, especially in medicine, with sobering accounts specifically among medical trainees. One quote from a resident in the NASEM report was particularly disturbing as it described sexual harassment as part of a continuum of what residents were expected to endure: “But the thing is about residency training is everyone is having human right violations. So it’s just like tolerable sexual harassment (sic)”. The NASEM report spurred a cluster of articles on this problem across various fields of medicine, not including EM. The NASEM report did single out surgery and EM as potentially problematic areas during residency due to their isolated “hierarchical and authoritative workplaces” although this claim was based on a report from 1998. In that study, women physicians who had completed residency training and were in practice reported the same rates of harassment as in other fields, although none of it is acceptable. As well, the National Institutes of Health (NIH) recently apologized for its failure to address sexual harassment. All these morally repugnant acts run counter to the moral norms to which modern Western societies subscribe.

Increasing national attention to the pervasiveness of sexual harassment in the workplace led to the “MeToo” movement, originally founded in 2006, and gained momentum through social media in recent months with the popular “#MeToo hashtag.” The vision of this movement is “...to address both the dearth in resources for survivors of sexual violence and to build a community of advocates, driven by survivors, who will be at the forefront of creating solutions to interrupt sexual violence in their communities.” A recent study identified that Internet searches for sexual harassment or assault were 86% higher than expected from October 2017 through June 2018, reaching record highs. Recently, TIME’S UP Healthcare was founded to address harassment and protect from retaliation and to promote equitable compensation for women and men at every level in the healthcare industry, specifically asking organizations to confront obstacles to prevent and solve these problems.

**THE ED ENVIRONMENT**

The ED environment is distinctive among practice settings, including elsewhere in the hospital, clinics and offices although there are some similarities to operating rooms and catheterization labs. The ED is characterized by large staffs, often working side by side; not uncommonly a stressful environment and sometimes a casual atmosphere, it can be described as a “fishbowl” of openness but is often practiced behind closed drapes or doors. Emergency physicians and coworkers work shoulder to shoulder and conversations often go on into the late hours of night or early morning, when one’s guard may be down, perhaps due to fatigue, stress, or the unique ED environment. Physicians and staff may at times use humor, inappropriate language, or inappropriate behavior as a coping mechanism. A recent study of New York EDs found that most residents had experienced verbal harassment (97%), verbal threats (78%), and/or sexual harassment (52%). Another recent study of 10 EM residencies found that most residents had experienced verbal abuse (86%) and verbal threats (68%) and significant numbers had experienced physical threats (50%), physical attacks (26%), sexual harassment (23%), or racial harassment (26%). In this study, women were more likely to encounter sexual harassment (37% compared to 8%, p < 0.001). A recent study identified significant issues of disruptive behaviors in the ED. In this study over 50% of respondents had witnessed disruptive behavior by physicians and/or nurses. In addition, respondents believed that these behaviors could be linked to...
adverse events, medical errors, poor quality, compromises in patient safety, and patient mortality. Another recent study demonstrated significant incidence of horizontal violence (malicious behavior of health care workers against each other), ranging from 1% to 34%.19

This phenomenon is not limited to emergency physicians. Emergency medical services (EMS) providers have also experienced harassing behaviors. A 1999 study of EMS personnel found 69% of women in EMS reported workplace sexual harassment by supervisors, coworkers, and patients.20

An attorney-EMT-P reported in 2018 that his law firm has been dealing with EMS harassment complaints for 30 years and has seen a recent increase in harassment complaints.21 News accounts of EMS personnel filing workplace harassment complaints show a public discussion of concern.22–24

The ED environment also includes many patients who are intoxicated by drugs or alcohol.25,26 Mental health conditions, addiction, and social issues are also common and may in some cases lead to disrespectful, abusive, or disinhibited behavior. This behavior may be more common in the ED where preexisting and ongoing relationships usually don’t exist. A recent study demonstrated that emergency physicians in Michigan have experienced verbal threats (75%) or physical assault (28%), and women physicians were more likely to have experienced physical assault.27 Another study identified a high incidence of workplace violence against ED workers, with sexual harassment the only type of violence more prevalent against females.28

In summary, the ED has the potential for harassment or abuse of providers as well as by providers. Educational, institutional, and departmental efforts must address both situations.

INAPPROPRIATE BEHAVIOR

Sexual harassment can be categorized as gender-based harassment (including use of suggestive language, crude jokes, or sexist comments or sharing sexist imagery); unwanted sexual attention (verbal or physical); and sexual coercion (whereby career advancement is contingent upon sexual favors).29,30 Research reveals that more women than men report having experienced sexual harassment when provided the full definition which includes an array of verbal and nonverbal behaviors that “convey hostility, objectification, exclusion, or second-class status about members of one gender.”31

One form of sexual harassment is “inappropriate touching.” Some may disagree about the definitions of appropriate and inappropriate touching in the workplace. Because of wide variability in what physical contact is welcomed or unwelcome, most physical contact should be avoided or limited in the workplace. Handshakes are an exception, as commonly accepted and in many cases expected. In general, personal space should be respected, and in most cases, physical contact with coworkers should be avoided.

Hugging is a complicated issue. One source states that hugging should be reserved for “family time.”32 In the authors’ opinion, depending on specific relationships and cultures, hugging between two old friends or colleagues may be allowable if welcomed by both parties and occasional. Contacts that should be avoided include such things as touching a pregnant woman’s gravid abdomen; buttocks pats; shoulder and neck massages; or the caressing of heads, ears, or hair. Although pats to the head, back, or shoulder may seem innocuous, the physical proximity can be misconstrued and has resulted in sexual harassment claims in healthcare settings.33 In general, it is also wise to recognize that the use of words and kindness can make an impact without touching and may be the safer strategy to deliver care and empathy in workplace.

PERSONAL RELATIONSHIPS AMONG STAFF

Personal relationships among staff members are controversial. Some believe that policies that discourage or disallow fraternization and personal relationships can reduce sexual harassment in the workplace, while others believe that such restrictions on relationships violate personal freedom. Many do not object to personal relationships between two single individuals but object if one of the parties is married to another person or in a committed relationship. Relationships based on a power mismatch can be fraught with personal and professional harm. Close personal relationships can be a distraction in the workplace, especially in the ED where teamwork is essential. Although some institutions prohibit romantic relationships between staff,34,35 such policies may be unrealistic and are often violated. When such relationships do occur, it should be required that personal relationships
remain professional in the workplace and do not interfere with patient care and teamwork. Some institutions have specific policies to govern relationships in the workplace. For example, the NIH has such a policy.\textsuperscript{36}

Personal relationships (including romantic and/or sexual) between individuals in inherently unequal positions, where one party has real or perceived authority over the other in their professional roles, may be inappropriate in the workplace and are strongly discouraged. If such a relationship exists or develops, it must be disclosed. This applies to all individuals in the NIH community, including employees, contractors, students, trainees, and fellows and includes anyone who holds a position of authority or perceived authority over another individual from a scientific or administrative perspective.

Personal relationships with patients or former patients are susceptible to negative consequences. Relationships with patients clearly involve a mismatch of power and vulnerability and may compromise the patient–physician relationship. The American Medical Association Code of Ethics provides guidance on professional self-regulation, including the following statement regarding relationships with patients:

Romantic or sexual interactions detract from the goals of the patient–physician relationship and may exploit the vulnerability of the patient, compromise the physician’s ability to make objective judgments about the patient’s health care and ultimately be detrimental to the patient’s well-being. A physician must terminate the patient–physician relationship before initiating a dating, romantic, or sexual relationship with a patient.\textsuperscript{37}

**TEACHER AND LEARNER RELATIONSHIPS**

Relationships that have a clear conflict of interest or power mismatch should be avoided. Teacher–learner personal relationships are not appropriate because of the inherent vulnerability of the learner. Supervisor and subordinate relationships are similarly generally inappropriate. Such inappropriate relationships may compromise the role of educator or supervisor, may harm the vulnerable individual, and may negatively affect the workplace.

**ORGANIZATIONAL POLICIES AND GUIDELINES**

Ensuring a professional and respectful working environment in medicine is crucial to patient care; the learning environment; and the success, health, and well-being of health care providers.\textsuperscript{10,20,38} The medical profession has an obligation to patients and providers to recognize and eliminate sexual harassment.\textsuperscript{39} Several organizations have provided guidance on the topic of sexual harassment including The American College of Emergency Physicians has stated in policy:\textsuperscript{40}

The American College of Emergency Physicians advocates tolerance and respect for the dignity of each individual and opposes all forms of discrimination against and harassment of patients and emergency medicine staff on the basis of an individual’s race, age, religion, creed, color, ancestry, citizenship, national or ethnic origin, language preference, immigration status, disability, medical condition, military or veteran status, social or socioeconomic status or condition, sex, gender identity or expression, sexual orientation, or any other classification protected by local, state, or federal law.

The American Medical Association has a policy on the issue that includes reporting, investigations, disciplinary actions, and confidentiality:

The AMA is committed to a zero tolerance for harassing conduct at all locations where AMA delegates and staff are conducting AMA business. This zero tolerance policy also applies to meetings of all AMA sections, councils, committees, task forces, and other leadership entities (each, an “AMA Entity”), as well as other AMA-sponsored events.\textsuperscript{41}

The Liaison Committee on Medical Education has published a general standard regarding the learning environment.\textsuperscript{42} However, despite widespread education and standards to ensure a professional learning environment, breaches still occur. The American Association of Medical Colleges graduation questionnaire includes assessment of the learning environment. In 2018, 5% of graduating U.S. medical students reported that they have been subjected to unwanted sexual advances (once, occasionally, or frequently).
Some (16.5%) have been subjected to offensive sexist remarks/names.\textsuperscript{43} Among academic faculty, women are more likely to report having experienced sexual harassment (30%, compared to 4% of men).\textsuperscript{44}

The Society for Academic Emergency Medicine (SAEM) has stated in policy:

Academic faculty in emergency medicine, as elsewhere, have an obligation to maintain appropriate boundaries in their relationships with those they supervise and teach, in particular, residents, fellows, and medical students, but also the many others who come to the emergency department to learn. Even the appearance of impropriety should be avoided.

This policy also suggests a balanced approach to management of such issues:

The Society for Academic Emergency Medicine expects each department to have a mechanism to address concerns and complaints about violations of appropriate boundaries by their faculty. Often, such a mechanism will be provided by the residency’s parent institution. When no such institutional mechanism exists, the residency should implement one, with clear channels for the learner to safely lodge the complaint, with no fear of retribution, and a clear method for assessing and responding to such complaints.\textsuperscript{45}

**LEGAL ISSUES**

To better understand sexual harassment in EM, it is important to review the foundation of the law of sexual assault and the development of the law of sexual harassment. American law has long recognized several classes of unwanted intrusions as violations of law. An act of unconsented restraint or offensive touching can be a form of battery,\textsuperscript{46} and an intentional act that creates an apprehension in another of an imminent, harmful, or offensive contact can be deemed an assault.\textsuperscript{47} Sexual assault is any nonconsensual sexual act proscribed by federal, tribal, or state law and includes when the victim lacks capacity to consent.\textsuperscript{48}

State laws commonly define various degrees of sexual assault ranging from first to fourth degree, inversely ordered in degree of severity. For instance, in California, fourth-degree sexual assault is defined as sexual contact and with one of the following: 1) force and/or coercion, 2) the victim incapacitated in some form, or 3) the perpetrator in a position of power over the victim.\textsuperscript{49} Sexual contact is defined as intentional touching of another person's intimate parts (groin or buttocks or the breasts of a female) or the clothing covering the intimate parts of another person with the aim of sexual gratification or arousal. Punishment for the crime of sexual assault may result in imprisonment and fines,\textsuperscript{38} and potential civil judgments for assault, battery,\textsuperscript{50} and intentional infliction of emotional distress.\textsuperscript{51} Sexual assault also encompasses sexual abuse consisting of verbal, visual, or noncontact behavior that forces a person to join in unwanted sexual activities or attention. Examples include voyeurism or peeping (viewing private sexual acts without consent), exhibitionism (public exposure), sexual harassment or threats, forcing someone to pose for sexual pictures, and sending someone unwanted texts or “sexts” (texting sexual photos or messages).\textsuperscript{52} Sexual assault is most commonly a state crime that must be proven by prosecutors beyond a reasonable doubt.\textsuperscript{53}

Sexual harassment is a more recently developed legal concept delineated as a form of discrimination. This has been elaborated through rulings based on Title VII of the Civil Rights Act of 1964.\textsuperscript{54} Sexual harassment consists of unwelcome sexual advances; requests for sexual favors; and other verbal or physical conduct of a sexual nature where the conduct explicitly or implicitly affects an individual’s employment, unreasonably interferes with an individual’s work performance or creates an intimidating, hostile, or offensive work environment. It is important to note that even though the most common sexual harassment may be by male supervisors of women employees, harassers and victims may be women or men, and both may be of the same sex.\textsuperscript{55} Harassers can be any coworker, or even agents or nonemployees, of a “covered entity.”\textsuperscript{44} The victim can be anyone affected by the offensive conduct, and the law can be violated even if there is no economic injury or resulting harm to employment.\textsuperscript{44}

The U.S. Supreme Court first recognized sexual harassment that results in a hostile work environment as a form of discrimination under Title VII in Meritor Savings Bank v. Vinson in 1986.\textsuperscript{56} Since sexual harassment is a federal violation, charges are made to the U.S. Equal Employment Opportunity Commission (EEOC) that has authority to investigate, using subpoenas, if necessary, if there is reasonable cause to believe harassment occurred. In contrast to the evidentiary
standard for sexual assault (beyond a reasonable doubt), the evidentiary standard for sexual harassment is the preponderance of the evidence (i.e., more likely than not). Plaintiffs may be awarded reinstatement, promotion, lost wages, damages for emotional distress, medical expenses, punitive damages, and attorney fees.57

Physicians can be civilly liable for harassment of employees, fellow health care workers, trainees, or patients.58,59 These infractions and subsequent fines and damage judgments, if any, are generally not covered by malpractice insurance. Sexual harassment is a form of sexual misconduct subject to scrutiny and sanction by state medical boards and other credentialing bodies that may result in disciplinary action, limitation of license and subsequent loss of board certification by the American Board of Emergency Medicine.60

MAINTAINING A SAFE WORKPLACE: PREVENTION, INVESTIGATION, REMEDIATION, AND NONRETALIATION

The hospital, the medical staff, and the departmental leadership should take steps to prevent harassment and should have zero tolerance if it occurs. Keys to doing this are clear policies that define a harassment free workplace, clear procedures for addressing policy transgressions, and education and training on harassment. These elements should focus on the cultivation, maintenance, and growth of a positive, safe, efficient, and harassment free workplace.

Despite the often chaotic conditions of the ED work environment (sometimes portrayed on television as a turbulent, jungle-like atmosphere), the tone should remain serious with the focus on maintaining the welfare of patients and coworkers. This may help reduce or eliminate inappropriate jokes or behaviors, which may constitute harassment or contribute to an environment where harassment occurs.

Once a complaint is filed, it must be fully investigated.61,62 A full discussion of the institutional process is beyond the scope of this paper but much of it is prescribed by federal and state laws and should be contained in institutional polices. Complaints should promptly be passed along to the human resources (HR) and legal departments who should engage in fact finding, usually through interviews. This should be done by professionals in the field. False allegations are reprehensible; can needlessly damage reputations, often permanently; and may also undermine effective responses to legitimate complaints. Because of the subjective nature of these allegations, direct proof of events may be difficult to establish. It is best to gather facts as close to the time of the alleged event as possible.

WHEN THE ALLEGED VICTIM IS A PHYSICIAN, ALLIED HEALTH PROFESSIONAL, OR OTHER AGENT OF THE HOSPITAL

If the situation involves a supervisor, a coworker, or a workplace issue, the process described above should be initiated. In some situations, the accused party may be put on leave and in general, if possible, it is best to limit interactions between the parties (in a nonpunitive fashion) at least until the matter is resolved. As mentioned, the HR department, legal department, and other professionals should be involved. Department leaders should not do this on their own. In some cases, the EEOC will be involved.63 This organization offers free mediation services. Ideally, both parties will agree to a remediation plan. A safe reporting environment is essential and it must be assured that retaliation is not allowed no matter how the investigation comes out.

As noted earlier, it is not uncommon that a physician or other provider will experience sexual harassment or abuse by a patient or family member who may be intoxicated, psychiatrically disturbed, a sociopath, or simply just abusive. In such cases, physicians should seek protection from physical harm from security officers or others. The provider may ask another provider to take over the care of the patient and once EMTALA requirements are satisfied the patient or other party may be removed from the premises.64

WHEN THE ALLEGED PERPETRATOR IS A PHYSICIAN, ALLIED HEALTH PROFESSIONAL, OR OTHER AGENT OF THE HOSPITAL

If a claim is brought against a physician, the same process will occur. While the law provides certain automatic protections for alleged victims, the same is not true for alleged perpetrators. Accused perpetrators may wish to seek their own independent legal advice. Parallel processes may occur at the level of the hospital, the medical staff, the employer or contracting entity, and
in the legal system. If inappropriate behavior has been found to occur, there may be mandatory reporting to state medical boards. For all of these reasons, physicians would be wise to avoid even the suspicion of inappropriate behavior. The authors strongly recommend the use and documentation of chaperones for the examination of any intimate body parts of both male and female patients, as well as documentation of the clinical reason for the examination.

If assault is alleged, and there is credible evidence, law enforcement may need to be involved and in some cases the event must be reported to Centers for Medicare and Medicaid Services as a “never” event. If a felony is established, physicians can expect to lose their licenses, medical staff privileges, and board certification.

**FUTURE DIRECTIONS**

Future directions to reduce or eliminate inappropriate behavior should include measurement of the problem through carefully constructed surveys, increased awareness through mandatory education, and more published reports and recommendations by those working in the field. Future research should better define the scope of the problem in the ED environment, including both witnessed and perceived incidence of harassment, types of harassment, and approaches to incidents. Additionally, solutions to the existing problems should be developed and investigated. Educational sessions on the subject by EM organizations may be helpful in raising awareness and proposing solutions.

**CONCLUSIONS**

Ensuring a professional and respectful working environment in medicine is crucial to patient care; to the learning environment; and to the success, health, and well-being of health care providers. Actions to promote a professional workplace in the ED should include professional behavior, appropriate communication and reporting of infractions, and ensuring that learners and other vulnerable staff feel comfortable reporting inappropriate behavior (Table I). Allegations must be investigated and appropriate actions taken in a fair and consistent application of justice.

**References**


20. Honeycutt L. Girl talk: an EMS educator reports her findings on the state of women in the industry. JEMS 1999;24:50–53.


42. Vigilant EM, Oliverio AL, Meeks LM. Sexual harassment and abuse: when the patient is the perpetrator. Lancet 2018;392:368–70.


52. NIH Policy Statement: Personal Relationships in the Workplace. Available at: https://hr.nih.gov/working-nih/


38. Freischlap JA, Faria P. It is time for women (and men) to be brave: a consequence of the #MeToo movement. JAMA 2018;319:1761–2.


