EDITOR’S PICK

Opioid Use and Misuse Three Months After Emergency Department Visit for Acute Pain
Raoul Daoust, Jean Paquet, Sophie Gosselin et al. 847

ORIGINAL CONTRIBUTIONS

Do Financial Incentives Change Length-of-stay Performance in Emergency Departments? A Retrospective Study of the Pay-for-performance Program in Metro Vancouver
Yuren Wang, Yichuan Ding, Eric Park et al. 856

Randomized Placebo-controlled Trial of Droperidol and Ondansetron for Adult Emergency Department Patients With Nausea
Robert Meek, Michaela J. Mee, Diana Egerton-Warburton et al. 867

Depression in Emergency Department Patients and Association With Health Care Utilization
David G. Beiser, Charlotte E. Ward, Milkie Vu et al. 878

Psychiatric Outcomes of Patients With Severe Agitation Following Administration of Prehospital Ketamine
Jacob A. Lebin, Arvin R. Akhavan, Daniel S. Hippe et al. 889

Drinking and Intimate Partner Violence Severity Levels Among U.S. Ethnic Groups in an Urban Emergency Department
Raul Caetano, Carol B. Cunradi, Harrison J. Alter et al. 897

Variation in Emergency Department Adherence to Treatment Guidelines for Inpatient Pneumonia and Sepsis: A Retrospective Cohort Study
Stacy A. Trent, Zachary J. Jarou, Edward P. Havranek et al. 908

Contents continued inside.
A Prospective, Multicenter Evaluation of Point-of-care Ultrasound for Small-bowel Obstruction in the Emergency Department
Brent A. Becker, Shadi Lahham, Mark A. Gonzales et al. 921

SYSTEMATIC REVIEWS (WITH OR WITHOUT META-ANALYSES)

Point-of-Care Ocular Ultrasound for the Diagnosis of Retinal Detachment: A Systematic Review and Meta-Analysis
Michael Gottlieb, Dallas Holladay, Gary D. Peksa 931

RESEARCH LETTERS

Underdosing of Benzodiazepines in Patients With Status Epilepticus Enrolled in Established Status Epilepticus Treatment Trial
Abhishek G. Sathe, Holly Tillman, Lisa D. Coles et al. 940

Emergency Department Risk Stratification After Opiate Overdose Is Just the Beginning
Brian M. Clemency, Joshua J. Lynch, Terrance Creighton et al. 944

HOT OFF THE PRESS

Hot Off the Press: A Systematic Review And Meta-analysis of Ketamine as an Alternative to Opioids for Acute Pain in the Emergency Department
Corey Heitz, Justin Morgenstern, Christopher Bond et al. 946

THE BRASS TACKS: CONCISE REVIEWS OF PUBLISHED EVIDENCE

Aspirin for Preventing a First Heart Attack or Stroke
Kristopher Roach, Michael Ritchie, Shahriar Zehtabchi 949

Early Endovascular Thrombectomy for Large-vessel Ischemic Stroke Reduces Disability at 90 Days
Mona Al Banna, Christopher D. Streib 953

COMMENTARY—UN SOLICITED

The Folly of the R: A Case Study
Robert M. Rodriguez 956

MEDIA REVIEW

Sura Khuder, Bethany J. Figg, Mary Jo Wagner 959

REFLECTIONS

An Old Man’s Love
Siu Fai Li 962
CME Information: Opioid Use and Misuse Three Months After Emergency Department Visit for Acute Pain

CME Editor: Corey Heitz, MD

Authors: Raoul Daoust, MD, MSc, Jean Paquet, PhD, Sophie Gosselin, MD, Gilles Lavigne, DMD, PhD, Alexis Cournoyer, MD, Eric Piette, MD, MSc, Judy Morris, MD, MSc, Véronique Castonguay, MD, Justine Lessard, MD, and Jean-Marc Chauny, MD, MSc

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

Educational Objectives
After reading the article, participants should be able to discuss the incidence of opioid use and the reasons for consuming these opioids three months after being discharged from the ED with an opioid prescription.

