EDITORS’ PICK

Mortality and Complication Rates in Adult Trauma Patients Receiving Tranexamic Acid: A Single-center Experience in the Post–CRASH-2 Era
Pablo Joaquin Erramouspe, Maria Florencia García-Pintos, Simranjeet Benipal et al. 358

ORIGINAL CONTRIBUTIONS

Outcomes With the Use of Bag–Valve–Mask Ventilation During Out-of-hospital Cardiac Arrest in the Pragmatic Airway Resuscitation Trial
Joshua R. Lupton, Robert H. Schmicker, Shannon Stephens et al. 366

Early Self-Proning in Awake, Non-intubated Patients in the Emergency Department: A Single ED's Experience During the COVID-19 Pandemic
Nicholas D. Caputo, Reuben J. Strayer, Richard Levitan 375

Containing COVID-19 in the Emergency Department: The Role of Improved Case Detection and Segregation of Suspect Cases
Liang E. Wee, Tzay-Ping Fua, Ying Y. Chua et al. 379

The Impact of Selecting Specific Cohorts for Benchmarking and Interpretation of Emergency Department Patient Satisfaction Scores
Gregory M. Archual, Ashish R. Panchal, Mark G. Angelos et al. 388

Video Versus Direct and Augmented Direct Laryngoscopy in Pediatric Tracheal Intubations
Amy H. Kaji, Carolyn Shover, Jennifer Lee et al. 394

Contents continued inside.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESEARCH METHODS AND STATISTICS</td>
<td>Patient-reported Outcome Measures in Emergency Care Research: A Primer for Researchers, Peer Reviewers, and Readers</td>
<td>Howard S. Kim, D. Mark Courtney, Danielle M. McCarthy et al.</td>
<td>403</td>
</tr>
<tr>
<td>RESEARCH LETTER</td>
<td>Inter-rater Agreement Between Self-rated and Staff-rated Clinical Frailty Scale Scores in Older Emergency Department Patients: A Prospective Observational Study</td>
<td>Thom Ringer, Cameron Thompson, Shelley McLeod et al.</td>
<td>419</td>
</tr>
<tr>
<td>HOT OFF THE PRESS</td>
<td>Hot Off the Press: The Effect of Financial Incentives on Patient Decisions to Undergo Low-value Head CT Scans</td>
<td>Justin Morgenstern, Corey Heitz, Chris Bond et al.</td>
<td>423</td>
</tr>
<tr>
<td>THE BRASS TACKS: CONCISE REVIEWS OF PUBLISHED EVIDENCE</td>
<td>Adjuvant Antibiotic Therapy After Incision and Drainage of Cutaneous Abscesses</td>
<td>John Conway, Benjamin Friedman</td>
<td>427</td>
</tr>
<tr>
<td></td>
<td>Accuracy of Point-of-care Ultrasound for Diagnosing Soft Tissue Abscess</td>
<td>Brit Long, Alex Koyfman, Michael Gottlieb</td>
<td>429</td>
</tr>
<tr>
<td>COMMENTARY-INVITED</td>
<td>Poison, Pixie Dust, and Prehospital Airway Management</td>
<td>Jason T. McMullan, Darren A. Braude</td>
<td>431</td>
</tr>
<tr>
<td>COMMENTARY-UN SOLICITED</td>
<td>Doctoring While Woman</td>
<td>Rebekah Mannix, Lois K. Lee</td>
<td>434</td>
</tr>
<tr>
<td></td>
<td>Routine Opt-out Syphilis Screening in the Emergency Department: A Public Health Imperative</td>
<td>Kimberly A. Stanford, Aniruddha Hazra, John Schneider</td>
<td>437</td>
</tr>
<tr>
<td>REFLECTIONS</td>
<td>Quarantined</td>
<td>Moon O. Lee</td>
<td>439</td>
</tr>
<tr>
<td></td>
<td>Pardon Me for Being a Wallflower</td>
<td>Janice Blanchard</td>
<td>440</td>
</tr>
</tbody>
</table>
Copyright and Copying (in any format)
Copyright © 2020 Society for Academic Emergency Medicine. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA, www.copyright.com, provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising or promotional purposes, for republication, for creating new collective works or for resale. Permissions for such reuse can be obtained using the RightsLink “Request Permissions” link on Wiley Online Library. Special requests should be addressed to permissions@wiley.com

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

For submission instructions, subscription and all other information visit http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1553-2712
CME Information: Mortality and Complication Rates in Adult Trauma Patients Receiving Tranexamic Acid: A Single-center Experience in the Post-CRASH-2 Era

CME Editor: Corey Heitz, MD

Authors: Pablo Joaquin Erramouspe, MD, María Florencia García-Pintos, MD, Simranjeet Benipal, Martin A. C. Manoukian, MD, John-Lloyd Santamarina, MD, Hiwote G. Shawagga, Linda L. Vo, Joseph M. Galante, MD, and Daniel Nishijima, MD, MAS

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 one hospital’s results using TXA for traumatic hemorrhage.

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. is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

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.
Mortality and Complication Rates in Adult Trauma Patients Receiving Tranexamic Acid: A Single-center Experience in the Post–CRASH-2 Era

Pablo Joaquin Erramouspe, MD\textsuperscript{1,2}, María Florencia García-Pintos, MD\textsuperscript{1}, Simranjeet Benipal\textsuperscript{1}, Martin A. C. Manoukian, MD\textsuperscript{1}, John-Lloyd Santamarina, MD\textsuperscript{1}, Hiwote G. Shawagga\textsuperscript{1}, Linda L. Vo\textsuperscript{1}, Joseph M. Galante, MD\textsuperscript{3}, and Daniel Nishijima, MD, MAS\textsuperscript{1}

ABSTRACT

Objectives: The CRASH-2 trial demonstrated that tranexamic acid (TXA) in adults with significant traumatic hemorrhage safely reduces mortality. Given that the CRASH-2 trial did not include U.S. sites, our objective was to evaluate patient characteristics, TXA dosing strategies, and the incidence of mortality and adverse events in adult trauma patients receiving TXA at a U.S. Level I trauma center in the post–CRASH-2 era.

Methods: We conducted a retrospective study that included patients aged 18 years or older who received TXA after an acute injury from July 2014 to June 2017. We excluded patients who received TXA orally, patients who received TXA for elective surgical procedures or nontrauma indications, patients who received it 8 hours or longer after the time of injury, and patients with cardiac arrest at time of emergency department arrival. Trained abstractors collected data from the trauma registry and hospital electronic medical records. Our primary outcome measures were in-hospital death and acute thromboembolic events within 28 days from injury.

Results: We included 273 patients with a mean (±SD) age of 43.8 (±18.7) years. The mean (±SD) time of administration of TXA from time of injury was 1.55 (±1.2) hours with 229 patients (83.9%) receiving TXA within 3 hours. The overall mortality within 28 days from injury was 12.8% (95% confidence interval [CI] = 8.9% to 16.7%), which was similar compared to that in the CRASH-2 trial (14.5%, 95% CI = 13.9% to 15.2%). The incidence of acute thromboembolic events was 6.6% (95% CI = 3.7% to 9.5%), which was higher than that in the CRASH-2 trial (2.0%, 95% CI = 1.73% to 2.27%). Patients in our cohort also received surgery (64.8% vs. 47.9%) and blood transfusions (74.0% vs. 50.4%) more frequently than those in the CRASH-2 cohort.

Conclusions: Adult trauma patients receiving TXA had similar incidences of death but higher incidences of thromboembolic events compared to the CRASH-2 trial. Variation in patient characteristics, injury severity, TXA dosing, and surgery and transfusion rates could explain these observed differences. Further research is necessary to provide additional insight into the incidence and risk factors of thromboembolic events in TXA use.

From the \textsuperscript{1}Department of Emergency Medicine; and the \textsuperscript{2}Faculty of Health, Queensland University of Technology, Translational Research Institute, Brisbane, QLD, Australia; and the \textsuperscript{3}Department of Surgery, UC Davis School of Medicine, Sacramento, CA.

Received July 17, 2019; revision received October 1, 2019; accepted October 3, 2019.


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

Author contributions: PJE and MFGP—study design, acquisition of data, analysis, interpretation of data, and drafting of manuscript; SB, MACM, JLS, HGS, LLV, and JMG—acquisition of data and critical revision of manuscript; DN—senior researcher and group leader, study conception, and critical revision of manuscript.

Supervising Editor: Michael S. Runyon, MD.

Address for correspondence and reprints: Pablo Joaquin Erramouspe, MD; e-mail: joaquinerramouspe@hotmail.com.


doi: 10.1111/acem.13883
In the United States, traumatic injury is the leading cause of death among individuals aged 1 to 44 years old. Hemorrhage is the primary cause of death in the first 24 hours after trauma and accounts for 30% to 40% of all trauma-related deaths. Traumatic hemorrhage is often exacerbated by trauma-induced coagulopathy, which is defined as a multifactorial pathology consisting of excessive fibrinolysis, coagulation factor consumption, and clotting dysfunction stimulated by the traumatic event. Severely injured trauma patients frequently demonstrate abnormal coagulation profiles within a few hours after the injury. Patients who develop acute coagulopathy are much more likely to die and die early.

Tranexamic acid (TXA) is a synthetic analog of the amino acid lysine used to attenuate hemorrhage by blocking plasmin-mediated fibrin clot breakdown. Developed in 1962 to treat postpartum hemorrhage, TXA was first approved by the U.S. Food and Drug Administration (FDA) in 1986 for short-term use to reduce or prevent bleeding during tooth extraction in hemophilic patients and in patients with severe menorrhagia. In 2010, the CRASH-2 trial demonstrated that the risk of hemorrhagic death in trauma patients is significantly reduced if TXA is administered 1 hour or less from the time of injury (5.3% in TXA group vs. 7.7% in placebo group) and between 1 and 3 hours (4.8% in TXA group vs. 6.1% in placebo group).

Currently, TXA is estimated to save 112,000 lives per year worldwide and is considered standard treatment in adults with traumatic bleeding. Given that the CRASH-2 trial was primarily conducted in developing countries where transfusion practices and identification of adverse events may differ compared to the United States, the true incidence of mortality and thrombotic events at U.S. trauma centers in the post-CRASH-2 era is largely unknown. Our objective was to evaluate the patient characteristics, TXA dosing strategies, and the incidence of mortality and adverse events in adult trauma patients receiving TXA in the post-CRASH-2 era.

METHODS

Study Design
We conducted a retrospective, observational, single-center study at a U.S. Level I trauma center. The study was approved by the local institutional review board. The design of this study has followed the STROBE reporting guidelines and is a historical cohort study with a Level of Evidence IV.

Study Setting and Population
The University of California Davis Medical Center is one of only three trauma centers in California with Level I certification in both pediatric and adult trauma care. In addition to providing primary injury management, the Trauma Service of the Department of Surgery cares for the vast majority of trauma victims admitted to the medical center and coordinates the care of surgical subspecialists. The trauma team is activated prior to or upon arrival of injured patients to the emergency department (ED). TXA is ordered by the trauma service and administered during the initial resuscitation for all patients who present within 3 hours of injury and meet at least one of the inclusion criteria: 1) systolic blood pressure (SBP) < 90 mm Hg, 2) initiation of the massive transfusion guideline, or 3) transport directly from the ED resuscitation to the operating room. In some cases where the exact time of injury in the resuscitation bay is unknown and the patient meets at least one of the inclusion criteria, TXA is administered. At the study center, trauma patients receive diagnostic testing for thromboembolic disease as clinically indicated but are not routinely screened without symptoms or signs (e.g., routine lower-extremity ultrasounds). However, all admitted trauma patients receive thromboembolic prophylaxis. Trained trauma registrars maintain the trauma registry with data variables, and abstractions are completed in accordance with the American College of Surgery National Trauma Data Standard.

We included patients who were 18 years and older and received TXA for an acute injury from July 1, 2014, to June 30, 2017. TXA in all the cases was ordered according to the international guidelines, bolus dose being 1 g of TXA administered over 10 minutes intravenously and maintenance infusion being 1 g administered over 8 hours intravenously. We excluded patients who received TXA orally, patients who received TXA for elective surgical procedures or nontrauma indications, patients who received TXA 8 hours or longer after the time of injury, and patients with cardiac arrest at the time of ED arrival. We first searched hospital electronic medical records (EMRs) for patients with orders of intravenous TXA that occurred during the study period and had a trauma team activation initiated. We then cross-matched these patients with the trauma registry using...
medical record numbers. Individual patient charts were then reviewed to further assess eligibility based on our study inclusion and exclusion criteria.

Study Protocol
Trauma registry variables included injury information, demographics, prehospital data, ED clinical variables, laboratory and radiology variables, diagnoses and procedures, injury severity, discharge notes, and death information. We reviewed the EMR charts of individual patients to abstract additional variables that were not included in the trauma registry. We selected these variables based on the data that were collected for the CRASH-2 trial. These abstracted variables included dosing of TXA (bolus and maintenance infusions), timing of TXA administration from injury time (minutes), cause of death (bleeding, head injury, myocardial infarction [MI], stroke, pulmonary embolism [PE], multiorgan failure, or other), type of surgical intervention (neurosurgical, chest, abdominal, pelvis), and complications (PE, deep vein thrombosis [DVT], stroke, operation for bleeding, MI, and gastrointestinal bleeding). Data abstraction procedures followed prior recommendations for retrospective studies.18 We developed a standardized data abstraction worksheet that was pilot tested on several patients. All data abstractors were trained on study procedures. To evaluate inter-rater agreement of data abstraction, a second abstractor independently reviewed 10% of randomly selected charts.

Measures
Our primary outcome measures were selected and defined based on the CRASH-2 trial. Our primary outcome measures were in-hospital death and acute thromboembolic events (defined as MI, stroke, PE, and DVT) within 28 days from the injury.

Data Analysis
We formatted the data and recoded the variables using STATA 13.1 statistical software. Descriptive statistics were used to characterize the study population overall. Nonnormal interval data were reported with medians and interquartile ranges. Patient characteristics were reported alongside the patient characteristics from the CRASH-2 trial. However, we did not conduct any comparisons between our study and the CRASH-2 trial as we did not have patient-level data from the CRASH-2 trial.

To assess inter-rater agreement, we calculated percent agreement and the kappa statistic (with 95% confidence intervals [CIs]) for binary variables and Pearson’s correlation coefficient for continuous variables.19,20 We considered variables with a kappa statistic of 0.60 or greater to have acceptable inter-rater reliability.21 Based on our database and hospital records of trauma patients admissions we calculated that 36 months of data acquisition was adequate to provide a sample size sufficient to ensure a narrow CI for the point estimates of mortality and thrombosis seen in the CRASH-2 trial.

RESULTS
Characteristics of the Study Subjects
Of the 321 EMR and trauma registry records with an order for TXA administration, we excluded a total of 48 records for a final study population of 273 unique patients treated in our center (Figure 1). Our study cohort included 202 males (74%) and the mean age of the cohort was 43.8 years, with a higher percentage of patients over 44 years of age than the CRASH-2 trial (44.7% vs. 23.0%). On arrival, our patients showed similar clinical vital signs (e.g., blood pressure, heart and respiratory rate) in comparison to the patients in the CRASH-2 and a similar proportion of penetrating injuries (36.3% vs. 32.5%). The mean (±SD) time of administration of TXA from time of injury in our study was 1.55 (±1.2) hours with 229 patients (83.9%) receiving TXA within 3 hours from the time of injury versus a mean time of administration of 2.8 hours in the CRASH-2 trial and only 67.4% of the patients receiving TXA within 3 hours. While 248 (90.8%) patients received the complete bolus dose, 183 (67%) received the complete maintenance dose, and 167 (61.1%) patients received both doses completely. Of the 106 patients who had incomplete TXA dosing, the total dose of TXA received and the reason for incomplete TXA dosing were not stated in EMR. Eight of these patients with incomplete TXA dosing died on the day of admission and an additional 29 patients did not undergo surgical intervention. Complete patient characteristics are reported in Table 1. Inter-rater agreement was acceptable for all variables evaluated (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.13883/full).
**Main Results**

The overall mortality within 28 days from injury in our study was 12.8% (95% CI = 8.9% to 16.7%). This incidence of mortality was similar to that in the CRASH-2 trial, which reported a mortality of 14.5% (95% CI = 13.9% to 15.2%). However, the incidence of acute thromboembolic events in our study was 6.6% (95% CI = 3.7% to 9.5%), which was higher than that to the CRASH-2 trial, which reported 2.0% (95% CI = 1.73% to 2.27%).

Sixteen patients had 18 thromboembolic events (Table 2). Four patients with a thromboembolic event died, including three patients with an acute ischemic stroke and one patient with a pulmonary embolism. Of these 16 patients, 14 received complete bolus dose of TXA, nine received complete maintenance dose and seven patients received the complete bolus and maintenance infusions. Thirteen patients received TXA within 3 hours from the time of injury, and three patients did not have the exact time of injury recorded. Regarding the severity of the injuries, only one patient had an ISS score less than 16 while the remaining patients suffered major injuries, and 11 patients had to undergo surgery for their injuries.

Patients in our study cohort also received surgery more frequently than in the CRASH-2 cohort (64.8% [95% CI = 58.9% to 70.5%] vs. 47.9% [95% CI = 46.9% to 48.8%]). More patients in our study cohort also received at least one blood product transfusion (74.0% [95% CI = 68.4% to 79.1%]) compared to the CRASH-2 cohort (50.4% [95% CI = 49.4% to 51.3%]). Other complications and descriptions of patient management are described in Table 3.

**DISCUSSION**

Our study provides a single-center perspective on the characteristics and outcomes of adult trauma patients receiving TXA in the post–CRASH-2 era. Furthermore, this study describes the dosage strategies and screening practices used in our center. The landmark CRASH-2 trial did not include sites from the United States; thus our study provides a real-world perspective into the use of TXA at a U.S. Level I trauma center.

Our study demonstrated some interesting observations. Although we were unable to statistically compare patient characteristics between our study and the CRASH-2 trial, our study cohort appears to be more female and older compared to the CRASH-2 cohort. Our findings are consistent with other studies also demonstrating an older trauma population in the United States compared to other countries.22–26 Administration of TXA also appeared to be sooner after the time of injury in our study cohort (mean ± SD time of TXA administration 1.55 ± 1.2 hours) compared to the CRASH-2 trial (mean ± SD time of TXA administration 2.8 ± 2.2 hours). In addition, a higher proportion of patients received TXA within 3 hours from injury in our study cohort compared to the CRASH-2 trial (83.9% vs. 67.4%). This observation is
likely due to a secondary analysis demonstrating that, in comparison to placebo, there was decreased death due to bleeding if TXA was given either within 1 hour of the time injury (RR = 0.87, 95% CI = 0.76 to 0.97) or 1 to 3 hours from the time of injury (RR = 0.87, 95% CI = 0.77 to 0.97). There was no statistical difference when TXA was administered more than 3 hours from the time of injury (RR = 1.00, 95% CI = 0.90 to 1.13). Additionally, it is also interesting to note that only 61% of patients in our study cohort received both bolus and maintenance TXA doses, while 94% received both TXA doses in the CRASH-2 trial. Theoretically, this would bias toward less thrombotic risk for our study cohort (less TXA exposure). However, our results demonstrated higher thrombosis in our study cohort compared to the CRASH-2 trial. The low rate of completion of both bolus and maintenance in our study cohort compared to the CRASH-2 trial cohort is important as it potentially reflects current U.S. practice. There are some thoughts that bolus dosing of TXA may be sufficient

### Table 1

**Characteristics of Patients Included in the Current Study and Patients Included in the CRASH-2 Trial**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Current Study, n = 273 (%)</th>
<th>CRASH-2 Trial, n = 10,093 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>202 (74.0%)</td>
<td>8,439 (83.6%)</td>
</tr>
<tr>
<td>Female</td>
<td>71 (26.0%)</td>
<td>1,654 (16.4%)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>43.8 (±18.7)</td>
<td>34.6 (±14.1)</td>
</tr>
<tr>
<td>&lt;25</td>
<td>46 (16.8%)</td>
<td>2,783 (27.6%)</td>
</tr>
<tr>
<td>25-34</td>
<td>62 (22.7%)</td>
<td>3,012 (29.8%)</td>
</tr>
<tr>
<td>35-44</td>
<td>43 (15.7%)</td>
<td>1,975 (19.6%)</td>
</tr>
<tr>
<td>&gt;44</td>
<td>122 (44.7%)</td>
<td>2,321 (23.0%)</td>
</tr>
<tr>
<td>Not known</td>
<td>0 (0)</td>
<td>2 (0.02%)</td>
</tr>
<tr>
<td><strong>Time since injury (hours)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>1.55 (±1.2)</td>
<td>2.8 (±2.2)</td>
</tr>
<tr>
<td>&lt;1</td>
<td>113 (41.4%)</td>
<td>3,756 (37.2%)</td>
</tr>
<tr>
<td>1 to &lt;3</td>
<td>116 (42.5%)</td>
<td>3,045 (30.2%)</td>
</tr>
<tr>
<td>&gt;3</td>
<td>28 (10.3%)</td>
<td>3,287 (32.6%)</td>
</tr>
<tr>
<td>Precise time not known but administered within 8 hours</td>
<td>16 (5.9%)</td>
<td>5 (0.05%)</td>
</tr>
<tr>
<td><strong>Complete TXA administration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bolus dose</td>
<td>248 (90.8%)</td>
<td>NR</td>
</tr>
<tr>
<td>Maintenance dose</td>
<td>183 (67.0%)</td>
<td></td>
</tr>
<tr>
<td>Bolus and maintenance dose</td>
<td>167 (61.1%)</td>
<td></td>
</tr>
<tr>
<td><strong>Type of injury</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penetrating</td>
<td>99 (36.3%)</td>
<td>3,281 (32.5%)</td>
</tr>
<tr>
<td>SBP (mm Hg)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤75</td>
<td>46 (16.8%)</td>
<td>1,566 (15.5%)</td>
</tr>
<tr>
<td>76-89</td>
<td>43 (15.7%)</td>
<td>1,615 (16.0%)</td>
</tr>
<tr>
<td>&gt;89</td>
<td>183 (67.0%)</td>
<td>6,901 (68.4%)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (0.37%)</td>
<td>11 (0.11%)</td>
</tr>
<tr>
<td><strong>Respiratory rate (/min)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>2 (0.73%)</td>
<td>160 (1.6%)</td>
</tr>
<tr>
<td>10-29</td>
<td>241 (88.3%)</td>
<td>8,355 (82.8%)</td>
</tr>
<tr>
<td>&gt;29</td>
<td>29 (10.6%)</td>
<td>1,491 (14.8%)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (0.37%)</td>
<td>87 (0.86%)</td>
</tr>
<tr>
<td><strong>Heart rate (beats/min)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;77</td>
<td>38 (13.9%)</td>
<td>875 (8.7%)</td>
</tr>
<tr>
<td>77-91</td>
<td>49 (17.9%)</td>
<td>1,727 (17.1%)</td>
</tr>
<tr>
<td>92-107</td>
<td>69 (25.3%)</td>
<td>2,556 (25.3%)</td>
</tr>
<tr>
<td>&gt;107</td>
<td>116 (42.5%)</td>
<td>4,872 (48.3%)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (0.37%)</td>
<td>63 (0.62%)</td>
</tr>
<tr>
<td><strong>Glasgow Coma Scale score (total)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-8</td>
<td>63 (23.1%)</td>
<td>1,799 (17.8%)</td>
</tr>
<tr>
<td>9-12</td>
<td>28 (10.2%)</td>
<td>1,353 (13.4%)</td>
</tr>
<tr>
<td>13-15</td>
<td>182 (66.7%)</td>
<td>6,934 (68.7%)</td>
</tr>
<tr>
<td>Not known</td>
<td>0 (0.0%)</td>
<td>7 (0.07%)</td>
</tr>
</tbody>
</table>

*Includes myocardial infarction, stroke and pulmonary embolism.*
to attenuate fibrinolysis and there are ongoing TXA trials evaluating bolus dosing alone versus bolus dosing with maintenance dosing.27 Our rates of maintenance dosing may reflect these considerations.

As previously mentioned, when compared to the CRASH-2 trial, our study cohort demonstrated a similar incidence of 28-day in-hospital mortality, although the incidence of acute thromboembolic events was higher. However, we should not conclude from our findings that the benefit-to-harm ratio of TXA use in the United States is different from that demonstrated in the CRASH-2 trial, as our study cohort is an observational study. Instead, our results simply suggest that the number of acute thromboembolic events observed in settings outside of the CRASH-2 trial may differ from what was observed in that trial.

The increased incidence of thromboembolic events observed in our study cohort compared to the CRASH-2 trial has multiple potential explanations. Our study cohort could have a higher baseline risk for thromboembolic disease due to either patient characteristics (e.g., age, comorbidities) or injury severity. Our study cohort had a higher proportion of older patients, a higher number of surgical interventions, and more blood transfusions in comparison to the CRASH-2 cohort. Older trauma patients are at a significantly higher risk for thromboembolic events compared to younger patients.28,29 Furthermore, the peak of DVT incidence has been reported at ages between 45 and 59 years, an age group that was much more frequent in our study cohort compared to the CRASH-2 cohort.30 In recent studies, it was demonstrated that in hospitalized older patients with traumatic brain injury (TBI) and in patients with greater TBI severity, incidence of acute thromboembolic events was significantly higher.31,32 It is possible that patients with TBI and intracranial hemorrhage may have prophylaxis withheld, thereby exposing these patients to an increased risk for thromboembolic disease. However, it is noted in our study that 120 patients sustained a head injury and only seven (5.8%) had a thromboembolic event which is a similar rate to the overall thromboembolic event rate of 6.6%. Additionally, surgery is a well-known risk factor for the development of thromboembolic events by producing changes in coagulation and fibrinolytic systems with a surge in circulating cytokines, and depending on the type of surgery (laparotomy vs. laparoscopic) and surgical site the risk could be even higher.33,34

Prior studies have also shown that blood product transfusion is associated with an increased risk of thrombotic events in a dose-dependent relationship between number of units transfused and thrombosis.35 It is also possible that sites included in the CRASH-2 trial screened less for thromboembolic events than compared to our study site where trauma patients receive diagnostic testing as clinically indicated. Other large trauma clinical trials enrolling similarly injured populations have also reported higher thrombotic event rates compared to the CRASH-2 trial.36–38 Several ongoing trauma clinical trials evaluating TXA should provide additional insight into the incidence of thrombotic events.39

**LIMITATIONS**

Our results should be interpreted in the context of several limitations. This study was conducted in a...
single U.S. trauma center; thus the results may not be
generalizable to other U.S. trauma centers with differ-
ent patient populations and resources. Compared to
the CRASH-2 trial, our simple size is small, but we
that calculated that 36 months of data acquisition was
adequate to provide a sample size sufficient to ensure
accuracy and precision during the analysis.

This study is subject to the inherent limitations of
using retrospective data. However, we followed the rec-
commended guidelines for retrospective reviews to min-
imize any bias.18

We have stated in the study population section that
all admitted trauma patients receive thromboembolic
prophylaxis as per policy. However, we were unable to
extract more nuanced description of this such as compli-
cance with compression stockings or early mobilization.