58. Vance v. Ball State University, 133 S. Ct. 2434 (2013)


64. Emergency Medical Treatment and Active Labor Act. 42 USC 1395dd et seq.

Integrated Use of Conventional Chest Radiography Cannot Rule Out Acute Aortic Syndromes in Emergency Department Patients at Low Clinical Probability

Peiman Nazerian, MD, Emanuele Pivetta, MD, PhD, Simona Veglia, MD, Edoardo Cavigli, MD, Christian Mueller, MD, Alexandre de Matos Soeiro, MD, Bernd A. Leidel, MD, Enrico Lupia, MD, PhD, Claudia Rutigliano, MD, Desiree Wussler, MD, Stefano Grifoni, MD, and Fulvio Morello, MD, PhD, on behalf of the ADvISED Investigators

ABSTRACT

Objectives: Guidelines recommend chest radiography (CR) in the workup of suspected acute aortic syndromes (AASs) if the pretest clinical probability is low. However, the diagnostic impact of CR integration for the rule-in and rule-out of AASs is unknown.

Methods: We performed a secondary analysis of the ADvISED multicenter study. Emergency department outpatients were eligible if an AAS was clinically suspected. Clinical probability was defined with the aortic dissection detection risk score (ADD-RS). CR was evaluated blindly by a radiologist, who judged on mediastinum enlargement (ME) and other signs.

Results: In 2014 through 2016, a total of 1,129 patients were enrolled and 1,030 were analyzed, including 48 (4.7%) with AASs. ADD-RS/ME and ADD-RS/any CR sign (aCRs) integration were more accurate than ADD-RS alone (area under the curve = 0.8 and 0.78 vs. 0.66, p < 0.001). The sensitivity and specificity of the integrated strategies were 66.7% (95% confidence interval [CI] = 51.5% to 79.9%) and 82.5% (95% CI = 79.9% to 84.8%) for ADD-RS/ME and 68.8% (95% CI = 53.6% to 80.9%) and 76.5% (95% CI = 73.7% to 79.1%) for ADD-RS/aCRs, respectively. The sensitivity and specificity of CR per se were 54.2% (95% CI = 39.2% to 68.6%) and 92.4% (95% CI = 90.5% to 93.9%) for ME and 60.4% (95% CI = 45.3% to 74.2%) and 85.2% (95% CI = 82.9% to 87.4%) for aCRs. The agreement (κ) between attending physicians and radiologists for ME was 0.44 (95% CI = 0.35 to 0.54). ADD-RS/ME rule-in (ADD-RS ≤ 1 and ME-present, or ADD-RS > 1) applied to 204 versus 130 patients with ADD-RS > 1, including 14 with AAS and 60 false-positives (FP). ADD-RS/aCRs rule-in (ADD-RS ≤ 1 and aCRs-present, or ADD-RS > 1) applied to 264 patients, including 15 with AAS and 119 FP. ADD-RS/ME rule-out (ADD-RS ≤ 1 and ME-absent) applied to 826 (80.2%) patients, including 16 with AAS (33.3% of cases). ADD-
Acute aortic syndromes (AASs), including aortic dissection, intramural aortic hematoma (IMH), penetrating aortic ulcer (PAU), and spontaneous aortic rupture, are deadly cardiovascular emergencies. They are relatively rare conditions presenting with common and unspecific symptoms such as truncal pain, syncope, and neurologic deficit. Conclusive diagnosis of AASs requires urgent computed tomography angiography (CTA) of the chest and abdomen or transesophageal echocardiography (TEE), but pretest selection of patients is needed owing to the risks, costs, and hassles of advanced aortic imaging. Proper selection for CTA/TEE is particularly cumbersome in stable patients with nonsevere presentations. This is evident from the high rate of misdiagnosis and delayed recognition of AASs, in spite of increasing use of CT in the emergency department (ED). In this challenging scenario, first-level imaging of the thoracic aorta represents a key element for physicians when evaluating patients with AAS-compatible symptoms.

Conventional chest radiography (CR) allows partial visualization of the thoracic aortic contour and can detect pathologic findings in some cases of AAS. Most commonly, aortic dilatation results in mediastinum enlargement (ME) at the level of the aortic knob, but also additional signs of AASs can be observed on a CR. The test characteristics of CR for the diagnosis of AASs have been estimated mostly by retrospective case-control studies. Only one relatively small single-center prospective study was performed more than a decade ago, reporting modest sensitivity and specificity (64 and 86%). American and European guidelines recommend CR use for the diagnostic work-up of patients at low clinical probability of AAS (class of recommendation IIb, level of evidence C), i.e., when the pretest probability is not high enough to request immediate advanced imaging.

Despite the low level of evidence supporting its use and exposure of patients to a small but not negligible dose of ionizing radiations, CR is routinely applied to screen for aortic enlargement in ED patients with truncal pain. The aim of this study was to prospectively evaluate the diagnostic performance of CR for suspected AAS when integrated with standardized risk-stratification of patients. In particular, the study sought to measure the additive value of CR in patients at low clinical probability of AAS. Two clinical questions were specifically addressed: 1) can a negative CR support rule-out of AASs in patients at low clinical probability? and 2) how accurate and efficient is CR as a rule-in criterion for advanced aortic imaging in patients at low clinical probability?

METHODS

Study Design and Setting

This was a predefined secondary analysis of the ADVISED prospective multicenter diagnostic accuracy study (ClinicalTrials.gov, No. NCT02086136), on data from five centers (all tertiary hospitals) in four countries. The study complied with the Declaration of Helsinki and was approved by the local ethics committees. Written informed consent of participants was obtained.

Selection of Participants

From September 2014 to December 2016, consecutive outpatients aged ≥ 18 years presenting to the ED were eligible if they experienced one or more of the following symptoms dating ≤ 14 days: chest pain, abdominal pain, back pain, syncope, signs or symptoms of perfusion deficit and if an AAS was considered in differential diagnosis by the attending physician. Patients were included in the present cohort only if a CR was requested by the attending physician for aortic evaluation before the availability of advanced imaging or surgery. Exclusion criteria were primary trauma or unwillingness to participate in the study.

Index Visit

During the index visit in the ED, patients were evaluated and managed by one or more emergency physician(s). Patient eligibility was established by the attending physician during his/her standard evaluation, which included a complete physical examination and electrocardiogram recording. Structured data collection for pretest probability assessment was performed prospectively during the index ED visit by the
attending physician or a medical researcher. For each patient, a standardized case report form (CRF) was filled, to record prespecified variables from medical history, presenting signs/symptoms, and risk factors of AASs. If CRF data could not be recorded (i.e., non-verbal or critically ill patient status), the patient was excluded from the study. If limited pieces of data were unavailable, these were defaulted to negative.

The tool used to assess the pretest probability of AAS was the aortic dissection detection risk score (ADD-RS), endorsed by international guidelines on aortic diseases.\(^7\),\(^9\),\(^11\) Briefly, the ADD-RS is based on presence/absence of 12 risk factors classified in three categories. The ADD-RS of each patient was automatically calculated as the number of categories (0 to 3) where at least one risk factor was present. In compliance with the European Society of Cardiology guidelines, patients with one or more risk factors in 0 or 1 risk category (ADD-RS \(\leq 1\)) were classified at low clinical probability of AAS, while patients with one or more risk factors in more than category (ADD-RS > 1) were classified at high clinical probability of AAS.\(^9\)

Subsequent diagnostic and clinical decisions, including the decision to perform conclusive imaging methods and disposition for the patient’s hospitalization or ED dismissal, were determined by the attending physicians based on their clinical judgment and in compliance with local protocols after evaluation of all clinical data including CR.

**Chest Radiography**

The CR was performed by posterior–anterior and lateral views in upright patients and by anterior–posterior view in supine patients. The following machines were used: Practix 300 Bucky diagnostic equipment, Practix 33 plus, Digital Diagnost VM, Bucky Diagnost TH X-Ray and Optimus (Philips Medical Systems), Mobilett XP Hybrid and Ysio fully digital system (Siemens), and MobileArt plus Agfa CR-85 digitizer system and MobileDaRt Evolution with Canon CXDI 701C wireless detector (Shimadzu).

All CR films were later reviewed by a single board-certified radiologist from each center, blinded to clinical data, advanced imaging test results, and final diagnosis. The radiologists were different from those who interpreted the CR films for clinical purposes and used a uniform questionnaire to document the presence or absence of ME and other eight additional signs of AASs: poor definition or irregularity of the aortic contour, double aortic knob sign, inward displacement of aortic wall calcification by more than 10 mm, tracheal displacement to the right, displacement of a nasogastric tube, left-sided pleural effusion, suspected pericardial effusion, and left apical opacity. ME was defined as a maximum width of mediastinum \(\geq 80\) mm at the level of the aortic knob or a ratio of mediastinum to chest width \(> 0.25\) or ME based on subjective evaluation (Figure 1). The presence/absence of ME was also independently evaluated by the attending physician during the ED visit. This evaluation was reported on the CRF.

**Final Diagnosis**

The primary imaging method allowing conclusive diagnosis of AASs was chest and abdomen contrast-enhanced multidetector CTA (\(\geq 64\) row-detectors). Other imaging methods accepted for a conclusive diagnosis of AAS were TEE and magnetic resonance imaging.

---

**Figure 1** Representative radiographic findings in a patient with acute type A aortic dissection. (A) CR in the supine position showing ME, poor definition/irregularity of the aortic contour, and tracheal displacement to the right. (B) Contrast-enhanced CT image of the same patient showing double lumen, intimal flap, and dilatation of the descending thoracic aorta indicating dissection and a large periaortic hematoma. CR = chest radiography; ME = mediastinum enlargement.
angiography (MRA). Instruments used for imaging were site-specific. These examinations were performed and interpreted by certified radiologists, cardiologists, or cardiac surgeons not involved in the present study.

Case Definition
The following etiological entities were considered in the definition of AASs: Stanford type A or B aortic dissection (AD), IMH, PAU, and spontaneous aortic rupture (SAR). Case adjudication was performed by two expert physicians who independently reviewed the diagnostic data obtained during the index ED visit and during the 14-day follow-up period. For all patients admitted to the hospital after the ED visit or with novel ED visits, medical records with diagnostic data were reviewed. Case adjudication was binary: AAS present or absent. In case of discordance, the case was adjudicated after discussion.

A case was predefined by evidence of AAS on conclusive imaging (CTA, TEE, MRA), surgery, or autopsy. For deaths occurring in patients without conclusive diagnostic data by advanced imaging, surgery or autopsy, adjudication was clinical, based on all available premortem data. In patients where adjudication was AAS absent, alternative diagnoses (AltD) to AASs were also indicated based on available data. Pre-specified AltD were the following: acute coronary syndrome, gastrointestinal disease, pleuritis or pneumonia, pericarditis, pulmonary embolism, stroke not related to AAS, limb ischemia not related to AAS, syncope not related to AAS, uncomplicated aortic aneurysm, muscle-skeletal pain, and other diagnoses.

Case Adjudication
The criterion standard was the case adjudication by two researchers who independently reviewed the cases, as previously described. Given the severity of AASs in untreated patients, we assumed that individuals with undiagnosed AAS would experience major events leading to repeated medical evaluation and to conclusive diagnosis within 14 days from the ED visit. Therefore, all patients not subjected to CTA, TEE, MRA, or surgery during the index ED visit, entered a 14-day follow-up to allow accurate case adjudication. The timeline of the follow-up was tailored on the acute phase of AASs per classic definition and guidelines.

Patients dismissed from the ED visit without conclusive diagnostic data were instructed to return to the ED in case of new, worsening, or recurrent symptoms. Patients or family members were interviewed by telephone using a structured questionnaire or underwent an outpatient visit after 14 days. The following health-related events since ED discharge were queried for all patients in follow-up: diagnosis of AAS or any aortic disease, subsequent ED visit, subsequent admission to hospital, or death. Hospital charts and dismissal documents of all enrolled patients were acquired and reviewed for final case adjudication.

Outcome Measures
The primary outcome was the diagnostic test characteristics of strategies integrating CR and ADD-RS for suspected AAS. In particular, we sought to define the test characteristics of the following integrated diagnostic strategies: rule out if (ADD-RS ≤ 1 and CR normal) or rule in if (ADD-RS > 1 and CR abnormal, or ADD-RS > 1). Secondary outcomes were: 1) diagnostic test characteristics of CR for AASs, irrespective of ADD-RS classification and 2) interobserver agreement between attending emergency physician and specialized radiologist for evaluation of ME on CR.

Sample Size
We aimed at including enough patients to provide accurate estimates of the diagnostic test characteristics of CR for AASs. In a previous study, an integrated rule-out strategy based on ADD-RS ≤ 1 and D-dimer (<500 ng/mL) had a sensitivity of 98.7% for AASs. Considering the reported sensitivity of CR (64%), this study was powered to test the null hypothesis that the sensitivity of a rule-out strategy combining CR with a low clinical probability of AAS exceeds 97.7%. Using a Type I error of 0.05 and a Type II error of 0.2, we needed to include about 976 participants to reject the null hypothesis.

Data Analysis
Dichotomous data were expressed as proportions and compared with Pearson’s chi-square test. Continuous data were expressed as median and interquartile range (25th to 75th percentile values) and were compared with nonparametric Mann-Whitney’s U-test. The diagnostic performance was assessed by computing sensitivity, specificity, and negative/positive likelihood ratios with their 95% confidence interval (CI). McNemar’s test was used to evaluate differences in sensitivity/specificity. Inter-rater agreement statistic κ was calculated with its 95% CI.
Radiologic signs and ADD-RS were used in univariate logistic models to select potential predictors of AAS diagnosis. The four strongest radiologic predictors of AAS were next introduced in the multivariable binary logistic regression analysis with the ADD-RS, to obtain one covariate for every eight to 10 events. For receiver operating characteristic (ROC) curve analysis, the area under the curve (AUC) was computed and compared per DeLong. A Fagan nomogram was used to visualize the effect of CR findings on the probability of AAS.

A p value < 0.05 was considered statistically significant. The analysis was performed with SPSS statistical package version 25.0 (SPSS Inc.) and MedCalc Statistical Software version 19.0.4 (MedCalc Software bvba).

RESULTS

Characteristics of Study Subjects

In the study period, 1,129 enrolled patients with suspected AAS underwent CR, and 1,030 (91.2%) were analyzed (Figure 2). The presenting symptoms were the following: anterior chest pain (n = 823, 79.9%), posterior chest pain (n = 283, 27.5%), lumbar pain (n = 50, 4.9%), abdominal pain (n = 149, 14.5%), syncope (n = 100, 9.7%), and symptoms of perfusion deficit (n = 63, 6.1%). Table 1 reports the clinical characteristics of study patients.

Computed tomography angiography was performed in 317 (30.8%) patients, TEE in 10, (1%) and MRA in 1 (0.1%). Two (0.2%) patients underwent two advanced imaging examinations, and 120 (11.7%) underwent coronary angiography. An AAS was adjudicated in 48 (4.7%) patients: type A AD in 19 (two during follow-up, 1.8%), type B AD in 11 (one during follow-up, 1.1%), IMH in 11 (1.1%), SAR in 3 (0.3%) and PAU in 4 (0.4%). In 982 (95.3%) patients, adjudication was AAS-absent. The AltD were muscle-skeletal chest pain (n = 358 patients, 34.8%), gastrointestinal disease (n = 122, 11.8%), acute coronary syndrome (n = 131, 12.7%), syncope (n = 46, excluded patients: n=3 did not consent n=11 lost at follow-up n=85 CR not available for radiologist’s evaluation

Figure 2 Flow diagram of the study. (%) refers to included patients. AAS = acute aortic syndrome; ADD-RS = aortic dissection detection risk score; AltD = alternative diagnosis; CR = chest radiography; CRF = case report form.
4.5%), pericarditis (n = 39, 3.8%), pleuritis or pneumonia (n = 35, 3.4%), uncomplicated aortic aneurysm (n = 18, 1.7%), pulmonary embolism (n = 14, 1.4%), stroke (n = 10, 1%), limb ischemia (n = 1, 0.1%), and other diagnoses (n = 208, 20.2%).

Chest Radiography Findings
Presence of ME on CR was detected blindly by a radiologist in 101 (9.8%) patients, including 26 (54.2%) of 48 with AAS: 11 of 19 (57.9%) patients with type A AD, five of 11 (45.5%) patients with type B AD, six of 11 (54.5%) patients with IMH, three of three (100%) patients with AR, and one of four (25%) patients with PAU. Any CR sign (aCRs) of AAS was detected in 174 (16.9%) patients. CR findings and the detailed diagnostic accuracy variables are reported in Table 2. Presence of ME had a sensitivity of 54.2% (95% CI = 39.2 to 68.6%, p = 0.25 vs. presence of aCRs) and a specificity of 92.4% (95% CI = 90.5 to 93.9%, p < 0.001 vs. presence of aCRs) for AAS.

The CR was performed in the upright position in 534 (51.8%) patients and in the supine position in 496 (48.2%) patients. The sensitivity of ME was similar in upright patients (50%, 95% CI = 15.7 to 84.3%) or supine patients (55%, 95% CI = 38.5–70.7%, p = 0.11), while the specificity was superior in upright patients (94.1%, 95% CI = 91.7 to 95.9% vs 90.3%, 95% CI = 87.3%–92.9%, p = 0.02).

Independent evaluation of CR for presence of ME by the attending physician (during the ED visit) was recorded for 1,001 (97.2%) patients. The sensitivity and specificity of attending physician-defined ME for AAS were 42.5% (95% CI = 27% to 59.1%) and 93% (95% CI = 91.2% to 94.6%), respectively (p = 0.09 and p = 0.5 vs. radiologist’s evaluation). The inter-observer agreement (k) between attending physicians and radiologists was 0.44 (95% CI = 0.35 to 0.54).

Additive Diagnostic Value of Chest Radiography
In both univariate and multivariate analysis, ME and inward displacement of intimal calcifications were predictors of AAS, in addition to the ADD-RS (Data Supplement S1, Tables S1 and S2 [available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13819/full] and Table 3).