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.
Opioid Use and Misuse Three Months After Emergency Department Visit for Acute Pain

Raoul Daoust, MD, MSc, Jean Paquet, PhD, Sophie Gosselin, MD, Gilles Lavigne, DMD, PhD, Alexis Cournoyer, MD, Eric Piette, MD, MSc, Judy Morris, MD, MSc, Véronique Castonguay, MD, Justine Lessard, MD, and Jean-Marc Chauny, MD, MSc

ABSTRACT

Background: Studies evaluating long-term prescription opioid use are retrospective and based on filled opioid prescriptions from governmental databases. These studies cannot evaluate if opioids were really consumed and are unable to differentiate if they were used for a new pain or chronic pain or were misused. The aim of this study was to assess opioid use rate and reasons for consuming 3 months after being discharged from the emergency department (ED) with an opioid prescription.

Methods: This is a prospective cohort study conducted in the ED of a tertiary care urban center with a convenience sample of discharged patients ≥ 18 years who consulted for an acute pain condition (≤ 2 weeks). Three months post-ED visit, participants were interviewed by phone on their past 2-week opioid consumption and their reasons for consuming: a) for pain related to the initial ED visit, b) for a new unrelated pain, or c) for another reason.

Results: Of the 524 participants questioned at 3 months (mean ± SD age = 51 ± 16 years, 47% women), 47 patients (9%, 95% confidence interval [CI] = 7%–12%) reported consuming opioids in the previous 2 weeks. Among those, 34 (72%) reported using opioids for their initial pain, nine (19%) for a new unrelated pain and four (9%) for another reason (0.8%, 95% CI = 0.3%–2.0%, of the whole cohort). Patients who used opioids during the 2 weeks after the ED visit were 3.8 (95% CI = 1.2–12.7) times more likely to consume opioids at 3 months.

Conclusion: Opioid use at the 3-month follow-up in ED patients discharged with an opioid prescription for an acute pain condition is not necessarily associated with opioid misuse; 91% of those patients consumed opioids to treat pain. Of the whole cohort, less than 1% reported using opioids for reasons other than pain. The rate of long-term opioid use reported by prescription-filling database studies should not be viewed as a proxy for incidence of opioid misuse.

Opioids are often a significant part of the management of moderate to severe acute pain after an emergency department (ED) visit. Since prescription opioid misuse, dependence, overdose, and death have all increased to epidemic proportions in the past 15 years, ED practitioners are now faced with the

From the Department of Emergency Medicine, Research Centre (RD, JP, AC, EP, JM, VC, JL, JMC), the Département de Médecine Familiale et Médecine d’Urgence, Faculté de Médecine (RD, AC, EP, JM, VC, JL, JMC), and the Faculties of Dental Medicine and Medicine (GL), Université de Montréal; the Department of Emergency Medicine, McGill University Health Centre, McGill University (SG); Trauma, Research Centre (GL), and the Centre for Advanced Research in Sleep Medicine (GL), Hôpital du Sacré-Coeur de Montréal (CIUSSS du Nord de-l’Île-de-Montréal), Montréal, Québec, Canada.

Received August 16, 2018; revision received September 26, 2018; accepted September 30, 2018.
Presented at the Society for Academic Emergency Medicine Annual Meeting, Indianapolis, IN, May 2018.
This study was supported by the Sacré-Coeur Hospital’s emergency medicine research funds. Clinical trial registration number: NCT02799004, ClinicalTrials.gov PRS.

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

Author contributions: RD and JMC conceived the study and obtained research funding; all authors contributed to the final protocol and data interpretation; JP was responsible for data management and statistical analyses; RD drafted the manuscript; AL, EP, JM, SG, GL, JL, and VC contributed substantially to its revision; and all authors approved the final manuscript as submitted and have agreed to be accountable for all aspects of the work.

Supervising Editor: Mark B. Mycyk, MD.
Address for correspondence and reprints: Raoul Daoust, MD, MSc; e-mail: raoul.daoust@videotron.ca.

ACADEMIC EMERGENCY MEDICINE 2019;26:848–855.