As mentioned in the results and discussion sections,
the reasons why not all our patients received complete
bolus and maintenance doses were not specified in the
EMR and remain unclear. Further research should be
conducted to compare different doses strategies.

CONCLUSIONS

In our single-center study, adult trauma patients receiv-
ing tranexamic acid had similar incidences of death
but higher incidences of thromboembolic events com-
pared to the CRASH-2 trial. Differences in patient
characteristics, injury severity, dosing of tranexamic
acid, and surgery and transfusion rates could explain
these observed differences. Further research is neces-
sary to provide additional insight into the incidence
and risk factors of thromboembolic events with TXA in
tranexamic acid use.

References

1. Centers for Disease Control and Prevention. Ten Leading
Causes of Death and Injury – Unintentional Injury. Injury
Prevention & Control: Data & Statistics. Available at:
http://www.cdc.gov/injury/wisqars/LeadingCauses_image
2. Gruen RL, Brohi K, Schreiber M, et al. Haemorrhage con-
4. Dutton RP, Stansbury LG, Leone S, et al. Trauma mortal-
ity in mature trauma systems: are we doing better? An
Trauma 2010;69:620–6.
5. Grecott R. The lethal triad. Hypothermia, acidosis &
coagulopathy create a deadly cycle for trauma patients.
6. Peng N, Su L. Progresses in understanding trauma-induced
cogulopathy and the underlying mechanism. Chin J Trau-
matol 2017;20:133–6.
in multiple injury: an analysis from the German Trauma
of trauma: hypoperfusion induces systemic anticoagulation
and hyperfibrinolysis. J Trauma 2008;64:1211–7; discus-
sion 1217.
10. Mitra B, Cameron PA, Mori A, Fitzgerald M. Acute coag-
ulopathy and early deaths post major trauma. Injury
stereo-isomer (trans-form) of amcha and its antifibrinolytic
(antiplasminic) action in vitro and in vivo. Keio J Med
12. Shakur H, Roberts I, Bautista R, et al. Effects of tranex-
amic acid on death, vascular occlusive events, and blood
transfusion in trauma patients with significant haemor-
rhage (CRASH-2): a randomized, placebo-controlled trial.
Lancet 2010;376:23–32.
13. Roberts I, Shakur H, Afolabi A, et al. The importance of
early treatment with tranexamic acid in bleeding trauma
patients: an exploratory analysis of the CRASH-2 ran-
domised controlled trial. Lancet 2011;377:1096–101,
101.e1–2.
Avoidable mortality from giving tranexamic acid to bleed-
ing trauma patients: an estimation based on WHO mor-
tality data, a systematic literature review and data from the
15. Napolitano LM, Cohen MJ, Cotton BA, Schreiber MA,
Moore EE. Tranexamic acid in trauma: how should we use
16. Burns PB, Rohrich RJ, Chung KC. The levels of evidence
and their role in evidence-based medicine. Plast Reconstr
17. American College Of Surgeons – The National Trauma
Data Bank (NTDB). Available at: https://www.facs.org/
quality-programs/trauma/tpc/group-centers/ntdb/ntds.
18. Kaji AH, Schriger D, Green S. Looking through the retro-
spectoscope: reducing bias in emergency medicine chart
19. Graham P, Bull B. Approximate standard errors and con-
fidence intervals for indices of positive and negative agree-


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.13883/full

Table S1. Interrater agreement conducted on 10% of the abstracted charts.
Outcomes With the Use of Bag–Valve–Mask Ventilation During Out-of-hospital Cardiac Arrest in the Pragmatic Airway Resuscitation Trial

Joshua R. Lupton, MD, MPH1, Robert H. Schmicker, MS2, Shannon Stephens3, Jestin N. Carlson, MD4, Clifton Callaway, MD, PhD5, Heather Herren, MPH2, Ahamed H. Idris, MD5, George Sopko, MD, MPH6, Juan C. J. Puyana4, Mohamud R. Daya, MD, MS1, Henry Wang, MD, MS7, and Matt Hansen, MD, MCR1

A related article appears on page 431.

ABSTRACT

Background: While emergency medical services (EMS) often use endotracheal intubation (ETI) or supraglottic airways (SGA), some patients receive only bag–valve–mask (BVM) ventilation during out-of-hospital cardiac arrests (OHCA). Our objective was to compare patient characteristics and outcomes for BVM ventilation to advanced airway management (AAM) in adults with OHCA.

Methods: Using data from the Pragmatic Airway Resuscitation Trial, we identified patients receiving AAM (ETI or a SGA), BVM ventilation only (BVM-only), and BVM ventilation as a rescue after at least one failed attempt at advanced airway placement (BVM-rescue). The outcomes were return of spontaneous circulation (ROSC), 72-hour survival, survival to hospital discharge, neurologically intact survival (Modified Rankin Scale ≤ 3), and the presence of aspiration on a chest radiograph. Comparisons were made using generalized mixed-effects models while adjusting for age, sex, initial rhythm, EMS-witnessed status, bystander cardiopulmonary resuscitation, response time, study cluster, and advanced life support first on scene.

Results: Of 3,004 patients enrolled, there were 282 BVM-only, 2,129 AAM, and 156 BVM-rescue patients with complete covariates. Shockable initial rhythms (34% vs. 18.6%) and EMS-witnessed arrests (21.6% vs. 11.3%)
Oxygenation in out-of-hospital cardiac arrest (OHCA) can be accomplished using passive insufflation, bag–valve–mask (BVM) ventilation, or ventilation following placement of advanced airways such as supraglottic airways (SGAs) or endotracheal intubation (ETI). The optimal strategy for prehospital airway and ventilation management, however, remains unclear. Two recent trials found that compared with ETI, SGA resulted in similar or better survival in OHCA than ETI.\textsuperscript{1,2} In a recently published randomized controlled trial (RCT) conducted in France and Belgium, Jabre et al.\textsuperscript{3} were unable to determine if a strategy of BVM-only ventilation was noninferior to a strategy of initial BVM ventilation followed by ETI in OHCA. An important limitation of this RCT is that prehospital care physicians performed ETI, which is uncommon in the United States. Since studies have demonstrated lower rates of first-pass success with paramedic initiated ETI during OHCA,\textsuperscript{1,2} the generalizability of the findings from Jabre et al. remains unclear.

Observational studies conducted in emergency medical services (EMS) systems outside the United States have reported both improved\textsuperscript{4,5} and decreased patient survival\textsuperscript{6} with BVM compared to ventilation through advanced airway devices such as a SGA or endotracheal tube. In the United States, observational studies of pediatric\textsuperscript{7} and adult OHCA\textsuperscript{8-11} have consistently found improved survival with BVM compared to advanced airway management (AAM). However, it is unclear if these findings are due to differences in patient characteristics of those treated with BVM compared to AAM or intrinsic benefits of BVM. Furthermore, BVM may also be used as a rescue device after failed attempts to place an advanced airway. The clinical course of these BVM-rescue patients may be significantly different and has not been clearly examined.

The Pragmatic Airway Resuscitation Trial (PART) examined the outcomes of OHCA patients in the United States cluster-randomized to receive either an SGA or ETI as the first airway management strategy.\textsuperscript{2} In PART, a sizeable number of patients received BVM-only ventilation despite the availability of advanced airway devices, which may be due to patients obtaining return of spontaneous circulation (ROSC) quickly or being do not attempt resuscitation (DNAR) or do not intubate, among other possible reasons. In addition, some patients also received BVM-rescue after failed attempts to place an advanced airway. The purpose of this study is to examine the characteristics and outcomes of patients that received BVM-only or BVM-rescue ventilation compared to those who were ventilated following successful placement of an advanced airway (SGA or ETI).

**METHODS**

**Design and Setting**

This is a secondary analysis of data from PART, which was conducted by 27 EMS agencies in Birmingham, Alabama; Dallas-Fort Worth, Texas; Milwaukee, Wisconsin; Pittsburgh, Pennsylvania; and Portland, Oregon. The trial was conducted under exception from informed consent (EFIC) and utilized cluster randomization with crossover at 3- to 5-month intervals. The institutional review boards of the participating institutions approved the trial under federal rules for conduct of emergency research under EFIC. The detailed methods of PART have previously been reported.\textsuperscript{12} This post hoc analysis was planned and implemented after the investigators knew the results of PART.

**Population**

Eligible subjects were adults (age \( \geq \) 18 years old) with nontraumatic OHCA treated by the above EMS agencies. Patients with preexisting tracheostomy, DNAR orders, an advanced airway present prior to EMS arrival, left-ventricular assist devices, a do-not-enroll bracelet, major bleeding or exsanguination, obvious asphyxia cardiac arrest, interfacility transports, and...
traumatic etiology of the arrest were excluded from the study.

**Intervention**

The study intervention in the PART was an initial airway management strategy using the laryngeal tube (LT) or ETI. Local agency protocols dictated the allowed number of LT or ETI insertion attempts as well as rescue airway strategies if the initial strategy was not successful.

**Data Quality**

All variables used in the analysis were abstracted and entered into an electronic database by trained research assistants. Each site underwent periodic data audits and the data entry forms also had multiple integrated logic checks to ensure high-quality data.

**Data Analysis**

The primary and secondary outcomes for this analysis match those of the parent trial: 72-hour survival after OHCA with secondary outcomes including ROSC, survival to hospital discharge, and survival with a favorable neurologic status defined as a Modified Rankin Scale score of ≤3 at hospital discharge. The novel 72-hour survival outcome was used in the parent trial due to the pragmatic design of the trial. We additionally evaluated the complication of pulmonary aspiration by reviewing radiograph reports of admitted patients that had a chest X-ray performed within 24 hours of ROSC.

The independent variable used in the analysis was airway type: BVM-only and BVM-rescue compared to AAM. Each EMS agency recorded the airway type, timing of placement, and number of attempts at advanced airway placement among other variables.

The association between the above primary and secondary outcomes and airway type were examined using generalized mixed-effects models. All cases included in the PART were eligible for inclusion and excluded only if they did not have complete independent, dependent, and covariate data available. We adjusted estimates for the following confounders: age, sex, shockable initial rhythm, EMS-witnessed arrest, bystander cardiopulmonary resuscitation (CPR), dispatch to arrival time, advanced life support versus basic life support first on scene, and study cluster. Models evaluated odds ratios (ORs) for each outcome (ROSC, 72-hour survival, survival to hospital discharge, and neurologic intact survival) when comparing BVM-only and BVM-rescue to the reference of successful AAM. We conducted a sensitivity analysis using LT and ETI as the references in our multivariable model with independent variables of interest of BVM-only and BVM-rescue. In this analysis, BVM-rescue was separated as BVM-rescue after failed LT as first attempt and BVM-rescue after failed ETI as first attempt to facilitate accurate comparisons against the referenced advanced airway.

**RESULTS**

There were 3,004 patients enrolled in PART, of whom 352 received BVM-only, 2,463 had AAM (LT or ETI), and 189 received BVM-rescue ventilation after failed advanced airway placement (Figure 1). There were 282 BVM-only, 2,129 AAM, and 156 BVM-rescue cases with complete covariate data (Table 1). Compared to those with AAM, BVM-only patients had higher proportions of initial shockable rhythms (34.0% vs. 18.6%), EMS-witnessed arrests (21.6% vs. 11.3%), and public location of arrest (21.3% vs. 11.3%). BVM-only patients had a lower proportion of epinephrine administration compared to AAM or BVM-rescue (60.6% vs. 97.6 and 94.2%, respectively). Total resuscitation time was also shorter for BVM-only patients (minutes [IQR]) compared to AAM or BVM-rescue patients (23.3 [17.1 to 28.6] minutes vs. 33.2 [26.3 to 41.0] and 29.9 [23.3 to 35.2] minutes, respectively). BVM-rescue ventilation and AAM groups had similar proportions of initial shockable rhythms (17.3% vs. 18.6%), EMS-witnessed arrests (13.5% vs. 11.3%), and public location of arrest (7.7% vs. 11.3%). Response times, proportion of successful intravenous (IV) access, and proportion receiving defibrillation via EMS were similar across the three groups.

Emergency medical services cited a specific reason for not attempting an advanced airway in 273 of 352 BVM-only cases (77.6%). ROSC was the most common reason cited (28.7%), followed by determination of death prior to advanced airway attempt (16.0%), jaw clenching with inability to insert advanced airway (12.8%), adequate ventilation achieved with BVM (12.1%), and arrival at ED before advanced airway attempt (8.2%). In 22.4% of BVM-only cases, the reasons for not attempting AAM were listed as other, unable to determine, or information not available. The BVM-rescue group had a higher proportion of two or more attempts (48.7%) compared to the
Figure 1 Study population

### Table 1
Patient and EMS Response Characteristics

<table>
<thead>
<tr>
<th></th>
<th>BVM-only Ventilation</th>
<th>BVM Ventilation, After Failed AAM</th>
<th>AAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>282</td>
<td>156</td>
<td>2129</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>63.0 (49.3, 74)</td>
<td>65.0 (52, 76)</td>
<td>65.0 (54, 76)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>172 (61.0)</td>
<td>87 (55.8)</td>
<td>1320 (62.0)</td>
</tr>
<tr>
<td>Initial rhythm, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VT/VF</td>
<td>96 (34.0)</td>
<td>27 (17.3)</td>
<td>395 (18.6)</td>
</tr>
<tr>
<td>PEA</td>
<td>52 (18.4)</td>
<td>33 (21.2)</td>
<td>463 (21.7)</td>
</tr>
<tr>
<td>Asystole</td>
<td>114 (40.4)</td>
<td>83 (53.2)</td>
<td>1069 (50.2)</td>
</tr>
<tr>
<td>No shock advised</td>
<td>20 (7.1)</td>
<td>13 (8.3)</td>
<td>202 (9.5)</td>
</tr>
<tr>
<td>Witness status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMS</td>
<td>61 (21.6)</td>
<td>21 (13.5)</td>
<td>241 (11.3)</td>
</tr>
<tr>
<td>Bystander</td>
<td>106 (37.6)</td>
<td>53 (34.0)</td>
<td>807 (37.9)</td>
</tr>
<tr>
<td>None</td>
<td>115 (40.8)</td>
<td>82 (52.6)</td>
<td>1081 (50.8)</td>
</tr>
<tr>
<td>Bystander resuscitation, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No bystander CPR</td>
<td>139 (49.3)</td>
<td>72 (46.2)</td>
<td>1011 (47.5)</td>
</tr>
<tr>
<td>Bystander CPR, no AED</td>
<td>108 (38.3)</td>
<td>64 (41.0)</td>
<td>817 (38.4)</td>
</tr>
<tr>
<td>Bystander CPR and AED</td>
<td>30 (10.6)</td>
<td>14 (9.0)</td>
<td>231 (10.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (1.8)</td>
<td>6 (3.8)</td>
<td>70 (3.3)</td>
</tr>
<tr>
<td>Public location</td>
<td>60 (21.3)</td>
<td>12 (7.7)</td>
<td>241 (11.3)</td>
</tr>
<tr>
<td>Response time (min), median (IQR)</td>
<td>5.3 (4.1–6.9)</td>
<td>5.2 (4.1–7.0)</td>
<td>5.1 (4.0–6.6)</td>
</tr>
<tr>
<td>ALS first responder</td>
<td>222 (78.7)</td>
<td>111 (71.2)</td>
<td>1415 (66.5)</td>
</tr>
<tr>
<td>Epinephrine administration, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administered</td>
<td>171 (60.6)</td>
<td>147 (94.2)</td>
<td>2078 (97.6)</td>
</tr>
<tr>
<td>Arrival to epinephrine</td>
<td>16.0 (13.5–18.9)</td>
<td>9.9 (9.0–10.9)</td>
<td>9.5 (9.2–9.8)</td>
</tr>
<tr>
<td>EMS shock administration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administered, n (%)</td>
<td>84 (29.8)</td>
<td>39 (25.0)</td>
<td>650 (30.5)</td>
</tr>
<tr>
<td>Arrival to first EMS shock</td>
<td>4.3 (3.5–5.1)</td>
<td>5.5 (5.0–7.4)</td>
<td>8.0 (7.3–8.9)</td>
</tr>
<tr>
<td>Total resuscitation time</td>
<td>23.3 (17.1–28.6)</td>
<td>29.9 (23.3–35.2)</td>
<td>33.2 (26.3–41.0)</td>
</tr>
<tr>
<td>IV/IO access, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful IV access</td>
<td>144 (51.1)</td>
<td>78 (50.0)</td>
<td>1023 (48.1)</td>
</tr>
<tr>
<td>Successful IO access</td>
<td>103 (36.5)</td>
<td>81 (51.9)</td>
<td>1198 (56.3)</td>
</tr>
</tbody>
</table>

*AAM = advanced airway management; ALS = advanced life support; BVM = bag-valve-mask; CPR = cardiopulmonary resuscitation; EMS = emergency medical services; IO = intraosseous; PEA = pulseless electrical activity; VT/VF = pulseless ventricular tachycardia/ventricular fibrillation.*
successful advanced airway ventilation group (30.0%; Table 2). There were similar numbers of attempts regardless of whether BVM-rescue was performed after an initial attempt at LT or ETI placement (Table 2).

Compared to AAM, BVM-only patients had higher field ROSC (37.9% vs. 34.7%), survival to hospital admission (35.8% vs. 24.9%), 72-hour survival (31.6% vs. 14.7%), survival to hospital discharge (28.0% vs. 6.8%), and neurologically intact survival (21.6% vs. 3.6%; Table 3). Compared to AAM, BVM-rescue patients had lower ROSC (28.2% vs. 34.7%), survival to hospital admission (24.4% vs. 24.9%) and 72-hour survival (14.7% vs. 14.7%), but higher survival to hospital discharge (10.9% vs. 6.8%) and neurologically intact survival (6.4% vs. 3.6%). The reported rate of aspiration on chest X-ray was similar between BVM-only (5.1%), BVM-rescue (1.1%), and AAM (4.0%) groups. These trends were similar when comparing BVM-rescue by first attempted airway (failed LT or ETI) to successfully placed advanced airways (LT or ETI). Although we did not have covariate data on all 3,004 PART cases for this analysis, the outcomes for this entire cohort were similar to those of this smaller subset with complete covariate data (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.13927/full).

After excluding cases in whom BVM-only was used due to rapid ROSC, the patient determined to be obviously dead, or a patient having a clenched jaw, we were left with 120 patients. Within this BVM-only subset, while rates of ROSC (16.7% vs. 34.7%) and survival to admission (20.0% vs. 24.9%) were lower compared to successful AAM, 72-hour survival was similar (14.2% vs. 14.7%) and there was higher survival to hospital discharge (10.8% vs. 6.8%) and neurologically intact survival (7.5% vs. 3.6%), respectively.

Of the BVM-only, BVM-rescue, and AAM patients who survived to hospital admission, 30.7% (31/101), 52.6% (20/38), and 57.6% (306/531) underwent therapeutic hypothermia, a potential surrogate indicator of their neurologic status at time of admission. Among BVM-only, BVM-rescue, and AAM patients receiving therapeutic hypothermia, 45.2%

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Airway Attempt Characteristics by First Attempted Airway</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First attempt</strong></td>
<td>BVM-rescue</td>
</tr>
<tr>
<td></td>
<td>LT</td>
</tr>
<tr>
<td>No.</td>
<td>67</td>
</tr>
<tr>
<td>No. of airway attempts, n (%)</td>
<td></td>
</tr>
<tr>
<td>One</td>
<td>34 (50.7)</td>
</tr>
<tr>
<td>Two</td>
<td>17 (25.4)</td>
</tr>
<tr>
<td>Three or more</td>
<td>16 (23.9)</td>
</tr>
<tr>
<td>Number of attempts, mean (±SD)</td>
<td>1.8 (±1.0)</td>
</tr>
</tbody>
</table>

AAM = advanced airway management; BVM = bag-valve-mask; ETI = endotracheal intubation; LT = King laryngeal tube.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Patient Outcomes by Airway Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First attempt</strong></td>
<td>BVM-only None</td>
</tr>
<tr>
<td></td>
<td>LT</td>
</tr>
<tr>
<td>No.</td>
<td>282</td>
</tr>
<tr>
<td>ROSC in field</td>
<td>107 (37.9%)</td>
</tr>
<tr>
<td>Admitted to hospital</td>
<td>101 (35.8%)</td>
</tr>
<tr>
<td>Survived to 72 hours</td>
<td>89 (31.6%)</td>
</tr>
<tr>
<td>Survived to hospital discharge</td>
<td>79 (28.0%)</td>
</tr>
<tr>
<td>MRS ≤ 3*</td>
<td>61 (21.6%)</td>
</tr>
<tr>
<td>Aspiration</td>
<td>9 (5.1%)</td>
</tr>
</tbody>
</table>

*Modified Rankin Scale, with 0 being no symptoms at all, 1-3 representing moderate, slight, or no significant disability, and 4-6 representing moderate severe disability, severe disability, or death.

Data are reported as n (%). AAM = advanced airway management; BVM = bag-valve-mask; ETI = endotracheal intubation; LT = King laryngeal tube; MRS = Modified Rankin Scale; ROSC = return of spontaneous circulation.
(14/31), 20.0% (4/20), and 18.3% (56/306) had neurologically intact survival, respectively. Survival outcomes were also better for those treated with BVM-only and BVM-rescue when separating groups by shockable versus nonshockable initial rhythms (Data Supplement S1, Tables S2 and S3).

In multivariable regression analysis (Figure 2) compared to AAM, BVM-only patients exhibited similar ROSC (OR = 1.29 [95% CI = 0.96 to 1.73]), but higher 72-hour survival (OR = 1.96 [95% CI = 1.42 to 2.69]), survival to hospital discharge (OR = 4.47 [95% CI = 3.03 to 6.59]), and neurologically intact survival (OR = 7.05 [95% CI = 4.40 to 11.3]). Compared to AAM, BVM-rescue patients exhibited similar ROSC (OR = 0.73 [95% CI = 0.47 to 1.12]) and 72-hour survival (OR = 1.08 [95% CI = 0.66 to 1.77]), but higher survival to hospital discharge (OR = 2.15 [95% CI = 1.17 to 3.95]) and neurologically intact survival (OR = 2.64 [95% CI = 1.20 to 5.81]).

In sensitivity analyses evaluating outcomes with LT and ETI as the references in multivariable regression, BVM-only had higher ROSC, survival to hospital discharge, and neurologically intact survival (Data Supplement S1, Table S4).

**DISCUSSION**

Patients who received BVM-only ventilation during OHCA when advanced airways were readily available appear to be distinctly different than those who received attempts at advanced airway placement. BVM-only patients were more likely to have had initially shockable rhythms, EMS-witnessed arrests, public location of arrests, and ROSC, explaining some of the improved 72-hour survival, survival to discharge, and neurologically intact survival of BVM-only patients observed in this study. The observed differences in baseline characteristics make it challenging to draw conclusions as to how BVM-only affects the outcomes for OHCA patients compared to AAM. The groups of patients receiving BVM-rescue were more similar to those receiving AAM, with the exception that they all had unsuccessful attempts at placing an advanced airway. Surprisingly, BVM-rescue had greater survival to discharge and neurologically intact survival than AAM despite similar initial rates of ROSC. This supports the notion that BVM may be associated with improved survival compared to AAM in OHCA.

![Graph showing ORs for outcomes by BVM-only and BVM-rescue ventilation compared to AAM. 95% CIs displayed. *Reference = successful AAM. AAM = advanced airway management; BVM = bag-valve-mask.](image-url)
There are many potential reasons for the outcome differences seen between BVM and AAM in this study. It may be that these differences are related to the type of advanced airway attempted, although our sensitivity analysis with LT and ETI as the reference still revealed significant improvement in outcomes for BVM-only compared to either of the AAM choices in multivariable analysis. BVM patients may be a unique group whereas BVM-only is a surrogate marker for better outcome, resulting in confounding by indication. However, BVM-rescue also had improved survival despite similar baseline characteristics to the AAM group. These outcome differences may be due to differences in ventilation characteristics with BVM compared to AAM, such as changes in the rate, volume, and pressure of ventilation. Finally, it may be that BVM as an airway strategy distracts less from other critical priorities early on in OHCA resuscitation, such as timely defibrillation and high-quality chest compressions. We emphasize that our findings are hypothesis generating only and that these results do not indicate a causal link between BVM and OHCA outcomes.

As previously mentioned, the recently published RCT by Jabre et al. comparing BVM-only to BVM followed by ETI found no evidence for noninferiority of BVM-only compared to ETI. However, in this study prehospital care physicians performed ETI in the majority of cases with 97.9% overall success, although first-pass success was not specifically reported. In our cohort and in most parts of the United States, a paramedic and not a physician performs ETI and thus may have lower rates of overall success. First-pass ETI success in our cohort was 51.6%, lower than the previously reported first-pass success of 68% in a European prehospital system with physician intubation. This may in part explain why in observational studies comparing BVM to AAM in United States OHCA patients, there is greater survival among those with BVM-only compared to AAM in both pediatric and adult populations. There have been no studies to date that have examined the outcome for BVM after failed advanced airway placement attempts in the United States. One prior study in Taiwan evaluated BVM compared to AAM (SGA and ETI) in a retrospective cohort where they also reported survival for patients with BVM-rescue after attempts at ETI. In that study, the survival for the BVM-rescue group after ETI attempt was worse than BVM-only or AAM, although the total number of patients in this group was 37, which is considerably smaller than the 189 patients in our study. In that study, BVM-only and BVM-rescue also had lower proportions of bystander CPR compared to successful ETI, a difference not seen in our study. Additionally, in that study, BVM-only was found to have lower ROSC and survival to hospital discharge, although unchanged neurologically intact survival compared to AAM with either a SGA or an ETI. These limitations make it difficult to compare the findings from the Taiwanese study with those from this study.

Differences in ventilation may explain the differences in outcomes between BVM and AAM. In a manikin trial comparing BVM ventilation and AAM, there were no differences in rates of breaths between modalities but BVM ventilation delivered lower mean tidal volumes compared to AAM. Future trials comparing airway strategies should monitor ventilation parameters (rate, volume, duration, and pressure) more closely to facilitate more meaningful comparisons. It is also important to note that unlike AAM, where after successful placement one rescuer can ventilate, the optimal technique of BVM requires two rescuers to be performed properly. This means for BVM-only to be effective it may require more team resources throughout the resuscitation compared to ventilation through a successfully placed advanced airway. There is observational evidence to suggest a relationship between number of rescuers on scene and survival. However, in a randomized, subjective comparison among emergency medical technicians in Austria using BVM or LT on 78 OHCA patients, there were no reported differences in the ease of ventilation between BVM and LT. In our study, we were unable to ascertain if one or two rescuers performed the BVM.