Table 1
Demographic and Clinical Characteristics of Enrolled Patients According to the Final Diagnosis

<table>
<thead>
<tr>
<th></th>
<th>Total (n = 1,030)</th>
<th>AAS (n = 48)</th>
<th>AltD (n = 982)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex</td>
<td>407 (39.5)</td>
<td>22 (45.8)</td>
<td>385 (39.2)</td>
<td>0.36</td>
</tr>
<tr>
<td>Age (years)</td>
<td>60 (47–73)</td>
<td>70 (63–81)</td>
<td>59 (47–72)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>sBP (mm Hg)</td>
<td>140 (125–155)</td>
<td>135 (120–160)</td>
<td>140 (125–155)</td>
<td>0.48</td>
</tr>
<tr>
<td>Heart rate (beats/min)</td>
<td>75 (67–86)</td>
<td>75 (60–85)</td>
<td>75 (67–86)</td>
<td>0.64</td>
</tr>
<tr>
<td>Marfan/Connective tissue disease</td>
<td>10 (1)</td>
<td>1 (2.1)</td>
<td>9 (0.9)</td>
<td>0.40</td>
</tr>
<tr>
<td>Family history of aortic disease</td>
<td>26 (2.5)</td>
<td>2 (4.2)</td>
<td>24 (2.4)</td>
<td>0.43</td>
</tr>
<tr>
<td>Known aortic valve disease</td>
<td>44 (4.3)</td>
<td>6 (12.5)</td>
<td>38 (3.9)</td>
<td>0.004</td>
</tr>
<tr>
<td>Recent aortic manipulation</td>
<td>11 (1.1)</td>
<td>1 (2.1)</td>
<td>10 (1)</td>
<td>0.47</td>
</tr>
<tr>
<td>Known thoracic aortic aneurysm</td>
<td>60 (5.8)</td>
<td>9 (18.8)</td>
<td>51 (5.2)</td>
<td>0.001</td>
</tr>
<tr>
<td>Abrupt onset of pain</td>
<td>394 (38.3)</td>
<td>25 (52.1)</td>
<td>369 (37.6)</td>
<td>0.044</td>
</tr>
<tr>
<td>Severe pain intensity</td>
<td>469 (45.5)</td>
<td>30 (62.5)</td>
<td>439 (44.7)</td>
<td>0.016</td>
</tr>
<tr>
<td>Ripping or tearing pain</td>
<td>166 (16.1)</td>
<td>10 (20.8)</td>
<td>156 (15.9)</td>
<td>0.37</td>
</tr>
<tr>
<td>Pulse deficit or sBP differential</td>
<td>40 (3.9)</td>
<td>6 (12.5)</td>
<td>34 (3.5)</td>
<td>0.002</td>
</tr>
<tr>
<td>Focal neurologic deficit</td>
<td>42 (40.8)</td>
<td>1 (2.1)</td>
<td>41 (4.2)</td>
<td>0.47</td>
</tr>
<tr>
<td>Murmur of aortic insufficiency</td>
<td>8 (0.8)</td>
<td>3 (6.3)</td>
<td>5 (0.5)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Hypotension or shock state</td>
<td>20 (1.9)</td>
<td>7 (14.6)</td>
<td>13 (1.3)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>History of arterial hypertension</td>
<td>532 (51.7)</td>
<td>34 (70.8)</td>
<td>498 (50.7)</td>
<td>0.007</td>
</tr>
<tr>
<td>Diabetes</td>
<td>139 (13.5)</td>
<td>6 (12.5)</td>
<td>133 (13.6)</td>
<td>0.83</td>
</tr>
<tr>
<td>Cocaine abuse</td>
<td>2 (0.2)</td>
<td>0 (0)</td>
<td>2 (0.2)</td>
<td>0.76</td>
</tr>
<tr>
<td>ADD risk score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>282 (27.4)</td>
<td>7 (14.6)</td>
<td>275 (28)</td>
<td>0.04</td>
</tr>
<tr>
<td>≤1</td>
<td>900 (87.3)</td>
<td>30 (62.5)</td>
<td>870 (88.6)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Values are reported as absolute number (%) for categorical variables or as median (interquartile range) for continuous variables.
AAS = acute aortic syndrome; ADD = aortic dissection detection; AltD = alternative diagnosis; sBP = systolic blood pressure.
the posttest probability of AAS (post to ADD-RS. indicating a potentially additive diagnostic value of CR (95% CI ADD-RS/aCRs integration had an AUC of 0.78

**Table 3**

Multivariate Logistic Regression Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD-RS</td>
<td>2.43</td>
<td>1.48–3.97</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Inward displacement of aortic wall calcification</td>
<td>41.96</td>
<td>2–879.94</td>
<td>0.016</td>
</tr>
<tr>
<td>Left apical opacity</td>
<td>6.64</td>
<td>0.68–65.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Displacement of a nasogastric tube</td>
<td>0.21</td>
<td>0–7.11</td>
<td>0.39</td>
</tr>
<tr>
<td>ME</td>
<td>10.4</td>
<td>5.4–20</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Variables were selected based on the results of univariate analysis (see supplementary table 1). McFadden’s adjusted R2: 0.194. ADD-RS = aortic dissection detection risk score; ME = mediastinum enlargement.

Since inward displacement of intimal calcifications was observed in only four study patients and lacked sensitivity for AAS, we retained for further analyses ME and aCRs. In ROC curve analysis (Figure 3), ADD-RS had an AUC of 0.66 (95% CI = 0.58 to 0.74), ADD-RS/ME integration had an AUC of 0.8 (95% CI = 0.73 to 0.87; p < 0.001 vs. ADD-RS), and ADD-RS/aCRs integration had an AUC of 0.78 (95% CI = 0.72 to 0.85; p < 0.001 vs. ADD-RS), indicating a potentially additive diagnostic value of CR to ADD-RS.

A Fagan nomogram showed the effect of CR on the posttest probability of AAS (post $P_{AAS}$, Figure 4). Within ADD-RS ≤ 1, the pretest probability of AAS was 3.3% (30 cases of AAS in 900 patients), corresponding to approximately one case of AAS in 30 individuals. The presence of ME on CR led to a post $P_{AAS}$ of 15.7% (95% CI = 11.1% to 21.6%), corresponding to approximately one case of AAS in six individuals. The absence of ME led to a post $P_{AAS}$ of 2% (95% CI = 1.1% to 3.1%), corresponding to approximately one missed case of AAS in 52 individuals. The presence of aCRs led to a post $P_{AAS}$ of 12.5% (95% CI = 8.9% to 17.3%), corresponding to approximately one case of AAS in eight individuals. The absence of aCRs led to a post $P_{AAS}$ of 2% (95% CI = 1.1% to 3.2%), corresponding to approximately one missed case of AAS in 50 individuals.

**Table 2**

Diagnostic Performance of Chest Radiography Findings for Diagnosis of Acute Aortic Syndrome

<table>
<thead>
<tr>
<th>Variable</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mediastinum enlargement</td>
<td>26</td>
<td>75</td>
<td>907</td>
<td>22</td>
<td>54.2 (39.2–68.6)</td>
<td>92.4 (90.5–93.9)</td>
<td>7.09 (5.05–9.96)</td>
<td>0.5 (0.44–0.56)</td>
</tr>
<tr>
<td><strong>Additional chest radiography signs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor definition/irregularity of aortic contour</td>
<td>9</td>
<td>22</td>
<td>960</td>
<td>39</td>
<td>18.8 (9–32.6)</td>
<td>97.8 (96.6–98.6)</td>
<td>8.37 (4.08–17.18)</td>
<td>0.83 (0.66–1.04)</td>
</tr>
<tr>
<td>Double aortic knob sign</td>
<td>10</td>
<td>21</td>
<td>961</td>
<td>38</td>
<td>20.8 (10.5–35)</td>
<td>97.9 (96.7–98.7)</td>
<td>9.74 (4.86–19.52)</td>
<td>0.81 (0.63–1.03)</td>
</tr>
<tr>
<td>Inward displacement of aortic wall calcification &gt; 10 mm</td>
<td>3</td>
<td>1</td>
<td>981</td>
<td>45</td>
<td>6.3 (1.3–17.2)</td>
<td>99.9 (99.4–100)</td>
<td>61.38 (6.50–579.17)</td>
<td>0.94 (0.14–5.12)</td>
</tr>
<tr>
<td>Tracheal displacement to the right</td>
<td>13</td>
<td>32</td>
<td>950</td>
<td>35</td>
<td>27.1 (15.3–41.8)</td>
<td>96.7 (95.4–97.8)</td>
<td>8.31 (4.67–14.78)</td>
<td>0.75 (0.63–0.91)</td>
</tr>
<tr>
<td>Displacement of a nasogastric tube</td>
<td>2</td>
<td>2</td>
<td>980</td>
<td>46</td>
<td>4.2 (0.5–14.3)</td>
<td>99.8 (99.3–100)</td>
<td>20.46 (2.94–142.15)</td>
<td>0.96 (0.36–2.56)</td>
</tr>
<tr>
<td>Left side pleural effusion</td>
<td>8</td>
<td>37</td>
<td>945</td>
<td>40</td>
<td>16.7 (7.5–30.2)</td>
<td>96.2 (94.8–97.3)</td>
<td>4.42 (2.18–8.97)</td>
<td>0.87 (0.76–0.99)</td>
</tr>
<tr>
<td>Suspected pericardial effusion</td>
<td>3</td>
<td>12</td>
<td>970</td>
<td>45</td>
<td>6.3 (1.3–17.2)</td>
<td>98.8 (97.9–99.4)</td>
<td>5.11 (1.49–17.53)</td>
<td>0.95 (0.74–1.22)</td>
</tr>
<tr>
<td>Left apical opacity</td>
<td>3</td>
<td>3</td>
<td>979</td>
<td>45</td>
<td>6.3 (1.3–17.2)</td>
<td>99.7 (99.1–99.9)</td>
<td>20.46 (4.24–98.72)</td>
<td>0.94 (0.42–2.09)</td>
</tr>
<tr>
<td>Any chest radiography sign</td>
<td>29</td>
<td>145</td>
<td>837</td>
<td>19</td>
<td>60.4 (45.3–74.2)</td>
<td>85.2 (82.9–87.4)</td>
<td>4.09 (3.11–5.38)</td>
<td>0.46 (0.43–0.5)</td>
</tr>
</tbody>
</table>

AAS = acute aortic syndrome; aCRs = any CR sign; CR = chest radiography; FN = false negative; FP = false positive; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; ME = mediastinum enlargement; TN = true negative; TP = true positive.
The additional 134 patients ruled in with integration of aCRs were 15 TP patients with AAS (31.3% of all AASs) and 119 FP patients with AltD. The ADD-RS/ME and ADD-RS/aCRs integrated strategies had similar sensitivity (Table 4; 66.7% vs 68.8%, p = 0.83), while the specificity was higher for ADD-RS/ME (82.5% vs 76.5%, p = 0.001).

ADD-RS/ME integration (ADD-RS ≤ 1 and absence of ME) ruled out 826 (80.2%) patients: 16 false-negative (FN) patients with AAS (33.3% of all AASs) and 810 true-negative (TN) patients with AltD. ADD-RS/aCRs integration (ADD-RS ≤ 1 and absence of aCRs) ruled out 766 (74.4%) patients: 15 FN patients with AAS (31.3% of AASs) and 751 TN patients with AltD. The diagnostic test characteristics of the CR-integrated strategies are summarized in Table 4.

**DISCUSSION**

To our knowledge, this is the largest prospective study evaluating CR for diagnostic workup of AASs. It is also the first study assessing the utility of CR in conjunction with standardized clinical probability assessment, as suggested by guidelines. The general diagnostic accuracy of CR for AASs was modest and lower than predicted, in spite of specialized, focused, and nonurgent evaluation of radiologic films by dedicated radiologists. In previous studies, the sensitivity of CR for AASs was highly variable (11%–94%), with a pooled sensitivity of 64% in a metaanalysis of 13 studies comprising 1,337 patients. Most were retrospective case-control studies comparing AASs with selected AltD and were poorly representative of ED case mixes. In the prospective study by von Kodolitsch et al., the sensitivity was 67%. The higher sensitivity found in the study by von Kodolitsch et al. compared with our study can be explained by the fact that in the study by von Kodolitsch et al., 58% of patients diagnosed with an AAS had a nondissecting aortic aneurysm, defined as a nonruptured fusiform or saccular dilatation (>50 mm) of the aorta. This condition, where ME is common, was not considered as an AAS in this study, in compliance with current definitions and clinical guidelines.

**Chest Radiography Integration for Rule In/Rule Out**

ADD-RS/ME integration (ADD-RS ≤ 1 and presence of ME, or ADD-RS > 1) ruled in 204 patients versus 130 with ADD-RS > 1 alone (Δ = 7.2%, 95% CI = 3.9% to 10.4%). The additional 74 patients ruled in with integration of ME were 14 true-positive (TP) patients with AAS (29.2% of all AASs) and 60 false-positive (FP) patients with AltD. ADD-RS/aCRs integration (ADD-RS ≤ 1 and presence of aCRs, or ADD-RS > 1) ruled in 264 patients (Δ = 13%, 95% CI = 9.6% to 16.4% vs. ADD-RS > 1 alone). The additional 134 patients ruled in with integration of aCRs were 15 TP patients with AAS (31.3% of all AASs) and 119 FP patients with AltD. The ADD-RS/ME and ADD-RS/aCRs integrated strategies had similar sensitivity (Table 4; 66.7% vs 68.8%, p = 0.83), while the specificity was higher for ADD-RS/ME (82.5% vs 76.5%, p = 0.001).

**Table 4.** Test Characteristics of Strategies Integrating the ADD-RD and CR Findings, for Diagnosis of AASs

<table>
<thead>
<tr>
<th>Rule out if:</th>
<th>Rule out if:</th>
<th>Rule out if:</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD-RS ≤ 1 and mediastinum not enlarged</td>
<td>ADD-RS ≤ 1 and any chest ray sign present or ADD-RS &gt; 1</td>
<td>ADD-RS ≤ 1 and any chest ray sign absent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>172</td>
<td>810</td>
<td>16</td>
</tr>
<tr>
<td>33</td>
<td>231</td>
<td>751</td>
<td>15</td>
</tr>
</tbody>
</table>

Sensitivity % (95% CI): 66.7 (51.5–79.9) vs 68.8 (53.6–80.9), Specificity % (95% CI): 82.5 (79.9–84.8) vs 76.5 (73.7–79.1), LR+ (95% CI): 3.81 (2.99–4.85) vs 2.92 (2.34–3.65), LR– (95% CI): 0.4 (0.27–0.6) vs 0.41 (0.27–0.62)

ADD-RS = Aortic Dissection Detection Risk Score; CR = Chest Radiography; FN = False Negative; FP = False Positive; LR+ = Positive Likelihood Ratio; LR– = Negative Likelihood Ratio; TN = True Negative; TP = True Positive.

Figure 4: Fagan nomogram showing the additive effect of CR to clinical probability assessment. The clinical probability of AAS is displayed on the left as Prior P (%). The middle line represents the result of CR, as presence (+) or absence (–) of any compatible sign. When a straight line is drawn through the prior P and CR result, the posttest probability of AAS is found on the right line (post P, %). The representative dotted lines represent the effect of CR findings for patients at low clinical probability of AAS; CR = chest radiography.
Execution of CR in the supine (via portable CR) or upright position potentially affects the diagnostic accuracy of CR for AAS, as supine CR artificially increases mediastinum width. Many previous studies did not report the number of portable versus nonportable CR examinations. In von Kodolitsch et al., CR examinations were performed at the bedside in 19% of patients. In this study, supine CR examinations were almost half of the total, which is expected to overestimate the sensitivity of CR. Furthermore, in our study, the sensitivity of ME was similar in patients evaluated in the supine or upright position. Taken together, these data indicate that the lower sensitivity estimated herein for CR is not related to the rate of CR examinations performed in the supine position.

Another factor potentially affecting the accuracy of CR for AASs is the expertise of the interpreting physicians. Two previous studies have found significant inter- and intraobserver variability between radiologists for detection of potential CR signs of AAS, with only fair agreement ($\kappa = 0.25$ for suspicion of aortic dissection, $\kappa = 0.23–0.33$ for presence of ME, irregularities of the aortic contour and pleural effusion). In this study, the interobserver agreement between attending physicians and radiologists for ME was also fair to moderate. However, the sensitivity of ME was lower than previously reported even when the CR was interpreted by specialized radiologists focusing their attention on potential aortic signs.

In this study, integration of CR findings improved the diagnostic accuracy over ADD-RS alone in AUC-ROC analysis. However, the additive diagnostic effect of CR to clinical pretest probability assessment should not be overestimated, because evidence is also provided that the pragmatic impact of CR integration on rule-in and rule-out algorithms of AASs was modest. First, prospective ED data shows that CR, even if interpreted by expert radiologists evaluating all potential radiologic signs of AAS, is unsuitable to support conclusive rule-out of AASs. Indeed, the posterior probability of AAS in patients with ADD-RS $\leq 1$ and a normal CR was higher than clinically accepted for other cardiovascular emergencies (e.g., acute coronary syndromes). Hence, clinicians requesting a CR in a stable patient at low disease probability should be aware that up to one-third of the AAS cases with these clinical characteristics may not show any CR abnormality. Instead, a rule-out strategy based on a biomarker (e.g., D-dimer) has shown higher sensitivity and lower failure rate, although the net effect on the number of CTA examinations requested in clinical practice for suspected AAS is yet to be determined. Second, with respect to CR as an additive rule-in tool, current results indicate that CR can help identify few additional AAS cases over ADD-RS alone. However, most patients at low clinical probability red-flagged by CR will be FP.

**LIMITATIONS**

The current study has limitations. First, decision to perform CR as the first-line imaging examination was left to the treating physicians. This contributed to focus our study on a low clinical probability cohort and likely introduced some degree of spectrum bias related to the variability of clinical gestalt, setting, and case mix. Therefore, we cannot exclude that, when applied to a patient population at higher probability of AASs, the sensitivity of CR for AASs may be different. Since patients with a history of thoracic aortic aneurysm, in whom ME is more frequent, were less likely to receive a CR and more frequently proceeded to advanced aortic imaging (selection bias), we speculate that CR may show higher sensitivity and lower specificity when applied to patients at higher probability of AAS. In 85 (7.5%) patients, CR data were not available for review, leading to patient exclusion. However, lack of CR data for posterior evaluation was random and is unlikely to have affected study results.

Second, incorporation bias can be predicted, because clinical findings, CR, and D-dimer results influenced the decision to order definitive testing. Incorporation bias also affected outcome determination. Indeed, criterion standard imaging was only performed in approximately one-third of subjects, while in the other patients, adjudication was based on a 14-day follow-up. Therefore, we cannot exclude that a small number of AAS cases with a nonsevere course have been missed. However, overlook of additional patients with AAS and normal CR findings would further reduce estimates of the diagnostic accuracy of CR, alone and in combination with pretest clinical probability assessment. Finally, statistical models exploring the additive effect of CR to ADD-RS have inherent limits due the limited number of AAS cases in our cohort.

**CONCLUSIONS**

This large multicenter study assessed the diagnostic impact of chest radiography in the workup of acute
aortic syndromes. Chest radiography integration improved the diagnostic accuracy over aortic dissection detection risk score alone, leading to moderate sensitivity. The specificity was higher with mediastinum enlargement than with any chest radiography signs. However, the efficiency of chest radiography integration for acute aortic syndrome rule-in was modest, and up to one-third of study patients with an acute aortic syndrome satisfied rule-out criteria. These results indicate that aortic dissection detection risk score/chest radiography integration cannot be used to define conclusive and safe rule-out strategies. The net pragmatic impact of chest radiography on the workup of acute aortic syndromes therefore appears questionable.

The authors thank Dr. Andrea Palazzo and Marco Gambassi (Careggi University Hospital, Firenze, Italy) for their collaboration in data analysis.

References

21. Luker GD, Glazer HS, Eagar G, Gutierrez FR, Sagel SS. Aortic dissection: effect of prospective chest radiographic

APPENDIX A

List of the ADvISED Study (Chest Radiography Subanalysis) Co-Investigators

Elisa Capretti,1 Gabriele Cerini,1 Andrea Palazzo,1 Federica Trausi,1 Maddalena Ottaviani,1 Paolo Baron,2 Paolo Bima,2 Paolo Fascio,2 Domenica Garabello,3 Jasper Boeddinghaus,4 Thomas Nestelberger,4 Gregor Sommer,5 Raphael Twerenbold,4 Wolfgang Bauerf,6 Anneke Dambergf,6 Lukas Porallaf,6 Matthias Tau- pitzi,7 and Múcio Tavares de Oliveira, Jr.8

1Department of Emergency Medicine, Careggi University Hospital, Firenze, Italy; 2S.C.U. Medicina d’Urgenza, Department of Emergency Medicine, A.O.U. Città della Salute e della Scienza, Molinette Hospital, Torino, Italy; 3Radiology Department, A.O.U. Città della Salute e della Scienza, Molinette Hospital, Torino, Italy; 4Cardiovascular Research Institute Basel (CRIB) and Department of Cardiology, University Hospital Basel, University of Basel, Switzerland; 5Clinic of Radiology and Nuclear Medicine, University Hospital of Basel, Basel, Switzerland; 6Department of Emergency Medicine, Campus Benjamin Franklin, Charité–Universitätsmedizin, Berlin, Germany; 7Institute of Radiology, Campus Benjamin Franklin, Charité–Universitätsmedizin Berlin, Germany; 8Heart Institute, University of São Paulo, São Paulo, Brazil.

Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13819/full

Data Supplement S1. Supplemental material.
Exploring Gender Bias in Nursing Evaluations of Emergency Medicine Residents

Krista Brucker, MD1, Nash Whitaker, MD1, Zachary S. Morgan, PhD2, Katie Pettit, MD1, Erynn Thinnes, MD1, Alison M. Banta1, and Megan M. Palmer, PhD1

ABSTRACT

Objectives: Nursing evaluations are an important component of residents’ professional development as nurses are present for interactions with patients and nonphysician providers. Despite this, there have been few prior studies on the benefits, harms, or effectiveness of using nursing evaluations to help guide emergency medicine residents’ development. We hypothesized that gender bias exists in nursing evaluations and that female residents, compared to their male counterparts, would receive more negative feedback on the perception of their interpersonal communication skills.

Methods: Data were drawn from nursing evaluations of residents between March 2013 and April 2016. All comments were coded if they contained words falling into four main categories: standout, ability, grindstone, and interpersonal. This methodology and the list of words that guided coding were based on the work of prior scholars. Names and gendered pronouns were obscured and each comment was manually reviewed and coded for valence (positive, neutral, negative) and strength (certain or tentative) by at least two members of the research team. Following the qualitative coding, quantitative analysis was performed to test for differences. To evaluate whether any measurable differences in ability between male and female residents existed, we compiled and compared American Board of Emergency Medicine in-training examination scores and relevant milestone evaluations between female and male residents from the same period in which the residents were evaluated by nursing staff.

Results: Of 1,112 nursing evaluations, 30% contained comments. Chi-square tests on the distribution of valence (positive, neutral, or negative) indicated statistically significant differences in ability and grindstone categories based on the gender of the resident. A total of 51% of ability comments about female residents were negative compared to 20% of those about male residents ($\chi^2 = 11.83, p < 0.01$). A total of 57% of grindstone comments about female residents were negative as opposed 24% of those about male residents ($\chi^2 = 6.03, p < 0.01$).

Conclusions: Our findings demonstrate that, despite the lack of difference in ability or competence as measured by in-service examination scores and milestone evaluations, nurses evaluate female residents lower in their abilities and work ethic compared to male residents.
Evaluations are a commonly implemented tool for feedback in graduate medical education. Faculty evaluations provide important feedback on resident physician performance to guide improvement during training. Studies have demonstrated that multidisciplinary feedback can be useful and reliable. One prospective study demonstrated that multidisciplinary evaluations improved performance of residents compared to faculty feedback alone. Nursing staff are thought to be an important component of a resident 360° evaluation as they are often present for resident interactions with patients, families, and other medical personnel. Effective collaboration and teamwork are essential skills for emergency medicine (EM) residents as is evidenced by their inclusion in the Accreditation Council for Graduate Medical Education (ACGME) milestones.

Several studies have attempted to explore the dynamic relationships between genders in leadership positions in medicine. Keck-McNulty and Wear reported that female residents most commonly expressed “excessive self-monitoring of communication style due to fears of being perceived as too demanding and not friendly enough . . . having to justify their orders more than their male peers . . . and receiving less assistance than their male peers.” The study of female leadership roles during resuscitations by Linden and colleagues also revealed gender discrepancies, stating that “female residents had to earn the trust and respect of the nurses more than their male counterparts.” These prior studies suggest that female residents continue to face challenges in their training program that their male counterparts do not.

Furthermore, recent research has revealed the presence of a gender bias in faculty evaluations of EM residents. However, few studies have sought to examine for the presence of gender bias in 360° evaluations, and results are conflicting. Early literature found that female residents received more favorable evaluations from nursing staff, whereas a more recent study in 2015 found the opposite; women received harsher feedback from nursing staff. The purpose of our study was to determine if gender bias exists in nursing evaluations of our EM residents.

METHODS

This is a retrospective study at a single ACGME-accredited EM residency program in Indiana University. The nursing evaluations at our institution are used to assess professionalism, interpersonal skills, and communication. In addition, the evaluation form includes a free-text box where nurses comment on any aspect of resident performance not strictly limited to communication and professionalism. The residency program supports 69 residents working in three urban emergency departments with a combined annual patient volume of over 250,000 visits. This study was reviewed by the institutional review board and was deemed to be exempt research.

To evaluate if any measurable differences in ability between male and female residents existed, we compiled and compared American Board of Emergency Medicine (ABEM) in-training examination scores and relevant milestone evaluations between female and male residents from the same period when the residents were evaluated by nursing staff. Milestones included in our evaluation included: Systems-based Practice 2 (participates in strategies to improve healthcare delivery and flow, demonstrates an awareness of and responsiveness to the larger context and system of health care); Professionalism (PROF) 1 (demonstrates compassion, integrity, and respect for others as well as adherence to the ethical principles relevant to the practice of medicine.); PROF 2 (demonstrates accountability to patients, society, profession, and self); Interpersonal and Communication Skills (ICS) 1 (demonstrates interpersonal and communication skills that result in the effective exchange of information and collaboration with patients and their families); and ICS 2 (leads patient-centered care teams, ensuring effective communication, and mutual respect among members of the team).

To evaluate if gender bias was present in nursing evaluations, we reviewed nursing evaluations completed between March 2013 and April 2016. On a biannual basis, all nurses working at each of the clinical sites were sent an electronic standard evaluation form for 10 assigned residents (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/10.1111/acem.13843/full). A list of nurses was provided by the departments’ nursing leadership. Forms were sent and completed via the residency’s online evaluation platform (MedHub). Assignments of nurses to particular residents and distribution of evaluations was completed by residency administration staff. Approximately 40 requests per resident were made, and the range of completed evaluations each resident received was 10 to 18.
To ensure blinding of resident gender for the reviewers, one member of the research team obscured the names and gendered pronouns from the comments. After blinding the comments were distributed equally to two independent reviewers. The reviewers were authors and were not blinded to the hypothesis of the study, but were blinded to the gender of the resident associated with each comment. The author responsible for blinding did not participate in coding.

The coding scheme used for this study was based on prior research (see Data Supplement S1, Appendix S2). Trix and Penska and Schmader et al. created word lists to perform a comparison of letters of recommendation based on gender. In our study, we used the grindstone, ability, and standout categories developed by Schmader and colleagues. Schmader and colleagues also coded for communication but a list of those words was not published in their 2007 study. Thus, based on our initial hypothesis that female residents would receive more comments related to interpersonal skills, we created a fourth word list. Our category of interpersonal closely matches the concepts noted in Madera et al. communal category but since a specific word list was not published we generated our own by reviewing a subset of evaluations completed by our faculty.

Comments were coded with standout if they distinguished the resident from his or her peers, regardless of whether it was in a positive or negative manner. Comments were coded ability when the comment spoke to the technical skill, knowledge, and competence of the resident. For grindstone the comment spoke to the resident’s work ethic, effort, or efficiency. Finally, comments were coded interpersonal if they described the type or quality of communication residents had with patients, nurses, families, or other caregivers. Comments could be coded in more than one category. An example of comments representative of these categories is shown in Table 1.

Members of the research team then determined the valence and strength of the comment. Options for the valence of a category included positive, neutral, or negative based on how favorably or unfavorably the evaluator described the resident on that topic. The strength of each category was coded as certain or tentative based on strength of conviction the commenter had on the topic. Discrepancies were resolved by distributing the comments coded differently by the two reviewers to a third member to make a determination based on their interpretation of the comment, which was informed by the comments of the other two members.

To test for differences in the responses to the quantitative questions on the nursing 360 evaluation we

<table>
<thead>
<tr>
<th>Category</th>
<th>Positive Example</th>
<th>Negative Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standout</td>
<td>Dr [NAME] is one of our favorites! [They] is great to work with, always staying calm in high stress situations—which of course we see a lot of those. [Their] time management is something other residents should strive for. [They] always keeps everyone up to date on the plan of care and is truly a joy to work with. Would love to see Dr [NAME] become a permanent member of our family when [their] residency is complete. [They] would be a huge asset to our team.</td>
<td>Extraordinarily dismissive and condescending toward nursing and support staff such as unit secretaries and not much better with patients and families.</td>
</tr>
<tr>
<td>Grindstone</td>
<td>I appreciate [NAME]’s desire to jump in and help with new patients. [They] also is willing to help fill the gaps on sick patients when the patient’s primary resident is tied up.</td>
<td>I think that [NAME] does a great job explaining things to patients and addressing their concerns. I do feel that at times, from my perspective, [they] can get easily overwhelmed and get behind a bit during a busy shift. I feel like in these circumstances that [they] can get a little behind and is not always able to keep up with updating the nurses and or patients about the next steps in their care.</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>Dr. [NAME] is extremely professional and energetic. [Their] bedside manner is phenomenal. [They] does a great job at taking the time to talk to patients even when they are difficult. [They] is great at approaching them in a very empathetic fashion.</td>
<td>Dr. [NAME] lacks communication skills, [they] is very condescending to RN’s and ancillary staff, I don’t think [they] intends for it to be that way but [their] tone and behavior comes across that way which many nurses find offensive. [They] is not very sympathetic to patient family members, [they] can be abrasive.</td>
</tr>
<tr>
<td>Ability</td>
<td>Dr. [NAME] is a strong resident. [They] seems very knowledgeable about cases and handles them well.</td>
<td>Dr. [NAME] is a very nice person and pleasant and polite in interactions with nursing and patients/families. There are many times however, that it appears [they] becomes overwhelmed easily and is not as efficient and confident in [their] decision making.</td>
</tr>
</tbody>
</table>
used the Mann-Whitney U-test, which is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous and not normally distributed. Once qualitative coding was complete, we tested for statistically significant differences between male and female residents. To test for differences in distributions of these codes between male and female residents, due to the dichotomous nature of the present/absent codes and the bimodal nature of the other codes, chi-square tests were used using SPSS version 24 (IBM). First analyses looked for differences in the presence of each of the categories. Subsequent analyses only looked at comments with the topics present to explore differences in the valence and strength of each of the categories based on the gender of the residents. Significance was determined at a p-value < 0.05.

### RESULTS

We did not find significant differences in skills or abilities between male and female residents on ABEM in-training examinations or selected milestone evaluations (Table 2). Reviewing the ordinal scale data available on 1,112 nursing evaluations reveals that female residents are reported to be less professional in their interactions with nurses ($p = 0.041$) and have less effective team leadership skills ($p = 0.019$) when compared to their male counterparts. Further, nurses are less likely to report being comfortable with female residents taking care of their family members ($p = 0.013$; Table 3).

Of the 1,112 completed evaluations, 332 (30%) contained text in the open-ended qualitative comments section. The proportion of evaluations which included free-text comments was not significantly different.

### Table 2

<table>
<thead>
<tr>
<th>Score</th>
<th>Female ($n = 33)^a$</th>
<th>Male ($n = 50)^a$</th>
<th>Mean Difference</th>
<th>$t$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-Service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 2</td>
<td>75.64 (±19.00)</td>
<td>79.90 (±9.29)</td>
<td>4.27</td>
<td>1.37</td>
<td>0.174</td>
</tr>
<tr>
<td>Year 3</td>
<td>75.42 (±19.02)</td>
<td>81.44 (±8.65)</td>
<td>6.02</td>
<td>1.96</td>
<td>0.096</td>
</tr>
<tr>
<td>Milestone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBP 2</td>
<td>3.86 (±0.33)</td>
<td>3.77 (±0.29)</td>
<td>0.92</td>
<td>1.34</td>
<td>0.184</td>
</tr>
<tr>
<td>PROF 1</td>
<td>3.77 (±0.28)</td>
<td>3.78 (±0.30)</td>
<td>0.01</td>
<td>0.16</td>
<td>0.876</td>
</tr>
<tr>
<td>PROF 2</td>
<td>3.75 (±0.26)</td>
<td>3.75 (±0.32)</td>
<td>0.01</td>
<td>0.08</td>
<td>0.940</td>
</tr>
<tr>
<td>ICS 1</td>
<td>3.85 (±0.30)</td>
<td>3.80 (±0.27)</td>
<td>0.04</td>
<td>0.66</td>
<td>0.509</td>
</tr>
<tr>
<td>ICS 2</td>
<td>3.79 (±0.28)</td>
<td>3.76 (±0.29)</td>
<td>0.02</td>
<td>0.35</td>
<td>0.726</td>
</tr>
</tbody>
</table>

*Data are reported as mean (±SD).

ICS = Interpersonal and Communication Skills; PROF = Professionalism; SBP = Systems-based Practice.

### Table 3

<table>
<thead>
<tr>
<th>Evaluation Item</th>
<th>Female</th>
<th></th>
<th>Median</th>
<th>Male</th>
<th></th>
<th>Median</th>
<th>$Z$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>Mean (±SD)</td>
<td></td>
<td>$N$</td>
<td>Mean (±SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the resident responsive to patient and family needs/questions?</td>
<td>443</td>
<td>8.34 (±1.69)</td>
<td>9</td>
<td>663</td>
<td>8.46 (±1.60)</td>
<td>9</td>
<td>−0.95</td>
<td>0.34</td>
</tr>
<tr>
<td>Does the resident effectively communicate with you?</td>
<td>442</td>
<td>8.05 (±1.98)</td>
<td>9</td>
<td>663</td>
<td>8.18 (±1.90)</td>
<td>9</td>
<td>−1.01</td>
<td>0.31</td>
</tr>
<tr>
<td>Does the resident behave professionally in their interactions with you?</td>
<td>442</td>
<td>8.49 (±1.85)</td>
<td>9</td>
<td>664</td>
<td>8.69 (±1.73)</td>
<td>9</td>
<td>−2.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Does the resident behave professionally in their interactions with patients and/or families?</td>
<td>442</td>
<td>8.61 (±1.65)</td>
<td>9</td>
<td>661</td>
<td>8.76 (±1.57)</td>
<td>9</td>
<td>−1.62</td>
<td>0.11</td>
</tr>
<tr>
<td>Does the resident effectively demonstrate team leadership skills?</td>
<td>442</td>
<td>7.83 (±2.02)</td>
<td>8</td>
<td>660</td>
<td>8.08 (±1.97)</td>
<td>8</td>
<td>−2.35</td>
<td>0.02</td>
</tr>
<tr>
<td>Does the resident respond in a reasonable and timely fashion to your questions and concerns about patient care/needs?</td>
<td>442</td>
<td>8.07 (±1.93)</td>
<td>9</td>
<td>661</td>
<td>8.29 (±1.74)</td>
<td>9</td>
<td>−1.68</td>
<td>0.09</td>
</tr>
<tr>
<td>Would you be comfortable with this resident’s care for you or a family member?</td>
<td>441</td>
<td>7.76 (±2.31)</td>
<td>8</td>
<td>662</td>
<td>8.09 (2.14)</td>
<td>9</td>
<td>−2.47</td>
<td>0.01</td>
</tr>
</tbody>
</table>
between female and male residents (33% vs 27%, \( \chi^2 = 3.425, p = 0.06 \)). The length of the comments also did not differ significantly between female and male residents (medians = 23 words vs. 19 words, \( p = 0.14 \)).

Upon review of the chi-square results, we did not find statistically significant differences between nurses’ comments about male or female residents in terms of whether any of the four categories were present (Table 4). Both standout and grindstone language was relatively rare in comments of both male and female residents, being present in only about one-sixth of the coded comments. Ability language was more common, which was coded in one out of every three comments for both male and female residents. By far the most common category coded was interpersonal language; approximately four-fifths of comments spoke of residents’ interpersonal skills (Table 4).

### Valence of Language

Chi-square tests on the distribution of valence within each category revealed statistically significant differences based on gender in the use of ability and grindstone language, as shown in Table 5. Regarding ability language, 51% of female residents had negative ability comments, while only 20% of male residents had negative ability comments (\( p < 0.01 \)).

Similarly, 57% of the grindstone comments about female residents were negative compared to (24%) of male residents, while more than three-quarters (76%) of male residents received positive grindstone language (\( p = 0.01 \)). The most parity in valence between male and female residents was in interpersonal language. Both female and male residents had nearly three-quarters (72%) of their interpersonal comments coded as positive (\( p = 0.92 \)).

### Strength of Language

Chi-square tests on the distributions of strength of language for each category only found a statistically significant for gender in the strength of the ability language. These results are shown in Data Supplement S2. Here almost one-third (30%) of ability comments about female residents appeared tentative while only one-seventh (14%) of those comments about male residents were tentative. Additionally, statistically significant yet moderate correlations between the valence and strength of standout, ability, and interpersonal language appear to show that nurses use more tentative language when giving criticism and more certain language when giving praise.

### DISCUSSION

Although there is now strong evidence that gender bias exists across many areas of academic medicine,\(^8,12\) the extent of the impact of those biases on daily professional interactions and professional training programs remains unknown. We designed this study to ask a simple question “is there gender bias in the way nurses evaluate residents?” Our study suggests gender bias in nursing evaluation of residents. Specifically, in
their written comments nurses evaluated female residents lower than their male counterparts in terms of ability and work ethic (grindstone). Although the discrepancies between male and female resident evaluations may be small and of unclear significance, they are concerning given the lack of gender differences in ability or competence as measured by in-service scores and milestone evaluations. This finding is similar to Mueller and colleagues who examined the differences in qualitative feedback that male and female residents received from attendings. Interestingly, we did not find differences in the interpersonal domain.

Initially, we found our sample had similar themes for male and female residents throughout the comments. It was only in coding each to a positive or negative valence that we began to notice the differences in the nursing evaluation of male and female residents. As such, merely reporting the absence or presence of words or phrases is not enough for a study to truly evaluate if there is bias. It is important to get the qualitative nature of the comment.

LIMITATIONS

There are several limitations to consider when interpreting the results of this study. The first is that our study is limited to one residency program and its three clinical training sites. As such, the results of this study could be a result of our training environment and may not be found in other programs.

Another limitation is related to the evaluations and how they are completed. First, the nature of the relationship between the nurse and the resident they are evaluating could impact the results. Because evaluations are sent to nurses at random, there is no minimum amount of exposure to a resident required before a nurse has the ability to evaluate the resident. The evaluation provided to the nurses (Appendix S1) provides a Likert scale from 0 to 10 without specific anchors, which can lead to variability in evaluation. Further, the evaluations were constructed as a measure to get feedback for the residents and not for the purpose of this study. Therefore, the results of this study may be due to variability in nurses’ exposure to the residents they are evaluating and/or their interpretation of the form. Further, surveys are completed anonymously so the gender of the nurse was not obtainable. Future work might explore the interactions between nursing gender and resident physician gender in influencing evaluations.

Finally, although we compared objective measures (milestones assessing communication and professionalism skills) between male and female residents as a surrogate marker for performance, this has not been proven to correlate to bedside performance. As a result, we cannot say with certainty that there are not differences in the abilities of our residents.

Another consideration is that coders were not blinded to the hypothesis of the study, but were blinded to the gender of the resident. This could have skewed the results toward finding bias. In addition the coding process, while based on previous work, required the combination of existing lists with those created specifically for this project, which may have affected our results. Finally, in our study design, individual comments could be in multiple categories, which would give greater weight to those comments and the respective nurses than comments falling into single categories.

CONCLUSION

The data presented here suggest gender bias in nursing evaluations of residents. We undertook this systematic study as a starting point in the design of a proactive effort to mitigate gender bias and bolster support for our female residents. More work is necessary to further understand the impact these differential evaluations have on the training experience of our female residents and what role they might play in our ability to recruit and retain women in academic emergency medicine.