ISSN 1553-2712 © 2019 by the Society for Academic Emergency Medicine
doi: 10.1111/acem.13628
dilemma of managing acute pain without contributing to the opioid misuse epidemic.\textsuperscript{6,7}

Although short-term opioid treatment for pain was previously not believed to cause addiction in itself,\textsuperscript{8} this belief has recently been challenged. For example, Butler et al.\textsuperscript{9} found that 11 of 59 (19\%) heroin and nonmedical opioid users were initiated to opioids by prescriptions that originated from an ED visit. However, that study examined a very small convenience sample, of which 80\% were already suffering from substance use disorder prior to the initial opioid exposure, suggesting that these patients were already at high risk of opioid use disorder.

Rates of long-term opioid use after ED visits for acute pain were evaluated in three studies. Hoppe et al.\textsuperscript{10} found that 12\% of patients who received an opioid prescription at ED discharge were recurrent opioid users at the 1-year follow-up. Barnett et al.\textsuperscript{11} showed that opioid use at 12 months was higher among patients treated by high-intensity prescribers compared to low-intensity prescribers. Another study evaluating long-term opioid use (>10 fills or 120 days supplied over 1 year) in ED opioid-naive patients showed variations in rates according to the insurance status: 1.1\% for commercial insurance, 3.1\% for “aged” Medicare, and 6.2\% for “disabled” Medicare.\textsuperscript{12}

In contexts other than the ED, a study of workers’ compensation claims for acute low back pain revealed that 10.4\% of them received five or more opioid prescriptions during the 30-day to 2-year follow-up.\textsuperscript{13} In a previous study on an elderly trauma population, we also found an incidence of 10.9\% of patients still using opioids after 1 year.\textsuperscript{14} In a follow-up of postoperative pain, 3.1\% of the population continued to receive opioids 3 months after major surgery.\textsuperscript{15} However, all these previous studies evaluating rates of long-term opioid use were retrospective and based on filled opioid prescriptions from pharmaceutical/governmental databases. Therefore, they cannot evaluate if opioids were really consumed and are unable to differentiate whether they were used for a new pain or chronic pain from the original event or were misused. Misuse has often been described as the use of opioids in a way that is divergent from the prescriber’s directions or in an inappropriate way; however, this is still a matter of debate.\textsuperscript{16} In this study, we defined opioid misuse as use of prescription opioids for another reason than pain.\textsuperscript{17}

Considering that recurrent or long-term opioid use established from filled prescription databases studies is sometimes used as a surrogate to opioid misuse,\textsuperscript{10} we sought to determine in a prospective study the incidence of opioid use and the reasons for consuming these opioids 3 months after being discharged from the ED with an opioid prescription.

\section*{METHODS}

\subsection*{Study Design and Setting}

This is a prospective cohort study conducted in the ED of a Canadian academic Level I trauma center with an affiliated emergency medicine residency program and an annual census of approximately 65,000 ED visits (mostly adults). This research is a planned follow-up study of an observational cohort study\textsuperscript{18} of patients who received an opioid prescription for acute pain after an ED visit. It involved separate follow-up interviews to obtain additional data on opioid use and misuse at 3 months. Approval was obtained from the local institutional ethics review board.

\subsection*{Selection of Participants}

We recruited patients aged 18 years and older treated in the ED during the period of June 2016 to July 2017 in a 24/7 manner. We included all patients with a pain condition present for less than 2 weeks (usual definition for acute pain)\textsuperscript{19} and discharged from the ED with an opioid prescription. The selection of patients to receive an opioid prescription was not predetermined; usual pain management (e.g., medications, immobilization, physiotherapy, follow-up visits) was left to the treating physician’s preference. It is usual practice in Canada to prescribe acetaminophen and opioids separately so patients can easily optimize daily acetaminophen dosage before adding an opioid. However, some physicians still prescribe combinations of opioids and acetaminophen. Patients who received an opioid prescription were identified by all ED physicians and then a research nurse verified patients’ eligibility, explained the study and obtained informed consent. This was a convenience sample as we were not able to determine the number of patients not identified by ED physicians (no electronic tracking system for outpatient prescriptions). We excluded only patients who did not speak French or English, were using opioid medication prior to the ED visit (past 2 weeks), stayed in the ED for more than 48 hours (patients with ED stay > 48 hours and discharged home represented less than 0.5\% of the population),
or were suffering from cancer or were treated for chronic pain.