LIMITATIONS

As emphasized, the primary limitation of our study is that the BVM-only patients were not randomized to that strategy, although we attempted to adjust for possible confounders in the analysis. We additionally did not have detailed explanations for the reasons why the BVM-only approach was used in all instances and why advanced airway placement was deemed to be unsuccessful in the BVM-rescue group. Our primary analysis combined ETI and LT into one AAM category and BVM-rescue after LT and after ETI, respectively, into one BVM-rescue category, which may have missed
differences when LT or ETI were used as primary comparisons. However, our sensitivity analysis utilizing LT or ETI as the reference groups showed unchanged trends in improved survival with BVM-only and although with wider CIs, similar trends in survival with BVM-rescue after LT and ETI, respectively. This study also suffers from similar limitations to those in the parent trial, namely that the SGA used was the LT and so our findings may not generalize to other SGA devices such as the i-gel. Our results are specific to the OHCA setting and not generalizable to the in-hospital cardiac arrest environment. In addition, we did not have access to CPR process data (chest compression rate, depth, perishock pause, CPR fraction) and so could not adjust for differences related to these variables. We also do not have data on rate, volume, duration, and pressure of ventilation, all of which have been shown to affect survival in other studies. Our aspiration complication data were based on chest X-rays obtained within 24 hours of ROSC for admitted patients and, although clinically relevant, may have missed episodes of visible aspiration during intubation that did not result in radiographic findings or aspiration events that occurred in patients that were not admitted. Finally, although larger than some other studies, the small sample size of BVM-rescue group raises the risk of a type II error in our observed differences.

CONCLUSION

Compared with advanced airway management, bag–valve–mask-only ventilation is associated with improved out-of-hospital cardiac arrest outcomes. Despite similar return of spontaneous circulation and 72-hour survival, bag–valve–mask-rescue ventilation was associated with improved survival to hospital discharge and neurologically intact survival compared to successful advanced airway management in the Pragmatic Airway Resuscitation Trial. Given the retrospective nature of this study, a cause-and-effect relationship cannot be established. Thus, a prospective trial is needed to explore the potential benefits of bag–valve–mask ventilation compared to advanced airway management for out-of-hospital cardiac arrest in the United States.

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.13927/full

Data Supplement S1. Supplemental material.
Early Self-Proning in Awake, Non-intubated Patients in the Emergency Department: A Single ED’s Experience During the COVID-19 Pandemic

Nicholas D. Caputo, MD, MSc1, Reuben J. Strayer, MD2, and Richard Levitan, MD3

ABSTRACT

Objective: Prolonged and unaddressed hypoxia can lead to poor patient outcomes. Proning has become a standard treatment in the management of patients with ARDS who have difficulty achieving adequate oxygen saturation. The purpose of this study was to describe the use of early proning of awake, non-intubated patients in the emergency department (ED) during the COVID-19 pandemic.

Methods: This pilot study was carried out in a single urban ED in New York City. We included patients suspected of having COVID-19 with hypoxia on arrival. A standard pulse oximeter was used to measure SpO2. SpO2 measurements were recorded at triage and after 5 minutes of proning. Supplemental oxygenation methods included non-rebreather mask (NRB) and nasal cannula. We also characterized post-proning failure rates of intubation within the first 24 hours of arrival to the ED.

Results: Fifty patients were included. Overall, the median SpO2 at triage was 80% (IQR 69 to 85). After application of supplemental oxygen was given to patients on room air it was 84% (IQR 75 to 90). After 5 minutes of proning was added SpO2 improved to 94% (IQR 90 to 95). Comparison of the pre- to post-median by the Wilcoxon Rank-sum test yielded P = 0.001. Thirteen patients (24%) failed to improve or maintain their oxygen saturations and required endotracheal intubation within 24 hours of arrival to the ED.

Conclusion: Awake early self-proning in the emergency department demonstrated improved oxygen saturation in our COVID-19 positive patients. Further studies are needed to support causality and determine the effect of proning on disease severity and mortality.

BACKGROUND

Prolonged and unaddressed hypoxia can lead to poor outcomes in patients with respiratory compromise.1 Boosting inspired oxygen (FiO2) is an effective therapy in many hypoxic patients; however, in patients with significant physiologic shunting, positive pressure may be required.2 This is usually delivered by invasive or non-invasive ventilation (NIV). These types of interventions require resources that under normal circumstances are generally available, however become quickly limited in times of surge. Awake proning has been demonstrated to decrease intubation and improve outcomes in ARDS patients.3

In New York City, during the early stages of the COVID-19 pandemic, patients presented en masse with moderate to severe hypoxia. Some of these patients were distressed, quickly deteriorated and required endotracheal intubation. However, COVID-19 produced
another group of patients whose pathophysiology confounded existing disease patterns. These patients had low oxygen saturations (SpO₂ < 90%), but were not in significant respiratory distress and often appeared clinically well; this group has been informally referred to as happy hypoxemics. Because many of these patients were markedly tachypneic, had chest radiographic findings similar to acute respiratory distress syndrome (ARDS), had hypoxemia not responsive to supplemental oxygen, and because of infectious aerosolization fears around alternative oxygenation modalities, many of them were intubated early in their hospital course. Ventilator stockpiles and critical care resources were quickly depleted the result of widespread early intubation of patients with COVID-19 lung disease. Based on prior literature, with other causes of ARDS it was speculated that proning of awake patients would improve patient’s oxygenation and prevent or delay intubation.⁴ We sought to describe our preliminary experience with the use of early proning of awake, non-intubated patients with suspected or confirmed COVID-19 disease and its impact on oxygenation in the ED. Our primary outcome was median SpO₂ after supplemental oxygen and proning were applied in tandem.

METHODS

Study design and setting
We conducted an observational cohort study of a convenience sample of patients at an urban, academic ED in New York City, USA between March 1st and April 1st of 2020. This study was approved by the Lincoln hospital institutional review board and ethics board.

The average annual volumes of the ED is approximately 175,000. The department generally performs about 40 to 50 intubations a month with the majority of intubations being performed by EM trainees under the direct supervision of an EM attending.

Selection of Participants
We included the first fifty adult patients (age ≥18 years old) who presented to the ED with hypoxia (SpO₂ <90%) and without resolution (SpO₂ >93%) despite supplemental oxygen and who were capable of self-proning during the early stages of the COVID-19 pandemic in March to April 2020. Patients were asked to self-prone/change position. We excluded patients with DNR/DNI code status, in cardiac arrest, receiving non-invasive ventilation (NIV) or those who were intubated in the prehospital setting. All patients had documented SARS-CoV-2 infection, confirmed by nasal/oropharyngeal swab followed by positive reverse transcriptase polymerase chain reaction detection of viral nucleic acid.

Methods of Measurement
Vital signs were obtained from the cardiac monitor (Philips IntelliVue, Philips USA) in real time. SpO₂ was measured through standard finger oximeters (Covidien Oximax, Covidien, USA). Hypoxemia was defined as a SpO₂ <90%.

Outcomes
The primary outcome was the change in SpO₂, determined prior to proning, after application of supplemental oxygen and after 5 minutes of proning without change in inspired oxygen. The secondary outcome was rate of patients who were proned but then required intubation within 24 hours of presentation to the ED. A patient was deemed to have failed proning if they showed respiratory failure defined as persistent SpO₂ < 90% in the setting of unresolved or worsening tachypnea with either accessory muscle use, altered mental status or hypercarbia on blood gas.

Analysis
The primary dependent variable was the SpO₂ which was not normally distributed (P> 0.1 by Shapiro–Wilk), necessitating the reporting of median values. For the clinical series, we analyzed the data using descriptive techniques. We determined median SpO₂ prior to proning and after proning. We determined the proportion of patients achieving SpO₂ >93% with proning. We compared the pre to post median values using the Wilcoxon Rank Sum test. We also determined the proportion of patients that failed proning (using the explicit definition of respiratory failure) and required intubation. All analyses were performed using XLStat (Addinsoft, New York, NY).

RESULTS
We included 50 patients in this convenience sample cohort, most with respiratory complaints leading to their visit to the ED. All patients were observed in the ED until admission to the floors. The median time observation of the cohort in the ED was 29 minutes (range 63 to 1620).

The median age of the cohort was 59 (IQR 50 to 68) with 60% of the group being male. Eighty percent
of the cohort were tachypneic on arrival (RR > 20). On arrival to the ER, over half of this cohort, 56% (28), had no supplemental oxygen being delivered (e.g., were on room air). Eighty percent of these patients arrived as “walk-ins” and 20% arrived by EMS. The remaining 44% (22) of these patients arrived to the ER with supplemental oxygen being provided, usually non-rebreather mask (n = 8) or nasal cannula at approximately 5 liters per minute (n = 14). The median SpO2 of patients who arrived without supplemental oxygen was 75% (IQR 62 to 82) and for those patients with supplemental oxygen in place was 82% (IQR 72 to 85). Overall, the median SpO2 at triage was 80% (IQR 69 to 85). This improved to 84% (IQR 75 to 90) after application of supplemental oxygen (non-rebreather mask [n = 38] or nasal cannula at approximately 5 liters per minute [n = 12]). After 5 minutes of proning was added, the media SpO2 increased to 94% (IQR 90 to 95). Comparison of the pre- to post-median by the Wilcoxon Rank–sum test yielded P = 0.001.

Thirteen patients (24%, 95% CI 14.6 to 40.3%) met the definition of respiratory failure plus clinical signs of respiratory distress within 24 hours of presenting to the ED and required endotracheal intubation. Of these 13 patients who required intubation, four patients were intubated within 30 minutes of proning, three patients were intubated between 30 and 60 minutes after proning and the remaining six were intubated after 60 minutes of initiation of proning but within 24 hours. Of those patients who were not intubated within 24 hours (n = 37), five were subsequently intubated (three between 24 and 48 hours and two after 72 hours) as inpatients.

**LIMITATIONS**

This study is a non-experimental sequential case series that reports an association between proning patients with COVID-19 and improvement in oxygen saturation. Though the effect size is significant and consistent with existing models of physiologic shunt, causal inferences arising from descriptive studies can only be hypothesized, not concluded. The patients described come from a convenience sample presenting to a single hospital and therefore may not represent other populations or the population at large. All aspects of care were uncontrolled; therefore the effect seen may be due not to proning, but to an unrecognized alternative treatment. In order to make a strong claim to causality, proning should be studied in a prospective trial that randomizes similar patients to proning or not, and where other aspects of care are congruent in both arms. Lastly, though oxygen saturation contributes to patient-oriented outcomes such as endotracheal intubation, vital signs are themselves a disease-oriented endpoint; attributing value to the treatment requires that it be measured against more important consequences such as duration of hospitalization or death.

**DISCUSSION**

COVID-19 is a novel disease arising from a novel pathogen, SARS-CoV-2. Frontline physicians working in New York City have been confronted with unprecedented challenges around resource scarcity and disease infectivity; however, the most enduring tribulation may be caring for patients who become critically ill and succumb to an illness that does not fit into existing models, does not respond to usual therapies, and for which there are no treatments established by rigorous science.

Clinicians managing the earliest cases of COVID-19 in China and Italy were faced with extraordinary levels of hypoxemia, and serious concerns that viral particles would be aerosolized during oxygenation therapies such as noninvasive ventilation and high flow nasal cannula. This led to a recommendation that patients who do not adequately respond to low-flow oxygen therapies (such as conventional nasal cannula or venturi mask) be intubated without the usual trial of pressurized oxygen modalities.

The *intubate early* approach was adopted in the first wave of critically ill COVID-19 patients seen in New York City hospitals, but early outcomes data from overseas demonstrated shockingly high mortality for intubated patients, and the inevitability of resource scarcity, if early intubation was continued, caused clinicians to seek strategies to delay or prevent the initiation of mechanical ventilation in COVID-19 patients.

Little was known of the pathophysiology of COVID-19 disease in the early days of the pandemic. An Italian described two patient subtypes that has framed management approaches across different phases of illness.

The conventional alternatives to mechanical ventilation—NIV and HFNC—have been used successfully in COVID-19 but their implementation is hindered by several factors in addition to the aforementioned aerosolization concerns. For reasons presently not understood, COVID-19 lung disease patients frequently...
demonstrate hypoxia out of proportion to dyspnea or distress, diminishing the utility of perhaps the most important indicator of respiratory function: pulse oximetry. Furthermore, COVID-19 patients requiring hospitalization often have huge oxygenation deficits, requiring very high oxygen flows that are difficult to maintain on awake patients who do not tolerate staying in one position and may inadvertently knock off their oxygen masks. Awake patients who are very ill with COVID therefore in some respects require a higher level of care than those on mechanical ventilation.

Maneuvers that can safely improve oxygenation without the need for additional resources are thus of immense value during a surge of COVID-19 patients. Our experience suggests that the use of rotating or proning is a valuable tool in improving oxygenation and decreasing respiratory effort in many patients with moderate or severe COVID-19. Proning is simple (many patients can rotate or prone themselves, without assistance, is without cost, and utilizes no additional personnel or departmental resources. Some patients, when attempting to prone, benefit from the strategic placement of blankets or pillows.

Any COVID-19 patient with respiratory embarrassment severe enough to be admitted to the hospital should be considered for rotation and proning. Care must be taken to not disrupt the flow of oxygen during patient rotation, but we recommend proning regardless of oxygenation modality. Typical protocols include 30–120 minutes in prone position, followed by 30–120 minutes in left lateral decubitus, right lateral decubitus, and upright sitting position. Positioning is guided by patient wishes—salutary effects are generally noticed within 5–10 minutes in a new position; do not maintain a position that does not improve the patient’s breathing and comfort. Healthcare providers that may be otherwise less active during the pandemic, such as physical medicine clinicians, may be mobilized to do “proning rounds” to great effect.

In conclusion, our series of patients with moderate to severe hypoxemia related to COVID-19 lung disease demonstrated an improvement in their SpO₂ after being placed in prone position. Until further studies indicate alternative oxygenation strategies or specific treatments that address the underlying hypoxic insult, we recommend early and frequent use of patient proning, with the hope that it will delay or prevent intubation.

It is critical to re-emphasize that patients with COVID-19 may desaturate precipitously and dangerously when disconnected from their oxygen source; patients with high oxygen requirements who are managed with alternatives to mechanical ventilation require vigilant monitoring and frequent, careful reassessment.

References
CONTAINING COVID-19 IN THE EMERGENCY DEPARTMENT: THE ROLE OF IMPROVED CASE DETECTION AND SEGREGATION OF SUSPECT CASES

Liang E. Wee, MPH1,2, Tzay-Ping Fua, MRCSEd (A&E)3, Ying Y. Chua, MRCP2, Andrew F. W. Ho, MMSc3, Xiang Y. J. Sim, MRCP2,4, Edwin P. Conceicao, BSc. (Nursing)4, Indumathi Venkatachalam, MPH2,4, Kenneth B.-K. Tan, MCEM3, and Ban H. Tan, MRCP2

ABSTRACT

OBJECTIVES: Patients with COVID-19 may present with respiratory syndromes indistinguishable from common viruses. This poses a challenge for early detection during triage in the emergency department (ED). Over a 3-month period, our ED aimed to minimize nosocomial transmission by using broader suspect case criteria for better detection and using appropriate personal protective equipment (PPE) for health care workers (HCWs).

METHODS: All ED admissions with respiratory syndromes over a 3-month period were tested for COVID-19. The sensitivity and specificity of screening criteria in detecting COVID-19 were assessed. A risk-stratified approach was adopted for PPE usage in the ED, based on high-risk “fever areas” and lower-risk zones. When a case of COVID-19 was confirmed, surveillance was conducted for potentially exposed patients and HCWs.

RESULTS: A total of 1,841 cases presenting with respiratory syndromes required admission over the study period. Among these, 70 cases of COVID-19 were subsequently confirmed. The majority (84.2%, 59/70) were detected at ED triage because they fulfilled suspect case criteria. Of these, 34 met the official screening criteria; an additional 25 were detected by the broader internal screening criteria. Over the 12-week period, the cumulative sensitivity of internal screening criteria was 84.3% (95% confidence interval [CI] = 73.6% to 91.9%), whereas the sensitivity of the official screening criteria was 48.6% (95% CI = 36.4% to 60.8%). Given the broadened internal criteria, the preexisting ED “fever area” was insufficient and had to be expanded. However, there were no cases of nosocomial transmission from intra-ED exposure, despite extensive surveillance.

CONCLUSION: Frontline physicians need to be given leeway to decide on the disposition of cases based on clinical suspicion during an ongoing outbreak of COVID-19. If a broader criterion is used at ED triage, ED facilities and isolation facilities need to be readied to accommodate a surge of suspect cases. Usage of appropriate PPE is essential in minimizing nosocomial transmission.

In late December 2019, a novel pathogen, SARS-CoV-2, first emerged in mainland China. Since then, an outbreak of acute respiratory disease (ARD) caused by this novel coronavirus (COVID-19) has evolved into a global pandemic, with cases of local transmission being reported soon after the detection of imported cases in affected countries. While heightened vigilance is necessary to prevent sustained...
transmission in new locations, challenges exist. To contain an outbreak of COVID-19, the role of a large hospital in this effort pivots on its frontline doctors, namely, emergency department (ED) physicians, assisted by infectious diseases (ID) specialists, to ensure that potential cases of COVID-19 are identified early and isolated upon arrival. This is challenging as individuals with COVID-19 may be relatively asymptomatic in the early stages of illness and may present with atypical manifestations; early stages of COVID-19 may thus be indistinguishable from ARDs caused by common respiratory viruses. Case definitions are hence essential in guiding the ED physician with regard to triage of potential suspect cases. The role of the ED as a line of defense in ensuring that suspected COVID-19 cases are contained and isolated from arrival is crucial, given the significance of nosocomial transmission.

In Singapore, a globalized Asian city-state with close travel links to mainland China, the risk of imported cases was recognized early. The local Ministry of Health (MOH) published initial suspect case criteria on January 2, 2020, that, in line with the initial case criteria published by the World Health Organization (WHO), focused on patients with pneumonia and a recent travel history to Wuhan, based on what was known about the outbreak at that time. Subsequently, suspect case criteria from both the WHO and our MOH changed to keep up with the evolution of the COVID-19 pandemic. The first imported case of COVID-19 in Singapore was reported in end-January 2020, followed by the first documented case of local transmission in early February 2020. As of end-February 2020, less than one-third of cases in Singapore were imported, and the rest were locally transmitted. Previous analyses have suggested that the surveillance system in Singapore is robust, with a high sensitivity for case detection. Here, we report our institution’s ED experience during the initial outbreak of COVID-19 over a 3-month period that saw a gradual shift from imported cases to locally transmitted ones, followed by a second wave of imported cases with onward local transmission. In line with the national strategy of containment, our ED aimed to minimize the risk of nosocomial transmission by utilizing a broader set of suspect case criteria to screen all attendees with respiratory symptoms for better detection; enforcing strict segregation of patients with respiratory syndromes and/or fever from the rest of the ED, with infrastructural modifications to accommodate the surge of patients and improve infection control; and adopting a risk-stratified approach to the usage of personal protective equipment (PPE). The impact of this strategy in case detection, resource utilization and infection control was assessed. Uniquely, the sensitivity and specificity of our suspect case criteria could be assessed, given that all admitted patients with respiratory symptoms were tested for COVID-19.

METHODS

Institutional Setting

Our institution, Singapore General Hospital is the largest public tertiary hospital in Singapore, with 1,785 beds. Our institution accounts for about a quarter of the total acute hospital beds in the public sector and about one-fifth of acute beds nationwide. Our institution’s isolation ward (IW) was used to nurse confirmed and suspected cases of COVID-19. From February 5, 2020, after our institution detected the first case of local transmission, all patients presenting with respiratory symptoms without an obvious history of contact with travelers or known COVID-19 cases were admitted to designated inpatient cohorted wards (respiratory surveillance wards [RSWs]). In the RSW, staff wore N95 masks and eye protection and SARS-CoV-2 was tested for prior to transfer or discharge via reverse transcriptase polymerase chain reaction (rtPCR) of respiratory samples. The minimum criteria for transfer/discharge was two negative COVID-19 samples taken 24 hours apart.

Broadening Case Definitions for Suspected COVID-19 Cases at ED Triage

On January 2, 2020, our local MOH issued a set of suspect case criteria for COVID-19, mandating admission and isolation for every person meeting the case criteria. The criteria closely matched those released by the WHO and U.S. Centers for Disease Control and Prevention. In this study, the WHO’s suspect case criteria for COVID-19, with close reference to our own MOH’s suspect case criteria to provide additional local context for relevance, was defined as the “official case criteria.” From February 27, 2020, the WHO case criteria were broadened to include patients presenting with ARD with no alternative etiology and a history of residence in any country reporting ongoing local transmission. Thereafter our own MOH’s suspect case criteria were utilized as the official case criteria, because the WHO case criteria would have
automatically included all patients with ARD present-
ing to our institution, given local transmission in Sin-
gapore. In our institution, however, apart from the
official case criteria, a broadened set of internal screen-
ing criteria was used to improve case detection. Both
the official case criteria and the internal screening cri-
teria were used at ED triage to decide on isolation of
patients presenting with either pneumonia or ARD of
any severity, based on various epidemiologic risk fac-
tors; patients who had the relevant risk factors were
considered as meeting case criteria. The changes in
both our internal screening criteria and the official
case criteria over a 3-month period are reflected in
Data Supplement S1, Table S1, available as supporting
information in the online version of this paper, which
is available at http://onlinelibrary.wiley.com/doi/
10.1111/acem.13984/full (for pneumonia) and Data
Supplement S1, Table S2 (for ARD of any severity).
Both sets of criteria underwent multiple changes dur-
ing this period, reflecting the fast-evolving situation.

ED Workflow During COVID-19 Outbreak:
Efforts to Protect Health Care Workers
As patients with COVID-19 may present with respira-
tory syndromes indistinguishable from those caused by
common respiratory viruses, a risk-stratified approach
was adopted for PPE usage in the ED. During the
study duration, all patients presenting to the ED with
respiratory syndromes or undifferentiated fever were
deemed to be at higher risk and were managed in seg-
grated areas of the ED (“fever areas”), where health
care workers (HCWs) used full PPE comprising N95
masks, eye protection (face shields), and disposable
gown and gloves, similar to the disposition and man-
agement of suspected COVID-19 cases. Outside of the
fever areas where the bulk of patients were classified
as lower risk, ED staff wore N95 masks for extended
periods in areas where they might potentially come
into contact with patients with respiratory symptoms,
such as at all triage areas, corridors of fever areas, in
the observation ward where patients with respiratory
symptoms might be held while awaiting admission,
and in the critical care area where patients with respi-
atory symptoms might need urgent resuscitation. In
other low-risk areas of the ED, usage of a surgical
mask was made the mandatory minimum standard.
Mask usage was made mandatory throughout the ED
as asymptomatic and presymptomatic patients might
also potentially transmit COVID-19.12 If masks were
used for extended duration, they could only be used
up to 4 to 6 hours, with N95 masks being placed in a
new and clean zip-lock bag each time the mask was
removed. Masks would be changed if soiled. This was
in line with local studies that did not detect contami-
nation of PPE despite extended use of N95 masks and
goggles with strict adherence to environmental and
hand hygiene.13 Full PPE was used for any aerosol-gen-
erating procedures (e.g., intubation) throughout the
whole ED, as this was deemed to be a high-risk proce-
dure.

Criteria for Inclusion and Exclusion of
COVID-19
All patients admitted via the ED and who presented
with respiratory symptoms at ED triage over a 3-
month period from January 1, 2020, to April 1, 2020,
were tested for COVID-19 via rtPCR of respiratory
samples for SARS-CoV-2. A confirmed case of
COVID-19 was defined as a positive test for SARS-
CoV-2 via rtPCR testing,10,11 while patients were con-
sidered negative for COVID-19 at the point of testing
if they had two negative COVID-19 samples taken
24 hours apart.9 To evaluate case detection at ED
triage, the sensitivity and specificity of our internal
screening criteria in detecting COVID-19 were calcu-
lated and compared to the official case criteria. When-
ever a case of COVID-19 was confirmed, activity
mapping and contact tracing were conducted retrospec-
tively by our hospital’s epidemiology team to deter-
mine whether any patients or staff in the ED had
been exposed, and surveillance was conducted for
potentially exposed patients and staff.

Ethics Approval
Because this was a descriptive study based on surveil-
lance data collected by the hospital’s Department of
Infection Prevention and Epidemiology and only aggre-
gate data were collected without patient identifiers,
ethics approval was not required under our hospital’s
institutional review board guidelines.

RESULTS
Case Detection for COVID-19 Using Broader
Internal Screening Criteria
From January 1, 2020, to April 1, 2020, over a 3-
month period, a total of 1,841 cases presented to our
ED with respiratory syndromes requiring admission or
fulfilling suspect criteria for COVID-19, and all were
tested for COVID-19. Among these, 70 cases tested
positive for COVID-19. Over the study period, there was a clear shift from imported cases to locally transmitted ones, followed by a successive wave of imported cases (Figure 1A). Of note, our institution picked up the first locally transmitted case in Singapore, which formed part of a cluster (Case 4, linked to Case 5, Figure 1B); these cases did not fulfill official case criteria but were detected by our internal screening criteria. The majority of COVID-19 cases (84.2%, 59/70) were detected at ED triage because they fulfilled suspect case criteria. Of these, 34 met the official screening criteria, an additional 25 were additionally picked up by the broader internal screening criteria, and 11 cases did not fulfill either set of criteria. In total, 695 patients met internal screening criteria, and 218 met official case criteria for suspected COVID-19. Over the 3-month period, the cumulative sensitivity of internal screening criteria in detecting COVID-19 cases for isolation at ED triage was 84.3% (95% confidence interval [CI] = 73.6% to 91.9%), with a specificity of 64.8% (95% CI = 62.5% to 67.0%), whereas the sensitivity of the official screening criteria was 48.6% (95% CI = 36.4% to 60.8%), with a specificity of 89.6% (95% CI = 88.1% to 91.0%; Table 1).

Our internal screening criteria picked up the additional 25 cases through maintaining higher vigilance; the details are provided in Table 2. Of the 11 cases that did not fulfill either official or internal screening
criteria, all were locally transmitted cases that did not have history of travel and did not have links with a confirmed COVID-19 case or cluster.

**Infrastructural Modifications in the ED During an Ongoing COVID-19 Outbreak**

To accommodate the large number of patients presenting to the ED with respiratory symptoms/undiifferentiated fever during an ongoing COVID-19 outbreak, the designated “fever area” was expanded by taking over the adjacent ambulatory surgery center and converting it into an expanded fever area, in which staff wore full PPE. Within the fever area, partitions 2 meters high were set up between trolleys to construct temporary cubicles and trolleys were spaced 2 meters apart, to reduce the risk of droplet spread; partitions and trolleys were wiped down after each patient. Traffic flows for patients managed in the fever areas were separated from the rest of the ED. Subsequently, to accommodate the rising number of patients, a sheltered carpark off-site was modified into an additional fever screening area for well patients with upper respiratory tract symptoms.

**Increased Inpatient Resource Utilization Arising From Broader Screening Criteria**

Broadening the screening criteria tripled the number of suspect cases, resulting in increased pressure on...
IW beds. If only cases who fulfilled official case criteria had been admitted to isolation, the bed occupancy rate of our IW would have remained at less than ~50% during the first 2 months of the outbreak, whereas by using our internal screening criteria, bed occupancy of the IW reached almost 100% on January 28 and again on the February 6. To accommodate this spillover and the large number of suspect COVID-19 cases requiring admission to IW, our institution converted 40 single rooms scattered throughout the hospital with attached toilets into extensions of the IW. The single rooms were used for lower-risk suspect cases, whereas negative-pressure rooms were used for patients requiring aerosol-generating procedures and for confirmed COVID-19 cases.