References

5. Berger JS, Pan E, Thomas J. A randomized, controlled crossover study to discern the value of 360-degree versus


10. Dayal A, O’Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. JAMA 2017;177:651.


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13843/full

Data Supplement S1. Supplemental material.

Data Supplement S2. Supplemental material.
Avoiding Misdiagnosis in Patients With Posterior Circulation Ischemia: A Narrative Review

Kiersten L. Gurley, MD1,2,3 and Jonathan A. Edlow, MD1,2

ABSTRACT

Posterior circulation strokes represent 20% of all acute ischemic strokes. Posterior circulation stroke patients are misdiagnosed twice as often compared to those with anterior events. Misdiagnosed patients likely have worse outcomes than correctly diagnosed patients because they are at risk for complications of the initial stroke as well as recurrent events due to lack of secondary stroke prevention and failure to treat the underlying vascular pathology. Understanding important anatomic variants, the clinical presentations, relevant physical examination findings, and the limitations of acute brain imaging may help reduce misdiagnosis. We present a symptom-based review of posterior circulation ischemia focusing on the subtler presentations with a brief discussion of basilar stroke, both of which can be missed by the emergency physician. Strategies to avoid misdiagnosis include establishing an abrupt onset of symptoms, awareness of the nonspecific presentations, consideration of basilar stroke in altered patients and using a modern approach to diagnosis of the acutely dizzy patient.

Posterior circulation strokes represent 20% of all ischemic strokes.1-4 Common posterior circulation stroke symptoms, including dizziness, clumsiness or imbalance, visual symptoms (diplopia, field cuts, or blurred vision), anisocoria, confusion and altered mental status, vomiting, headache and neck pain, problems with speech and swallowing, and decreased hearing,5 are less specific than typical anterior circulation stroke symptoms.6 The physical examination to detect them is more nuanced; the National Institutes of Health Stroke Scale (NIHSS) is weighted toward the anterior circulation7 and patients with posterior circulation strokes can have a NIHSS of zero with disabling deficits.8 These patients are misdiagnosed twice as often as those with anterior events. Misdiagnosed patients are at higher risk for complications of the initial stroke as well as recurrent events due to lack of secondary stroke prevention and lack of specific treatment of the underlying vascular pathology. Computed tomography (CT) of the brain, often the default brain imaging test in the emergency department (ED), has low sensitivity for posterior circulation infarction.9-11 Even magnetic resonance imaging (MRI) with diffusion-weighted imaging (DWI) is imperfect, especially in acute posterior circulation strokes.12

Strategies to avoid misdiagnosis include establishing an abrupt onset of symptoms, awareness of the nonspecific presentations, consideration of basilar stroke in altered patients, and using a modern approach to diagnosis of the acutely dizzy patient. We will emphasize the subtler, more easily missed presentations rather than the clinically overt ones.
DISCUSSION/OBSERVATIONS

Relevant Anatomy and Vascular Pathology

Classically, paired vertebral arteries arise from the subclavian arteries, ascend in the neck through the transverse foraamina of the cervical vertebrae, and then fuse intracranially to form the basilar artery whose branches supply the posterior circulation and join the anterior circulation via the posterior communicating arteries (Figure 1 and Table 1). This classic anatomy is present in only 50% of individuals. Most variations are clinically insignificant, and often posterior strokes do not strictly follow a specific vascular territory. Some variations including the fetal posterior cerebral artery (f-PCA), the artery of Percheron, and vertebral artery hypoplasia (Figures 1A–1C) are frequently mentioned in radiology reports and may be clinically significant. The f-PCA arises from the anterior circulation so that ipsilateral carotid disease usually associated with anterior circulation symptoms can cause a posterior circulation stroke, affecting the interpretation of vascular imaging (Figures 2A–2C). The artery of Percheron is a single unilateral vessel that supplies both sides of the medial thalamus, ischemia of which results in a bilateral thalamic stroke with significant mental status, memory, psychiatric, speech, and ocular symptoms (Figure 2D). Vertebral artery hypoplasia is overrepresented in patients with posterior inferior cerebellar artery infarctions, which are usually ipsilateral to the hypoplastic vessel.

The vascular pathology of posterior and anterior circulation strokes is the same. The proximal vertebral artery is a common location for atherosclerotic occlusive disease and a frequent source of artery-to-artery embolism. Cardioembolism is overrepresented as a cause of posterior circulation stroke compared to anterior circulation events making telemetry a useful adjunct (Figure 2E).

Although uncommon, vertebral artery dissection (VAD) is an important mechanism in posterior circulation strokes accounting for up to 25% of strokes in younger patients (Figures 2F and 2G). Trauma associated with VAD can be trivial or absent. The headache characteristics are variable; thunderclap headache occurs in only 9% of VAD patients; however, the pain is usually unique for the patient and different from prior headaches.

In a consecutive series of 186 patients with first-ever VAD, 89% presented with cerebral ischemia and 11% presented with pain only. These proportions may be skewed (toward more ischemia) by referral bias; a systematic review of VAD reported that stroke occurred in 63% of cases. Importantly, this review reported that 25% of patients with VAD had no pain at any point in their course. Intracranial dissections are associated with subarachnoid hemorrhage (SAH). Horner’s syndrome is more common in carotid dissections (38%, from injury to the sympathetic fibers than run along the vessel) than in VAD (13% from a lateral medullary infarct).

Unusual etiologies in posterior circulation ischemia are rotational vertebral artery compression (usually associated with a contralateral hypoplastic vertebral artery), vertebrobasilar dolichoectasia (a dilatative arteriopathy associated with elongated, tortuous vessels), migraine-related stroke, reversible cerebral vasospastic syndrome, central nervous system vasculitis, and in elderly patients, giant cell arteritis.

Misdiagnosis

Diagnosis of posterior circulation strokes is often missed or delayed. Between 28 and 59% of cerebellar strokes are initially misdiagnosed. In one study of 465 stroke patients, those with posterior strokes were 2.5 times more likely to be missed compared to anterior circulation strokes (38% vs. 16%).

Clinical characteristics are often nonspecific and overlap with those of anterior stroke and stroke mimics. Patients endorse dizziness, altered cognition including psychiatric and memory-related symptoms, nausea and vomiting, headache, blurred vision, visual field abnormalities, dysphagia, hearing loss, clumsiness, and ataxia. In a study of 611 stroke patients, 61 (10%) were initially misdiagnosed. Factors associated with misdiagnosis included visual or gait disturbances, presentation of dizziness, sensory symptoms, and nausea. A recent meta-analysis of stroke misdiagnosis found that mild, nonspecific, and transient symptoms were all associated with misdiagnosis. A diagnostic trap is to not consider stroke in patients without lateralizing symptoms or signs, especially in younger patients. To avoid that trap, consider stroke as the likeliest explanation for the abrupt onset of any neurologic symptoms.

Of a consecutive series of 240 cerebellar stroke patients, 25 (10%) presented with isolated dizziness. Risk factors for misdiagnosis included young age, VAD as a cause, and a presentation of dizziness. Misdiagnosis of posterior circulation stroke is a needle-and-haystack phenomenon; the presenting symptoms are common in ED patients, most of whom are...
not having strokes. However, because these symptoms are so common, even the small fraction of misdiagnosed patients translates into a large absolute number. For patients presenting with dizziness alone, the number potentially harmed has been estimated at between 45,000 and 75,000 patients per year in the United States.

Although no prospective data exist, some findings suggest worse clinical outcomes in misdiagnosed patients compared to correctly diagnosed ones.
<table>
<thead>
<tr>
<th>Vascular Territory</th>
<th>Location</th>
<th>Area Supplied</th>
<th>Syndrome</th>
<th>Clinical Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertebrobasilar artery</td>
<td>Medulla and cervical spinal cord</td>
<td>Medial medullary syndrome (intracranial disease may lead to Wallenberg syndrome)</td>
<td>Contralateral arm and leg weakness, hemibody loss of tactile, vibration, position sense, ipsilateral tongue paralysis</td>
<td></td>
</tr>
<tr>
<td>Posterior inferior cerebellar artery</td>
<td>Inferior posterior cerebellar hemisphere, inferior vermis, lateral medulla</td>
<td>Lateral medullary or Wallenberg syndrome superior cerebellar artery syndrome</td>
<td>Vertigo, nausea, vomiting, ipsilateral facial numbness and dysmetria, Homer’s syndrome, dysphagia, and ataxia dysphonia contralateral hemisensory loss below the face</td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>Top</td>
<td>Midbrain, thalamus, mesial temporal lobes and occipital lobes</td>
<td>Top of the basilar syndrome</td>
<td>Somnolence, peduncular hallucinosis, convergence nystagmus, skew deviation, oscillatory eye movements, Colliers sign (retraction and elevation of eye lids), vertical gaze paralysis</td>
</tr>
<tr>
<td>Mid-BA</td>
<td>Lateral and medial pons</td>
<td>Lateral mid-pontine syndrome</td>
<td>Ipsilateral loss of facial sensation and motor function of the trigeminal nerve, ipsilateral dysmetria</td>
<td></td>
</tr>
<tr>
<td>Pontine paramedian penetrators</td>
<td>Anteromedial pons</td>
<td>Dorsal mid-pontine syndrome</td>
<td>Ipsilateral nuclear facial palsy, horizontal gaze palsy, and contralateral arm and leg weakness</td>
<td></td>
</tr>
<tr>
<td>Short pontine circumferential arteries</td>
<td>Anterolateral pons</td>
<td>Superior medial pontine syndrome</td>
<td>Ipsilateral intranuclear ophthalmoplegia, palatal, facial, pharyngeal and/or ocular myoclonus, dysmetria, and contralateral arm and leg weakness, ocular bobbing</td>
<td></td>
</tr>
<tr>
<td>Proximal basilar</td>
<td>Lower pons</td>
<td>Locked-in syndrome</td>
<td>Quadriplegia, horizontal gaze paralysis, bifacial paralysis, tongue and mandibular weakness, awareness is spared</td>
<td></td>
</tr>
<tr>
<td>Anterior inferior cerebellar artery</td>
<td>Ipsilateral labyrinth, lateral pontine tegmentum, and brachium pontis, ICP</td>
<td>Lateral pontine syndrome</td>
<td>Ipsilateral dysmetria, hearing loss, Homer’s syndrome, choreiform dyskinesia, contralateral thermoanalgesia</td>
<td></td>
</tr>
<tr>
<td>Superior cerebellar artery</td>
<td>Dorsolateral upper brainstem and cerebellum and superior cerebellar peduncle</td>
<td>Superior cerebellar artery syndrome</td>
<td>Ipsilateral limb ataxia, vertigo, nystagmus, dysarthria, and gait ataxia</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>Thalamus and occipital cortex</td>
<td>Contralateral homonymous hemianopsia</td>
<td>Homonymous hemianopsia with macular sparing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Occipital lobe (dominant) plus splenium</td>
<td>Alexia without agraphia</td>
<td>Homonymous hemianopsia and alexia without agraphia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ventral occipital cortex; infralocalcine</td>
<td>Achromatopsia</td>
<td>Loss of color differentiation contralateral to the side of the lesion, can have quadrantanopsia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optic radiation or supralocalcine</td>
<td>Inferior quadrantanopsia</td>
<td>Inferior quadrantanopsia</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Myers loop (temporal lobe) or infralocalcine</td>
<td>Superior quadrantanopsia</td>
<td>Superior quadrantanopsia</td>
<td></td>
</tr>
<tr>
<td>Bilateral PCA</td>
<td>Both occipital lobes</td>
<td>Cortical blindness</td>
<td>Bilateral cortical blindness (often with denial), normal ophthalmologic examination, confabulations, visual hallucinations</td>
<td></td>
</tr>
<tr>
<td>Anterior spinal artery</td>
<td></td>
<td>Anterior spinal artery syndrome</td>
<td>Quadraparesis, bilateral pain and temperature loss, decreased sphincter tone, autonomic instability, and hyperreflexia</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Nouh et al.²
BA = basilar artery; PCA = posterior cerebellar artery.
Reasons for clinical deterioration include extension of the stroke, brain stem compression from posterior fossa edema, and recurrent stroke.43

Another misconception likely resulting in misdiagnosis is that isolated transient dizziness is never due to a transient ischemic attack (TIA), a false notion promulgated by a National Institutes of Health (NIH) consensus conference in 1975.44 In a study of 1,141 stroke patients, of 59 transient neurologic events preceding posterior circulation stroke, only five (8%) fulfilled the NIH criteria for TIA.45 The other 54 cases were isolated vertigo (n = 23); binocular visual disturbance (n = 9); vertigo with other nonfocal symptoms (n = 10); isolated slurred speech, hemisensory tingling,
or diplopia (n = 8) and nonfocal events (n = 4). Of the posterior strokes, over 8% had transient episodes of isolated vertigo in the 48 hours preceding the stroke, almost certainly, due to transient ischemia. Finally, the limitation of imaging, including DWI-MRI, likely contributes to misdiagnosis.12

Can these diagnostic traps be mitigated?

Neurologists try to match symptoms and signs with the culprit artery. For nonneurologists, distinguishing a lateral medullary from a lateral pontine infarct is less important than distinguishing a stroke from a nonstroke. The presence of robust collateral circulation and vascular anatomic variations can alter the classic posterior circulation syndromes. Therefore, we will focus the discussion on a symptoms-based approach.5

**Dizziness/Vertigo/Lightheadedness**

New evidence suggests that the time-honored “symptom quality” approach to dizziness (i.e., “what do you mean, dizzy?”) is flawed.46 Despite a weak evidence base, this approach has been the predominant paradigm for decades; an approach based on “timing and triggers” is more evidence-based.47,48 The descriptive word that patients use to describe their dizziness is diagnostically meaningless.49 Thus, we use the generic term “dizziness” to mean vertigo, unsteadiness, lightheadedness, and imbalance. Of note, the “timing and triggers” method is no different than taking a history from any other patient: How did the symptoms start? How have they evolved? Are other symptoms present? What is the medical and epidemiologic context?48

Acutely dizzy patients have one of three syndromes (Figure 3). For two of them, the acute vestibular syndrome (AVS—acute onset of persistent dizziness) and the triggered episodic vestibular syndrome (t-EVS—brief episodes of dizziness triggered by something), physical examination often supports a confident and specific diagnosis.50 The third, the spontaneous episodic vestibular syndrome (episodes of dizziness that are not triggered by anything), is most commonly caused by vestibular migraine but the most worrisome cause is posterior circulation TIA. By definition, patients are

---

**Figure 3** Diagnostic approach to the acutely dizzy patient. ATTEST = A, associated symptoms; TT, timing and triggers; ES, examination signs; and T, additional testing as needed. The first step is to take a history focused on associated symptoms, timing and triggers of the dizziness, and the overall context. Many patients’ histories will suggest a general medical cause (various toxic, metabolic, infectious, or cardiovascular causes). In this group of patients, we recommend a very brief diagnostic “stop” to reduce misdiagnosis. As part of this stop, first make sure there are no suspicious neurovestibular signs (nystagmus, limb ataxia, or gait/truncal ataxia). If a general medical cause still seems likely, evaluate and treat for the presumed diagnosis or diagnoses. For patients with a positive stop or whose history does not suggest a general medical cause, ask questions aimed at timing and triggers to place the patient into one of three categories. For patients in the AVS and t-EVS, physical examination will often allow a specific diagnosis to be made. For patients with the s-EVS, use history to try to distinguish vestibular migraine from TIA or other causes since, by definition, these patients will no longer have symptoms and their dizziness cannot be triggered at the bedside. For each vestibular syndrome, only the most common benign and dangerous diagnosis is listed. AVS = acute vestibular syndrome; BPPV = benign paroxysmal positional vertigo; CPPV = central paroxysmal positional vertigo; s-EVS = spontaneous episodic vestibular syndrome; t-EVS = triggered episodic vestibular syndrome; TIA = transient ischemic attack.
asymptomatic (otherwise they would have an AVS) and the dizziness is not triggerable so the decisions are based entirely on history; physical examination is unhelpful.

Patients with dizziness from posterior circulation stroke usually present with an AVS. Although other conditions can cause an AVS, three causes predominate—vestibular neuritis, posterior circulation stroke, and multiple sclerosis. In one observational study, a new diagnosis of multiple sclerosis accounted for less than 2% (3/170) of cases. Therefore, the major differential is neuritis versus stroke.

In 2009, a study showed that a specialized three component ocular motor examination (head impulse test, nystagmus and test of skew—HINTS) was 100% sensitive in differentiating neuritis from stroke, outperforming MRI (88% sensitive when done within the first 48 hours). The examinations were performed by neurootologists. Studies have shown that stroke neurologists and emergency physicians (EPs) with special training and using Frenzel lenses can also accurately perform the HINTS examination. Our experience is that EPs can learn to perform and interpret these tests without specialized equipment, but prospective data are lacking. Therefore, we recommend two additional components of the examination be routinely performed—a targeted neurologic examination of visual fields, cranial nerves, and the cerebellum as well as evaluation for truncal ataxia and gait.

Occasional patients with benign paroxysmal positional vertigo will endorse constant dizziness (usually with acute intermittent worsening) superficially mimicking an AVS, when in fact they have a t-EVS. In this situation, performing a Dix-Hallpike (or other provocative) maneuvers may disclose the true nature of the problem. Confidently diagnosing a common peripheral vestibular process is one way to rule out a worrisome central one. One key misconception is that dizziness exacerbated by movement means that the process is peripheral. One must distinguish dizziness that is triggered by movement, no dizziness at baseline but dizziness starts with some obligate movement (suggestive a peripheral etiology) from dizziness that is exacerbated by movement, dizzy at baseline but made worse by movement (which can be due to a central process).

**Headache and Neck Pain**

Headaches are common and mostly caused by migraine and tension-type headache. Although more common in hemorrhagic strokes, headache still occurs in between 8 and 27% of cases of ischemic stroke and is even more common in patients with posterior circulation strokes, especially cerebellar. Headache and neck pain may also indicate VAD or SAH requiring emergency evaluation and treatment. Questions aimed at distinguishing the current episode of pain from prior ones may help to distinguish migraine from another process. Systematically performing a neurologic examination that targets the visual fields, cranial nerves, and cerebellar function including gait is one strategy to reduce misdiagnosis in these patients. Absence of new neurologic deficits would make a stroke very unlikely although it is important to recall that VAD can present with isolated pain.

**Sensory Symptoms**

Because sensory symptoms are inherently subjective, patients often have difficulty describing them. A history of sudden onset of lateralized numbness suggest an ischemic etiology. Distinguishing positive symptoms from negative ones can provide additional clues (Table 2). Positive sensory symptoms (tingling, pain, dysesthesias, or an aura) are more often due to migraines, seizure disorders, and peripheral neuropathy, whereas negative sensory symptoms (numbness or loss of sensation, sight or hearing) are more often associated with ischemia.