**Measurements**

Patients’ demographic information, pain intensity at triage, arrival mode, triage priority, and length of ED stay were extracted from our computerized medical system. ED physicians entered the final diagnosis, pain intensity at discharge, and which pain medications were prescribed. Two weeks after the initial ED visit, patients were asked by phone if they had filled their initial opioid prescription and if they had consumed opioids in the previous 2 weeks. Three months after the ED visit, patients were contacted (at least three times) by phone again and were asked questions about their pain medication use in the previous 2 weeks. The 3-month follow-up time was chosen since it generally defines the start of chronic pain. After patients were presented with a list of all the prescription opioids available in the market, they were asked (in English or French at patient’s preference): “Have you used one of these opioids in the last two weeks?” If yes, patients were then asked: “In the following choices, which one better describes the reason why you took these opioids?": a) for pain related to the initial ED visit, b) for a new unrelated pain, or c) for another reason. Misuse has often been described as the use of opioids in a way that is divergent from the prescriber’s directions or in an inappropriate way, but it is still a matter of debate. In this study, we defined opioid misuse as “use of prescription opioids for a reason other than pain,” because it is less susceptible to interpretation bias. We did not ask to further elaborate on the other reason for consuming opioids to avoid the social desirability bias that could arise from that type of question. Study data were collected and managed using REDCap (Research Electronic Data Capture), a secure, Web-based application tool hosted in the hospital.

**Data Analysis**

To compare the different opioid forms, each initial opioid prescription was transformed into an oral morphine 5-mg tablet equivalent, using the method of Berdine and Nesbit. Dosages of 3.33 mg of oxycodone and 1.25 mg of hydromorphone were considered equipotent to one morphine 5-mg tablet.

The study sample size was estimated based on the 12% rates of opioid use observed at the 1-year follow-up by Hoppe et al. To have a precise estimate of the incidence of opioid use at a 3-month follow-up with a margin of error of 3% in determining a confidence interval (CI) of 95% for an estimate rate of 12%, a minimum of 451 patients was required.

Incidences of opioid use, and misuse (not using opioids for pain) 3 months after the ED visit, were calculated with the 95% CI. One-way ANOVA, chi-square, and Kruskal-Wallis tests were used depending on the normality of variables, to compare baseline characteristics of included, refusing to participate, and lost to follow-up patients. Post hoc tests adjusting for multiple comparisons were performed when main effects were significant. t-test, chi-square, and Mann-Whitney U-tests were used to compare baseline characteristics between patients who consumed opioids at 3 months and those who did not. For this comparison, Cohen’s effect size were also reported; small, medium, and large effect sizes for chi-square and Mann-Whitney U-tests are 0.1, 0.3, and 0.5, respectively, and 0.2, 0.5, and 0.8, respectively, for the t-test statistic. Alpha level was set at 0.05, and all statistics were performed using SPSS version 23 (IBM Corp.).

**RESULTS**

All ED physicians participated in the identification of eligible patients. A total of 1,316 patients meeting the inclusion criteria were initially contacted. Of these, 29% had exclusion criteria, 13% declined to participate, and 18% could not be reached for the 3-month follow-up, leaving 524 participants (Figure 1). Included patients and those who refused to participate or were lost at 3 months were almost similar on all baseline characteristics (Table 1). Patients who refused to participate were older than those lost to follow-up, but neither group was significantly different from included patients. Patients’ mean (±SD) age was 51 (±16) years, 47% were female, and mean pain intensity at triage was 7.7, decreasing to 5.2 at ED discharge. At discharge, patients received a prescription for a median of 30 tablets of 5 mg of morphine (or equivalent), 94% of them filled it, and 79% of them consumed opioids during the first 2-week period after the ED visit.

In the 524 participants questioned at 3 months, 47 patients (9%, 95% CI = 7%–12%) reported consuming opioids in the previous 2 weeks. Baseline characteristics were similar among patients who consumed opioids at the 3-month follow-up and those who did not. However, patients who used opioids during the 2 weeks after the ED visit were 3.8 (95% CI = 1.2–
12.7) times more likely to consume opioids at 3 months compared to those who did not (Table 2). Among the 47 opioid users at 3 months, 34 (72%) said that they consumed opioids for their initial pain, nine (19%) for a new unrelated pain, and four (9%) for another reason. In these four patients (0.8% of the whole cohort, 95% CI = 0.3%–2.0%) who consumed opioids for reasons other than pain, all of them had consumed opioids during the 2 weeks following the initial ED visit (Table 3).