Impact of Improved Case Detection at ED Triage on Minimizing Nosocomial Spread

Most cases fulfilled suspect case criteria and were managed in designated fever areas within ED and admitted to the IW. Although 11 unlinked cases of COVID-19 were not detected at ED triage, because all patients presenting to the ED with respiratory syndromes or undifferentiated fever were managed similar to suspect COVID-19 cases and triaged into fever areas where full PPE was used, and patients were spaced farther apart, intra-ED exposure was minimized. Within the ED, only one case of COVID-19 was managed outside of a fever area, because the initial respiratory symptoms reported at triage were fairly mild. Because the patient stayed ≥ 12 hours in the ED, a total of 43 patients were deemed to have potentially significant unprotected exposure and required quarantine. A total of 20 staff in the ED were potentially exposed; however, because all ED staff used N95 masks, none of the staff were deemed to have significant unprotected exposure requiring quarantine. All patients and staff were followed up for 14 days post-exposure; none developed symptoms compatible with COVID-19. Swabs from the patient’s room, call bell, and trolley in the ED were tested for SARS-CoV-2 (taken ≥ 24 hours post-exposure) and were all negative. The majority of the 11 cases not fulfilling suspect case criteria were triaged into the RSW where enhanced PPE was used, hence minimizing inpatient exposure. Only one case of undifferentiated fever was initially triaged to the general ward; the patient remained in the general ward for 18 hours before being shifted to an RSW where the diagnosis of COVID-19 was confirmed. To date, despite extensive surveillance and monitoring of potentially exposed staff and patients, no documented cases of nosocomial transmission from intra-ED exposure have been identified. Given that staff in the ED managed these unlinked cases with full PPE and ED staff wore N95 masks for extended periods in areas where they might potentially come into contact with patients with respiratory symptoms, the number of staff requiring quarantine as a result of unprotected exposure was minimal and the ED was kept fully operational.

DISCUSSION

Case definitions are important in the early stages of an infectious diseases outbreak, by helping to ensure appropriate triage and isolation and by rationing testing resources, which may be scarce especially in outbreaks caused by a novel pathogen. Given close travel links between Singapore and China, with almost 300,000 visitor arrivals a day,14 our country was at higher risk of imported cases15 and our institution had maintained vigilance for potential cases of COVID-19 since the start of January 2020. While using a case definition based on travel history and contact with confirmed cases of COVID-19 is possible in the early phases of an outbreak when cases are mostly imported, during ongoing community transmission, distinguishing between cases of COVID-19 and ordinary pneumonia becomes difficult. Our institution had previously experienced an outbreak of severe acute respiratory syndrome (SARS), in 2003; and appropriate triaging, cohorting, and selective isolation was found to be an effective and practical model of intervention in cohorts exposed to a SARS outbreak.16 However, to achieve appropriate triaging, case definitions are crucial. During SARS, early studies showed low sensitivity and potential undertriage at ED, when the WHO case criteria for SARS were used for clinical assessment, in areas with established local transmission.17–19 Furthermore, distinguishing COVID-19 posed its own set of challenges. In SARS, fever was a predominant feature on initial presentation, and few patients were asymptomatic,20 allowing for fever and severity of respiratory disease to form part of case criteria.17–21 However, it appears that fever, though common, may not occur in all patients with COVID-19 on initial presentation, and individuals may not present with severe respiratory disease. In a large study of more than 1000 patients with COVID-19, fever occurred in only 43.8% of patients on presentation...
but developed in 87.9% following hospitalization.\textsuperscript{22} Thus, admission strategies adopted during SARS outbreaks that focused on isolating all patients with febrile pneumonia\textsuperscript{23} may not be feasible in detecting all patients with COVID-19 ARD on initial presentation to hospital. Admission strategies adopted for MERS that focused on isolating and cohorting patients with significant travel history and a compatible clinical syndrome\textsuperscript{24} may not be suitable during the later phases of an outbreak of COVID-19 with ongoing local transmission. While official case criteria were fairly sensitive, more than half of confirmed COVID-19 cases did not fulfill the official case definitions. The effectiveness of case definitions in practice depends on the context and the user.\textsuperscript{25} The importance of early detection of COVID-19 at ED triage cannot be understated, given that a single case of COVID-19 managed without appropriate PPE can result in the quarantine of large numbers of HCWs and disrupt hospital operations during an ongoing outbreak.\textsuperscript{26}

As such, to improve case detection at ED triage, our institution allowed physicians on the ground more leeway in determining who to test/isolate, via the use of a broader set of internal screening criteria. This improved sensitivity and the pickup rate for COVID-19 at ED triage. Usage of a looser set of criteria and flexibility afforded to frontline staff likely enabled our institution to detect several of the country’s “firsts,” such as the first imported case and the first case of local transmission. While official case definitions are important to appropriately target testing, there must be room for flexibility based on the latest available information and clinical suspicion. In China, laboratory tests for COVID-19 were originally requested according to the case definitions, which included an epidemiologic link to Hubei or other confirmed cases; subsequently, a more liberal clinical testing regimen allowed clinicians to test at lower thresholds of suspicion.\textsuperscript{27} Similarly, the initial cases of COVID-19 in other countries were only detected because of a higher index of suspicion by frontline clinicians, because they did not meet official case criteria.\textsuperscript{28,29} However, increased sensitivity comes at the expense of specificity, and a fine balance needs to be struck, because an overly broad definition will strain hospital resources, especially in institutions with limited isolation beds. Infrastructural modifications are also necessary to accommodate the increased number of suspected COVID-19 cases identified using a broader set of internal screening criteria. By partitioning the ED into higher risk and lower-risk areas, and setting up different traffic flows, the risk of cross-infection can be mitigated.\textsuperscript{30} Finally, while most COVID-19 cases presenting to our ED over a 12-week period were successfully isolated, one-fifth of cases did not have any known epidemiologic links that allowed detection based on case criteria alone. This suggests that, as community penetration deepens, likely through mildly symptomatic cases, it will be more and more difficult to use epidemiologic criteria to pick out potential cases at triage in the ED. Within the ED, managing all patients presenting with respiratory syndromes in designated fever areas and the creation of inpatient RSWs, with upgraded PPE and better bed spacing of patients, would appear to be the last line of defense against patient-to-staff transmission during an outbreak of COVID-19 with ongoing community transmission. This is essential, given that even a single patient in an overcrowded ED can ignite a nosocomial outbreak of respiratory disease caused by a novel pathogen, with devastating consequences.\textsuperscript{31}

**LIMITATIONS**

The limitations of our study are as follows. The diagnostic yield of PCR testing for SARS-CoV-2 would likely be dependent on the quality and type of respiratory tract sample, and the sensitivity and specificity of the assay are unknown at present. COVID-19 cases may thus have been missed due to sampling issues with the tests available at the time. Our hospital was not the designated hospital for receiving COVID-19 suspect cases from primary care clinics and hence the majority of COVID-19 patients in Singapore were not managed here. Hence, our experience would be more relevant to the majority of health care institutions that are not designated centers for the management of COVID-19 cases, especially in countries and regions in the earlier stages of an outbreak.

**CONCLUSIONS**

In conclusion, during an ongoing outbreak of COVID-19 with progression from imported cases to locally transmitted ones, our ED managed to detect and appropriately isolate the majority of confirmed COVID-19 cases upon triage at the ED, over a 3-month period. Our institution’s internal screening criteria had higher sensitivity in deciding on appropriate ED triage of suspect cases, compared to the published
official case criteria. However, if a broader criterion is used for the triage of suspected COVID-19 cases, ED facilities and isolation facilities need to be readied to accommodate a larger number of cases. Frontline physicians need to be given leeway to decide on the disposition of cases based on clinical suspicion and be afforded appropriate PPE. Continued vigilance at the frontline as part of an overall containment strategy may reduce the likelihood of nosocomial transmission and buy precious time for hospital preparedness during an outbreak of a novel pathogen.

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.13984/full
Data Supplement S1. Supplemental material.
The Impact of Selecting Specific Cohorts for Benchmarking and Interpretation of Emergency Department Patient Satisfaction Scores

Gregory M. Archual, MBA, Ashish R. Panchal, MD, PhD, Mark G. Angelos, MD, and David P. Way, MEd

ABSTRACT

Objectives: Emergency departments (EDs) patient satisfaction metrics are highly valued by hospitals, health systems, and payers, yet these metrics are challenging to analyze and interpret. Accurate interpretation involves selection of the most appropriate peer group for benchmark comparisons. We hypothesized that the selection of different benchmark peer groups would yield different interpretations of Press Ganey (PG) patient satisfaction scores.

Methods: Emergency department PG summary ratings of “doctors section” and “likelihood-to-recommend” raw scores and corresponding percentiles were derived for three benchmark peer groups from three academic years (2016, 2017, and 2018). The three benchmarks are: 1) the PG Large database; 2) the PG University HealthSystem Consortium (UHC) database; and 3) the Academy of Administrators in Academic Emergency Medicine (AAAEM) database, which is composed only of EDs from academic health centers with emergency medicine residency training programs. Raw scores were converted to percentile ranks for each distribution and then compared using Welch’s ANOVA and Games-Howell pairwise comparisons.

Results: For both patient satisfaction raw scores evaluated, the AAAEM database was noted to have significantly higher percentile ranks when compared to the PG Large and PG UHC databases. These results were consistent for all three time frames assessed.

Conclusions: Benchmarking with different peer groups provides different results, with similar patient satisfaction raw scores resulting in higher percentile ranks using the AAAEM database compared to the two PG databases. The AAAEM database should be considered the most appropriate peer group for benchmarking academic EDs.

Patient satisfaction is considered a critical evaluative metric of acute care delivery, widely used by hospitals and EDs throughout the United States. Competition in the health care market has led to a greater interest in patient experience surveys and their use in rewarding health systems and providers for high satisfaction score performance. However, accurately evaluating patient satisfaction scores has been a challenge...
due to the high variability in scores across facilities. Interpretation is further confounded by the benchmark cohort selected for comparison. Large academic health systems tend to be engaged in the tripartite mission of providing patient care to the community, training future health care providers and conducting medical research. The broader mission of the academic health center (AHC) places additional demands on providers, which likely impacts patient perception of their care in the emergency care setting. Accurate interpretation of patient satisfaction scores is contingent upon selecting organizations with similar missions as the appropriate cohort for benchmarking one’s organizational performance.

For emergency medicine departments who belong to large AHCs, finding a benchmarking cohort requires a source of benchmarking data composed of emergency departments (EDs) that most accurately resemble those within large AHCs. The logical alternative benchmarking cohort to the Press Ganey (PG) Large database should be the PG University HealthSystem Consortium (PG UHC) database, since the UHC is an alliance of AHCs in the United States. Another option is the database created by the Academy of Administrators in Academic Emergency Medicine (AAAEM), who have been producing a benchmarking survey since 2010. The AAAEM membership is composed of EDs housed in AHCs (as previously described by Reznik et al.). The data set created by the AAAEM includes information on their member’s clinical, education, and research operations, which is reflective of the large academic emergency medicine department’s tripartite missions.

The purpose of this project was to compare three databases of ED patient satisfaction data and evaluate them as peer groups for benchmarking: PG Large database, PG UHC database, and the AAAEM database. The quarterly PG Large database report typically contains patient satisfaction scores derived from over 300,000 patients from over 1,000 facilities across the United States. These institutions in the database represent hospitals and medical centers of all types and sizes. The quarterly PG UHC database contains patient satisfaction scores from approximately 50,000 patients from 160 institutions representing facilities closely associated with universities (AHCs and their primary affiliates). The annual AAAEM database is derived from patient satisfaction scores from an estimated 50,400 patients from 42 (on average) primary academic sites of departments or divisions of emergency medicine with university-based residency training programs.

Emergency department PG summary ratings of “doctor section” and “likelihood to recommend” raw scores and corresponding percentiles were derived for three benchmark peer groups from three academic years (2016, 2017, and 2018). We hypothesized that the selection of different benchmark peer groups would yield different interpretations of PG patient satisfaction scores.

**METHODS**

**Assembling Data Sets**

The data for this study were derived from three sources across three academic years (2016, 2017, and 2018): 1) the third quarter reports for the corresponding academic years (January–March) entitled PG ER/ED means and ranks reports from the PG Large database (Press Ganey Associates Inc., South Bend, IN); 2) similar third-quarter reports derived from the PG UHC database (Press Ganey Associates Inc.); and 3) reports from the databases created annually (2016, 2017, and 2018) by members of the AAAEM. From each of the first two reports, two data tables were used to create the datasets for analysis. The first table was drawn from the means and ranks of facility and doctors section scores and the second table from the question on likelihood of recommending (“What is your likelihood of recommending this ED?”). Each table reported the mean score derived from national norms at each corresponding percentile rank, 1 to 99.

Academy of Administrators in Academic Emergency Medicine member institutions supply their PG scores on these two questions, which are stored on a proprietary database platform (Roundtable Analytics, Research Triangle Park, NC). We extracted the doctor section scores and likelihood-of-recommending scores for the AAAEM primary academic EDs and calculated the corresponding percentile ranks for each item using the frequencies utility in IBM-SPSS, Version 25.0. The final data sets included the three associated PG mean scores (PG Large database, PG UHC database, and AAAEM) for each point on the percentile rank (1–99) distribution. A separate data set for analysis was created for each item (doctor and likelihood) for each year (2016, 2017, and 2018).

**Primary Data Analysis**

The three score distributions were plotted using a line graph with the percentile rank on the abscissa and the
mean PG score on the ordinate. This provided a visual comparison of the percentile ranks at any point on the mean score distributions.

A second analysis involved Welch’s analysis of variance (ANOVA) for each PG question at each time point across the 3 years. Welch’s ANOVA is a useful alternative to the traditional one-way ANOVA when the assumptions of normality are violated. The Games-Howell pairwise comparisons were used as post hoc tests when statistically significant differences between the three distributions were observed. Cohen’s D effect sizes and corresponding confidence intervals were calculated using the resource provided by Lenhard and Lenhard. Effect size informs us as to the distance between group (distribution) means in standard deviation units.

RESULTS

There were no significant differences observed between the distributions generated by the PG Large and PG UHC databases on either question (doctor section or likelihood of recommending) at the end of quarter 1 for any of the 3 years (Figures 1–3). The distributions generated by the AAAEM database, however, were significantly different from the other two databases for all six comparisons (see Figures 1–3). The resulting pattern is similar for both PG survey items and from year to year.

The statistical difference between the distributions are demonstrated by the results of the Welch’s ANOVA and Games-Howell tests. Consistent for both questions across all 3 years, the AAAEM average scores were significantly lower than the average scores from either of the PG databases. Furthermore, the effect sizes for these differences as measured by Cohen’s D indicate large effects (see Figures 1–3).

DISCUSSION

Comparison of patient satisfaction scores from the three different databases highlights the significant differences in overall percentiles based on the choice of cohort as the comparison group for benchmarking. As an example, selecting a PG doctor section mean rating score of 82.5 is associated with a point at the 16th percentile in the PG Large database, at the 22nd percentile in the PG UHC database, and at the 54th percentile in the AAAEM database. These large differences in percentile scores across data sources
Figure 2 2017 PG patient satisfaction scores and corresponding percentile ranks comparing distributions from PG Large and UHC database reports to AAAEM member scores. Welch’s ANOVA and Games-Howell post hoc tests are shown in the tables. Doctor’s summary scores are shown on the left and likelihood-to-recommend scores on the right. AAAEM = Academy of Administrators in Academic Emergency Medicine; Db = database; PG = Press Ganey; UHC = University HealthSystem Consortium.

Figure 3 2018 PG patient satisfaction scores and corresponding percentile ranks comparing distributions from PG Large and UHC database reports to AAAEM member scores. Welch’s ANOVA and Games-Howell post hoc tests are shown in the tables. Doctor’s summary scores are shown on the left, and likelihood-to-recommend scores on the right. AAAEM = Academy of Administrators in Academic Emergency Medicine; Db = database; PG = Press Ganey; UHC = University HealthSystem Consortium.
illustrate the fundamental challenges associated with interpreting data, which has been aggregated for benchmarking purposes, and highlights the importance of understanding the characteristics of the members of each comparison cohort.

Practitioners commonly aggregate individual-level data into a value, such as a mean or median to represent a workgroup, or an institution. The health care enterprise is the summary product of numerous individual professionals who contribute to the large-scale system. However, the measurement properties related to the original instrument used to collect data at the individual level, such as evidence of validity or reliability, do not necessarily extend to the aggregated data that represents a workgroup or institution.

Furthermore, high levels of aggregation assume that the individual organizations summarized in the resulting norms are homogeneous. Overly aggregated data conceal potentially important differences between and among individual members of the aggregation group. For instance, the aggregation of PG satisfaction surveys across all the members in the Large database has the effect of hiding important differences between hospitals from different settings (urban, rural, suburban); ownership types (community, noncommunity); missions (public, private not for profit, private for profit); structure (system, independent, network); or activities (health care, research, education). Since there is a lot of heterogeneity across members of the PG Large database and their EDs in the United States, the PG Large database report should be considered overly aggregated and an inaccurate representation of EDs within AHCs.

The solution to overly aggregated data is to disaggregate the data and reaggregate it using homogeneous subgroups. Unfortunately, when dealing with PG data, disaggregation becomes a problem because of its proprietary nature. For our purposes, let us assume that the UHC represents a disaggregation and reaggregation of PG data into a meaningful subgroup who are all affiliates of a university and who carry out all or parts of an academic mission (patient care, education, and research). The question we need to ask now is: Are the hospitals in the UHC sufficiently homogeneous? The UHC is an alliance of 107 AHCs and 338 of their affiliated hospitals, representing more than 90% of the nation’s nonprofit AHCs. These include AHCs, community hospitals, integrated health delivery networks, and non–acute health care providers, a group seemingly more diverse than the AAAEM survey group.

Our analysis suggests that the UHC database is not sufficiently different with regard to homogeneity than the PG Large database. In fact, these two databases yield almost identical distributions of mean scores and associated percentile ranks. Since the UHC members vary by the degree in which they have close affiliations with medical schools, they also vary by the degree in which they emphasize the tripartite mission of the AHC: patient care, education, and research.

The AAAEM is an organization of emergency medicine department administrators at AHCs. They have been gathering the performance data of their member’s EDs, including PG data for almost 10 years. We found the AAAEM’s primary AHCs to be sufficiently homogenous in their tripartite missions to justify aggregation across members, providing us with the best picture of patient satisfaction with emergency medicine physicians practicing at AHCs, which include residency training programs. The reaggregation of PG patient satisfaction data across AAAEM primary academic affiliates differs significantly from the norms provided by either the PG Large or PG UHC database reports and provide an alternative perspective to aid academic EDs in interpreting their PG patient satisfaction scores.

LIMITATIONS

One of the limitations of this study involves the nature of the data itself. Because the patient experience data generated from the PG Large and UHC databases is proprietary, we do not have granular information as to the individual EDs included in these databases. This makes it more difficult to explain our findings in terms of the differences in characteristics between EDs housed within AHCs and those that are not.

We are also forced to assume, but cannot definitively explain, that the consistently observed differences between the EDs from the AAAEM database and the EDs from the PG Large and UHC databases are due to the differences in their functions as academic institutions. In other words, the AAAEM cohort may score significantly lower on patient satisfaction metrics simply because the competing missions of education and research that take place in an academic ED alters the delivery of care. Further work is needed to determine if this is a correct interpretation or whether alternative factors, such as those observed by Pines et al.2 (patient volume, teaching status, and length of stay) better explain this finding.
CONCLUSION

Patient satisfaction is an important metric for emergency medicine groups across the United States, but is not useful without a relevant benchmark for interpretation. This study demonstrates the importance of identifying the appropriate peer group to employ when interpreting patient satisfaction scores against benchmarks. Overall, there is a combination of academic emergency medicine characteristics that are different from University HealthSystem Consortium and Large database cohorts, which do not appear to be granular enough to provide a homogenous population to benchmark academic EDs. Different organizations have different missions and therefore patient satisfaction scores can only be evaluated in the context of their total mission using benchmarks derived from homogenous peer groups. Future directions should include examination of other benchmarked metrics commonly used in current at risk contracts to determine best homogeneity groups for comparison.

REFERENCES

Video Versus Direct and Augmented Direct Laryngoscopy in Pediatric Tracheal Intubations

Amy H. Kaji, MD, PhD, Carolyn Shover, MD, Jennifer Lee, MD, Lisa Yee, MD, Daniel J. Pallin, MD, MPH, Michael D. April, MD, DPhil, Jestin N. Carlson, MD, MS, Andrea Fantegrossi, MPH, and Calvin A. Brown III, MD

ABSTRACT

Objectives: With respect to first-attempt intubation success, the pediatric literature demonstrates either clinical equipoise or superiority of direct laryngoscopy (DL) when compared to video laryngoscopy (VL). Furthermore, it is unknown how VL compares to DL, when DL is “augmented” by maneuvers, such as optimal external laryngeal manipulation (OELM), upright or ramped positioning, or the use of the bougie. The objective of our study was to compare first-attempt success between VL and all DL, including “augmented DL” for pediatric intubations.

Methods: We analyzed the National Emergency Airway Registry database of intubations of patients < 16 years. Variables collected included patient demographics, body habitus, impression of airway difficulty, intubating position, reduced neck mobility, airway characteristics, device, medications, and operator characteristics, adjusted for clustering by center. Primary outcome was the difference in first-attempt success for DL and augmented DL versus VL. Secondary outcomes included adverse events. In a planned sensitivity analysis, a propensity-adjusted analysis for first-attempt success and a subgroup analysis of children < 2 years was also performed.

Results: Of 625 analyzable pediatric encounters, 294 (47.0%, 95% confidence interval [CI] = 25.1% to 69.0%) were DL; 332 (53.1%, 95% CI = 31.0% to 74.9%) were VL. Median age was 4 years (interquartile range = 1 to 10 years); 225 (36.0%, 95% CI = 30.8% to 41.2%) were < 2 years. Overall first-pass success was 79.6% (95% CI = 74.1% to 84.9%). VL first-pass success was 278/331 (84.0%) versus 219/294 for DL (74.5%), adjusted for clustering (odds ratio [OR] = 1.7, 95% CI = 1.3 to 2.5). Multivariable regression showed that VL yielded a higher odds of first-attempt success than DL augmented by OELM or use of a bougie (adjusted OR = 5.5, 95% CI = 1.7 to 18.1). Propensity-adjusted analyses supported the main results. Subgroup analysis of age < 2 years also demonstrated VL superiority (OR = 2.0, 95% CI = 1.1 to 3.3) compared with DL. Adverse events were comparable in both univariate and multivariable analysis.

Conclusions: When compared to DL, VL was associated with higher first-pass success in this pediatric population, even in the subgroup of patients < 2 years, as well as when DL was augmented. There were no differences in adverse effects between DL and VL.

From the 1Department of Emergency Medicine, Harbor–University of California Los Angeles Medical Center, Torrance, CA; the 2Department of Emergency Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA; the 3Department of Emergency Medicine, San Antonio Uniformed Services Health Education Consortium (SAUSHEC), Fort Sam Houston, TX; and the 4Department of Emergency Medicine, St. Vincent Hospital, Allegheny Health Network, Erie, PA.

Received June 18, 2019; revision received September 25, 2019; accepted October 13, 2019.

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

Author contributions: CAB, JNC, MDA, and DJP conceptualized and designed the study and critically reviewed and revised the manuscript; AF coordinated and supervised data collection and reviewed and revised the manuscript; APK carried out the initial analyses, drafted the initial manuscript, and critically reviewed the manuscript; CS, LY, and JL collected data and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Supervising Editor: Rob Reardon, MD.

Address for correspondence for reprints: Amy H. Kaji, MD, PhD; e-mail: akaji@emedharbor.edu.


ISSN 1553-2712 © 2019 by the Society for Academic Emergency Medicine doi: 10.1111/acem.13869
Video laryngoscopy (VL) is frequently used for emergency airway management. In the adult population, multiple emergency department (ED) observational studies have suggested superiority of VL over direct laryngoscopy (DL). However, the pediatric literature demonstrates either clinical equipoise or superiority of DL. In the most recent guidelines released by the Difficult Airway Society (DAS), VL is not part of the pediatric difficult airway algorithm, as no consensus emerged from the Delphi process on the use of VL. The authors of a recent Cochrane meta-analysis that included 12 randomized controlled trials and 803 children concluded that VL was associated with higher intubation failure and longer intubation times. Similarly, a separate meta-analysis of 14 randomized controlled trials also concluded that VL was associated with a higher incidence of failed airways.

Unfortunately, the composite body of evidence comparing VL to DL has been rated by the Cochrane Review as low quality. Marked heterogeneity between studies and the rarity of documented events have precluded authors of the Cochrane reviews from drawing any conclusions with regard to the association between VL use and peri-intubation adverse events. Multiple techniques for augmenting DL exist, including minimally invasive and inexpensive maneuvers such as deliberate optimal external laryngeal manipulation (OELM), intubation position (patient is seated upright or patient’s head is ramped vs. neutral head or cervical spine extension), and use of a bougie. Although there is ample adult literature supporting the use of such augmentation maneuvers, there is less supportive evidence in the pediatric literature. Rather than a “ramped” position, during the development of the DAS difficult airway algorithm, a Delphi consensus was reached for the use of a shoulder roll in children < 2 years of age, while a neutral or “sniffing” position of the head without elevation of the head, was favored for those who were older. Additionally, cricoid pressure in the child can distort or occlude the laryngeal inlet. Still, the DAS recommends that OELM “be attempted in the first intubation attempt, where the laryngeal view is impaired.” The DAS further supported the use of the bougie when there were poor laryngoscope views, although blind intubation with a bougie was rejected. Given the conflicting evidence and clinical equipoise with respect to augmentation maneuvers and VL versus DL, our study objective was to compare first-attempt success between DL to VL, including the impact of “augmented” DL, for pediatric ED intubations.

PATIENTS AND METHODS

We performed a secondary analysis of prospectively collected multicenter data from the National Emergency Airway Registry (NEAR), an international network of 25 academic and community hospitals. We obtained institutional review board approval at each site prior to data collection and analysis.

Intubating providers entered data into a secure Web-based data collection form requiring institution-specific login credentials and passwords (StudyTRAX, version 3.47.0011, ScienceTRAX). Variables collected included patient demographics, body habitus and estimated weight, preintubation hemodynamics, methods of preoxygenation, impression of airway difficulty, reduced neck mobility (e.g., presence of cervical collar), airway characteristics (e.g., mouth opening, Mallampati score), intubation position (seated upright or patient’s head is ramped e.g., with head of the bed elevated) versus neutral head or cervical spine extension), device (straight, curved, video, etc.), medications and doses, operator characteristics (postgraduate year [PGY] level, type of specialty including anesthesia, otolaryngology, pediatric emergency medicine, respiratory therapist, and nurse practitioner), first-pass intubation success or failure, and adverse events. Additionally, the operator was asked to assess “initial airway difficulty,” based upon gestalt, prior to attempted intubation. After data upload, study investigators reviewed all data, using quality assurance algorithms to identify and correct data entry errors. Each participating center had a designated site investigator ensuring entry of greater than or equal to 90% of all ED intubations. A study coordinator monitored data compliance (AF, Brigham and Women’s Hospital, Boston MA) by cross-referencing intubations reported by each site with online entries. We reported all data in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.