An exception to this “rule,” thalamic strokes can present with pain (a positive sensory symptom) or hemiballismus (a positive motor symptom; Figure 2D). Pure thalamic infarcts were found in up to 11% of posterior strokes, with some thalamic involvement in up to 27% of cases. Although thalamic stroke can present with isolated sensory symptoms, associated motor symptoms, aphasia, psychiatric complaints, pain

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Positive Versus Negative Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Symptoms</strong></td>
<td><strong>Negative Symptoms</strong></td>
</tr>
<tr>
<td>Visual disturbance</td>
<td></td>
</tr>
<tr>
<td>Flashing lights</td>
<td>Field deficit</td>
</tr>
<tr>
<td>Floaters</td>
<td>Loss of conjugate gaze</td>
</tr>
<tr>
<td>Geometric shapes</td>
<td></td>
</tr>
<tr>
<td>Formed hallucinations</td>
<td></td>
</tr>
<tr>
<td>Somatosensory</td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>Hearing loss</td>
</tr>
<tr>
<td>Paresthesia</td>
<td>Numbness</td>
</tr>
<tr>
<td>Aura</td>
<td>Neglect</td>
</tr>
<tr>
<td>Tingling</td>
<td></td>
</tr>
<tr>
<td>Motor</td>
<td></td>
</tr>
<tr>
<td>Involuntary jerking movements</td>
<td>Paralysis or decreased strength</td>
</tr>
<tr>
<td>Language</td>
<td></td>
</tr>
<tr>
<td>Speaking gibberish, extra words</td>
<td>Decreased speech</td>
</tr>
</tbody>
</table>
syndromes, and decreased level of consciousness also occur. Sensory findings from PCA territory stroke are usually due to thalamic infarction related to small vessel disease, but are also found in superficial cortical PCA strokes.

**Altered Mental Status**

Altered mental status in the setting of an acute posterior stroke encompasses a broad spectrum—subtle or even transient neuropsychiatric disturbances to coma. Artery of Percheron strokes are rare and cause variable symptoms usually including decreased mental alertness, confusion, and psychiatric symptoms such as confabulation, perseveration, apathy or aggression, and hallucinations. An important clinical clue is a limitation of vertical gaze. Basilar stroke patients are obviously ill and are generally admitted; however, because timely endovascular interventions can be life-saving, early diagnosis is important. Prodromal headache and dizziness and presence of anisocoria and gaze palsies are common. Some basilar stroke patients will have involuntary movements that mimic seizure, leading to misdiagnosis of status epilepticus. Abrupt onset of altered mental status and any of these findings should prompt CT angiogram.

**Nausea and Vomiting**

Nausea or vomiting occurred in 27% of the 407 patients in the New England posterior circulation stroke registry. Vomiting, an independent risk factor for misdiagnosis, may be so severe as to divert attention from other symptoms such as headache or dizziness and rarely occurs in isolation. In a case series of 18 patients with strokes involving the lower brainstem (nucleus prepositus hypoglossi) presenting with dizziness and prominent vomiting, all had some other finding, most commonly trimal ataxia, nystagmus, or facial palsy.

**Visual Symptoms**

Examination of visual fields, pupils, and eye movements may help reduce misdiagnosis. Pertinent physical findings include a Horner's syndrome, nystagmus, diplopia, and decreased vision (field cut or blurred vision). The ptosis from a Horner's syndrome is mild and can mimic enophthalmos. The small pupil may be overlooked and is easier seen in a dark room (accentuating the difference in size). The ptosis is ipsilateral to the small pupil and less apparent than in a third nerve palsy (where the ptosis is more apparent and ipsilateral to the large pupil). Patients with a Horner's syndrome from a lateral medullary stroke usually have other findings as well such as ataxia, diplopia, and sensory changes and often are due to VAD.

Although nystagmus is a physical finding, some patients will spontaneously complain that their eyes or that their visual world are moving (oscillopsia). The details of nystagmus, not just the presence or absence, are diagnostically useful. Persistent nystagmus that is direction-changing, vertical, or torsional should be considered central. A sudden-onset internuclear ophthalmoplegia, caused by a lesion in the medical longitudinal fasciculus characterized by an impairment of adduction of the affected eye on conjugate lateral gaze and nystagmus, is also often due to stroke. Diplopia is another symptom/sign of posterior circulation stroke.

Unilateral PCA infarcts can cause a contralateral homonymous hemianopsia or quadrantanopsia. Patients may be unaware of the field cut and often describe decreased or poor vision in one eye or simply bumping into things on the side of the field cut. In the Lausanne stroke registry of the nearly 3,400 patients, 117 were found to have a pure superficial PCA territory stroke. Of the 117, a total of 78 (67%) had a hemianopia, 26 (22%) had a quadrantanopia, and eight (7%) had bilateral field abnormalities.

Assessment of visual fields by bedside confrontation methods is not as accurate as formal perimetry but useful when positive. Hallucinations can occur in thalamic and superficial PCA infarcts, another exception to the general rule about positive (hallucinations) versus negative (lack of vision) symptoms. In superficial PCA strokes, hallucinations, both simple and complex, were found in 10% of 117 patients. This is similar to the Charles Bonnet syndrome, which follows an acute visual loss.

**Language and Speech Deficits**

Dysarthria, clumsy or slurred speech, occurs with strokes that interfere with the process of forming words and are more common with cortical PCA infarcts. Patients describe “thick” or “heavy” speech. Of 62 consecutive patients with dysarthria from stroke, 61% were due to posterior circulation strokes, mostly in the brain stem and cerebellum. Overall, dysarthria occurs in 29% of cerebellar strokes (as a symptom) and 46% (as a physical finding). Dysarthria was found in nearly half of patients with thalamic
Having the patient say the word “Pawtucket” which tests the “pa,” “ta,” and “ka” sounds (which are made using three different parts of the mouth and tongue) is a quick way to test for dysarthria.

Various types of vascular aphasias also occur.77 Having a patient say, write, and then read a phrase or short sentence is a simple bedside test of these functions.

**Cranial Neuropathy-related Symptoms**

Small brain stem strokes can present with prominent cranial nerve–related symptoms, due to damage of the nuclei themselves or to ischemia of the postnuclear fibers as they traverse the brain stem. These patients can present with abnormal eye movements, facial hypesthesia or weakness, acute dizziness or hearing loss, and difficulty phonating and swallowing. In lateral medullary infarct, the common sensory deficit is loss of pain and temperature, leaving light touch intact. Careful cranial nerve testing is important in all patients in whom posterior stroke is a consideration.

**Neuroimaging**

Some form of brain imaging is required in patients with acute stroke; however, physicians must understand the limitations of early imaging. Brain CT is generally the first test performed; however, its sensitivity for posterior circulation ischemic stroke ranges from 7% to 42%.9–11 For patients presenting with acute dizziness, the sensitivity is even lower. More than one-third of patients with acute PCA strokes will show a hyperdense PCA sign on noncontrast CT.78 DWI-MRI also has important limitations in early-presenting posterior circulation ischemic strokes, especially for those presenting with isolated dizziness. In this setting, the false-negative rate of DWI-MRI is after onset of symptoms ranges from 12% to 18%.53,79,80 For those with small strokes, the false-negative rate is 53%.79 Physical examination using HINTS is more accurate in these patients than MRI, at least when done by specialists and is well within the scope of practice for EPs.

Overall, for patients with acute ischemic stroke, the evidence shows that the false-negative rate of DWI-MRI is 6.8% with an odds ratio (OR) for having a falsely negative DWI-MRI strongly associated with posterior circulation location (OR = 5.1, 95% confidence interval = 2.3–11.6, p < 0.001).12 If there is still significant diagnostic ambiguity after physical examination and any imaging that might have been done, it is safest to admit the patient to the hospital for further testing and urgent neurologic consultation.

If VAD or basilar stroke are being considered, vascular imaging is indicated. Although catheter digital subtraction angiography (DSA) remains the criterion standard, magnetic resonance angiography (MRA) and CT angiography (CTA) are typically performed. A CTA has advantages over MRA and is often readily available in the ED.74,81 Discussing the differential diagnosis with the radiologist may help to focus their attention to the relevant anatomy and, if clinical suspicion is high and the first imaging study is negative, an alternate study, or occasionally, DSA may be diagnostic.

**CONCLUSIONS**

Patients with posterior circulation cerebrovascular events are often misdiagnosed. Strategies that may decrease misdiagnosis include:

- With any neurologic symptoms, always establish if the onset is abrupt or not because abrupt onset suggests ischemia.
- Be aware of the nonspecific symptoms of posterior circulation ischemia and that variant vascular anatomy can cause nonclassic findings.
- For dizzy patients with an AVS, use a timing and triggers diagnostic approach taking a history as one would with any other chief complaint instead of focusing the history on the response to the question, “What do you mean by dizzy?”
- Perform a brief posterior circulation physical examination that targets the brain stem, cerebellum, and occipital lobes.
- Be aware of the limitations of brain imaging and adopt a nuanced approach that recognizes that initial brain imaging is frequently nondiagnostic.

**References**

10. Kerber KA, Schweigler L, West BT, Fendrick AM, Mor-14. van Raamt AF, Mali WP, van Laar PJ, van der Graaf Y.

13. Alpers BJ, Berry RG, Paddison RM. Anatomical studies
12. Edlow BL, Hurwitz S, Edlow JA. Diagnosis of DWI-nega-

18. Thierfelder KM, Baumann AB, Sommer WH, et al. Ver-

tebral artery hypoplasia: frequency and effect on cerebellar


Vertebral artery hypoplasia and posterior circulation infarc-
tion in patients with isolated vertigo with stroke risk fac-

16. Jimenez Caballero PE. Bilateral paramedian thalamic

artery infarcts: report of 10 cases. J Stroke Cerebrovasc Dis
2010;19:283–9.


14. van Raamt AF, Mali WP, van Laar PJ, van der Graaf Y.
The fetal variant of the circle of Willis and its influence on the

13. Alpers BJ, Berry RG, Paddison RM. Anatomical studies of
the circle of Willis in normal brain. AMA Arch Neurol

the NIHSS does not equal the absence of stroke. Ann

9. Hwang DY, Silva GS, Furie KL, Greer DM. Comparative
sensitivity of computed tomography vs. magnetic resonance
imaging for detecting acute posterior fossa infarct. J Emerg

8. Edlow BL, Hurwitz S, Edlow JA. Diagnosis of DWI-nega-
tive acute ischemic stroke: a meta-analysis. Neurology

7. Inoa V, Aron AW, Staff I, Fortunato G, Sansing LH. Lower NIH stroke scale scores are required to accurately

6. Compter A, Kappelle LJ, Algra A, van der Worp HB. Nonfocal symptoms are more frequent in patients with
vertebral artery than carotid artery stenosis. Cerebrovasc

5. Searls ED, Pazdera L, Korbel E, Vysata O, Caplan LR.

Atelectasis – ARDS – adult respiratory distress syndrome


the NIHSS does not equal the absence of stroke. Ann

1. Kerber KA, Schweigler L, West BT, Fendrick AM, Morganstern LB. Value of computed tomography scans in ED
dizziness visits: analysis from a nationally representative

1. Kerber KA, Schweigler L, West BT, Fendrick AM, Morganstern LB. Value of computed tomography scans in ED
dizziness visits: analysis from a nationally representative

1. Kerber KA, Schweigler L, West BT, Fendrick AM, Morganstern LB. Value of computed tomography scans in ED
dizziness visits: analysis from a nationally representative

1. Kerber KA, Schweigler L, West BT, Fendrick AM, Morganstern LB. Value of computed tomography scans in ED
dizziness visits: analysis from a nationally representative
emergency department by emergency medicine and neurology services. Stroke 2016;47:668–73.


Every year, more than one-third of deaths in America occur in a hospital. More than half of patients who are admitted to the hospital present through the emergency department (ED). While the ED is the “front line” for the majority of admitted patients, it is also where over 137 million undifferentiated patients present every year, with just 14% admitted to the hospital. Acutely ill patients with hypotension, tachycardia, or hypoxemia are often easily identified and admitted to the hospital.

However, little is known regarding in-hospital death of ED patients who present to the hospital with normal vital signs. One systematic review concluded that abnormal vital signs alone fail to identify many patients at risk of adverse events. For instance, one study reported that only 28% of critical events (death, intensive care unit admission, or cardiac arrest) occurred in patients with two or more abnormal signs in a population of hospitalized veterans.

Despite the inadequacy of vital signs alone as a predictor of death, no published studies have specifically analyzed how often patients who visit the ED with normal vital signs die in the hospital and what factors are associated with these potentially unanticipated deaths. Given that the unexpected nature of these deaths impacts both clinicians and patients’ loved ones perhaps differently than others, we address this important gap by examining the incidence of these deaths among ED patients with normal vital signs and describing common visit reasons and diagnosis codes.

As vital signs are standard criteria for initial triage and treatment decisions, these data could help inform strategies to improve quality of care.

In detail, this study is a descriptive analysis comparing characteristics of deaths and nondeaths among patients who presented to the ED with normal initial vital signs. We used data collected in the 2007 to 2016 years of the National Hospital Ambulatory Medical Care Survey (NHAMCS), an annual data set from the Centers for Disease Control and Prevention. NHAMCS provides data on service provision and utilization from a national sample of visits to hospital emergency and outpatient departments and ambulatory surgery centers.

We selected only visits that originated in the ED for this study. Visits from patients younger than 9 years were excluded, as normal ranges for vitals differ for young children, and our sample lacked deaths in this age range. Visits without at least one diagnosis given were excluded. We excluded observations with missing or incomplete data in one or more included vital signs, which accounted for 36% of deaths and 18% of nondeaths.

Based on previous guidelines and clinical judgment, we defined normal initial vital signs as systolic blood pressure of 90 to 140 mm Hg, diastolic blood pressure of 60 to 90 mm Hg, pulse of 60 to 100 beats/min, respiratory rate of 12 to 20 breaths/min, and oxygen saturation of 94% to 100% from pulse oximetry. Observations meeting inclusion criteria were stratified into death and nondeath cohorts. Death
was defined as either “died in ED” or a hospital discharge status of “dead.”

The two outcomes of interest were: 1) the most common physician’s International Statistical Classification of Diseases and Related Health Problems (ICD) diagnoses as grouped into Clinical Classifications Software (CCS) codes and 2) patient-reported reasons for visit.

Clinical Classifications Software codes are a diagnosis and procedure categorization tool developed by the Agency for Healthcare Research and Quality (AHRQ) to aggregate the more than 14,000 ICD codes into 285 clinically meaningful groups for researchers to analyze patterns of diagnoses, utilization, and outcomes. Each record in NHAMCS listed up to three ICD diagnoses, and all were included for analysis. NHAMCS used ICD-9 diagnoses codes up to the 2015 survey. The 2015 CCS software applied only to ICD-9 codes, so our CCS analysis excluded data from 2016. We opted against using a beta version of the ICD-10 CCS software due to reliability concerns, because preliminary analyses suggested some discontinuities between the ICD-9 and ICD-10 tools.7 Regarding reasons for visit, data from all years and all reasons listed (up to five per record) were included for analysis.

We compared demographic composition by sex, age, race, and insurance payer type as well as median and average vital signs between deaths and nondeaths. For statistical analysis, we used the two-sided t-test to test for differences in means and the Pearson chi-square test for differences in proportions. We applied patient visit weights to generate national estimates. These weights account for probability of visit selection, nonresponse, and ratio of sampled hospitals to hospital universe.6 This study was deemed exempt from the University of California at San Francisco Committee on Human Research.

We performed a national study of ED visits with normal initial vitals using the NHAMCS data set, which uses weighted sampling methods to achieve nationally representative visit information. The sample in NHAMCS of patients with normal vital signs included 77,398 visits, of which 67 were deaths. This translated to 357,007,550 visits with normal vital signs and 255,751 deaths over a 10-year period. The average age of deaths was older than that of nondeaths (73 years vs. 33 years, p < 0.001), and there were accordingly more Medicare patients (30% vs. 12%, p < 0.001).

The most common CCS codes for deaths were septicemia, fluid and electrolyte disorders, residual codes, and acute and unspecified renal failure, which together comprised 29% of codes (95% confidence interval [CI] = 20.3% to 39.6%). In comparison, sprains and strains (5.2%, 95% CI = 4.9% to 5.4%) and superficial injury (5.1%, 95% CI = 4.9% to 5.3%) were the most common CCS codes among nondeaths (Table 1A). Of note, there is little overlap in the most common CCS codes with just superficial injury or contusion and nonspecific chest pain as commonalities in the two groups.

For patient-reported reasons for visit, patients who died most frequently reported psychosis-related symptoms, general weakness, accidents, abdominal pain/cramps/spasms, and tiredness/exhaustion (Table 1B). These top five reasons accounted for nearly a quarter of visit reasons (24.4%, 95% CI = 15.8% to 35.7%). Patients alive at discharge most commonly reported abdominal pain, chest pain, or headache. Of the top 10 reasons for visits for nondeaths, all except for two were in the top reasons for visits for deaths.

In summary, our national study of ED visits from 2007 to 2016 with normal vital signs found that a very small percentage (<1%) resulted in ED death or death during the hospital stay. Three main findings stand out. First, septicemia, fluid and electrolyte disorders, and acute renal failure accounted for the greatest number of deaths from patients with normal vitals. Second, common visit reasons of patients who died included psychosis-related problems and general weakness. Third, there was substantial overlap in visit reasons for deaths and nondeaths.

The small percentage of deaths supports normal vital signs at presentation as a strong predictor of survival. Still, some diagnoses associated with death often lack manifestations in initial blood pressure or other vital readings. Our data on diagnoses and visit reasons common to death offer insights for better identifying patients at risk, especially as extended wait times are associated with higher risk of mortality.8

As diagnoses are not often obvious or known at presentation, patients’ self-reported reason for visit is a critical data point in initial triage decision making. The most common visit reasons among deaths included broadly concerning symptoms like chest pain and abdominal pain but also reasons less likely to evoke urgent evaluation like psychosis-related symptoms and general weakness. Although it presents with relatively nonspecific signs, altered mental status, a
symptom of psychosis, is associated with high resource use and mortality, which potentially suggest inadequate risk stratification for these patients. Our findings also have important policy implications, given the increasing scrutiny and refusal of payment for patients who visit the ED for what has been deemed “nonemergency” or “inappropriate” visits. Most recently, Blue Cross & Blue Shield of Mississippi initiated a policy of adjusting reimbursements if the visit level is not in alignment with the diagnosis severity. While studies have shown the dangers of using a diagnosis-based denial of payment, our results show that using chief complaints could also be imprecise and potentially dangerous. While policymakers and health care providers alike might view psychosis-related complaints and general weakness as nonemergent, for example, our findings show that such assumptions would be seriously mistaken in certain cases.