**DISCUSSION**

This prospective study showed that 9% of patients discharged from the ED with an acute pain condition still consumed opioids 3 months later and 91% of them did so to manage pain (72% initial pain, 19% new pain). Furthermore, of the whole cohort, less than 1% consumed opioids for other reasons than pain, suggesting misuse.

The 9% incidence of opioid use at 3 months observed in this study is slightly lower than the long-term opioid use reported 1 year after ED discharge in another study (12%),\(^\text{10}\) in patients suffering from acute low back pain (10.4%),\(^\text{13}\) or in elderly trauma patients (10.9%).\(^\text{14}\) However, these studies used filled prescriptions databases that could overestimate opioid use since not all patients filling an opioid prescription consumed them. As a case in point, in this study, 21% of patients who filled their opioid prescription after the initial ED visit did not consume them.

Our results also show that patients who used opioids 2 weeks after their ED visit were more likely to consume opioids at 3 months. History of opioid use has also been associated with long-term opioid use in orthopedic trauma,\(^\text{26}\) surgery,\(^\text{27}\) and general trauma populations.\(^\text{14}\) However, this could be explained by ongoing pain that will eventually evolve to chronic pain.

Among the opioid users at 3 months, almost all patients (91%) consumed opioids for treating pain (initial or new pain). Considering the entire sample of 524 patients questioned at 3 months, nearly 7% of patients still took opioids at 3 months to treat their initial pain. This proportion could represent a population of patients with chronic pain. However, the prevalence of chronic pain reported in this study could very well be higher because some patients may have used medications other than opioids (or other methods altogether) to treat their ongoing pain. Indeed, postsurgical\(^\text{28}\) or traumatic pain studies\(^\text{29}\) usually report higher prevalence of chronic pain.

Less than 1% of patients reported consuming opioids for reasons other than pain, suggesting that these patients are possible opioid misusers. This proportion of misuse could be underestimated since we relied on self-report which are affected by social desirability and 18% of patients could not be reached for the 3-month follow-up. However, we asked if opioids were consumed for something other than pain, which should be less influenced by social desirability than asking directly if they abuse opioids. Interestingly, three of the four patients (75%) consuming opioids for reasons other than pain were aged over 50, while opioid abuse and overdose are generally observed in subjects under 50.\(^\text{30}\)

Recurrent or long-term opioid use rates established from filled prescription databases studies are sometimes used as a surrogate to opioid misuse.\(^\text{10}\) However, these rates included patients with chronic or new developing pain which constitutes, according to our results, the majority of long-term opioid users. Within the limit of our study, our results suggest that the risk of long-term opioid use for reasons other than pain is low for ED discharged patients with an opioid prescription treating an acute pain condition.

**LIMITATIONS**

This study has limitations. The convenience sample from one ED center limits the generalization of our
results as a selection bias could exist. However, patients were recruited 24/7, and consecutive recruitment was limited only by the fact that the investigators could not determine the number of patients missed by ED physicians (no electronic tracking system for outpatient prescriptions). The rate of patients who refused to participate and who were lost to the 3-month follow-up was important (44%) but the baseline characteristics of these patients were similar to those of patients included in the study. Self-reported opioid use could be biased by social desirability issues or by misunderstanding questions. However, questions were relatively easy to understand and studies have shown self-reports of illicit substance use to be valid relative to urine drug screen.31–33 Some Table 2 results should be interpreted with caution. The relatively small number of patients using opioid at 3 months limits our power to identify specific patient characteristics that could be significant. For example, the percentage of female, abdominal pain conditions and patients receiving prescriptions of acetaminophen or morphine seem to be higher for patients using opioids at 3 months. Using larger prospective samples and electronic questionnaires to estimate opioid use should be considered in the future since several reviews support the reliability and validity of self-reporting risk behaviors when privacy is assured and when assessments are self-administered and computerized.34–36