Selection of Participants

We included intubation encounters of patients younger than 16 years of age entered into the NEAR database from January 1, 2016, through December 31, 2018. Periods of participation vary for individual centers because facilities joined NEAR on a rolling basis. We included patients who underwent intubation using either VL or DL on the first attempt. We classified each of the laryngoscope categories included on the NEAR data form as either VL or DL. The C-MAC,
GlideScope, McGrath, King Vision, and Pentax airway scope (AWS) were all defined as VL, whereas DL was defined as a nonoptically or video-enhanced laryngoscope with either a straight or a curved blade. This included the Macintosh or Miller blade. We excluded encounters with missing data for device, position of intubation, or documentation of intubation success. We also excluded cases where a laryngoscope was not chosen as the first device (Figure 1).

Outcome Measures

The primary outcome was first attempt success for VL versus DL. This was further stratified by type of augmentation maneuver for each device category (bougie, OELM, and ramping). Additional calculations include multivariable and propensity scored-adjusted odds of first-attempt success between VL and DL including augmented DL. Secondary outcomes were adverse events, which included peri-intubation hypoxia (defined as oxygen saturation < 90%), esophageal intubation, vomiting, bradydysrhythmias, cardiac arrest, dental trauma, epistaxis, hypotension, laryngospasm, mainstem intubation, malignant hyperthermia, pneumothorax, tachydysrhythmias, tracheal tube cuff failure, medication errors, iatrogenic bleeding, and pharyngeal laceration.

Data Analysis

We exported all study data from StudyTRAX to SAS version 9.4 for statistical analysis. To account for within-site correlations, we performed a cluster analysis. We describe the binomial distributions of first-pass success, stratified by DL/VL plus bougie, DL/VL plus OELM, DL/VL plus bougie or DL/VL plus OELM, and DL/VL plus both bougie and OELM. Based on previously described predictors that affect first-pass success and adverse events,\(^6,7\) we then described the differences between the VL and DL cohorts with regard to body habitus, impression of airway difficulty, reduced neck mobility, airway characteristics (Mallampati score, mouth opening, thyromental distance), and intubator characteristics (PGY education level). To compare baseline characteristics classified in multiple levels (e.g., body habitus), we used the Rao-Scott chi-square test that accounts for clustering, and we calculated odds ratios (ORs) with 95% confidence intervals (CIs) to compare baseline characteristics coded in two levels (e.g., initial airway difficulty and neck immobility; see 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.13869/full, for further details regarding variable coding for the multivariable statistical models). In addition to these known baseline predictors for a difficult airway, we also included binary variables including intubation indication (trauma vs. medical) and intubation mechanism (rapid sequence intubation [RSI] versus no RSI) for exploratory multivariable models to determine whether VL was an independent predictor of first-pass success when compared to all DL or augmented DL.

We planned to evaluate seven exploratory multivariable models, adjusting for all of the difficult airway characteristics: 1) all VL versus all DL; 2) VL versus DL, additionally adjusting for use of bougie; 3) VL versus DL, additionally adjusting for the use of OELM; 4) VL versus DL, also adjusting for the use of the ramped intubation position; 5) VL versus DL, additionally adjusting...
for the use of both OELM plus bougie; 6) VL versus DL, additionally adjusting for the use of OELM or DL plus bougie; and 7) VL versus DL, additionally adjusting for the use of OELM plus bougie plus ramped intubation position. However, because there were no cases with documentation of DL plus OELM plus bougie plus ramped or DL plus OELM plus bougie and only three cases with documentation of DL plus bougie, there were insufficient cases for the last three maneuvers to perform a comparison in the multivariable models. Since these multivariable models were theory-driven, we included all of the known baseline predictors of first-pass success and adverse events, regardless of statistical significance of the univariate comparisons. After accounting for within-site correlations, we report adjusted ORs and their 95% CIs for the covariates. We performed diagnostic checks for outliers and assessed for multicollinearity using variance inflation factors (VIF) for each of the covariates. We considered a VIF of >5 to indicate potential multicollinearity. We assessed model-fit using the Hosmer-Lemeshow goodness-of-fit statistic. We did not adjust for multiple comparisons or assess effect modification with interaction terms, as there were limitations in the numbers with the outcomes of interest and we intended these multivariable models to be exploratory and hypothesis-generating.

Due to the potential for confounding by indication, a propensity score (for the use of VL) adjusted sensitivity analysis was also performed for the primary outcome of first-pass success. The variables that were included in the propensity score were selected specifically because these were felt to be potential predictors of a difficult airway, and these would thereby predispose the operator to select VL instead of DL. The propensity score accounted for age < 2 years, method of intubation, body habitus, trauma, Mallampati score, mouth opening, thyromental distance, whether the airway was initially suspected to be difficult, and intubator level.

Although no sample size calculation was performed prior to data analysis, a post-hoc power calculation demonstrated a power of 95% to detect a 10% difference (which would be considered to be a clinically significant difference in first-pass success rates). We assumed a first-pass success of 0.8 for DL and respective sample sizes of 294 and 331 in the DL and VL cohorts. We also performed a subgroup analysis of children < 2 years of age, as their anatomy and physiology may systematically differ from the older cohort and constitute a more homogeneous population.

RESULTS

Only 23 of the 25 sites had documented intubations to include in the pediatric data set. The range in number of intubations at the 23 sites was one to 116, with a median of 20 (interquartile range = 5 to 47).

There were 625 pediatric encounters, 331 (53.0%, 95% CI = 31.0 to 74.9) VL and 294 (47.0%, 95% CI = 25.1 to 69.0) DL. The overall median age was 4 years, with an interquartile range of 1 to 10 years, and 225 (35.9%) of the patients were < 2 years of age. RSI was utilized for 452 encounters (72.2%, 95% CI = 68.3% to 76.1%; Table 1). With respect to baseline demographics, the DL cohort was younger with a median difference of 1.5 years. Baseline difficult airway characteristics differed between the two cohorts for body habitus, Mallampati classification, thyromental distance, and intubator level (Table 1). Additionally, there was a lower proportion of PGY-1 to 3 intubators in the DL group, when compared to the VL cohort, suggesting that more experienced intubators used DL. In the overall cohort (including both VL and DL first attempts), there were 21 (3.4%) that incorporated the ramped position, 186 (29.8) that either added OELM or used the bougie, 141 (22.6%) that added only OELM, 51 (8.2%) that added only the bougie, and six (1.0%) that simultaneously used the OELM with the bougie. There were no cases with DL that featured the simultaneous use of either OELM and bougie or combined ramping, OELM, and bougie (Table 1). Overall first-pass success occurred in 498 attempts (79.6%, 74.2% to 84.9%) and overall adverse events occurred in 89 cases (14.3%, 10.3% to 18.3%; Table 2). First-pass success for all VL was 279/331 (84.0%, 78.0% to 91.3%) versus 19/294 (74.5%, 70.0% to 79.9%) for all DL, for a crude OR, adjusted for clustering within center of 1.8 (95% CI = 1.2 to 2.7). Additionally, VL had better first-pass success when examining only the subgroup (n = 411) where no augmentation maneuvers were used (OR = 1.6, 95% CI = 1.0 to 2.6) and when examining the subgroup (n = 214) where any augmentation maneuver was used (OR = 2.4, 95% CI = 1.2 to 4.7). Table 3 describes the first-pass success, stratified by each of the augmentation maneuvers. Table 4 shows both the unadjusted and the multivariable adjusted OR for first attempt success stratified by our defined exploratory models. Our first model comparing all VL to all DL demonstrated superiority of VL for first-pass success (adjusted OR = 1.8, 95%
### Table 1
Overall Descriptive Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Total population*</th>
<th>VL ($n = 331$, 53.0%, 95% CI = 31.0%–74.9%)</th>
<th>DL ($n = 294$, 47.0%, 95% CI = 25.1%–69.0%)</th>
<th>Adjusted for clustering within sites, p-values†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex</td>
<td>350 (56, 49.5–62.6)</td>
<td>186 (56.2, 48.4–64.0)</td>
<td>130 (44.2, 35.1–53.4)</td>
<td>OR = 1.0, 95% CI = 0.7–1.4, p = 0.9</td>
</tr>
<tr>
<td>Age (years)</td>
<td>4 (1–10)</td>
<td>5 (2–12)</td>
<td>2 (0.8–7)</td>
<td>Median diff = 1.4, 95% CI = 0.8–2.0, p &lt; 0.0001</td>
</tr>
<tr>
<td>Age &lt; 2 years</td>
<td>225 (36.0, 30.8–41.2)</td>
<td>108 (32.5, 23.5–41.5)</td>
<td>117 (39.8, 34.7–44.9)</td>
<td>OR = 0.7, 95% CI = 0.5–1.0, p = 0.06</td>
</tr>
<tr>
<td>Oxygen saturation at start of intubation (%)</td>
<td>100 (99–100)</td>
<td>100 (98–100)</td>
<td>100 (98–100)</td>
<td>Median diff = 0, 95% CI = 0–0, p = 0.1</td>
</tr>
<tr>
<td>Intubation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSI</td>
<td>452 (72.2, 68.4–76.3)</td>
<td>241 (72.6, 67.7–77.5)</td>
<td>211 (71.8, 66.4–77.1)</td>
<td>Rao-Scott p = 0.003</td>
</tr>
<tr>
<td>Sedation only</td>
<td>13 (2.1, 0.7–3.4)</td>
<td>8 (2.4, 0.3–4.5)</td>
<td>5 (1.7, 0–3.7)</td>
<td></td>
</tr>
<tr>
<td>Paralysis only</td>
<td>35 (5.6, 0–11.9)</td>
<td>26 (7.8, 0–15.7)</td>
<td>9 (3.1, &lt;0–7.1)</td>
<td></td>
</tr>
<tr>
<td>Topical anesthesia</td>
<td>1 (0.2, 0–0.4)</td>
<td>1 (0.3, 0–0.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topical anesthesia with sedation</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No medications</td>
<td>125 (20.0, 16.8–23.1)</td>
<td>57 (17.2, 11.8–22.5)</td>
<td>68 (23.1, 19.8–26.5)</td>
<td></td>
</tr>
<tr>
<td>Trauma mechanism</td>
<td>217 (34.7, 22.6–46.8)</td>
<td>115 (34.6, 21.5–47.8)</td>
<td>102 (34.7, 21.2–48.1)</td>
<td>OR = 1.0, 95% CI = 0.7–1.4, p = 1.0</td>
</tr>
<tr>
<td>Body habitus</td>
<td>Out of 486 (missing 139)</td>
<td>Out of 272 (missing 59)</td>
<td>Out of 214 (missing 80)</td>
<td>Rao-Scott p = 0.01</td>
</tr>
<tr>
<td>Very thin</td>
<td>17 (3.5, 1.5–5.5)</td>
<td>10 (3.7, 1.4–5.9)</td>
<td>7 (3.3, 0.8–5.8)</td>
<td></td>
</tr>
<tr>
<td>Thin</td>
<td>113 (23.2, 16.6–29.8)</td>
<td>76 (27.8, 22.5–33.2)</td>
<td>37 (17.3, 9.5–25.1)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>311 (63.9, 55.1–72.6)</td>
<td>157 (57.5, 51.6–63.5)</td>
<td>154 (72.0, 60.0–83.9)</td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td>43 (8.8, 5.4–12.2)</td>
<td>29 (10.6, 5.9–15.4)</td>
<td>14 (6.5, 2.8–10.3)</td>
<td></td>
</tr>
<tr>
<td>Morbidly obese</td>
<td>3 (0.6, 0–1.3)</td>
<td>1 (0.4, 0–1.1)</td>
<td>2 (0.9, 0–2.4)</td>
<td></td>
</tr>
<tr>
<td>Initial airway difficulty</td>
<td>160 (25.6, 19.2–31.9)</td>
<td>95 (28.6, 19.6–37.6)</td>
<td>65 (22.1, 16.8–27.4)</td>
<td>OR = 1.4, 95% CI = 1.0–2.0, p = 0.06</td>
</tr>
<tr>
<td>Neck immobility</td>
<td>181 (28.9, 18.3–39.6)</td>
<td>101 (30.4, 17.6–43.2)</td>
<td>80 (27.2, 15.6–38.8)</td>
<td>OR = 1.1, 95% CI = 0.8–1.7, p = 0.4</td>
</tr>
<tr>
<td>Mallampati score (No. missing)</td>
<td>Out of 624</td>
<td>Out of 330 (missing 1)</td>
<td>Rao-Scott p = 0.0004</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>108 (17.3, 9.5–25.0)</td>
<td>73 (22.1, 14.3–29.8)</td>
<td>35 (11.9, 5.9–17.9)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>52 (8.3, 3.7–12.9)</td>
<td>35 (10.6, 5.0–16.1)</td>
<td>17 (5.8, 0–11.2)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16 (2.5, 0.5–4.6)</td>
<td>10 (3.0, 0.6–4.3)</td>
<td>6 (2.0, 0–5.9)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>13 (2.1, 1.1–3.0)</td>
<td>8 (2.4, 0–6.4)</td>
<td>5 (1.7, 0–3.1)</td>
<td></td>
</tr>
<tr>
<td>Not assessed</td>
<td>436 (68.9, 59.1–80.5)</td>
<td>205 (61.9, 54.9–69.0)</td>
<td>231 (78.6, 66.8–90.4)</td>
<td></td>
</tr>
<tr>
<td>Mouth opening</td>
<td></td>
<td></td>
<td></td>
<td>Rao-Scott p = 0.1</td>
</tr>
<tr>
<td>Normal</td>
<td>190 (30.4, 22.5–38.2)</td>
<td>99 (29.8, 23.5–36.2)</td>
<td>91 (31.0, 19.4–42.5)</td>
<td></td>
</tr>
<tr>
<td>Reduced</td>
<td>72 (11.5, 8.4–14.6)</td>
<td>46 (13.9, 9.2–18.6)</td>
<td>26 (8.8, 6.4–11.2)</td>
<td></td>
</tr>
<tr>
<td>Not assessed</td>
<td>364 (58.1, 52.2–64.0)</td>
<td>187 (56.3, &lt;51.1–61.6)</td>
<td>177 (60.2, 50.1–70.3)</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
Table 1. (continued)

<table>
<thead>
<tr>
<th>Thyromental distance</th>
<th>Total population*</th>
<th>VL (n = 331 53.0%, 95% CI = 31.0%–74.9%)</th>
<th>DL (n = 294, 47.0%, 95% CI = 25.1%–69.0%)</th>
<th>Adjusted for clustering within sites, p-values†</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 finger</td>
<td>17 (2.7, 0.9–4.5)</td>
<td>12 (3.6, 1.0–6.2)</td>
<td>5 (1.7, 0–3.7)</td>
<td></td>
</tr>
<tr>
<td>2 fingers</td>
<td>69 (11.0, 6.0–16.1)</td>
<td>43 (13.0, 9.4–16.5)</td>
<td>26 (8.8, 1.1–16.6)</td>
<td></td>
</tr>
<tr>
<td>3 fingers</td>
<td>57 (9.1, 4.4–13.8)</td>
<td>37 (11.1, 5.6–16.7)</td>
<td>20 (6.8, 1.4–12.2)</td>
<td></td>
</tr>
<tr>
<td>4 fingers</td>
<td>45 (1.5, 0–3.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not assessed</td>
<td>477 (76.2, 66.0–86.4)</td>
<td>235 (70.8, 62.8–78.8)</td>
<td>242 (82.3, &lt;68.8–95.9)</td>
<td></td>
</tr>
</tbody>
</table>

Intubator level (missing 13)

| PGY-1                     | 44 (7.2, 2.7–11.6) | 30 (9.3, 3.3–15.2) | 14 (4.8, 0.4–9.3) |                                                |
| PGY-2                     | 138 (22.5, 14.7–30.3) | 63 (19.5, 10.3–28.8) | 75 (25.9, 16.2–35.6) |                                                |
| PGY-3                     | 233 (38.0, 23.5–52.6) | 140 (43.3, 21.3–65.4) | 93 (32.1, 22.0–42.1) |                                                |
| PGY-4                     | 102 (16.6, 5.2028.1) | 35 (10.8, 3.4–18.2) | 67 (23.1, 7.0–39.2) |                                                |
| Fellow                    | 61 (10.0, 0–21.7)  | 42 (13.0, 0–29.2)  | 19 (6.6, 0–13.2)  |                                                |
| Attending                 | 35 (5.7, 2.8–8.6)  | 13 (4.0, 1.7–6.3)  | 22 (7.6, 0.6–14.5) |                                                |

Augmentation maneuvers

<table>
<thead>
<tr>
<th>Bougie</th>
<th>Out of 623 (missing 2)</th>
<th>Out of 330 (missing 1)</th>
<th>Out of 293 (missing 1)</th>
<th>OR = 16.5, 95% CI = 5.4–50.4, Fisher’s p &lt; 0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>OELM</td>
<td>Out of 624 (missing 1)</td>
<td>Out of 330 (missing 1)</td>
<td>81 = 27.6, 13.4–41.9</td>
<td>OR = 0.6, 95% CI = 0.4–0.8, p = 0.004</td>
</tr>
<tr>
<td>Ramped</td>
<td>21 (3.4, 1.0–5.7)</td>
<td>12 (3.6, 1.0–6.8)</td>
<td>8 (2.7, 0–5.9)</td>
<td>OR = 1.3, 95% CI = 0.5–3.3, p = 0.4</td>
</tr>
<tr>
<td>Bougie or OELM</td>
<td>186 (29.8, 16.7–42.8)</td>
<td>102 (30.7, 10.4–51.0)</td>
<td>84 (28.6, 13.7–43.5)</td>
<td>OR = 1.1, 95% CI = 0.8–1.7, p = 0.6</td>
</tr>
<tr>
<td>Bougie and OELM</td>
<td>6 (1.0, 0–2.9)</td>
<td>6 (1.8, 0–5.1)</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Bougie + OELM + ramped</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
</tbody>
</table>

Data are reported as n (%, 95% CI) or median (IQR).

IQR = interquartile range; OELM = optimized external laryngeal manipulation; PGY = postgraduate year; RSI = rapid sequence intubation.

*95% CI adjusted for clustering (out of N = 625 unless otherwise specified).
†Rao-Scott p-value for difference in categorical variables; Wilcoxon rank-sum test p-value for difference in continuous variables.
The model adjusting for use of either OELM or the bougie also demonstrated superiority of VL for first-pass success (OR = 5.5, 95% CI = 1.7 to 18.1). However, VL did not demonstrate improved first-pass success in the model adjusting only for augmentation with OELM.

Adverse events were relatively uncommon (vomiting = 4 [0.6%], bradycardia = 11 [1.8%], cardiac arrest = 5 [0.8%], hypotension = 9 [1.4%], and mainstem intubation = 10 [1.6%]). There were no cases of dental trauma, epistaxis, lip laceration, laryngospasm, malignant hyperthermia, pneumothorax, tachycardia, iatrogenic bleeding, pharyngeal laceration, or laryngoscope failure. Thus, we counted any adverse event when comparing the two cohorts. The proportions of adverse events and hypoxia in each of the cohorts were comparable both in univariate analysis and in the exploratory multivariable analysis (Tables 2 and 5).

The propensity score–adjusted sensitivity analysis supported our main results: the OR for first-pass success when using VL was 1.8 (95% CI = 1.1 to 3.3). The subgroup analysis of those less than 2 years of age comprised 225 children, of whom 117 (52.0%) underwent intubation by VL and 108 (48.0%) underwent intubation by DL (Data Supplement S1, Table S1). First-pass success was higher when comparing all VL to all DL (OR = 2.0, 95% CI = 1.1 to 3.3), after adjustment for difficult airway characteristics with no differences in overall adverse events or hypoxia. When stratifying by the use of different augmentation maneuvers (bougie, OELM, ramped), there were no statistically significant differences in first-pass success for VL versus DL, either when examined individually (Data Supplement S1, Table S2) or when all augmentation maneuvers are classified as a group (DL plus any augmentation vs VL plus any augmentation and DL without augmentation and VL without augmentation) (Data Supplement S1, Table S3).

**DISCUSSION**

In the adult literature, VL is touted as offering a superior view of the glottis, when compared to DL. When
intubations. Optimal positioning of a young child for glottis visualization can be difficult and the alignment of the anatomic axes will depend on the age and size of the patient. In addition, provider experience with younger children may also be limited, especially in a nonpediatric hospital. Furthermore, pediatric cardiorespiratory physiology is associated with the more rapid onset of oxyhemoglobin desaturation during apnea. Thus, the termination of an intubation attempt may not be due to technical difficulty with visualization, but rather due to the rapid desaturation that occurs in pediatrics. Many of the prior studies were also performed in the operating room or on infant or neonatal manikins. Finally, the studies were conducted several years ago when there was less experience with VL. Providers have become more facile with VL (even if used in the adult population) with its increasing popularity, thereby narrowing the experience gap between VL and DL.

**LIMITATIONS**

Our study has several limitations. First, there is the possibility of misclassification. As discussed in the DAS pediatric difficult airway algorithm, for children less than 2 years of age, a position that is opposite of ramping (e.g., shoulder rolls) is recommended. In our sample of children less than 2 years of age, “ramping” was checked off as having been performed in 11 subjects, yet it is possible that the providers checked off “ramping” when they were actually placing shoulder rolls or placing a child in the “sniffing” position. Second, misclassification of DL versus VL is possible,
since we were unable to discern from the documentation whether a CMAC laryngoscope was being used, how the video component was being used. Clearly, if only the supervisor was viewing the video screen, this should have been classified as a DL attempt. Third, our sample size of 625 patients is relatively small and no power calculation was performed prior to data analysis. Finally, confounding is a risk in any self-reported observational study. Although we controlled for measured confounders such as age, operator experience, method, indication, patient demographics, and difficult airway characteristics, it is impossible to control for all possible unmeasured clinical variables that may have impacted the devices chosen and intubation outcomes. Recall deterioration and recording bias may have influenced results as well, but we believe that this would be evenly distributed across sites and encounters. We strived to minimize such errors by requiring entry of 90% or more of all intubation encounters for site inclusion into the study and by encouraging completion of data forms in real time.

Other limitations relate to the possibility that our study population may have limited external generalizability. The median age of our overall study population was 4, and many would argue that there are systematic differences in the neonatal and infant group, when compared to adolescents and school age children. However, even in our neonatal and infant subgroup analysis, VL demonstrated superior first-pass success. The proportion of patients suspected to initially have a difficult airway (160/625, or 25.6%) is relatively high, and our study population may thereby not be representative of that in other ED pediatric populations. We sought to control for difficult airway characteristics as a potential confounder through regression analyses, and we also performed a propensity score–adjusted sensitivity analysis. In addition, we do not know why providers elected to use augmented maneuvers such as OELM. It may be that these cases were anticipated to be difficult and the first-attempt success may have been different, had these techniques not been employed.

CONCLUSION

In contrast to prior literature, video laryngoscopy was associated with higher first-pass intubation success in our pediatric population, including those under 2 years of age, when compared to direct laryngoscopy with or without augmentation maneuvers.

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.13869/fullData Supplement S1. Supplemental material.
Patient-reported Outcome Measures in Emergency Care Research: A Primer for Researchers, Peer Reviewers, and Readers

Howard S. Kim, MD, MS\(^1\,^2\), D. Mark Courtney, MD, MSCI\(^3\), Danielle M. McCarthy, MD, MS\(^1\,^2\), and David Cella, PhD\(^4\,^5\)

ABSTRACT

Patient-reported outcomes (PROs) are of increasing importance in clinical research because they capture patients’ experience with well-being, illness, and their interactions with health care. Because PROs tend to focus on specific symptoms (e.g., pain, anxiety) or general assessments of patient functioning and quality of life that offer unique advantages compared to traditional clinical outcomes (e.g., mortality, emergency department revisits), emergency care researchers may benefit from incorporation of PRO measures into their research design as a primary or secondary outcome. Patients may also benefit from the ability of PROs to inform clinical practice and facilitate patient decision making, as PROs are obtained directly from the lived experience of other patients with similar conditions or health status. This review article introduces and defines key terminology relating to PROs, discusses reasons for utilizing PROs in clinical research, outlines basic psychometric and practical assessments that can be used to select a specific PRO measure, and highlights examples of commonly utilized PRO measures in emergency care research.

Patient-reported outcomes (PROs) are increasingly utilized in clinical research to capture patient symptoms, functioning, well-being, and their experience with the health care system. Because PROs tend to focus on patient functioning and quality of life, they are frequently used as endpoints in clinical trials of chronic conditions such as cancer,\(^1\) heart failure,\(^2\) and chronic obstructive pulmonary disease.\(^3\) There is also a growing movement to incorporate routine PRO collection into clinical care to better track patients’ longitudinal health status and response to treatment\(^4\) and to monitor the quality of care delivery.

Although PROs have not been historically associated with emergency care research, there is ample opportunity for emergency care researchers to incorporate PROs into future work to augment research impact, respond to specific funding opportunities, and improve patient care. Focusing on PROs, as opposed to laboratory or imaging-based outcomes, may have significant potential benefits to both clinicians and patients, and they can be collected as primary or secondary outcomes in conjunction with clinical or health utilization outcomes. Although a researcher may choose to create a new PRO measure to suit their needs,\(^5\)\(^–\)\(^7\) the vast majority of research questions will have a number of existing PRO measures from which to choose.

This review article defines key terminology relating to PROs, describes reasons for using PROs in research, summarizes basic psychometric assessments...
of PRO measures, and highlights commonly utilized PRO measures applicable to emergency care research. This review can assist researchers in selecting an appropriate PRO measure and will aid clinicians in interpreting the rigor of any published research utilizing PROs.

**KEY DEFINITIONS: WHAT IS A PRO MEASURE?**

A PRO is “any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else.” Thus, the defining feature of a PRO is the patient’s agency in providing the response, rather than a clinician’s subjective assessment (e.g., Emergency Severity Index triage score) or objective measurement (e.g., oxygen saturation), or data abstracted from electronic medical record (e.g., emergency department [ED] length of stay) or administrative claims data (e.g., filled prescriptions). Patient responses can be collected verbally in person or over the telephone or via written or electronic response.

Research measures quantify a particular construct, or concept that is being studied. In the case of PRO measures, these constructs might be pain, physical functioning, or quality of life. An instrument is the specific apparatus for collecting data (i.e., test), and each instrument consists of a number of items (i.e., questions). Patient responses to individual items are totaled to produce a final score on a scale of possible scores. For example, the Oswestry Disability Index (ODI) is a 10-item instrument intended to measure the construct of low back pain–related disability; patient responses to individual items assessing pain intensity and daily functioning are totaled to produce a final score on a scale from 0 to 100. Some PRO instruments measure multiple domains—or aspects—of the construct of interest, such as the 36-item Short-Form Health Survey (SF-36). The SF-36 assesses patient quality of life and produces a separate summary score for the domains of physical health and mental health.