This study has several limitations. Because of missing data and strict inclusion criteria for normal vital signs, our sample size was small despite the 10-year

### Table 1
Most Common CCS Physician Diagnoses and Patient Reasons for Visit Among Deaths Versus Nondeaths for Visits With Normal Initial Vitals

<table>
<thead>
<tr>
<th>A. Diagnosis</th>
<th>Weighted No.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deaths (n = 580,251 weighted codes)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Septicemia (except in labor), fluid and electrolyte disorders, residual codes (unclassified), acute and unspecified renal failure</td>
<td>168,212</td>
<td>29.0</td>
</tr>
<tr>
<td>Pneumonia (except that caused by tuberculosis or sexually transmitted disease), gastrointestinal hemorrhage, secondary malignancies, superficial injury/contusion, other liver diseases, respiratory failure/insufficiency/arrest (adult), nonspecific chest pain, coma/stupor/brain damage</td>
<td>145,774</td>
<td>25.1</td>
</tr>
<tr>
<td><strong>Nondeaths (n = 517,091,755 weighted codes)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sprains and strains</td>
<td>26,776,301</td>
<td>5.2</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>26,280,524</td>
<td>5.1</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>25,930,819</td>
<td>5.0</td>
</tr>
<tr>
<td>Spondylosis/intervertebral disc disorders/other back problems</td>
<td>18,341,049</td>
<td>3.5</td>
</tr>
<tr>
<td>Other nervous system disorders</td>
<td>14,886,872</td>
<td>2.9</td>
</tr>
<tr>
<td>Other upper respiratory infections</td>
<td>14,518,397</td>
<td>2.8</td>
</tr>
<tr>
<td>Urinary tract infections</td>
<td>14,168,458</td>
<td>2.7</td>
</tr>
<tr>
<td>Other symptoms/problems related to psychosis, general weakness, accident (NOS), abdominal pain/ cramps/spasms (NOS), tiredness/exhaustion</td>
<td>116,226</td>
<td>24.4</td>
</tr>
<tr>
<td>Blood in stool (melena), urinary tract infection (NOS), chest pain, vomiting, fainting (syncope), difficulty in swallowing (dysphagia), shortness of breath, nausea, other malignant neoplasms, other specific therapeutic procedures</td>
<td>119,051</td>
<td>25.0</td>
</tr>
<tr>
<td><strong>B. Reason for Visit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal pain, cramps, spasms, NOS</td>
<td>34,101,320</td>
<td>5.1</td>
</tr>
<tr>
<td>Chest pain</td>
<td>22,402,580</td>
<td>3.3</td>
</tr>
<tr>
<td>Headache, pain in head</td>
<td>22,323,460</td>
<td>3.3</td>
</tr>
<tr>
<td>Nausea</td>
<td>19,700,293</td>
<td>2.9</td>
</tr>
<tr>
<td>Vomiting</td>
<td>19,651,206</td>
<td>2.9</td>
</tr>
<tr>
<td>Back pain, ache, soreness, discomfort</td>
<td>19,088,376</td>
<td>2.8</td>
</tr>
<tr>
<td>Accident, NOS</td>
<td>16,104,925</td>
<td>2.4</td>
</tr>
<tr>
<td>Cough</td>
<td>14,288,418</td>
<td>2.1</td>
</tr>
<tr>
<td>Vertigo—dizziness</td>
<td>11,718,312</td>
<td>1.7</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>10,707,201</td>
<td>1.6</td>
</tr>
</tbody>
</table>

CCS = Clinical Classifications Software; NHAMCS = National Hospital Ambulatory Medical Care Survey; NOS = not otherwise specified. Top reasons for deaths were grouped into cells with unweighted observations of at least 30 per NHAMCS guidelines. 
study period and should be considered with appropriate caution. Second, the vital signs recorded in NHAMCS reflect the initial set of vitals taken and does not reflect vital sign progression throughout the patient’s stay. It is possible that vital signs later became abnormal either in the ED or in the hospital (by definition, those who died would all eventually have had abnormal vitals). We cannot exclude some degree of measurement error from the variability of providers and techniques used to take vitals. NHAMCS also lacks underlying cause of death ICD codes and historical clinical data, so we are unable to determine the provider’s impression of cause of death or exclude cases such as terminal cancer where death may not be unexpected in the setting of normal initial vitals. Finally, the descriptive study design lacks statistical adjustment for other factors and precludes causal inference.

In conclusion, we identified septicemia, fluid and electrolyte disorders, acute renal failure, psychosis, general weakness, and more advanced age as factors common to patients presenting to the ED with normal vital signs who later died. With a substantial overlap of visit reasons between visits resulting in deaths and those that did not, our findings support a careful analysis of health plan initiatives that seek to decrease “avoidable visits,” especially those using chief complaint.

The authors thank Dr. Astha Singhal for her advice on this study.

References

Supporting Information
The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13833/full
BACKGROUND

Discussions about goals of care at the end of life are an essential component of emergency medicine. Such discussions are aided by accurate prognosis, so that lifesaving interventions can be provided to those in need, but hopefully without providing overly aggressive care to patients with little hope of recovering. Such prognosis is difficult, especially in the chaotic and time-limited environment of an emergency department (ED). An accurate decision tool would be welcomed. Unfortunately, available tools are only modestly accurate and have not been rigorously validated.1,2 Basic demographic data alone are insufficient to predict individual patient risk.3 Therefore, the purpose of this study was to validate a personalized risk score—the Criteria for Screening and Triaging to Appropriate alternative care (CriSTAL)—in older patients presenting to the ED.4

ARTICLE SUMMARY

This is a prospective observational study looking to validate the CriSTAL decision rule, designed to predict the short-term risk of death in an elderly population. They studied two separate cohorts of ED patients over the age of 65. The first consisted of 1,143 patients from five hospitals in Australia, and the second consisted of 349 patients from a single hospital in Ireland. The primary outcome was mortality at 3 months. Although there was a statistical difference in the scores between deceased and surviving patients, it is not clear that the difference is big enough to be clinically useful.

QUALITY ASSESSMENT

Validating decision tools in different clinical settings is essential, and this international prospective cohort is a welcome addition to the literature. Unfortunately, there are a number of issues that will limit clinical applicability. Most importantly, although the CriSTAL score was statistically higher in patients who died within 3 months, there was significant overlap between the groups that will limit usability for individual patients. The more familiar sensitivity, specificity, and likelihood ratios are not included in this article, but based on the numbers presented it seems unlikely that the results of this score will change pretest probability enough to help emergency physicians make accurate predictions about short-term death. Furthermore,
because different versions of the score were tested, and different cutoffs were used, it is likely that the score was overfit to these specific populations and will perform less well in external validations. Clinician accuracy with the score (or inter-rater reliability) was not tested. With only a 2-point difference between the groups, any variability in scoring between clinicians could significantly alter the results. Ideally, we would also like to see the score compared to clinician judgement before it is used clinically.

**KEY RESULTS**

They included 1,143 patients from five hospitals in Australia in the derivation cohort and 349 patients from a single hospital in Ireland in a validation cohort. The mean age was over 75 years and there was about a 50/50 split between males and females in both groups. Approximately 5% of both groups had do not attempt resuscitation orders or advanced directives. The overall 3-month mortality was 10.1% in the Australian Cohort and 12.9% in Ireland. The mean CriSTAL scores were statistically higher among patients who were deceased at 3 months than among survivors. In Australia the scores were 8.1 (95% confidence interval [CI] = 7.7 to 8.6) versus 5.7 (95% CI = 5.1 to 6.2) for deceased and survivors, respectively. In Ireland, the scores were 7.7 (95% CI = 6.9 to 8.5) versus 5.7 (95% CI = 5.5 to 6.0).

**AUTHORS’ COMMENTS**

An accurate tool for prognosis at the end of life would be incredibly useful in the ED. Unfortunately, based on this study, it does not look like the CriSTAL tool provides an accurate enough prognosis of short-term mortality to be useful in emergency care. It is worth reviewing the individual factors that increase an elderly patient’s short term risk of mortality, but we should be cautious about making any definitive predictions based on these factors.

**TOP SOCIAL MEDIA COMMENTARY**

ACEP Geriatric Sect: Question 1 for latest #SGEMHOP authors—[Cardona 2019]⁴ notes accuracy of ISAR to CriSTAL, but [Carpenter 2015]⁵ highlighted numerous #GeriED instruments to predict mortality yet none displayed LR+ > 10 or LR− < 0.10 (or came close). How does CriSTAL compare to all?

Lauren Southerland (@LSGeriatricEM): #SGEMHOP as any good skeptic, I would be interested in seeing a comparison of the predictive tool to clinical gestalt of physicians.

Jay Banerjee (@POBanerjee): An excellent critique of the tool has been presented. I would prefer the CFS, mean time of < 1 min. Boils down to degree of certainty a clinician is comfortable with to discuss EoL care; the AUC just props it up!

Simon Mooijaart (@DrSimonPM): Never too early to discuss EoL, provided done correctly…

**ACEP Geriatric Sect:** CriSTAL appears time-consuming and complex to compute. Beyond reliability, how would clinicians incorporate into busy ED practice? Does CriSTAL begin to address myriad challenges of #GeriED

Dr. Cardona’s response: CriSTAL is not time-consuming. Our experience with over 3,000 in 11 hospitals across 4 countries demonstrates that it takes less than 5 minutes to document form the clinical record, even if the data collector is a junior clinician who does not have prior knowledge of the patient. And it can save a week in intensive care, so it’s worth the investment. We have also computerised the tool so the calculation of the CriSTAL score is automated.

Ken Milne: We see multiple elderly patients every shift in the ED. Who here personally has 5 more minutes per elderly patients? Who else working in the ED has 5 more minutes? Do we have evidence that using
the tool does save a week in the ICU? Everything seems to be a priority in the ED. When we add one thing we take away from another thing. Even if it does not take much time, all those little things do add up.

**TAKE-TO-WORK POINTS**

Although accurate prognostication at the end of life would be incredibly valuable, the CriSTAL decision tool is unfortunately not accurate enough for clinical use at this time. However, you may wish to consider the components of this rule when discussing end of life care with your elderly patients.

**References**

Tranexamic Acid for the Treatment of Epistaxis

Michael Gottlieb, MD1, Alex Koyfman, MD2, and Brit Long, MD3

NARRATIVE

Epistaxis is a common reason for patients to present to the emergency department (ED), reflecting one of every 200 ED visits in the United States.1 While many cases of epistaxis are self-limiting, those requiring medical treatment can be associated with significant time and health care costs.2 Additionally, nasal packing and hemostatic matrices can be painful and require the patient to return for at least one follow-up visit. Therefore, identifying an effective and inexpensive treatment is of particular importance. Tranexamic acid is an anti-fibrinolytic agent that has been proposed as one potential modality for this.

The Cochrane Review discussed here included randomized controlled trials comparing tranexamic acid (TXA) in any formulation (e.g., delivered orally, intravenously, or topically) with usual care versus usual care with placebo, usual care with any other hemostatic agent, or usual care alone.3 The primary outcome was the proportion of patients with rebleeding within 10 days and significant adverse events (i.e., seizures, thromboembolic events).

Among the six trials (n = 692 patients), two studies used oral TXA,4,5 while the remaining four used topical TXA.6–9 For the primary outcome (n = 225 patients), TXA was associated with lower rates of rebleeding at 10 days (47% vs. 67%; relative risk [RR] = 0.71, 95% confidence interval = 0.56 to 0.90; absolute risk difference = 20%; number needed to treat = 5; moderate-quality evidence) compared to placebo. There were no significant differences between groups for adverse events, although only five of the trials reported adverse events.4,5,7–9 The included trials did not report outcomes requiring further intervention (e.g. repacking, surgery, embolization).

Another recent systematic review of topical TXA in epistaxis identified faster discharge rates, reduced rebleeding at 24 hours, and greater patient satisfaction.
with TXA, but no difference in rebleeding at 30 minutes. While the studies utilized different search strategies, both were informed by similar studies in their reviews. The current review further supports the potential value of this intervention.

**CAVEATS**

Interpreting the results of the systematic review and meta-analysis discussed here warrants some caution. First, there was significant clinical heterogeneity in the study populations, with differences in the routes of administration (i.e., oral vs. topical), comparator groups (placebo vs. anterior nasal packing), and primary outcomes. Additionally, only three studies assessed the primary outcome of rebleeding at 10 days, while several other trials had different individual study outcomes (e.g., bleeding control within 30 minutes). Moreover, there was poor reporting of adverse events in the included studies. However, no significant adverse events were reported and most events were considered minor in nature (e.g., nausea, vomiting). Further, while rebleeding at 10 days is a clinically significant outcome, there were limited data on other ED-relevant outcomes (e.g., time to bleeding cessation, time to discharge, return to ED rates). Anterior epistaxis might also respond differently to treatment than posterior epistaxis. Only three of the included trials assessed the location of bleeding and only enrolled patients with anterior epistaxis. Other trials did not specify the location of bleeding. Another important limitation is that a large number of patients in some of the trials were on antiplatelet agents (i.e., aspirin, clopidoogrel, or both); this could have affected the outcomes and contributed to the clinical heterogeneity in response to treatment. Future studies should identify what subgroup of patients (e.g., anterior vs. posterior epistaxis, antiplatelets use) are most likely to benefit from TXA, which delivery route is most effective, and how to better assess differences in adverse events.

The existing evidence supports the efficacy of TXA to reduce the risk of rebleeding at 10 days among adult patients. Despite inconsistent reporting of adverse events, the occurrence of such events appears to be unlikely, particularly with topical use. Therefore, we have assigned a color recommendation of green (benefit > harm) to the use of TXA for epistaxis.

Editor’s Note: Brass Tacks are concise reviews of published evidence. This series is a result of collaboration between Academic Emergency Medicine and the evidence-based medicine website, www.TheNNT.com. For inquiries please contact the section editor, Shahriar Zehtabchi, MD (Shahriar.Zehtabchi@downstate.edu).

**References**

Factors Predicting Difficult Endotracheal Intubation

Brit Long, MD1, Alex Koyfman, MD2, and Michael Gottlieb, MD3

Summary

Factors Predicting Difficult Endotracheal Intubation

Positive LR findings (LR+)

- History:
  - History of difficult intubation = 16–19
- Signs:
  - Upper lip bite test grade 3 = 14
  - Shorter hyomental distance = 6.4
  - Retrognathia = 6
  - Combination of findings on Wilson score = 9.1
  - Impaired neck mobility = 4.2
  - Modified Mallampati score > 3 = 4.1

Negative LR findings (LR-)

- History:
  - Absence of a history of difficult intubation = 0.72–0.82
- Signs (absence of):
  - Upper lip bite test grade 3 = 0.42
  - Shorter hyomental distance = 0.84
  - Retrognathia = 0.85
  - Combination of findings on Wilson score = 0.60
  - Impaired neck mobility = 0.77
  - Modified Mallampati score > 3 = 0.52

Who was in the studies

- 62 studies comprising 33,559 patients, with all intubations completed in the operating room

Editor’s Note: Brass Tacks are concise reviews of published evidence. This series is a result of collaboration between Academic Emergency Medicine and the evidence-based medicine website www.TheNNT.com. For inquiries please contact the section editor, Shahriar Zehtabchi, MD (e-mail: Shahriar.zehtabchi@downstate.edu).

NARRATIVE

Endotracheal intubation is a common procedure in emergency medicine, and recognizing a potentially difficult intubation is imperative in planning for the procedure. While the “can’t intubate, can’t ventilate” scenario is rare, it is catastrophic if the airway operator is not prepared.1–3 Thus, predicting factors associated with difficult endotracheal intubation is important for emergency clinicians, with consideration of airway adjuncts such as video laryngoscopy, supraglottic airway devices, and cricothyrotomy.4 Some of the factors associated with intubation failure (or difficult intubation) include a history of prior difficult intubation, limited upper lip bite test (the patient bites the upper lip with his/her lower incisors), retrognathia, short thyromental and hyomental distance, decreased cervical spinal motion, higher modified Mallampati classification (defined by visibility of oropharyngeal structures with maximal mouth opening and tongue protrusion), and composite scores such as the Wilson score (incorporating weight, mobility of the cervical spine and jaw, retrognathia, and incisor appearance).4–7

The systematic review discussed here included studies evaluating risk factors (based on medical history or physical examination) or clinical tests that could...
predict difficult intubation (outcome) in adults (>18 years) undergoing endotracheal intubation with direct laryngoscopy. Authors assessed the quality of the included trials using the Rational Clinical Examination series quality checklist. The authors of the meta-analysis identified 62 relevant studies (n = 33,559 patients), which were all performed in the operating room (OR). The overall prevalence of difficult intubation was 10% (95% confidence interval [CI] = 8.2%–12%), which was most commonly defined by Cormack-Lehane grade 3 or 4. Cormack-Lehane grade 3 is defined as only the epiglottis visualized and grade 4 by neither glottis nor epiglottis seen on direct laryngoscopy. Other definitions included combination of Cormack-Lehane grade with additional requirements such as number of intubation attempts, time, and use of bougie in six studies; percentage of glottic opening in one study; Intubation Difficulty Scale score > 5 in three studies; or minimum intubation time or number of attempts in five studies. History of prior difficult intubation was associated with an increased likelihood of difficult intubation (positive likelihood ratio [LR+] = 16–19). Clinical examination findings including upper lip bite test class 3, defined as inability to bite any part of the upper lip with lower incisors, was a strong predictor of difficult intubation (LR+ = 14, 95% CI = 8.9–22). Other findings, such as retrognathia (LR+ = 6.0, 95% CI = 3.1–11), hyomental distance < 3 to < 5.5 cm (LR+ = 6.4, 95% CI = 4.1–10), impaired neck mobility (LR+ = 4.2, 95% CI = 1.9–9.5), impaired mouth opening (LR+ = 3.6, 95% CI = 2.1–6.1), and the modified Mallampati score > 3 (LR+ = 4.1, 95% CI = 3.0–5.6) also predicted difficult intubation. The Wilson score was also a strong predictor of difficult intubation (LR+ = 9.1, 95% CI = 5.1–16). However, no clinical factor or composite score was useful in excluding difficult intubation. Sensitivity analyses did not change interpretation of results.

CAVEATS

The trials included in the systematic review (rated as high-quality) identified certain findings are associated with an increased risk of difficult intubation. However, none of the findings were sufficient to exclude this. There was some variability in the reference standard used among studies to define a difficult airway, although the majority of studies incorporated the Cormack-Lehane classification system. In addition, studies that used the time of intubation or number of intubation attempts to define a difficult airway might have been influenced by the individual clinician’s ability or experience in intubation. Several predictors such as impaired cervical motion and retrognathia are subjective and vulnerable to interobserver variability.

Authors of the systematic review limited their analysis to studies with independent assessments of predictors and outcomes in order to reduce bias. This led to exclusion of studies conducted in emergency settings. Therefore, all studies included in the systematic review were performed in the OR setting, limiting the applicability to the emergency department (ED) setting. Endotracheal intubation in the OR setting is more commonly associated with a nonemergent need for endotracheal intubation. While ED patients may differ with regard to mental and hemodynamic status, presence of gastric contents or vomiting, and ability to cooperate well with the assessments, knowledge of factors associated with difficult intubation and adequate preparation are still essential. Finally, this analysis evaluated only direct laryngoscopy. Therefore, the results of this review may not reflect current airway technology incorporating video laryngoscopy, extraglottic airway devices, and other advanced techniques.

In summary, the existing evidence indicates that several findings predict a difficult endotracheal intubation, but their absence cannot reliably exclude this scenario. The most accurate assessment was the upper lip bite test, followed by shorter hyomental distance, retrognathia, impaired neck mobility, modified Mallampati score > 3, and the Wilson score. Future studies should incorporate new airway technology such as video laryngoscopy and include emergency situations.

References

to predict difficult tracheal intubation: a prospective study.
6. Samsoon GL, Young JR. Difficult tracheal intubation: a
7. Wilson ME, Spiegelhalter D, Robertson JA, Lesser P. Pre-
8. Detsky ME, Jivraj N, Adhikari NK, et al. Wi-
ll this patient be difficult to intubate? The Rational Clin-
cal Examination Systematic Review. JAMA 2019;321:
493–503.
9. Simel D. Update: primer on precision and accuracy. In:
Simel D, editor. The Rational Clinical Examination. New
10. Cormack RS, Lehane J. Difficult tracheal intubation in
Over the past decade, there has been tremendous growth in freestanding emergency departments (ED) in the United States. While freestanding EDs have in fact been around since the 1970s, this recent growth has been spurred by a variety of factors including: overcrowded hospital-based EDs, expanding demands for emergency care, a desire by health care systems to create satellite locations to attract patients, health care entrepreneurship, and evolving state rules that have made it easier in some localities to open them.¹

Freestanding EDs are physically separate from hospitals. Freestandings may either be health system or hospital-affiliated, or independent, depending on ownership and local state laws. Yet like hospital-based EDs, freestanding EDs are open 24/7 and can evaluate and stabilize patients with resources similar to any ED: advanced diagnostic and laboratory facilities and staffing by emergency physicians, but in general do not have onsite, in-person specialist care like surgery or neurology.