<table>
<thead>
<tr>
<th>Baseline Characteristics</th>
<th>Included (N = 524)</th>
<th>Refused to Participate (n = 176)</th>
<th>Lost to Follow-up (n = 238)</th>
<th>p-value for Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (±SD)</td>
<td>50.8 (15.8)</td>
<td>53.1 (18.6)*</td>
<td>48.5 (16.5)</td>
<td>0.01†</td>
</tr>
<tr>
<td>Female (%)</td>
<td>46.9</td>
<td>53.4</td>
<td>47.5</td>
<td>0.32</td>
</tr>
<tr>
<td>ED arrival mode (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>By himself</td>
<td>80.5</td>
<td>75.6</td>
<td>78.6</td>
<td>0.37</td>
</tr>
<tr>
<td>By ambulance</td>
<td>19.5</td>
<td>24.4</td>
<td>21.4</td>
<td>0.78</td>
</tr>
<tr>
<td>High (level 1 or 2) triage priority (%)</td>
<td>43.6</td>
<td>45.5</td>
<td>42.0</td>
<td>0.78</td>
</tr>
<tr>
<td>Pain intensity (0–10 scale) at triage, mean (±SD)</td>
<td>7.7 (±2.0)</td>
<td>7.9 (±1.9)</td>
<td>8.1 (±1.9)</td>
<td>0.11§</td>
</tr>
<tr>
<td>ED treatment section (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambulatory</td>
<td>65.6</td>
<td>62.5</td>
<td>63.4</td>
<td>0.71</td>
</tr>
<tr>
<td>On stretcher</td>
<td>34.4</td>
<td>37.5</td>
<td>36.6</td>
<td>0.71</td>
</tr>
<tr>
<td>Type of pain conditions (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal pain</td>
<td>40.5</td>
<td>41.5</td>
<td>48.9</td>
<td>0.14</td>
</tr>
<tr>
<td>Fracture</td>
<td>18.5</td>
<td>23.3</td>
<td>18.3</td>
<td>0.14</td>
</tr>
<tr>
<td>Renal colic</td>
<td>17.0</td>
<td>18.8</td>
<td>16.6</td>
<td>0.14</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>6.7</td>
<td>3.4</td>
<td>5.2</td>
<td>0.14</td>
</tr>
<tr>
<td>Other</td>
<td>17.4</td>
<td>13.1</td>
<td>10.9</td>
<td>0.14</td>
</tr>
<tr>
<td>Acetaminophen† prescription at ED discharge (%)</td>
<td>71.4</td>
<td>69.3</td>
<td>72.3</td>
<td>0.80</td>
</tr>
<tr>
<td>NSAIDs prescription at ED discharge (%)</td>
<td>45.2</td>
<td>42.6</td>
<td>51.3</td>
<td>0.17</td>
</tr>
<tr>
<td>Opioid prescription type at ED discharge (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphine</td>
<td>43.1</td>
<td>42.3</td>
<td>45.0</td>
<td>0.67</td>
</tr>
<tr>
<td>Oxycodone</td>
<td>41.3</td>
<td>36.6</td>
<td>36.6</td>
<td>0.67</td>
</tr>
<tr>
<td>Hydromorphone</td>
<td>15.6</td>
<td>21.1</td>
<td>18.5</td>
<td>0.67</td>
</tr>
<tr>
<td>Morphine 5-mg equivalent pills prescription, median (Q1–Q3)</td>
<td>30 (20–48)</td>
<td>30 (20–45)</td>
<td>30 (19–48)</td>
<td>0.43§</td>
</tr>
<tr>
<td>ED stay (hours), median (Q1–Q3)</td>
<td>5.2 (3.6–7.5)</td>
<td>5.8 (3.9–8.2)</td>
<td>5.0 (3.4–7.9)</td>
<td>0.22§</td>
</tr>
<tr>
<td>Pain intensity (0–10 scale) at ED discharge, mean (±SD)</td>
<td>4.7 (±2.8)</td>
<td>4.4 (±2.8)</td>
<td>5.1 (±2.8)</td>
<td>0.08†</td>
</tr>
</tbody>
</table>