Although PROs are frequently used in patient-centered outcomes research (PCOR), these terms are fundamentally distinct. PCOR refers to a specific research approach that engages patients in research design and conduct, interpretation of results, and dissemination of findings. By engaging with patients throughout the entirety of the research process, PCOR focuses on the patient’s individual needs, goals, and values as defined by the patient. Although a study may collect and utilize PROs, the research is not considered patient-centered unless it treats patients as central stakeholders in the overall effort. Conversely, although a study is not explicitly required to utilize PROs to meet the definition of patient-centered research, it would be unusual for a patient-centered study not to utilize PROs given the central role of patients in designing and conducting the research.

Finally, the National Institutes of Health (NIH) recently sponsored development of a set of PRO measures for public use. The Patient-Reported Outcomes Measurement Information System (PROMIS) measures are the product of an NIH Roadmap for Medical Research (now NIH Commons) initiative to develop and evaluate a set of publicly available, efficient, and flexible PRO measures to be used in multiple clinical research contexts. PROMIS measures are increasingly utilized in clinical research due to their accessibility, efficiency, and ease of use; however, there are many non-PROMIS PRO measures that researchers may elect to use.

**WHY USE A PRO MEASURE?**

Researchers may choose to utilize PRO measures for several reasons. First, the researcher may be primarily interested in the patient’s experience of illness rather than the health care provider’s or health system’s perspective of illness. This use could reflect a growing viewpoint that traditional definitions of treatment success (e.g., reducing hospital readmissions or inhibiting tumor progression) may be shortsighted and in some cases less important to patients than their overall functioning or quality of life. It may even be the case that focusing on PROs may ultimately improve clinical outcomes, as an emerging body of work from the oncology literature indicates that PROs may have more prognostic significance than clinical factors. Funding agencies are also becoming increasingly aware of the importance of PROs in clinical research, and utilizing a PRO measure may highlight the thoroughness of the selected research approach.

Additionally, when PROs are utilized as secondary outcomes they may inform patient decision making should the intervention prove to be effective. This use for PROs is particularly true when the offered intervention is associated with significant side effects. To give a hypothetical example, if a new cancer treatment...
was found to extend disease-free survival by 3 months (primary clinical outcome) but was associated with significant reductions in quality of life (secondary PRO), patients might not elect to receive such an intervention. Similarly, PRO measures may provide clinicians with the ability to speak to patients in a common language when discussing the risks and benefits of a particular intervention or the natural history of a new diagnosis. While there is value in communicating clinical outcomes such as 30-day risk of major adverse cardiovascular event, PROs such as physical functioning—and the ability to walk to the mailbox or climb one flight of stairs—may be more easily understood by patients because of their relevance to everyday life.

Alternatively, PROs may offer the ability to detect meaningful clinical differences that are not easily detected in traditional clinical outcomes. This is especially relevant when the clinical outcome has an inherently low incidence rate, such as mortality or repeat ED visits, which requires a correspondingly large sample size to detect meaningful differences. For example, a study evaluating a new treatment for submassive pulmonary embolism might require enrolling multiple thousands of patients to detect a mortality difference but only hundreds of patients to detect a meaningful difference in physical functioning.

Other important outcomes, such as self-efficacy or fatigue, are not associated with observable physical manifestations and therefore can only be obtained by patient report. In research that seeks to understand the mechanism of benefit of a particular therapy (e.g., mechanistic studies of behavioral interventions), PRO measurement may be necessary to test important hypotheses about the role of psychological mediators such as pain catastrophizing or fear avoidance beliefs.

Finally, many PRO instrument scores are referenced to the population mean to facilitate interpretation of the results. For example, PROMIS scores are standardized to the U.S. population and are reported as T scores with a population mean of 50 and a standard deviation (SD) of 10. Thus, if a patient scores a 60 on the PROMIS-Pain Interference measure, that patient’s self-reported consequences of pain on relevant aspects of life is 1 SD higher (i.e., worse) than the average American. These research results can then be easily disseminated to and understood by a broad general audience. Table 1 illustrates several examples of published emergency care studies leveraging PRO measures to augment their overall research impact.

### MEASUREMENT PROPERTIES: WHAT MAKES A GOOD PRO MEASURE?

Patient-reported outcome instruments are developed de novo using a rigorous combination of qualitative and quantitative methods. The quality and performance of the resulting PRO instruments are then assessed by their psychometric properties, such as reliability and validity. Possessing a basic understanding of these terms is necessary to assess whether a PRO instrument has been properly selected to achieve the stated research objective.

Broadly speaking, reliability is the extent to which an instrument measures something in a reproducible or consistent fashion, while validity refers to whether the instrument measures what it purports to measure. Both reliability and validity are quantified using various coefficients (e.g., Pearson correlation, Cronbach’s alpha), however, all share a common scale of 0 to 1 with a general rule of thumb that coefficient values >0.7 to 0.8 are desired.

On a deeper level, reliability refers to two properties. The first property, test–retest reliability (i.e., test stability), is the extent to which a measure is reproducible and is tested with repeat administrations of the same PRO instrument in a short time interval during which change is not expected. This ensures that measured changes over time are in fact real (e.g., the result of on intervention or change in condition) and not the result of a poorly designed instrument.

The second reliability property, internal consistency (i.e., homogeneity), is the extent to which individual items within a multi-item PRO instrument are consistent with one another and therefore produce a coherent total score. For example, an instrument designed to assess depression might contain multiple questions about depressive symptoms; responses to a question about lack of interest in activities should be generally consistent with responses to a question about problems sleeping, given that both are related to depression. Emergency care researchers may already be familiar with other forms of reliability, such as inter-rater reliability, which assesses the degree of agreement between subjective ratings from multiple observers. It should be noted, however, that inter-rater reliability is more relevant to other settings in which clinician observers rate a patient’s clinical status rather than the patient providing self-report of their symptoms (e.g., Clinical Institute Withdrawal Assessment for Alcohol or CIWA Score).
Table 1
Examples of Emergency Care Research Utilizing PRO Measures and How PRO Measurement Adds Value

<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Study Design and Question</th>
<th>PRO Measure</th>
<th>Method and Timing (Response Rate) of Measurement</th>
<th>How PRO Measurement Adds Value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friedman et al. 72</td>
<td>Randomized controlled trial. What is the comparative effectiveness of three analgesic medication regimens used to treat ED patients with acute low back pain?</td>
<td>Roland-Morris Disability Questionnaire (RMDQ)</td>
<td>In-person and telephone; n = 323 participants Baseline, 1 week (96%), and 3 months (91%) after ED visit</td>
<td>Collecting outcome data on pain and functioning by patient self-report is more practical and has more face validity than other methods, such as direct observation or actigraphy monitoring. The study findings may also be easily communicated to patients when explaining the rationale for treatment selection.</td>
<td>This single-center randomized controlled trial found that adding cyclobenzaprine or oxycodone/acetaminophen to naproxen alone did not improve functional outcomes at 1-week follow-up.</td>
</tr>
<tr>
<td>Talan et al. 30</td>
<td>Pilot randomized controlled trial. Can ED patients with acute appendicitis be safely managed with antibiotics instead of appendectomy?</td>
<td>12-Item Short-Form Health Survey (SF-12)</td>
<td>In-person; n = 30 participants Baseline, 2 weeks (97%), and 1 month (97%) after ED visit</td>
<td>If an antibiotic-first strategy is found to be noninferior to an appendectomy strategy with respect to clinical outcomes—but results in better quality of life—then patients might prefer an antibiotic-first strategy.</td>
<td>This pilot randomized trial was intended to demonstrate feasibility of study methods. The following outcomes are descriptive only. Patients in the antibiotic-first group had lower complication rates at 1 month (14% vs. 6%) and higher quality of life on the SF-12 Physical Component Subscore (56 vs. 47) compared to the surgery group.*</td>
</tr>
<tr>
<td>Rising et al. 130</td>
<td>Prospective observational study. Is the level of ED patient uncertainty associated with a higher likelihood of a 30-day ED return visit?</td>
<td>Uncertainty Scale (U-Scale)</td>
<td>In person and telephone; n = 200 participants Baseline and 1 month (78%) after ED visit</td>
<td>Patient uncertainty can only be measured by self-report. Uncertainty may play an important role in a patient’s decision to visit the ED, both in the initial ED visit and at repeat visits. Thus, uncertainty is an important construct to measure in research focusing on ED utilization and recidivism.</td>
<td>In this observational study, total U-scale score was not associated with 30-day ED return visits; however, higher uncertainty on the Treatment Quality subscale was associated with increased odds of a 30-day ED revisit.</td>
</tr>
<tr>
<td>Pluymaekers et al. 36</td>
<td>Randomized controlled trial. Can ED patients with recent-onset stable atrial fibrillation safely undergo delayed cardioversion instead of early cardioversion?</td>
<td>Atrial Fibrillation Effect on Quality of Life (AFEQT)</td>
<td>In person; n = 437 participants in the Netherlands‡ Baseline and 1 month (98%) after ED visit</td>
<td>If delayed cardioversion was found to be noninferior to early cardioversion with respect to sinus restoration, then patients might elect to defer early cardioversion given the possibility of spontaneous sinus conversion while maintaining the same quality of life.</td>
<td>In this multicenter randomized controlled trial, delayed cardioversion was found to be noninferior to early cardioversion with respect to sinus restoration (1* outcome, 91% vs. 94%). Patients receiving delayed vs. early cardioversion had similar AFEQT scores (2* outcome, 72 vs. 73).</td>
</tr>
<tr>
<td>Lin et al. 125</td>
<td>Cross-sectional survey. What is the national prevalence of burnout among U.S. EM residents?</td>
<td>Maslach Burnout Inventory (MBI)</td>
<td>Online survey by e-mail solicitation; n = 1,522 respondents Cross-sectional survey (21%) of U.S. EM residents</td>
<td>Physician burnout is best measured by self-report. Burnout is an important aspect of physician wellness that can be selected as a primary outcome in future interventional studies.</td>
<td>In this prospective cross-sectional study, 76% of surveyed EM residents reported symptoms consistent with burnout.</td>
</tr>
</tbody>
</table>

1* = primary outcome; 2* = secondary outcome; PRO = patient-reported outcome.

*Because this was a pilot randomized study, the primary study objective was to demonstrate feasibility of study methods and not to conduct a comparison of outcomes. Therefore, these outcome data are descriptive and no formal hypothesis testing was conducted.

†This study was conducted in the Netherlands, where the ability to secure post-ED follow-up may differ from that in the United States.
Validity broadly refers to whether an instrument adequately reflects the construct to be measured. There are multiple types of validity that can be assessed. Content validity is the degree to which the content of an instrument adequately samples all the important domains of a construct. For example, a PRO measure of pain would optimally assess pain intensity, activity interference, and affective interference. One aspect of content validity is face validity, which describes the extent to which items appear to measure what they purport to measure; this determination is often adjudicated by content experts, such as patients or clinicians. Construct validity is the degree to which an instrument’s score is consistent with hypotheses associated with the intended construct. Construct validity accumulates over time, as evidence confirming predicted relationships with similar (and different) concepts accumulates and as the stability of the structure of the concept is demonstrated across different groups of people and over time. For example, one would expect that patients with more severe depression scores on an instrument will have higher suicide risk.

Related types of validity (e.g., criterion validity, cross-cultural validity, discriminant validity) address whether what is being measured correlates with similar measures of the same construct and, conversely, does not correlate with measures of different constructs. Given the many types of validity, we note that a blanket reference to any PRO questionnaire as “valid” is an oversimplification of the concept. A better approach would be to refer to an instrument in terms of the extent to which it has demonstrated validity or the populations for whom there is evidence of validity. Thus, validity data never stop accruing to an instrument, and one cannot have too much validity.

### HOW TO SELECT A PRO MEASURE

Selecting a PRO instrument with adequate reliability and validity can improve the scientific rigor of a research study, much in the same way that selecting an appropriate data collection method for a retrospective chart review may increase confidence in the study findings. Yet many available PRO measures have been shown to be reliable and valid, and selecting an instrument based on its psychometric properties alone might be disadvantageous if the instrument has other weaknesses. This section details a number of other important properties that researchers should consider when selecting a PRO instrument.

**Respondent Burden**

To evaluate a PRO measure, the data must first be successfully collected from patients. Thus, researchers should consider the total number of questions posed by a PRO instrument and whether the length of the survey might become so burdensome as to discourage further patient participation (or in the case of verbally administered surveys, the researcher’s motivation to continue the interview). While respondent burden is typically not a deciding factor in selecting a single PRO measure, it can become significant when a researcher attempts to collect multiple PRO measures in the same administration. For example, the 36 items of the SF-36 are not overly burdensome if administered in isolation; however, adding additional measures such as the 24-item Roland Morris Disability Questionnaire and the 13-item Patient Activation Measure might make the respondent burden prohibitively high. Researchers must therefore carefully select the PRO measures deemed to be most important to their research objective and avoid the temptation to add on measures simply because one is already collecting patient-reported data.

One strategy to minimize respondent burden is to use a computer-adaptive testing (CAT) format of existing PRO measures, if available. In CAT, a computer algorithm selects the next question to be administered from a pool of available questions based on the response to the previous question. Using an iterative process, the algorithm uses each new patient response to calibrate the patient’s final score on the measurement scale, ultimately requiring a smaller number of questions than the corresponding fixed-length version of the instrument. Many of the PROMIS measures come in CAT format and can calculate a final score in as few as four questions. Utilizing PRO measures with CAT formatting can substantially shorten the time to survey completion, even among older adults, and only requires 1 minute compared to more traditional full-length instruments, which may require 5 minutes or longer. PRO measures administered through REDCap survey integration can be completed by patients on a computer, tablet, or smartphone. For patients who do not have computer access or prefer not to provide responses electronically, researchers can collect verbal responses by telephone and input patient responses into the CAT interface.

Additionally, many lengthy PRO measures have been reduced to more efficient “short-form” versions to reduce overall respondent burden, such as the 12- and eight-item versions of the original 36-item Short-
Figure 1 Example of a PRO Measure. (A) Example of the PROMIS–Pain Interference (PROMIS-PI) Short Form, in which patients respond to all six items. The final score is calculated using item-level calibrations and can be performed via the HealthMeasures Scoring Service at https://www.assessmentcenter.net/ac_scoringservice. PROMIS instruments and scoring are publicly available at http://www.healthmeasures.net/. ©2008–2015 PROMIS Health Organization and PROMIS Cooperative Group. (B) Patient-facing PROMIS-PI CAT interface in REDCap. In this CAT format, the patient’s response to each question guides the computer to select the next question from a question bank. Each successive response increases the confidence of the cumulative score, and the survey concludes when the cumulative score’s standard error drops below a specified level. Thus, patients may be queried a total of four to 12 questions, depending on their responses. Many PRO measures, including PROMIS-PI, are available in the REDCap Shared Library and can be easily embedded into patient data collection tools. ©REDCap 8.10.20, Vanderbilt University. CAT = computer-adaptive testing; PRO = patient-reported outcome. [Color figure can be viewed at wileyonlinelibrary.com]
Form Health Survey (SF-12 and SF-8).\textsuperscript{10,20,21} PROMIS conveniently offers compilations of various short forms, such as the PROMIS-Global-10 (analogous to the SF-12) and the PROMIS-29 Profile, which contains the four-item short forms for anxiety, depression, fatigue, pain interference, pain intensity, physical function, sleep disturbance, and ability to participate in social roles and activities. Figure 1 demonstrates both the short-form and the CAT format of the PROMIS–Pain Interference instrument.

**Disease-specific Measure Versus Generic Measure**

Patient-reported outcome measures can be specific to a particular disease or therapy (e.g., atrial fibrillation effect on quality of life\textsuperscript{22}) or can be general assessments of health applicable to a wide range of patients with or without disease (e.g., SF-36). The decision to utilize a disease-specific or generic PRO measure will depend on the objective of the research study, although it is not uncommon for research studies to include both types of measures. Generic instruments have the advantage of being extensively validated and utilized in prior research, and they can be easily compared to studies conducted in other clinical contexts or diseases. However, because generic instruments measure general health rather than disease-specific symptoms, their scores may reflect the contributions of other factors such as patient demographics or medical comorbidities. In contrast, disease-specific instruments were explicitly developed to measure the disease of interest and may more accurately reflect relevant features of that disease—thereby resulting in greater power to discriminate between relevant groups of patients.\textsuperscript{23,24} We note, however, that this contrast between generic and disease-specific is, to a degree, a false distinction given the advent of item response theory which allows for highly precise assessment of common symptoms in a generic assessment framework.

**Pragmatic Considerations**

All other factors being equal, researchers may select a PRO measure that has been previously used by others or is readily available in existing survey administration software. Selecting a previously utilized PRO measure may facilitate comparison between studies or pooling of data between investigators. Many PRO measures are currently available in the REDCap Shared Library,\textsuperscript{25,26} allowing for quick and efficient incorporation into patient data collection procedures.

**Existing Data Sets**

Although researchers may choose any PRO measure when designing a prospective study, those conducting secondary analyses of existing data sets must select from the limited supply of PRO measures previously collected. Unfortunately, many commonly used data sets in emergency care research such as the National Hospital Ambulatory Medical Care Survey and Nationwide Emergency Department Sample do not contain PRO data. Other existing data sets have collected a limited number of PRO measures, such as the National Health and Nutrition Examination Survey and the Behavioral Risk Factor Surveillance System, which utilize the Centers for Disease Control Four-item Health-Related Quality of Life survey (CDC HRQOL-4)\textsuperscript{27} and the eight-item version of the Patient Health Questionnaire (PHQ-8) for depression.

**COMMONLY UTILIZED PRO MEASURES**

This section will provide a brief overview of existing PRO measures that may be applicable to emergency care research. Examples of commonly assessed outcomes and relevant PRO measures are presented in Table 2. This list is not meant to be exhaustive but is instead intended to illustrate examples of PRO measures to provide context for the selection of an appropriate measure. Regardless of the specific measure selected, researchers should ensure that they utilize the most recent version of a PRO measure given that many instruments have been modified and updated over time.

**General Health Status and Health-related Quality of Life**

Generic assessments of general health status and health-related quality of life differ from other PROs because of their multidimensional focus on patients’ overall well-being and functioning, as well as the ability to use these instruments in patients with or without the condition of interest. One of the most well-known PRO measures assessing health-related quality of life is the SF-36, which contains 36 questions evaluating eight scales—physical functioning, physical role, emotional role, bodily pain, general health, vitality, social functioning, and mental health—organized under the two domains of physical health and mental health.\textsuperscript{10,28} The SF-36 is also available in shorter formats (SF-12, SF-8),\textsuperscript{20,29,30} which may reduce respondent burden.
Table 2
Commonly Assessed PROs in Emergency Care Research and Corresponding Measures

<table>
<thead>
<tr>
<th>PRO Measure</th>
<th>Number of Items (Completion Time)</th>
<th>Accessibility</th>
<th>Benefits</th>
<th>Limitations</th>
<th>Similar Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of life (QoL)</td>
<td>Short-Form Health Survey (SF-36)</td>
<td>SF-36: 36 items (10 min) SF-12/SF-8: 12 or 8 items (3 min)</td>
<td>SF-36: free SF-12/SF-8: cost</td>
<td>Widely used and recognized SF-36: REDCap integrated</td>
<td>SF-36: high respondent burden SF-12/SF-8: no REDCap integration; cost</td>
</tr>
<tr>
<td>Disease-specific QoL</td>
<td>Atrial Fibrillation Effect on Quality of Life Questionnaire</td>
<td>20 items (5 min)</td>
<td>Cost ($500)</td>
<td>More specific to A-fib than generic measures</td>
<td>Moderate respondent burden; cost No REDCap integration</td>
</tr>
<tr>
<td>Physical function</td>
<td>PROMIS – Physical Function</td>
<td>Short form: 6, 8, 10, 12, or 20 items (1-5 min) CAT: 4-12 (1 min)</td>
<td>REDCap integrated; CAT format; automatic scoring; minimal respondent burden</td>
<td>SF-36/SF-12 Physical Function Subscore, PROMIS Upper Extremity</td>
<td></td>
</tr>
<tr>
<td>Discomfort (e.g., pain, nausea, itch)</td>
<td>Numeric Rating Scale</td>
<td>1 item (&lt;1 min)</td>
<td>Free</td>
<td>Minimal respondent burden; easily understood by patients, clinicians, and general public</td>
<td>Unidimensional: only measures intensity, does not measure activity or affective interference</td>
</tr>
<tr>
<td>Pain</td>
<td>Brief Pain Inventory</td>
<td>Long form: 32 items (10 min) Short form: 9 items (5 min)</td>
<td>Free for unfunded research; variable pricing for funded research</td>
<td>Measures multiple domains of pain (intensity and interference)</td>
<td>No REDCap integration; moderate respondent burden Question 2 requires the patient to illustrate an anatomic figure (limits telephone or electronic response)</td>
</tr>
<tr>
<td>Anatom-specific pain</td>
<td>Roland Morris Disability Questionnaire (low back pain)</td>
<td>24 items (5 min)</td>
<td>Free</td>
<td>Widely used and recognized; REDCap integrated</td>
<td>Moderate respondent burden Oswestry Disability Index (low back); Neck Disability Index; Disorders of the Arm, Shoulder, and Hand</td>
</tr>
<tr>
<td>Depression</td>
<td>Patient Health Questionnaire 9,99</td>
<td>9 items (3 min)</td>
<td>Free</td>
<td>Widely used and recognized; minimal respondent burden, REDCap integrated</td>
<td>Hospital Anxiety and Depression Scale; Center for Epidemiological Studies Depression Scale-10; PROMIS-Depression</td>
</tr>
<tr>
<td>Posttraumatic stress</td>
<td>PTSD Checklist for DSM-5 (PCL-5)</td>
<td>20 items (5 min)</td>
<td>Free</td>
<td>No REDCap integration; moderate respondent burden</td>
<td>Post-Traumatic Adjustment Scale; Peri-Traumatic Distress Inventory</td>
</tr>
<tr>
<td>Activation and self-efficacy</td>
<td>Patient Activation Measure</td>
<td>13 items (5 min)</td>
<td>Cost ($150)</td>
<td>No REDCap integration; moderate respondent burden</td>
<td>PROMIS–Self Efficacy, Patient Health Engagement Scale, Smoking Self-Efficacy Questionnaire</td>
</tr>
<tr>
<td>Physician burnout</td>
<td>Maslach Burnout Inventory</td>
<td>22 items (5 min)</td>
<td>Variable pricing</td>
<td>Widely used and recognized</td>
<td>No REDCap integration, moderate respondent burden; cost Professional Fulfillment Index, Compassion Fatigue and Satisfaction Self-Test for Helpers</td>
</tr>
</tbody>
</table>

Reliability and validity: Please note that all listed PRO measures in this table have evidence of reliability and validity in the literature. In general, researchers should select measures with demonstrated reliability and validity, rather than create ad hoc instruments (i.e., a customized survey) that have not been previously validated or found to be reliable.

REDCap integrated: These measures are approved for use and available in the REDCap Shared Library.

Automatic scoring: REDCap automatically calculates the final score, which can be downloaded by the researcher for each participant. Instruments that do not have automatic scoring require manual scoring by the researcher using recommended scoring sheets.

Licensing and purchase: Websites for licensing and purchase of each PRO are included in Data Supplement, Appendix S2.

CAT = computer-adaptive testing; PRO = patient-reported outcome.
Other frequently utilized health-related quality-of-life measures include the EuroQol five-dimensional questionnaire (EQ-5D), which assesses mobility, self-care, usual activities, pain/discomfort, and anxiety/depression;31–33 PROMIS–Global Health, which evaluates physical, mental, and social components of health;12 and the CDC HRQOL.34 The broad scope of these general health-related quality-of-life measures allows for a more global assessment of an intervention effect and facilitates comparison to the population norm or other known interventions.35 Assessments of general health status may also be disease-specific, such as the Atrial Fibrillation Effect on Quality-Life Questionnaire,22,36,37 the Airway Questionnaire,38,39 and the Quality of Life in Neurological Disorders (Neuro-QoL) measure.40

**Physical Functioning**

Physical functioning refers to a patient’s self-reported capability to perform functional activities associated with daily living, such as lifting, walking, or sleeping. Because physical functioning is an important component of general health, the SF-36 contains specific questions that map to a physical functioning scale—these questions can either be administered alone as a 10-item short form or be reported as a subscore of the complete SF-36.41,42 PROMIS–Physical Function is a separate instrument from PROMIS–Global Health and comes in both short form and CAT format.43,44 PROMIS also has a specific measure to assess physical functioning of the upper extremity, PROMIS–Physical Function–Upper Extremity, which includes activities such as writing or opening containers.

**Pain**

The experience of pain is a complex phenomenon, with an entire field of research devoted to describing various pain qualities, measuring pain, and differentiating individual responses to pain. Because pain is primarily reported by patients, rather than quantified with a laboratory assay, PRO measures are frequently used in pain management research. The majority of PRO measures focus on the domains of pain intensity and pain interference.

Pain intensity, or how much a patient hurts, is most frequently assessed using a numeric rating scale (NRS; 0 to 10)45–55 or visual analog scale (VAS, 0 to 100).55–62 Although these measures consist of a single question, they meet the definition of a PRO because patients provide direct report of their health status. The PROMIS–Pain Intensity short form contains three questions relating to pain over the past 7 days at its worst, on average, and currently; patients rate their pain on a scale from 1 (“had no pain”) to 5 (“very severe”).

**Pain interference** describes the consequences of pain on relevant aspects of the patient’s life, including social, cognitive, emotional, physical, and recreational activities. PROMIS–Pain Interference comes in both short form and CAT format and is widely applicable to various patient populations and conditions.60,63 The Brief Pain Inventory (BPI) is one of the earliest pain measures assessing the severity of pain and impact on daily functioning; the short-form version includes 15 items. Although the BPI was originally developed for cancer-related pain,64 it has since been widely validated and applied to noncancer pain65 and in ED patient populations.66–69

Several measures of pain interference describe pain-related impairment specific to an anatomic site. Although we have grouped these PROs with pain interference measures, they also have overlap with measures of physical functioning. The Roland Morris Disability Questionnaire is a 24-item instrument assessing how low back pain affects daily functional activities70 and has served as the primary outcome in several ED-based randomized controlled trials.71–74 The ODI is a 10-item instrument assessing how low back pain affects daily functioning; total scores are summed to categorize patients with minimal, moderate, severe disability, or crippled.9 The Neck Disability Index is a modification of the ODI evaluating how neck pain interferes with various functional tasks.75 The Disabilities of the Arm, Shoulder, and Hand Score contains 30 items assessing the patient’s ability to perform certain upper-extremity functions and the degree of painful symptoms.33,76

Finally, a patient’s internal thoughts and feelings due to pain can be assessed using the 13-item Pain Catastrophizing Scale.77 Pain catastrophizing—or the degree to which patients ruminate about pain, magnify pain, and feel helpless about managing their pain—is thought to be an important predictor of recovery and treatment success.68,78–80

**Nausea**

A number of emergency care studies have evaluated the comparative effectiveness of acute interventions for nausea. These studies have exclusively utilized single-item PRO measures, such as the NRS55,81,82 or VAS,55,56,61,82,83 due to the frequency of assessment
required to capture acute change during an ED visit. Research evaluating chronic conditions and nausea symptoms over a longer period of time can use disease-specific instruments, such as the nine-item Gastro-paresis Cardinal Symptom Index, or generic instruments, such as the four-item Nausea and Vomiting scale of PROMIS–Gastrointestinal.