The meteoric rise of freestanding EDs has generated considerable debate, and to put it bluntly, deep acrimony in some parts of the country. First, concerns have been raised about freestanding ED billing practices. Freestanding EDs charge both a provider fee and a facility fee—similar to hospital-based setting and have largely similar prices.² Critics contend that this is unjustified because they tend to see lower-acuity patients than EDs and are not exposed to the same Emergency Medical Treatment and Active Labor Act (EMTALA)-related obligations. However, freestanding ED operators argue that acuity is largely similar and that facilities fees are justified as considerable fixed investments (i.e., CT scanners) are required to offer ED-level services. Freestanding EDs also have higher staffing costs with board-certified emergency physicians, compared to urgent care centers, which are often staffed by nonboarded doctors or advanced practice providers. Data also show that freestanding EDs tend to preferentially locate in more affluent, richly populated areas and not in higher need, poorer communities.³⁴ However, this practice is similar to urgent care centers, which also locate in similar, wealthy areas.⁵

From the perspective of patients, there are also concerns over the high bills after freestanding ED care—particularly when the freestanding is out-of-network with local insurers. Billing is a particular challenge because many freestanding EDs have not been able to negotiate favorable rates with insurers or, in some cases, any rate for that matter. In fact, a strategy of many insurers has been to try to shut out freestanding EDs entirely and refuse to pay for care that has led to considerable litigation, as well as business problems that have led to many freestanding ED closures.

Proponents of freestanding EDs argue they offer a better product than many hospital-based EDs and certainly urgent care centers. They stress the vital importance of having access to board-certified emergency physicians. They tout shorter waiting times than hospital-based EDs and a more patient centered care experience. In fact, the bright, spiffy facilities are a more pleasant environment, rather than the chaotic, often hallway-based care of many inner-city EDs that lack natural light. In addition, some data show that freestanding EDs may reduce hospital admissions because they are physically separated from the hospital. And in the case of independent freestanding EDs, they don’t have the same financial incentive to admit.⁶

Nevertheless, the underlying concern by insurers as well as policymakers is that the entry of freestanding...
EDs—or other acute care alternatives for that matter—will raise the cost of providing emergency care. This is of particular concern today in an environment of rising health care costs. Freestanding EDs have gained so much attention that it has risen to the level of MEDPAC, a federal advisory committee, which recommended cutting Medicare payments to freestanding EDs by 30% that are within 6 miles of a hospital. However, MEDPAC also recommended to promote the “rural” version of freestanding EDs, allowing them to not only bill standard fees but also to receive annual fixed payments to keep them afloat.

In this issue of Academic Emergency Medicine, Ho et al. shed further light on the freestanding ED debate. Specifically, Ho and team assessed how the entry of freestanding emergency EDs impacted utilization and overall spending on emergency care in four states from 2013 to 2017: Arizona, Florida, North Carolina, and Texas. In the end, they found mixed effects. In Texas, Florida, and North Carolina, an extra freestanding ED increased emergency provider payments per beneficiary by 3.6% and out-of-pocket costs by $3.60 per person. But contrast, a similar, additional freestanding in Arizona did not impact overall payments. This occurred primarily because prices went down more than 10% and emergency care out-of-pocket costs even declined $15.30 per person. When it came to utilization, an extra freestanding resulted in one additional ED visit per 500 beneficiaries in Texas, Florida, and Arizona, yet there was no change in utilization in North Carolina.

So, what are we to make of these divergent effects? In my view, we can make several inferences. First, we should not be surprised that in health care—a supply-sensitive service—adding service locations would increase costs and utilization. At the margin, the presence of another facility should lead to greater demand for services, at least according to the literature. For readers of the Dartmouth Atlas, we know that more hospital beds are linearly associated with higher hospitalization rates—called Roemer’s Law by health care policymakers—more cardiologists lead to more heart procedures and more spine surgeons in a community leads to more fused spines.

But how about when a new alternative care site directly competes—at least conceptually—with another? The “We will replace ED visits and lower costs . . .” pitch is a frequent marketing tactic by acute care alternatives seeking coverage, yet in acute care, studies have consistently demonstrated this claim to be fictional. Opening retail clinics has essentially no impact on ED visits. Even the latest innovation—direct-to-consumer telemedicine—is almost entirely additive, with only 10% of visits replacing outpatient visits, primarily visits to clinics. In a recent study, the opening of an urgent care center was found to have a significant effect on ED visits at only one of two nearby academic sites. However, the effect was estimated to lower low-acuity ED visits by only 1% or 66 visits in a year to a busy academic department, a relatively meager effect.

What the study by Ho study and others demonstrate is that the effect of adding more competing health care capacity—be it freestanding EDs or other alternatives—depends greatly on existing players in the market (both acute and primary care), current prices and competition, and how the population wants to consume health care. In aggregate, these studies disconfirm the narrative that we can make health care cheaper by just increasing sites of care in an environment where people have the choice to go anywhere that is most convenient. It also does not corroborate uniform claims by insurers that freestanding EDs drive up costs and utilization nor the narrative by freestanding ED operators, telemedicine providers, and others that by offering more, attractive alternatives will draw people away from hospital-based EDs.

Stepping back, we need to consider whether we are even asking the right question when it comes to freestanding EDs. Specifically, is there a way to leverage what is good about freestanding EDs—speedy, comprehensive service by highly trained professionals in a pleasant environment—and separate that from the payment model (fee for service prices similar to hospital-based EDs), which has led to an anaphylactic, unwelcome reception by the insurance industry?

Perhaps the right question is how can freestanding EDs actually improve care and enhance value. The answer lies in models that leverage freestanding EDs in ways that are connected to the broader health system and longitudinal care. The archetypal example is Kaiser Permanente, which leverages freestanding centers for emergency and observation care in the greater Washington, DC, area and in other parts of the country. These centers are different because they are connected to the broader system of care. These centers are not used in isolation as (i.e., the traditional disconnected model) but are used by patients and providers as a thoughtful link in the longitudinal continuum of care that also includes a healthy dose of telemedicine.
Another promising model is the Culinary Care Center in Las Vegas, NV, which is operated by my employer US Acute Care Solutions. The Culinary offers the Culinary Union’s >90,000 employees zero copay emergency-level care and is well staffed by emergency physicians and advanced practice providers. Because the Union members have zero copay, they liberally use the Culinary for emergency care and tend to avoid other sites for emergency care unless absolutely necessary. They also have the ability to access on-site primary and dental care if needed. The model has saved the Culinary Union an estimated >$20 million per year, primarily in avoided costs of outside emergency care and sources of inpatient care.

Ultimately, perhaps we need to rethink freestanding EDs—and other alternatives. The question should not be how adding another disconnected service impacts costs and utilization. The answer from the literature is that it will in most cases add costs and utilization—as the Ho paper found in three of four states—and is the expected outcome. We need to shift the discourse toward connectivity and value-based care and figure out how patient-centered models like freestanding EDs can be integrated into the longitudinal care system, fill unmet public health needs (i.e., in rural settings), or both.16–17

Jesse M. Pines, MD, MBA, MSCE  
(pinesj@usacs.com)  
US Acute Care Solutions, Canton, OH

Supervising Editor: John H. Burton, MD

REFERENCES


Conflict of Interest Blind Spots in Emergency Medicine: Ideological and Financial Conflicts of Interest in the Gender Bias Literature

I read with interest the March issue devoted to gender issues in Emergency Medicine and was struck by what I see as a disturbing trend in the emerging literature on gender issues in Emergency Medicine: the lack of disclosure of ideological and financial conflicts of interest. Multiple authors in this issue have founded or currently work as advisors for an advocacy organization addressing gender issues in Emergency Medicine. While it is true that ideology is an inescapable entity in all human activities, including scientific research, researchers who are founders, leaders, or senior advisors to advocacy organizations ought to disclose those relationships as part of standard research ethics.

The topic of non-financial conflicts of interest has deep roots in the scientific community, taking form in linguistic,1 values-based,2 and even historical dialectical critiques.3 Ideology has been suggested to play as much a part as financial conflicts of interest in multiple facets of the scientific world, and perhaps more so.4 If it is true that even a pen provided for a drug manufacturer can affect behavior and attitudes towards a product,5 how much more so would the unconscious temptation to selectively collect data or selective interpret that data to serve one’s own deeply held beliefs be present, especially in the polarizing issue of gender and the origins of the gender pay and advancement gaps. Indeed, ideological conflicts, though largely understudied, likely extend to the full spectrum of what plagues financial ones, to include publication bias,6 framing research questions or comparisons to favor the outcome the authors would like to see, deplatforming or discrediting dissenters,7,8 and drawing overly strong conclusions to one’s data,9 some of which can even be seen in the research articles in this issue.

A good example is the Bennett et al.10 study in the March issue. This study showed that when accounting for a list of available factors that might account for gender differences in academic promotion, when comparing the individual academic ranks of professor and associate professor, no difference was seen, or as the article put it, the results “approached significance.” When they took the composite endpoint of both associate and full professor, with no indication of choosing that endpoint before data analysis and without noting the problems with composite endpoints,11 they were able to find a difference. And yet the conclusion trumpeted “there remains an obvious unmet need for the field of emergency medicine to encourage and promote equal advancement across gender.” The tone of conclusion which is so disparate from the actual data calls to mind the landmark study demonstrating that conflicts of interest affected the interpretation of the data on the efficacy of oseltamivir for influenza.9 Noting the comparison of individual academic ranks as approaching significance is also curious in light of another finding that also approached significance yet wasn’t commented on – differences in Medicare reimbursements between males and females.

It must also be noted that many of the nondisclosed conflicts of interest in this issue were not just...
ideological. Multiple authors in this issue had undisclosed financial conflicts of interest relating to gender issues as well. Dr. Kass leads an advocacy group organized as a limited liability corporation which receives funding from large medical associations and corporate groups, while Dr. Choo leads a consultancy group which measures and addresses issues of gender equity. In other words, authors in this issue published on topics which their other financial interests depend upon without disclosing these conflicts of interest.

The assumption that many businesses of this sort make is that gender inequities in medicine are largely based on historical and systematic discriminative factors. This assumption has significant implications for the medical literature. If the gender wage and advancement gap is instead explainable by one of the myriad social, personal, and productivity factors that have been suggested and demonstrated, and not based on bias or institutional aspects, then there is a danger of what may be called a “runaway phenomenon.” This runaway phenomenon is characterized by published research, originating from and interpreted heuristically based on inaccurate underlying assumptions, which posits or discovers an “indisputable” problem. The identification of the problem leads to demands for more research money, which begets research attempting to remedy the problem, which then reinforces the existence of the problem that was never conclusively established as a problem in first place. The addition of financial conflicts of interest only further strengthens the hold of the original underlying assumptions and as described earlier, affects the conception, execution, and interpretation of that research that is conducted.

In light of this, instead of hurrying to “move from descriptions of gender gaps to innovative interventions,” we really ought to be seeking to clarify this very difficult issue in determining to what extent the gap, which surely exists, is due to factors that are problematic. Indeed, if the gap is due to the factors inexhaustibly listed in Table 1, which have either inadequate data at this point or none at all, then the tone of the discussion about differences in gender equity becomes a very different one, and would steer further research to examine the relatively barren research areas of physician gender differences in values and behaviors.

Up to this point, the many potential differences between men and women that might account for salary or promotion differences have been examined piecemeal because of limitations of the available conglomerated data. For example, the study in this issue by Baird et al looking at salary differences did not address overtime or other aspects of work choice that might reflect a more income-focused physician, which, in other fields has largely been able to account for the full salary differential. Another study in JAMA on academic salaries, in addition to demonstrating the absence of an adjusted wage gap for emergency medicine, did not address part-time status, maternity leave, amount of NIH funding, or many other variables. More work needs to be done to get more granular detail into wage and academic promotion gender differences in both the academic and nonacademic environments to assess whether institutional factors exist which unfairly create these differences. It should be noted that in the general population, the more variables that are accounted for between men and women, the smaller the gap seems to become, and this may also be true in Emergency Medicine.

Time is not up on our attempting to understand the issue of gender disparities, and we need to understand that ideology will likely have effects on the research that is conducted and published - or not published - going forward. Ideology affects us all: we are all subject to cognitive biases which drive us to find or interpret data that aligns with our beliefs and values that are as deep as they are purely subjective. For this reason, we need to begin a thoughtful conversation on how to deal with ideological conflicts of interest, particularly with what should require reporting in the

<table>
<thead>
<tr>
<th>Table 1 Examples of Areas of Incomplete Research in Gender Pay and Academic Rank Inequities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifestyle factors in choosing Emergency Medicine</td>
</tr>
<tr>
<td>Incidence of extended leave</td>
</tr>
<tr>
<td>Amount of NIH funding ($)</td>
</tr>
<tr>
<td>Cited publications</td>
</tr>
<tr>
<td>Night and holiday shift behavior</td>
</tr>
<tr>
<td>Part time work13</td>
</tr>
<tr>
<td>Traveling locums tenens (aka “special ops”)</td>
</tr>
<tr>
<td>Working where domiciled</td>
</tr>
<tr>
<td>Characteristics of work location (rural, suburban, urban)</td>
</tr>
<tr>
<td>Shift productivity10,14</td>
</tr>
<tr>
<td>Compensated administrative responsibilities including leadership positions</td>
</tr>
<tr>
<td>Entrepreneurial behavior (free standing ED, urgent care ownership, etc.)</td>
</tr>
<tr>
<td>Validity of inter-specialty comparisons</td>
</tr>
</tbody>
</table>
medical literature. My own view is that membership in an advocacy organization, political party, or other groups that we align with in our personal lives as part of what Hannah Arendt called the participatory democracy need not be disclosed. However, for the sake of the integrity of the literature, and to prevent the runaway phenomenon, authors who serve in leadership roles in advocacy groups or who have consulting groups whose existence is dependent on the ontologic reality of the published topic ought to disclose these ties. Advocacy and research need not be mutually exclusive activities, but for the readership of the medical literature to best cast its critical eye, they deserve to know if these relationships exist.

Chase M. Donaldson, MD
(Chase.m.donaldson.mil@mail.mil)
131st Field Hospital, Baghdad, Iraq

References

We acknowledge that Dr. Dara Kass is the founder and CEO of FemInEM, LLC. The failure to disclose this was an oversight. This research study was not supported in any way financially by FemInEM; we as authors take joint responsibility for this oversight of nondisclosure. We also fully stand by the findings and conclusions of our study.

While we respect the commentary by Dr. Donaldson, we cannot agree with the validity of gender equity research being undermined writ large. The letter states “how much more so would the unconscious temptation to selectively collect data or selective [sic] interpret that data to serve one’s own deeply held beliefs be present, especially in the polarizing issue of gender and the origins of the gender pay and advancement gaps.”

If we were publishing on the effects of exercise on cardiovascular disease, and one of our authors failed to disclose ownership in a fitness facility, would a comment refer to exercise as a polarizing issue or cardiac disease as a deeply held belief? The premise that gender is a polarizing issue, rather than an objective matter, minimizes the scores of validated, peer-reviewed works on the subject and attempts to revert to the belief that the gender wage gap and other gender disparities are opinions rather than well-established facts that the scientific community knows them to be.

The letter further states “Businesses like those of these authors make the assumption that the reason for gender inequities in medicine are based on historical and systematic discriminative factors. If the gender wage and advancement gap is instead explainable by one of the myriad social, personal, and productivity factors that have been suggested and demonstrated, and not based on bias or institutional aspects, then the businesses are selling a product no one needs.” This, again, highlights a bias and has little to do with the objective data on gender inequities in medicine. Specifically, the statement that organizations dedicated to addressing gender gaps in medicine—whether by bringing women in medicine together as a community or consulting on methods to decrease bias—were created to convince people of the existence of a nonexistent problem further exemplifies the issue that the commentary seems to have with gender inequities in medicine.

Our fundamental concern with this commentary is the use of a real issue (the oversight of nondisclosure) as an opportunity to postulate that the effects of gender on the careers of women in medicine are orchestrated phenomena for the enrichment of those who are attempting to correct the “problem.” This is, quite simply, wrong.

Christopher L. Bennett, MD, MA
(Christopher.lee.bennett@gmail.com)
Ali S. Raja, MD, MPH
Dara Kass, MD
Nate Gross, MD
Angela M. Mills, MD

1 Department of Emergency Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA
2 Doximity, San Francisco, CA
3 Department of Emergency Medicine, Columbia University, Vagelos College of Physicians and Surgeons, New York, NY

Supervising Editor: John H. Burton, MD

References
“W hat do you want to do?” Matt asked.

I looked up from the blue, gasping baby to the physician asking the question. “Well, back home I would have intubated her already.”

“Then you should do that here,” he responded.

Intubating the newborn was easy, and as the adrenaline rush passed, I looked up and asked the obvious question, “Now what?”

“Now you bag her until you think she can be extubated . . . or you think it’s futile.” He headed for the door.

“I’ll be back. Emergency called with a critical patient.”

For the next half hour, I stood at the infant warmer, sweating in the tropical heat trapped so effectively by the concrete block, bagging the baby. I prayed, asking God for the child’s life. It shouldn’t be like this. There should be a NICU, a ventilator, a neonatologist. But this child had only me: a jetlagged visitor on his first day in Papua New Guinea (PNG), unsure of what to do next as the flies buzzed in and out of the open window behind me.

She had endured 12 hours of labor that stalled. When we delivered her by C-section a thick slime of meconium stained the amniotic fluid. Suctioning worked well enough and the nurse whisked her off to the nursery.

The same nurse returned shortly after. “The baby is not breathing well,” she said, understating the issue. Walking into the nursery, I spotted our recent delivery: ashen, blue-gray color with agonal respirations. I intubated her, the nurse moved on to attend to other newborns, and Matt went to emergency. I was left alone with the baby, the heat, and my prayers as I bagged.

Five years ago, I was asking Matt what he wanted to do. He was my intern then. “Let’s tap it,” he said without hesitation.

“You know how?” I asked. He replied confidently, prepped the patient and performed a good arthrocentesis. As we walked away, I asked Matt how many he had done before.

“None. I watched it on YouTube,” he replied.

“You’ve never done that procedure?”

“Nope.”

“But I asked if you knew how to do it.”

“I do. I just never had.”

The boldness he displayed in residency has made him an excellent physician in this resource-constrained environment.

The baby was pink now and breathing regularly on her own. Relieved, I extubated her shortly before Matt returned. Over the next week, her respirations improved steadily. When we saw her later that month in outpatient clinic she had become a happy, healthy baby. In PNG, children are not named until three months of age, given the prevalent infant mortality, but in a gesture of faith, her mother had already named her.

Returning to practice in the United States, I see that I provide much care. I order tests, call consultants, perform procedures, but none of that work requires me to care. Most of it keeps me from the patient’s bedside. In PNG, without those distractions, I had to provide care in an environment that exposed my limitations. The
acute knowledge of my limitations fostered greater empathy for the suffering patient and allowed me to care boldly for them. In the United States, I provide technologically superior care. In PNG, I was able to provide emotionally superior care precisely because the technology was absent. I couldn’t put a patient on a ventilator and walk away. I couldn’t create emotional distance between myself and the patient. I was tied to the bedside by the lack of technology. Within these limits, I found the freedom to be a physician.

Jonathan T. Miller, MD
(millerjt@upmc.edu)
UPMC Hamot Emergency Medicine Residency
UPMC Hamot Hospital
Erie, PA

Supervising Editor: Brian Zink, MD