NSAIDs = nonsteroidal anti-inflammatory drugs; Q1–Q3 = first and third quartiles.
*Patients who refused to participate were older than patients lost to follow-up.
†p-value for one-way ANOVA.
‡Acetaminophen was always prescribed separately from opioids.
§p-value for Kruskal-Wallis test.
Table 2
Baseline Characteristics of Opioid Users and Nonusers at the 3-month Follow-up

<table>
<thead>
<tr>
<th>Baseline Characteristics</th>
<th>Opioid Use at 3 Months (n = 47)</th>
<th>No Opioid Use at 3 Months (n = 477)</th>
<th>p-value for Chi-square</th>
<th>Cohen’s Effect Size*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (±SD)</td>
<td>50.3 (±13.8)</td>
<td>50.9 (±16.0)</td>
<td>0.82†</td>
<td>0.04</td>
</tr>
<tr>
<td>Female (%)</td>
<td>59.6</td>
<td>45.7</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>ED arrival mode (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>By himself</td>
<td>83.0</td>
<td>80.3</td>
<td>0.65</td>
<td>0.02</td>
</tr>
<tr>
<td>By ambulance</td>
<td>17.0</td>
<td>19.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (level 1 or 2) triage priority (%)</td>
<td>38.3</td>
<td>44.1</td>
<td>0.44</td>
<td>0.03</td>
</tr>
<tr>
<td>Pain intensity (0–10 scale) at triage, mean (±SD)</td>
<td>8.0 (±2.0)</td>
<td>7.7 (±2.0)</td>
<td>0.40‡</td>
<td>0.14</td>
</tr>
<tr>
<td>ED treatment section (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambulatory</td>
<td>70.2</td>
<td>65.1</td>
<td>0.48</td>
<td>0.03</td>
</tr>
<tr>
<td>On stretcher</td>
<td>29.8</td>
<td>34.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of pain conditions (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal pain</td>
<td>38.3</td>
<td>40.7</td>
<td>0.16</td>
<td>0.06</td>
</tr>
<tr>
<td>Fracture</td>
<td>19.1</td>
<td>18.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal colic</td>
<td>10.6</td>
<td>17.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>14.9</td>
<td>5.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>17.0</td>
<td>17.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetaminophen† prescription at ED discharge (%)</td>
<td>80.9</td>
<td>70.4</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td>NSAIDs prescription at ED discharge (%)</td>
<td>40.4</td>
<td>45.7</td>
<td>0.49</td>
<td>0.03</td>
</tr>
<tr>
<td>Opioid prescription type (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphine</td>
<td>55.3</td>
<td>41.9</td>
<td>0.45</td>
<td>0.06</td>
</tr>
<tr>
<td>Oxycodone</td>
<td>29.8</td>
<td>42.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydromorphone</td>
<td>14.9</td>
<td>15.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphine 5-mg equivalent pills prescription, median (Q1–Q3)</td>
<td>30 (20–60)</td>
<td>30 (20–45)</td>
<td>0.76‡</td>
<td>0.01</td>
</tr>
<tr>
<td>Filled their initial opioid prescription (%)</td>
<td>100</td>
<td>93.7</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>Consumed opioids during the first 2-week after ED visit (%)</td>
<td>93.0</td>
<td>77.7</td>
<td>0.02</td>
<td>0.11</td>
</tr>
<tr>
<td>ED stay (hours), median (Q1–Q3)</td>
<td>5.5 (4.2–7.7)</td>
<td>5.2 (3.6–7.5)</td>
<td>0.50§</td>
<td>0.03</td>
</tr>
<tr>
<td>Pain intensity (0–10 scale) at ED discharge, mean (±SD)</td>
<td>4.7 (±2.9)</td>
<td>4.7 (±3.0)</td>
<td>0.99†</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

NS = not significant; NSAIDs = nonsteroidal anti-inflammatory drug; Q1–Q3 = first and third quartiles.
*Small, medium, and large effect sizes for chi-square and Mann-Whitney U-tests are 0.1, 0.3, and 0.5, respectively, and for the t-test statistic, 0.2, 0.5, and 0.8, respectively, and for the t-test statistic, 0.2, 0.5, and 0.8, respectively.
†p-value for t-test.
‡Acetaminophen was always prescribed separately from opioids.
§p-value for Mann-Whitney U-test.