**Mental Health and Health Beliefs**

Many generic assessments of health status and health-related quality of life, such as the SF-36 and PROMIS–Global Health, report subscores for a mental health domain. Additional PRO measures are available for specific mental health outcomes; however, the range of available mental health measures is far too expansive for a complete detailing here. In this section, we will focus on PROs that have been previously utilized in emergency care research.

*Emotional distress* includes negative affective symptoms such as anxiety and depression. These constructs may be particularly relevant in emergency care given the association between anxiety and ED utilization. The Patient Health Questionnaire-9 (PHQ-9) contains nine items from the full Patient Health Questionnaire and is used to screen for depression and measure the severity of depressive symptoms. The seven-item Generalized Anxiety Disorder scale similarly screens for and measures the severity of anxiety. The Hospital Anxiety and Depression Scale is a 14-item instrument that measures both anxiety and depression in hospitalized patients, although it has since been more widely applied to the general population. PROMIS offers a number of measures assessing emotional distress, including PROMIS Anger, Anxiety, and Depression, in both short form and CAT format.

*Self-efficacy* refers to a patient’s confidence in successfully performing specific tasks or behaviors. PROMIS offers a general self-efficacy measure and a specific measure for self-efficacy in managing chronic conditions. Self-efficacy is closely related to the construct of patient activation or the beliefs that patients hold about their ability to self-manage their health. The 13-item Patient Activation Measure is the most well-known instrument for activation and assesses a patient’s knowledge, skill, and confidence for managing their own health and health care. Patient activation has been extensively evaluated in emergency care research, due to the inverse relationship between patient activation and ED utilization. Disease-specific measures of self-efficacy are also available for use, such as the 12-item Smoking Self-Efficacy Questionnaire, which measures patients’ confidence in their ability to refrain from smoking when faced by various internal and external stimuli.

**Other PROs Used for Common Emergency Care Scenarios**

A number of other PRO measures are available for specific areas of clinical care common to the practice of emergency medicine. The 22-item Post-Concussive Symptom Score measures the severity of acute symptoms after concussion. Research focusing on the pediatric emergency care population have utilized the Post-Concussion Symptom Inventory to assess concussion symptom burden in children and adolescents, with different formats of varying question length for each age group (13–26 items).

The ED is an increasingly common care venue for patients with substance use disorder. Patient-reported consequences of drug and alcohol abuse can be measured using the 15-item Short Index of Problems or the 16-item brief Rutgers Alcohol Problems Index. PROMIS also offers a number of assessments in both seven-item short form and CAT format: PROMIS Substance Use–Appeal, PROMIS Substance Use–Severity, PROMIS Substance Use–Prescription Pain Medication Misuse, PROMIS–Alcohol Use Negative Consequences, and PROMIS–Alcohol Use Negative Expectancies. Risk-taking behaviors relevant to substance use can be measured with specific instruments, such as the 11-item HIV risk-taking behavior scale, and health care utilization can be quantified using the Treatment Services Review.

Interest in the concept of shared decision making between patients and clinicians is growing rapidly, especially in patient-centered outcomes research. Studies of ED-shared decision-making interventions have utilized the single-item Control Preference Scale to assess a patient’s desire for involvement in decision making. The degree of shared decision making perceived by the patient in patient–clinician interactions can be measured using the nine-item Shared Decision-Making Questionnaire, or alternatively patient uncertainty...
Research Evaluating Clinician Wellness

A growing area of research focuses on clinician wellness in the practice of medicine. Although it sounds unusual to refer to clinician wellness as a patient-reported outcome, studies that seek to improve wellness or alleviate burnout in health care providers have merely reframed the clinician as the patient of interest. Several recent randomized trials have investigated the effect of flexible duty hour scheduling on satisfaction and well-being in general surgery and internal medicine resident physicians using an ad hoc PRO measure, and a number of observational studies have characterized burnout among residents from all specialties. Burnout, which is characterized by high levels of emotional exhaustion or depersonalization resulting from workplace stress, is frequently measured using the 22-item Maslach Burnout Inventory. The recently developed 16-item Professional Fulfillment Index is specific to professional fulfillment and burnout among clinicians and explicitly assesses interpersonal disengagement pertaining to patient care.

Reporting on PRO Measures in Scientific Writing

After selecting an appropriate PRO measure, researchers should adequately describe the measure in scientific writing. Complete reporting of PRO measures will improve the clarity of writing, substantiate the rigor of the research approach, and facilitate study replication and comparison. To provide guidance on PRO measure reporting, a “PRO Extension” to the Consolidated Standards of Reporting Trials (CONSORT) checklist has been developed for randomized trials in which PROs serve as primary or important secondary outcomes (Data Supplement, Appendix S1, available as supporting information in the online version of this paper, which is available at http://onlineibrary.wiley.com/doi/10.1111/acem.13918/full). This checklist provides important guidance that is also applicable to the conduct of observational research, such as identifying the PRO as a primary or secondary outcome, describing the relevant domains and hypothesis of the PRO, and citing the PRO instrument’s reliability and validity.

Additionally, we recommend that researchers explicitly state the directional vector of PRO measure scores (e.g., higher score = worse function) to maximize reader comprehension. When able, researchers should also include the PRO instrument in the manuscript appendix to improve the clarity of data collection methods and facilitate study replication.

CONCLUSIONS

Patient-reported outcome measures are a valuable tool for augmenting the impact of emergency care research. By focusing on the patient’s perspective of health and health care, patient-reported outcomes highlight outcomes that are most important to the lived experience of patients. This adds significant value and relevance to research findings and may improve communication between clinicians and patients regarding the risks and benefits of an intervention or the natural history of a new diagnosis. When selecting a patient-reported outcome measure for use, researchers should consider the reliability and validity of the patient-reported outcome instrument in addition to other important properties such as respondent burden and relevance to the population and disease of interest. Proper selection and application of a patient-reported outcome measure can significantly strengthen the scientific rigor of a research approach and facilitate overall research success.

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.13918/full

Data Supplement S1. Supplemental material.
Inter-rater Agreement Between Self-rated and Staff-rated Clinical Frailty Scale Scores in Older Emergency Department Patients: A Prospective Observational Study

Thom Ringer, MD, JD, Cameron Thompson, MSc, Shelley McLeod, PhD, and Don Melady, MD, MScED

Frailty is a state of vulnerability arising from multiple medical and psychosocial problems primarily affecting older people. It is associated with adverse outcomes including mortality, prolonged hospitalization, and functional dependence after discharge. Identifying frailty early may help concentrate resources on patients at high risk of iatrogenesis, functional decline, and death. Despite the growing number of emergency department (ED) visits by older people, frailty is relatively underexamined in the ED setting.

The Canadian Study of Health and Aging Clinical Frailty Scale (CFS) is a commonly used frailty assessment tool. Derived from a 5-year cohort study involving over 10,000 older Canadians, the CFS employs a 9-point scale based on clinical judgment. A CFS score of 1 to 3 represents nonfrailty; 4 represents vulnerability to frailty; 5 and 6 represent mild and moderate frailty, respectively; and a score of 7 or higher represents severe frailty. It is quick to administer and predicts patient-important outcomes including mortality, adverse discharge (e.g., long-term care), and functional decline. Dresden et al. found that MD-assigned CFS scores in the ED were associated with adverse discharge destination. Lewis et al. found that ED CFS scores were as accurate in predicting poor outcomes as more time-intensive instruments based on objective measures of physical frailty, but more practical and less disruptive.

As Dresden et al. note, the level of agreement between self-rated and provider-rated frailty sheds light on whether providers and patients interpret frailty in the same way, helping to inform research and choice of screening instrument. Acceptable agreement between provider CFS scores would justify the use of a single instrument by various categories of ED providers, where registered nurses (RNs) are responsible for initial evaluation and management before assessment by a physician or advanced practice provider (e.g., nurse practitioner or physician assistant [APP]).

Dresden et al. reported moderate agreement between patient and MD CFS scores. However, that study did not involve RNs or APPs, the latter of whom play a growing role in Canadian EDs. The objective of this study was to determine the level of agreement between self-rated and staff-rated (RN and MD/APP) CFS scores in older ED patients.

This was a prospective, observational study of patients aged 75 or older presenting to an urban Canadian academic ED (annual census 65,000)
between December 2018 and April 2019. At this institution, approximately 12.5% of the annual ED census are 75 years and older.

Research personnel obtained written informed consent to participate. Patients requiring resuscitation, those with a Glasgow Coma Scale score of <14, and patients who could not understand English were excluded. The study received institutional research ethics board approval (18-0271-E).

The CFS instrument consists of a single page divided into nine ratings, consisting of:

- Category name (e.g., CFS 1 is denominated as "very fit," CFS 8 as "very severely frail").
- Silhouette of a patient representing a particular functional status (e.g., "very fit" is depicted as a person jogging; "very severely frail" is depicted as a person lying in bed).
- Brief description (e.g., "very fit" people are "robust, active, energetic and motivated... They are among the fittest for their age;" "severely frail" people are "Completely dependent for personal care, from whatever cause.")

Patients received a copy of the CFS and were asked to circle the CFS score that best described themselves. Each patient’s RN and MD/APP were given a copy of the same instrument and asked to circle the frailty rating that best described the patient. All raters were blinded to each other’s scores.

To estimate our desired sample size, we posited the three categories of CFS scores used in the literature (nonfrail [CFS ≤ 4], mildly to moderately frail [CFS 5 or 6], and severely frail [CFS ≥ 7]) \( \text{9} \). We assumed that each of these categories would have a frequency of 0.33. To detect a kappa of 0.80 with 80% power, we estimated that 144 patient encounters would be required.\( \text{8} \) We added 10% to account for withdrawals and missing data, resulting in a final sample size of 160 patient encounters.

For the primary analysis, we estimated inter-rater agreement between ordinal (i.e., 1 to 9) CFS scores using quadratic-weighted kappa statistics with 95% confidence intervals (CIs) for these dyads: patient–RN, patient–MD/APP, and RN–MD/APP. In a secondary analysis, CFS scores were dichotomized as frail (CFS ≥ 5) or nonfrail (CFS ≤ 4; a cut-point prevalent in the literature). We then estimated intrarater agreement using Cohen’s kappa statistics with 95% CI.\( \text{9,6} \)

We interpreted kappa values as follows:\( \text{9} \)
- <0.20: Poor agreement
- 0.20 to 0.40: Fair agreement
- 0.41 to 0.60: Moderate agreement
- 0.61 to 0.80: Good agreement
- >0.80: Very good agreement

Over the 4-month study period, 159 of 160 patient encounters were included. One encounter was excluded due to a missing MD CFS score. Mean (±SD) age of the included patients was 82.9 (±6.0) and 86 (±54.1%) were female. Figure 1 displays distributions of CFS scores by rater category. A high proportion of patients (\( n = 31, 19.4\% \)) rated themselves as "very fit" (CFS = 1), whereas fewer patients were rated as “very fit” by RNs (\( n = 17, 10.6\% \)) and MD/APPs (\( n = 15, 9.4\% \)). Overall, very few patients (\( n = 5, 3.2\% \)), RNs (\( n = 7, 4.4\% \)), or MD/APPs (\( n = 8; 5.0\% \)) assigned ratings of "severely frail" (CFS ≥ 7). These patterns resemble those reported by Dresden et al.\( \text{7} \)

Inter-rater agreement between patient and RN scores was moderate (0.59, 95% CI = 0.46 to 0.71), as was the agreement between patient-rated frailty scores and those reported by MD/APPs (0.53, 95% CI = 0.42 to 0.64). Inter-rater agreement between RN and MD/APP scores was good (0.74, 95% CI = 0.67 to 0.81). When CFS scores were dichotomized as “nonfrail” (CFS < 5) or “frail” (CFS ≥ 5), patient–RN and patient–MD/APP inter-rater agreement was 0.51 (95% CI = 0.35 to 0.67) and 0.42 (95% CI = 0.30 to 0.59), respectively, and RN–MD/APP inter-rater agreement was 0.72 (95% CI = 0.60 to 0.83).

This was a small study performed in a single academic ED and may not be generalizable to other centers. Patients in this study were not consecutively enrolled, which may have introduced risk of selection bias. For instance, research and clinical staff may have tended not to approach patients who seemed unwell, agitated, or somnolent (e.g., because of dementia and/or delirium). This may have yielded a sample that was less frail than the population of interest (all older ED patients). It is possible that some RNs and MD/APPs were biased by their impression of the patient’s current health status (i.e., acute condition), rather than their baseline condition. We did not distinguish between different APP categories, nor assess staff’s years of experience. Finally, we did not evaluate patient characteristics such as level of education,
cognitive impairment, or functional status, which may represent potential confounders.

Our permissive inclusion criteria (all stable ED patients over the age of 75 without serious neurologic impairment) and the fact that we did not specially train or select participating ED staff enhance our study’s generalizability to other ED settings. However, by not including seriously ill or cognitively impaired older people, we may have excluded the frailest ED patients (e.g., with terminal illness or advanced dementia). Moreover, excluding non–English-speaking participants may affect generalizability of results to the actual ED population, in that culture may affect individuals’ understanding and interpretation of frailty.

The objective of this study was to determine the level of agreement between self-rated and ED staff-rated CFS scores in older ED patients. Patient-rated CFS scores showed moderate agreement with those of both RNs and MD/APPs. RNs and MD/APP CFS scores showed good inter-rater agreement.

The difference between provider–provider and provider–patient agreement is a novel finding. It may be explained partially by the higher propensity of some patients, also reported by Dresden et al., to rate themselves as “very fit” (CFS = 1) or “fit” (CFS = 2). This hypothesis is supported by a subgroup analysis of our 63 participants with self-rated CFS scores of ≤ 2. In this group, there was no statistically significant agreement between patient–RN (0.06, 95% CI = –0.05 to 0.17) or patient–MD/APP (0.07, 95% CI = –0.20 to 0.16) dyads, whereas RN–MD/APP agreement was moderate (0.59, 95% CI = 0.39 to 0.80).

Absent an objective criterion standard for ED frailty, we cannot conclude whether the phenomenon is one of underrating by patients, overrating by providers, or both. Contributing factors may include negative perceptions of the term “frailty” and the fact that it carries different meanings in the lay and medical contexts. Older adults’ inaccurate understanding of their objective health status, particularly among frail individuals, may play a role. Conversely, ED providers may tend focus on patients’ burden of comorbidities, while minimizing patient-important considerations such as functional status and mobility.

Our findings support further study of the CFS in the ED as an alternative to time-intensive
comprehensive geriatric assessments. Subsequent research should continue to explore both staff- and self-rated frailty and examine the relationship between CFS scores assigned at triage and a range of patient- and system-important outcomes, including resource utilization, length of hospital stay, ED revisits, functional changes, and survival.

We recognize the study participants for their contributions as well as the staff of the Schwartz/Reisman Emergency Department at Mount Sinai Hospital, Toronto, Ontario, Canada.

References

Hot Off the Press: The Effect of Financial Incentives on Patient Decisions to Undergo Low-value Head CT Scans

Justin Morgenstern, MD, Corey Heitz, MD, Chris Bond, MD, and William K. Milne, MD

BACKGROUND

The CT scan is arguably the most important piece of diagnostic technology that we use in emergency medicine. It allows for incredibly rapid identification of a myriad of life-threatening conditions. However, likely because it is such a valuable tool, there seems to be little doubt that we overuse it. For example, one study that looked retrospectively at all head CTs ordered for trauma concluded that more than one-third were unnecessary based on the Canadian CT Head Rule. Not only does unnecessary testing reduce efficiency and add costs, it also directly harms patients with unnecessary radiation. Many imaging decisions are obvious—the patient either clearly requires or clearly does not require imaging. However, there is a great deal of uncertainty in emergency medicine, which leaves a sizeable number of patients in a gray zone—where harms and benefits are closely matched, qualitatively different, or just unknown. For these patients, shared decision making is probably the best route forward. Furthermore, even when it seems clear to the physician that imaging is not required, we are often met with resistance from our patients. Thus, it is important to know what factors influence patients’ decisions to undergo CT. This study by Iyengar et al. examines the impacts of financial incentives, as well as varying levels of risk and benefit, on patient preference for CT imaging in the setting of low-risk head injury.

ARTICLE SUMMARY

This is a cross-sectional survey in which participants were presented with a hypothetical low-risk head injury scenario to assess their desire for diagnostic imaging. Participants were randomized to receive different estimates of risk and benefit (1% or 0.1%). They were also randomized to receive a hypothetical offer of either $0 or $100 to forgo imaging. They recruited 913 patients and overall 54% wanted imaging. Desire for CT decreased with lower benefit and higher risk and when money was offered to forgo the CT.

QUALITY ASSESSMENT

This is a clever study examining an interesting question. However, using hypothetical scenarios probably limits external validity, as decisions made while healthy will not necessarily mirror those made when stressed and unwell, especially after a head injury. That being said, it is not clear which represent better decisions: those made while healthy and stress-free or those made in the heat of the moment when facing a high-stress emergency. External validity is further limited in this single-center study.
study by the unique study population, with the majority of the participants being Caucasian and having at least some college education. Furthermore, it is interesting that nearly 25% of the participants worked in health care, which could significantly influence the results of this study. Finally, from the perspective of a physician who has only worked in Canada and New Zealand, the fact that more than half of patients wanted a CT despite the very-low-risk scenario seemed incredibly high, and we wonder whether these results would be replicated in other countries. Finally, it should be noted that the numbers used for harms and benefit in this study were hypothetical and designed to be easy to understand rather than to accurately represent the true harms and benefits of CT. Therefore, although the trends are likely true, the exact numbers would vary in real clinical scenarios.

KEY RESULTS

They enrolled a convenience sample of 913 patients. Overall, despite the low-risk scenario (in which the Canadian CT Head Rule would advise against imaging), 54% of patients stated that they would want a CT scan. A higher benefit of CT resulted in a greater desire for imaging, whereas a higher risk and the offer of a financial incentive decreased desire for imaging. Specifically, if the benefit was reported as 0.1% then 49.6% of people wanted a CT, whereas if it was 1% then 58.9% wanted a CT (odds ratio [OR] = 1.48, 95% confidence interval [CI] = 1.13 to 1.92). If the risk was reported as 0.1% then 59.3% of people wanted a CT, whereas if it was 1% then 49.1% wanted a CT (OR = 0.66, 95% CI = 0.51 to 0.86). Finally, if no cash incentive was offered then 60% of people wanted a CT, whereas if $100 was offered to forgo the CT then 48.3% of people wanted a CT (OR = 0.64, 95% CI = 0.49 to 0.83).

AUTHORS’ COMMENTS

One number really jumped out in this study. In the group of participants who were told that there was only a 0.1% benefit from CT but a 1% harm, half of people still wanted a CT. In other words, despite being explicitly told that the harms of CT were 10 times higher than the benefits, half still opted for the imaging. That is a shocking finding. It may be explained by qualitative differences in the harms and benefits (the harms are delayed, but the benefits are immediate). Or perhaps, despite the excellent efforts of the authors to display information in multiple ways, harms and benefits were simply misunderstood. Or perhaps the number represents informational bias, in which people assume more information is always better, and therefore will always prefer more tests. Whatever the reason, the desire for imaging even when the harms are known to outweigh the benefits is a fascinating finding in a study designed to examine excessive, unnecessary diagnostic imaging.

TOP SOCIAL MEDIA COMMENTARY

Minh Le Cong (@ketaminh)
I don’t agree with paying financial incentive to influence health care decisions. Ethically it’s similar to drug companies giving gifts to influence doctors decisions.

Ken Milne (@theSGEM) responds
We felt similar. Adding in $$$ incentives could create more health inequities and we had concerns about social justice aspect. Medicine is hard enough without having to consider these $$$ which could bias our management. Listen to the podcast.

Pik Mukherji (@ercowboy)
I work in NYC. My experience in a busy ED, with multiple referring services and urgent cares sending people for CT- is NOT that 50% still want one after we chat. 10-15% is generous. “It only takes a 1wk ICU stay (or a 3 hr head CT) to avoid a 15 min. convo.”

Tim Montrief (@EMinMiami)
Big thing that stood out to me (having lived in Ann Arbor for the first 20 some odd years of my life) The vast majority of these pts are highly educated and white. There was also a very high percentage (24%) that worked in healthcare. How might that affect external validity?

Youri Yordanov (@YordaYou)
This is so weird from our side of the atlantic . . . .

Ken Milne (@theSGEM) responds
Also on this side of the Atlantic but north of the US border.

Michael Schweitzer
Would there be a limit, like when the supermarket will only let you buy so many units of something on sale? Because I’d go twice a day to the ER to demand a CT head if this came true. Easy money. Imagine the hordes of folks who show up just to
say “Oh well, if you don’t think I need a CT for this large pimple on my forehead, I’ll just take the cheque and go now.”

Will Meurer responds
It was intended to be a reduction in ones expected copayment. Most US insurers have some form of patient financial contribution for emergency department visits. (For Medicare it is 20% of outpatient visits unless you have some sort of supplemental insurance.)

**TAKE-TO-WORK POINTS**

Even in very-low-risk scenarios, patients demonstrate a desire for advanced diagnostic imaging. When the harms clearly outweigh the risks, we have a responsibility to protect our patients. However, when decisions are not clear cut, it is important to understand the various factors that influence patients’ decisions, so that we are able to guide our patients through an effective shared decision-making process.
References


Adjuvant Antibiotic Therapy After Incision and Drainage of Cutaneous Abscesses

John Conway and Benjamin Friedman, MD

<table>
<thead>
<tr>
<th>NNT color recommendation</th>
<th>Yellow (benefits and harms should be individualized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary heading</td>
<td>Adjuvant antibiotic therapy after incision and drainage of cutaneous abscess is associated with an increased rate of clinical cure</td>
</tr>
</tbody>
</table>
| Benefits in NNT          | 1 in 14 were helped (treatment failure prevented)  
1 in 10 were helped (recurrence prevented) |
| Benefits in percentages  | 7.4% reduced risk of treatment failure  
10.0% reduced risk of recurrence |
| Harms in NNT (NNH)       | 1 in 23 were harmed (experienced adverse events) |
| Harms in percentages     | 4.4% higher risk of adverse events |
| Efficacy endpoints       | Treatment failure, recurrence |
| Harm endpoints           | Adverse events, diarrhea |
| Who was in the studies   | Four studies with 2,406 adults and children undergoing incision and drainage for cutaneous abscess |

NARRATIVE

Annually more than 3 million patients present to U.S. emergency departments (EDs) with cutaneous abscess, a number that has been increasing. Standard treatment involves incision and drainage (I&D), while routine use of systemic antibiotics after I&D is controversial. Recently, two large studies found increased cure rates with systemic antibiotics after I&D compared to placebo. The goal of the systematic review summarized here is to provide updated evidence on the efficacy of systemic antibiotics with activity against methicillin-resistant Staphylococcus aureus (MRSA) after I&D of cutaneous abscess.

The review identified four randomized trials comprised of 2,406 adult and pediatric subjects who presented with acute, simple, cutaneous abscesses that required I&D. Three took place exclusively in the ED and one in a mix of ED and outpatient settings. In three trials, participants were randomized to receive trimethoprim–sulfamethoxazole (TMP-SMX) or placebo while one trial randomized participants to receive TMP-SMX, clindamycin, or placebo. The primary outcome was treatment failure within 21 days based on clinical assessment and the need for further intervention. Secondary outcomes were recurrence, overall adverse events (gastrointestinal symptoms, rashes, and generalized symptoms), and diarrhea.

Antibiotic therapy was associated with an increased rate of clinical cure (absolute risk difference [ARD] = 7.4%, odds ratio [OR] = 2.3, 95% confidence interval [CI] = 1.8 to 3.1; NNT = 14) and a reduced risk of recurrence (ARD = 10%, OR = 0.3, 95% CI = 0.2 to 0.4; NNT = 10). Antibiotic therapy was also associated with an increase in adverse events (ARD = 4.4%, OR = 1.3, 95% CI = 1.1 to 1.6; NNH = 23) but no significant change in diarrhea.

CAVEATS

The quality of evidence was high, the risk of bias was low, and there was no significant heterogeneity. Additionally, another systematic review and meta-analysis,
which included randomized controlled trials of antibiotics without activity against MRSA, reached the same conclusions as the authors of this analysis. There are, however, limitations. One limitation is that I&D technique was not standardized in two of the studies. This is unlikely to have affected the outcome since I&D is a simple procedure and the two studies that did standardize the I&D technique both nonetheless demonstrated a benefit to antibiotics. Another is that two different antibiotics (TMP-SMX and clindamycin) and multiple dosing regimens were used, although the clinical cure rate between antibiotics was not different. Additionally, the studies were not powered to detect rare adverse events such as severe allergic reactions and Clostridium difficile infection. Finally, there was variation in follow-up period with three studies assessing patient outcomes at 7 to 10 days and the fourth study assessing outcomes at 14 to 21 days.

Notably, the clinical cure rate without antibiotics was 84% compared to 92% with antibiotics, and treatment failure rarely results in life-threatening complications or even hospitalization—usually just a return visit with an additional I&D and outpatient antibiotics. The slightly increased clinical cure rate must be balanced against the harms associated with antibiotic use including adverse events and antibiotic resistance. The harms that would be caused to the community by increasing antibiotic resistance may outweigh the benefits to the individual in many cases.

In summary, adjuvant antibiotics given routinely after I&D of cutaneous abscesses were associated with increased clinical cure, decreased recurrence, and increased adverse events in this review. The benefits should be weighed against the adverse events, the cost of treatment failure, and the impact on society of increasing antibiotic usage. Based on the continued necessity for clinicians to weigh the benefits and harms of adjuvant antibiotics, the most appropriate rating is yellow (benefits and harms should be individualized).

References
Accuracy of Point-of-care Ultrasound for Diagnosing Soft Tissue Abscess

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

Summary heading
Point-of-care ultrasound may be helpful in differentiating abscess from cellulitis when there is clinical uncertainty

Positive findings NA
Negative findings NA
Who was in the studies 747 emergency department patients (children and adults) with soft tissue infections

NARRATIVE

Over 3 million cases of skin and soft tissue infections (SSTIs) including cellulitis and abscesses are managed in U.S. emergency departments (EDs) each year.1-4 Overlap in presentations of cellulitis and abscess, which require different therapeutic approaches, has prompted increasing research into point-of-care ultrasound (POCUS) to help differentiate the two.2,5,6

The systematic review summarized here included prospective cohort studies evaluating POCUS for diagnosis of abscess in ED patients.7 The authors of the systematic review included patients with clinical evidence of SSTI. Reference standards varied, typically including draining purulent discharge, computed tomography scan, or clinical follow-up. There were no restrictions with regard to POCUS machine, transducer, protocol, or clinician background. The primary outcome was diagnostic accuracy for abscess in the ED.

The authors identified eight relevant studies (n = 747 patients), with three conducted in adult ED and five in pediatric ED. Calculation of the point estimates for the diagnostic accuracy of POCUS found a sensitivity of 95.5% (95% confidence interval [CI] = 88.9 to 98.3) and specificity of 80.3% (95% CI = 56.4 to 92.7).8

CAVEATS

There are important limitations to the validity of these data. First, patients with cellulitis but initially negative POCUS for abscess may develop abscess later, confounding the reported results. Second, the included studies incorporated various criterion standards for abscess diagnosis due to absence of a definitive criterion. Perhaps more importantly, this review included few studies, all with convenience samples, routine contamination between clinicians and sonographers (for both diagnosis and management decisions), and shifting reference standards.

These methodologic challenges tend to inflate sensitivity and specificity estimates, a concern highlighted by findings from both the largest study in the systematic review and a larger study published after the review.9,10 The largest study included in the analysis, comprising 25% of the review’s sample size, found that POCUS did not add to the diagnostic posttest probability (and may have lowered both sensitivity and specificity) when clinicians felt confident of the diagnosis before ultrasound (i.e., when pretest
probability was high). However, when the pretest prob-
ability was low or moderate, ultrasound was found to
be helpful in increasing the posttest probability.9 Simi-
larly, a large recent study reported that when clinicians
felt certain (>90% of cases) of the diagnosis, ultra-
sound was unhelpful, while in most uncertain cases it
improved accuracy.10

Based on this evidence, the accuracy numbers
reported in the systematic review do not appear reli-
ably valid for typical or common POCUS use in
SSTI. We believe that the diagnostic accuracy of
POCUS is dependent on the pre-test probability of
abscess. POCUS does appear, however, to be poten-
tially helpful in identifying abscess in ED patients
in cases of diagnostic uncertainty. Therefore, we have
assigned a color recommendation of yellow (unclear if
benefits), although we recognize that POCUS is help-
ful in cases with clinical uncertainty after clinical exa-
mination.

REFERENCES

1. Hersh AL, Chambers HF, Maselli JH, Gonzales R.
National trends in ambulatory visits and antibiotic pre-
scribing for skin and soft-tissue infections. Arch Intern
2. Gottlieb M, Schmitz G, Grock A, Mason J. What to do
after you cut: recommendations for abscess management
lines for the diagnosis and management of skin and soft
tissue infections: 2014 update by the Infectious Diseases
4. Taira BR, Singer AJ, Thode HC Jr, Lee CC. National epi-
5. Squire BT, Fox JC, Anderson C. ABSCESS: applied bed-
side sonography for convenient evaluation of superficial
6. Iverson K, Haritos D, Thomas R, Kannikeswaran N. The
effect of bedside ultrasound on diagnosis and management
of soft tissue infections in a paediatric ED. Am J Emerg
In patients presenting to the emergency department with
skin and soft tissue infections what is the diagnostic accu-
racy of point-of-care ultrasonography for the diagnosis of
abscess compared to the current standard of care? A sys-
tematic review and meta-analysis. BMJ Open 2017;7:
e013688.
8. Correction: in patients presenting to the emergency depart-
ment with skin and soft tissue infections what is the diag-
nostic accuracy of point-of-care ultrasonography for the
diagnosis of abscess compared to the current standard of
care? A systematic review and meta-analysis. BMJ Open
2017;7:e013688corr1.
NJ, Alpern ER. Emergency ultrasound assisted examina-
tion of skin and soft tissue infections in the paediatric
emergency department. Acad Emerg Med 2013;20:545–
53.
initial bedside ultrasonography on emergency department
skin and soft tissue infection management. Ann Emerg
Cardiac arrest resuscitations are, in their simplest form, a series of decisions made in rapid succession with little supporting information. If we make enough “right” decisions in the “right” amount of time in the “right” patients, the patient might survive the event. Many variables, like age and comorbidities, cannot be controlled; others, like time of day or being alone, are a matter of luck. But, we control what we can in an effort to maximize outcomes.

Historically, resuscitation science has focused on the cardiac aspects of “cardiopulmonary” resuscitation (CPR), and we resuscitate all pulseless patients in a similar manner. High-quality CPR is characterized by optimizing compression fraction; perfecting chest compression rate, depth, and recoil; and minimizing interruptions including perishock pauses. Dispatch prearrival instructions and community outreach programs have focused on compression-only CPR for adults. This has resulted in relative inattention to the pulmonary components of “cardiopulmonary” resuscitation. We believe that too much ventilation is bad\(^1\) and that it takes relatively little ventilation to match the low-perfusion state of cardiac arrest, but there is not good evidence to determine the best method of airway management; timing of airway management; or ventilation rate, volume, or pressure.\(^1\)

In this issue, Lupton et al.\(^2\) report their secondary analysis of data from the PART trial.\(^3\) The PART trial demonstrated improved survival of adult nontraumatic out-of-hospital cardiac arrest (OHCA) patients who received a blindly inserted laryngeal tube compared to those who underwent endotracheal intubation. In this secondary analysis of the data, the authors report that patients who received only bag–mask ventilation (BMV), rather than a laryngeal tube or endotracheal intubation, had higher survival and better neurologic outcomes. This was true whether BMV was performed primarily or following failed attempts at endotracheal intubation or the laryngeal tube placement.

This is not the first report of BMV versus other airway interventions for prehospital OHCA. Large retrospective studies of Japanese\(^5\) and U.S.\(^6\) cardiac arrest registries also found improved neurologic outcomes with mask ventilation when compared to any invasive method. In addition, a large retrospective study from Korea\(^7\) found equivalent or worse outcomes with invasive airways compared to prehospital BMV. The problem with all of these data (including PART data) is that it comes from trials in which patients were not randomized to BMV versus other interventions, and a randomized trial of intubation versus BMV in France and Belgium was inconclusive.\(^4\)

The difficulty with interpreting data from trials with no randomization to BMV is that OHCA patients preferentially managed by BMV are likely very different from those receiving an invasive airway. These patients may have had either early return of spontaneous circulation (ROSC), before any attempt at an invasive airway, or they had too much muscle tone to allow placement of an advanced airway, and both of these groups would have a higher likelihood of a good outcome. On the other hand, some may have had several failed attempts at invasive airway insertion, which could contribute to a worse outcome.\(^8\)

It is noteworthy that Lupton et al.\(^2\) and Hasegawa et al.\(^5\) both found improved outcomes in patients for whom BMV was used after failed invasive airway insertion compared to patients with successful invasive airway insertion. This raises the question of whether there is something inherently better about BMV for OHCA. As the authors state, “ventilation characteristics” may be responsible; in other words, the effectiveness of ventilation may impact outcomes as much as
the device chosen to provide ventilations. We believe that there is plausibility to such a theory; it is quite possible that an advanced airway is simply a better conduit to deliver volume, and therefore overventilation, compared to BMV. The inherent difficulty with BMV may ironically prove lifesaving.

The elephant in the room for many clinicians is the concern for aspiration, leading to prioritization of endotracheal tube (ETT) over extraglottic device (EGD) over BMV. Both PART and AIRWAYS-2 demonstrated equivalent aspiration rates among patients primarily managed with an EGD or an ETT, suggesting either that an EGD provides as much aspiration protection as an ETT or that patients had already aspirated prior to airway management. We have very little data comparing aspiration rates of patients managed with only BMV compared to advanced airways. In PART, the aspiration rate for patients with BMV only was lower than that for patients managed with a BMV after failed intubation attempts, but again patient characteristics in these two groups may have been very different.

We should acknowledge that every resuscitation poses unique challenges, and prehospital providers must individually decide between endotracheal intubation, BMV, and use of an EGD. The initial decision is multifactorial, as is the second (or third) choice if the initial strategy fails. Even in the setting of a clinical trial, paramedics in PART chose not to establish an advanced airway in 12% of enrolled subjects (352/3,004) and abandoned attempts in another 6% (189/3,004). Many subjects without an advanced airway attempt were noted to have ROSC or jaw clenching. Although we do not have specific information about why paramedics abandoned advanced airway attempts and reverted to BMV, the fact that these patients had better outcomes highlights the need to recognize patient-level differences and allow for provider decision making. Despite rigorous attempts to adjust for differences in subject populations in the PART trial, provider decision making remained an unmeasured/unmeasurable confounder.

In the end, as the authors note, unknown confounders of these data make the current findings best suited for hypothesis generation. Differences in subjects, differences in provider decision making about when to place the device and when to abort attempts, and the small number of patients managed with BMV limit generalizability of these data. Further, we know little about how patients were actually ventilated beyond what device was used. We believe that trials of airway management in OHCA focusing on any airway device or technique will continue to provide conflicting results until patient-centered physiologic variables are actually measured, especially tidal volume delivery, tissue oxygenation, and cerebral perfusion. We know that in other clinical settings bag masks, laryngeal tubes, laryngeal mask airways, and ETTs can all provide adequate ventilation and oxygenation, so there is really no reason to believe that in the setting of OHCA one of these devices is a poison (harmful or deadly) and the other is pixie dust (magical or lifesaving).

We should guard against changing our practice based on this type of preliminary data. The situation is reminiscent of preliminary data raising concerns about carotid artery and jugular vein compression by an EGD, which has since been debunked. What should EMS medical directors, administrators, providers and educators do with these results until further prospective studies are completed? For now, do not stop using advanced airways if they are part of your guidelines but do not get tunnel vision on them either. Such a strategy is consistent with the latest AHA update. The focus of cardiac arrest resuscitation should remain on early defibrillation, high-quality CPR, and probably early epinephrine, at least for non-shockable rhythms. The initial oxygenation strategy could either be passive, via a non-rebreather mask for suspected sudden cardiac events (aka cardio-cerebral resuscitation), or BMV for suspected primary respiratory arrests. While the timing of advanced airway placement is uncertain, once positive pressure ventilation is initiated, the focus should be on ways to limit harm from over-ventilation, whether via BMV, EGD or ETT. Current harm-reducing strategies include educational interventions, on-going quality assurance and continuous quality improvement, choice of self-inflating bag size, choice of self-inflating bag grip, and possibly use of a mechanical ventilator.

Jason T. McMullan, MD, MS
Darren A. Braude, MD
(dbraude@salud.unm.edu)
1 Department of Emergency Medicine, University of Cincinnati, Cincinnati, OH
2 Departments of Emergency Medicine and Anesthesiology, University of New Mexico Health Sciences Center, Albuquerque, NM

Supervising Editor: Rob Reardon, MD
References


Fifty-one percent of the U.S. population is female.\textsuperscript{1} Thirty-six percent of professionally active physicians are female.\textsuperscript{2} Twenty-seven percent of active emergency medicine physicians are female.\textsuperscript{3}

All of us graduated medical school with the hope our future career in medicine would be fulfilling. For many of us, that became true though for most of us, it has not been easy. We knew it was not going to be easy. But back in medical school, it was the medicine that seemed the most daunting. We did not fully realize the extramedical challenges that lay ahead.

Most of us have had our credentials questioned, either overtly or insidiously. “When will I see the doctor” ask our patients, expecting a male provider or, even more commonly, “Are you the nurse?” This occurs even after introducing ourselves as “doctor” while wearing our white coat with a badge saying “doctor.” These daily acts of mistaken identity serve as microaggressions, which can chip away at us, furthering a nascent inclination toward imposter syndrome. Though we are not imposters, our own colleagues sometimes further undermine our professional status. Women routinely are not afforded formal professional acknowledgement compared to their male colleagues. Even in the context of prestigious events such as grand rounds, women introduced by men are less likely to be addressed by their professional titles than men introduced by men.\textsuperscript{4}

Many of us are underpaid. Female academic physicians are paid nearly 10\% less than their male counterparts, even after controlling for clinical time worked and years in practice.\textsuperscript{5} Over time, this difference becomes accentuated. This occurs not just at an institutional level but at a systemic level and often despite awareness of the inequity. Despite previously published data showing an inappropriate gender salary gap in emergency medicine, this gap has remained essentially unchanged over the past 4 years.\textsuperscript{6} Women are literally valued less than men.

Many of us balance work and family.\textsuperscript{7} In fact, the “balancing” act begins well before the decision to have a family. Most of us have been indirectly or directly questioned regarding our plans for childbearing or childrearing, as if we are human resources ticking time bombs about to explode with fecundity. For those of us who decide to have a family, many of us lose career opportunities based on these decisions. “You have a lot on your plate right now” is one of many euphemisms used to suggest that a young family may not be compatible with a successful career, at least for a female partner.

Most of us do not enjoy the 12-week paid maternity leave recommended by the American Academy of Pediatrics. For female faculty at the top 12 U.S. medical schools, the mean paid childbearing leave was 8.6 weeks in 2015.\textsuperscript{8} And the challenges continue after returning to clinical work after childbearing leave. Accommodations for supporting breast-feeding are often inadequate—both in time set aside for pumping and in physical space. Paid family leave for spouses (e.g., paternity leave), for non–birth parents, for same-sex couples, and for adoptive parents is even more precarious. Most of us understand the irony that medical caregivers are not robustly supported in caregiving for their own families.

Many of us shoulder a disproportionate share of household duties. In one study, women spent 9 hours more per week than their male colleagues on domestic activities.\textsuperscript{9} Even in dual-physician couples, weekly professional hours worked by women with children are lower than those worked by women without children. Similar differences were not observed among men.\textsuperscript{9} For those of us in academic medicine, this likely contributes to a slower pace toward academic

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

Author contributions: RM and LL conceived of the presented idea; RM drafted the first draft which was revised by LL; and both authors contributed to the final manuscript.

ISSN 1553-2712 © 2019 by the Society for Academic Emergency Medicine doi: 10.1111/acem.13903
advancement and promotion. Women comprise 39% of full-time faculty in U.S. medical schools, but only 21% of full professors, 15% of department chairs, and 16% of deans. The proportion of women full professors increased from 14% in 2003 to 21% in 2013, but continues to be substantially lower than the 79% of male full professors. Emergency medicine is not immune to these realities. Female physicians comprise only 28% of the academic emergency medicine workforce and are significantly less likely to be associate or full professors when compared to their male colleagues, even after adjustment for factors associated with academic advancement.

Very few of us are women of color. As of 2014, only 7.9% of female medical school graduate are black and only 5.6% are Hispanic or Latino. Black, Hispanic, and Native American female doctors continue to be underrepresented in emergency medicine and their respective proportions of trainees have not changed substantially in the past 20 years. Why is that?

Some of us are at grave risk of self-harm. The suicide rates for female physicians are more than twice that of the general population. Female physicians report higher rates of burnout compared to their male colleagues. Depressive symptoms increase substantially during the internship year for both men and women, but this increase is greater for women.

All of us should ask ourselves, “What can we do?” None of us can afford to passively wait for gaps in pay, promotion, and mental health to close themselves.

We have to acknowledge women are being treated unequally in subtle and not so subtle ways. There is growing literature underscoring the challenges of doctoring while woman. While evidence suggests female physicians bring significant value to patient care, diversity in the emergency medicine workforce is important for our patients and our colleagues alike. Not only are female patients more likely to survive an acute myocardial infarction (AMI) when treated by a female physician, but male physicians are more effective at treating female AMI patients when they work with more female colleagues.

We have to promote the worth of female emergency medicine physicians. We have to change the culture allowing women to be undervalued. We have some suggestions.

- Increase institutional awareness of bias (conscious and unconscious), discrimination, and harassment to demonstrate support for female physicians in the workplace.
- Examine the salaries of males and females in the practice and equalize them based on their professional time and experience.
- Promote transparency in hiring guidelines, promotion guidelines, and salary guidelines to address disparities in these arenas.
- Enact mandatory paid parental leave for both parents to promote wider cultural changes in caregiving norms.
- Establish family-friendly policies, such as flexible hours and work-from-home options for nonclinical professional work. This could help address the disproportionate burden women physicians bear for childrearing.
- Adopt programs supporting domestic needs, including affordable, easily accessible childcare (e.g., daycare in the hospital or clinic), emergency childcare provisions, completing errands (e.g., picking up dry cleaning), and meal preparation and delivery programs.

None of us believes that these challenges are unique to emergency medicine but most of us believe that our specialty can and should lead the medical field in proactively addressing the challenges of doctoring while woman. We are a forward-thinking, innovative discipline accustomed to solving urgent problems. Half a decade ago, the American College of Emergency Physicians (ACEP) put forth a policy statement designed to maximize the potential of women in emergency medicine. Have you read the statement? It recommends that employers adopt policies and practices that will enable women to have productive and sustained careers. Does your workplace engage in the recommended “best practices” that will enable women to have productive and sustained careers in emergency medicine?

All of us, women and men, must work together to increase women in leadership and in mentorship roles to ultimately bring about a necessary culture shift in the field of emergency medicine in particular and medicine as a whole. No change is easy. But acknowledging the challenges and supporting policy changes is a step toward supporting women working in the field of medicine. Working toward a culture where women garner equal respect, equal compensation, and equal opportunities for advancing in our careers will not only benefit women, but will also benefit our patients and colleagues. All of us deserve the chance to have the fulfilling careers we envisioned in medical school.
References


Routine Opt-out Syphilis Screening in the Emergency Department: A Public Health Imperative

The United States is experiencing a resurgence in syphilis cases, with a 71% increase in new diagnoses in the past 5 years. Left untreated, syphilis can lead to a host of medical complications, as well as increase the risk of HIV infection. Congenital syphilis rates are rising even faster than adult syphilis rates, with a 185% increase from 2014 to 2018, a change that has paralleled increases in primary and secondary syphilis among reproductive-aged women. This dramatic rise in syphilis cases may be due in part to a “lack of screening by physicians [and] patients not asking for screening.”

Given this rapidly growing domestic syphilis epidemic, new strategies are required for identifying those at highest transmission risk. Populations at highest risk for syphilis are often also those who may not have access to primary care. The highest rates of syphilis, for example, are seen in young black men between the ages of 20 and 34. Congenital syphilis is found in black infants at more than six times the rate of white infants and almost twice the rate of Latinx. These populations often rely on the emergency department (ED) for their medical needs. Visits to the ED thus represent a key opportunity to stem the spread of syphilis through screening, treatment, and referral.

Screening programs in the ED work. Routinized HIV screening, for example, is increasingly a component of ED care and is supported by the CDC as a successful screening intervention strategy. With the appropriate infrastructure, these programs can effectively reach a large population that would otherwise be missed through traditional screening recommendations and can be implemented with minimal disruption to workflow. Our ED, for example, leverages the electronic medical record (EMR) to select patients who are appropriate candidates for screening and automatically prompt nurses in triage to order the screening tests for these patients. This type of screening infrastructure, which already exists in many EDs, can easily be modified to add syphilis screening with essentially no additional end-user effort required. As a quality improvement project in our ED, syphilis testing was bundled with the HIV screening prompt, leading to an increase from an average of 250 syphilis tests per month in the ED to more than 1200 monthly tests after implementation of the EMR-driven protocol, with high rates of positivity.

To further decrease resource utilization in an ED context, we recommend a reverse sequence testing algorithm for syphilis screening. Screening is performed using an antibody test that reflexes to confirmatory testing if positive, allowing for detection of all stages of infection and reducing false positives. Given the difficulties of locating patients once they leave the ED, particularly in the populations who utilize the ED as their primary source of care, point-of-care syphilis tests that result while the patient is in the ED are also showing promise as a way to ensure that patients are properly treated and not lost to follow-up.

Through a partnership with outpatient providers in a diversity of insurance contexts, patients testing positive for syphilis in the ED can be brought into clinic for treatment irrespective of ability to pay. In our hospital, a dedicated infectious diseases staff member reviews all positive tests and contacts patients to schedule outpatient treatment. During follow-up visits, these patients can be screened for HIV and other sexually transmitted infections, if this screening was not done.
during the ED visit, and they can be provided with counseling, including HIV preexposure prophylaxis (PrEP) initiation if appropriate. Equally important, these patients are directly linked to primary care during this visit. Through this multidisciplinary model, screening for syphilis during routine ED visits may lead not only to a reduction in syphilis rates but also to a reduction in new cases of HIV, as well as engagement in key primary care services.

Emergency departments in syphilis-prevalent regions should strongly consider implementing universal screening efforts. In the context of the current dramatic rise in the number of syphilis cases, it is imperative that we find novel means of reaching those at highest risk. As these populations disproportionately utilize the ED for medical care, ED screening programs are a logical space for engagement. With automated ordering algorithms and an infrastructure in place to ensure prompt follow-up of positive cases, screening of ED patients has the potential to significantly improve the diagnosis and treatment of syphilis in the most impacted communities with minimal disruption to ED workflow, potentially averting not just serious complications of the disease but also cases of congenital syphilis and HIV. Such an approach represents an important strategy to address a major public health issue that impacts all of our patients.

Kimberly A. Stanford, MD, MPH
(kstanford@medicine.bsd.uchicago.edu)
Aniruddha Hazra, MD
John Schneider, MD, MPH
University of Chicago, Chicago, IL

Supervising Editor: John H. Burton, MD

References

Quarantined

As an emergency medicine physician, I am used to being at work when most people are at home. At the end of February, the first patient in the United States to be diagnosed with COVID-19 without a travel history occurred in Sacramento, CA. Eighty-nine healthcare professionals who had contact with the patient were quarantined. At the beginning of March, I worked in the emergency department and treated a patient with viral symptoms without a travel history who was admitted to the hospital. It was also the first day the hospital implemented an internal laboratory test that was FDA approved to detect novel coronavirus.

The next morning, I checked the chart and saw that the patient tested positive for novel coronavirus with the new lab test. It confirmed the growing suspicion that the virus was being transmitted within the community. I immediately started thinking of what I had planned for the day, which included signing closing paperwork on our first house. I spoke with the county health department who said that CDC used a different test and they were not familiar with the hospital test. They said to wait for the official county result for the patient and that I did not need to be quarantined. I let the title company know I had a potential exposure and postponed the signing for the next day.

Over the next few hours, occupational health told me to be in a 14-day quarantine and to stay 6 feet away from my family; my family did not need to be quarantined. Then occupational health said my family needed to be quarantined. The CDC physician working with the county health department said I was under quarantine, but my family was not. Hospital infection control said I probably did not need quarantine since I was wearing a procedure mask when I saw the patient. When my kids came home, I asked them to keep 6 feet away from me. My 5-year old daughter thought I was playing a joke and came to me for a hug and I put out my hand and said, "Sorry, no hugs for now." My family ate dinner at the dining room table, while I ate standing in the kitchen.

The second morning after exposure, my 5-year old daughter asked me to play checkers with her. We found a way to play while staying 6 feet away from each other. She would move a piece and step away from the board, then I would move my piece and step away. After breakfast, I spoke with the county health department to clarify my quarantine status who said a physician would call me back. By noon, a CDC physician who specializes in healthcare professionals exposed to COVID-19 told me that the recommendations would be changing again. Given the high rate of community transmission in the area, I no longer needed to be quarantined. The title company also called the county health department and was told I was still under quarantine. After multiple phone calls, we designated power of attorney to my sister to sign the closing paperwork. The title company insisted on mailing the paperwork to my sister rather than have her come into the office to sign.

By late afternoon, I received a call from occupational health who confirmed I no longer needed to be quarantined. I did a risk assessment and decided it was probably safe to eat dinner with my family at the same dining table. I am off quarantine and my next shift in the emergency department will be in a few days. As I approach the 14-day mark, I realize that I may have not gotten COVID-19 from this exposure but eventually I probably will.

Moon O. Lee, MD, MPH
(moonlee@stanford.edu)
Department of Emergency Medicine
Stanford University School of Medicine
Stanford, CA

Supervising Editor: Brian J. Zink, MD
Emergency medicine physicians are supposed to be outgoing gregarious risk takers, right? That’s the perception of our specialty when you watch television shows like “ER” or “Grey’s Anatomy.” Maybe because we are always in the public’s eye at the workplace, we are expected to be interested in that same type of public engagement in other settings.

Having just returned from a large emergency medicine annual conference, where frequent social interactions with colleagues was the norm, I am reminded that I am an introvert in the often extroverted world of emergency medicine. I genuinely enjoyed the conference both for the learning opportunities and for the chance to meet friends and colleagues, both old and new. However, at the end of each day, I felt absolutely exhausted. It was as if I had run this mental marathon and I was one of the last people to cross the finish line, huffing and puffing along the way.

Introverts thrive from the energy within, whereas extroverts are fueled from the spark of others. Susan Cain’s book *Quiet: The Power of Introverts in a World That Can’t Stop Talking* notes that extroverts are action driven, excel when faced with conflicts, and do well among large groups of people. It seems like the extrovert is an ideal fit for the specialty of emergency medicine.

In medical school, extroverts generally do better than introverts during clinical rotations. Extroverts are more likely to speak in groups, which is helpful for problem-based learning sessions and rounds on the wards. In these settings, introverts may seem disinterested or unengaged. I struggled with this myself in medical school, always thinking for prolonged periods before the rare occasion in which I would nervously speak up in class.

Fortunately, books such as Cain’s have made it easier to admit that you are an introvert. There is not much written about physicians and introversion, particularly in the specialty of emergency medicine. Most physicians likely possess a combination of introversion and extroversion traits. According to Cain, both introverts and extroverts have positive leadership traits that can promote success in the academic and clinical settings. While extroverts are good at leading by inspiring a vision for a team, introverts’ leadership skills shine through the implementation of input gathered from each individual within that team.

Networking and peer interaction are important aspects of our specialty that can promote the attainment of leadership positions and advancement in academic emergency medicine. We must interact to form new partnerships and collaborations. For the introvert, achieving these networking goals is not always natural, but instead can involve baby steps.

I know that there are many more introverts out there in emergency medicine at all levels of training and in all practice environments. If we don’t speak up during grand rounds or at a departmental meeting, it doesn’t mean we are not interested, we are just taking a long time to think about what we want to say. At networking events, if you don’t see us actively meeting new people, it doesn’t mean we are arrogant or bored; it just means we feel much more comfortable with a one-on-one interaction. If we leave the event early, it just means we need to quietly recharge in our own comfortable space.

Because networking is not second nature, introverts are less likely than extroverts to actively seek mentors. Introverts may therefore benefit from more formally established mentorship relationships. Mentors can be valuable in helping introverts improve presentation and self-advocacy skills and in aiding career development. Since introverts may be relatively quiet in meetings, our voices may get lost. Mentors can give introverts a comfortable platform to express valuable opinions and to help us become more effective communicators in the group.
Mentors can also facilitate placement of introverts in roles that accentuate our unique strengths and help us showcase our skills for senior administrators to encourage promotion to leadership positions.6

I find extroverts delightful, especially their energy and ability to engage others. But introverts are pretty cool too, it just may take a little more time to get to know us. There is room for both in our specialty.

Janice Blanchard, MD, PhD
(jblanchard@mfa.gwu.edu)
Department of Emergency Medicine,
George Washington University,
Washington, DC

Supervising Editor: Brian Zink, MD

References