Table 3
Individual Data of the Four Patients Who Used Opioids at 3 Months for Reasons Other Than Pain

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age (Years)</th>
<th>Pain Condition</th>
<th>Pain at Triage</th>
<th>Pain at ED Discharge</th>
<th>Opioid* Type</th>
<th>Quantity* of M5E Pills Prescribed</th>
<th>Received* Acetaminophen Prescription</th>
<th>Received* NSAIDs Prescription</th>
<th>Filled Initial Opioid Prescription</th>
<th>Opioid use 2 Weeks After ED Visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>60</td>
<td>Zona</td>
<td>8</td>
<td>5</td>
<td>Morphine</td>
<td>30</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>F</td>
<td>53</td>
<td>Musculoskeletal</td>
<td>10</td>
<td>3</td>
<td>Hydromorphone</td>
<td>32</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>M</td>
<td>60</td>
<td>Abdominal pain</td>
<td>7</td>
<td>0</td>
<td>Morphine</td>
<td>40</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>F</td>
<td>37</td>
<td>Abscess</td>
<td>NA</td>
<td>0</td>
<td>Oxycodone</td>
<td>8</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

M5E = oral morphine 5-mg pill equivalent; NA = not available.
*At the initial ED visit.
CONCLUSION

In summary, opioid use at the 3-month follow-up in ED patients discharged with an opioid prescription for acute pain is relatively low (9%) and not necessarily synonym of opioid misuse; most of those patients (91%) still consumed opioids for pain. Of the whole cohort, less than 1% of the patients consumed opioids at 3 months for reasons other than pain. The rate of long-term opioid use reported by filled prescription database studies should not be used as a surrogate for opioid misuse.

The authors would like to thank Martin Marquis and Dominique Petit for their contributions to manuscript revision.

References


Do Financial Incentives Change Length-of-stay Performance in Emergency Departments? A Retrospective Study of the Pay-for-performance Program in Metro Vancouver

Yuren Wang, MS, Yichuan Ding, PhD, Eric Park, PhD and Garth Hunte, MD, PhD

ABSTRACT

Background: Pay-for-performance (P4P) programs have been implemented in various forms to reduce emergency department (ED) patient length of stay (LOS). This retrospective study investigated to what extent the timing of patient disposition in Metro Vancouver EDs was influenced by a LOS-based P4P program.

Methods: We analyzed ED visit records of four major hospitals in Metro Vancouver, Canada. For each ED, we individually tested whether LOS was distributed discontinuously at the LOS target before and after the P4P program was terminated. For the P4P effective period, we examined whether patients discharged just prior to the LOS target had a higher 7-day return-and-admission (RA) rate—the probability that a patient, after being discharged home, returned to any ED within 7 days and was admitted to an inpatient unit—than patients discharged just after the target.

Results: Prior to the termination of the P4P program, in all four EDs, the LOS density of admitted patients was discontinuous and had a significant drop at the P4P 10-hours admission LOS target; a similar phenomenon was observed among discharged patients at the 4-hours discharge LOS target, but only in the two lower-volume EDs. Furthermore, in a lower-volume ED, patients who were discharged right before the 4-hours P4P LOS target had a higher 7-day RA rate than patients discharged right after the LOS target. After the termination of the discharge incentive, the discontinuity at the discharge LOS target became less evident, but patients were still more frequently admitted just before 10 hours in three of the four EDs as the local health authority continued to support the admission incentive scheme after the government terminated the P4P program.

Conclusions: The LOS-based financial incentive scheme appears to have influenced the timing of ED patient dispositions. The results suggest mixed consequences of the P4P program—it can reduce access block for admitted patients but may also lead to discharges associated with return visits and admissions.

In recent years, emergency department (ED) overcrowding has presented a big challenge in many countries. Longer length of stay (LOS), particularly longer ED boarding time, has been associated with higher mortality and worse health outcomes. In recent years, pay-for-performance (P4P) programs in...
various forms have been implemented in several regions and countries as part of an effort to reduce ED LOS. However, the extent to which these LOS-based P4P programs affect the LOS of individual ED patients is not well understood. The existing literature builds analysis on aggregate performance metrics such as mean or median LOS and thus provides limited insights into the particulars of how EDs react to P4P programs. For example, Vermeulen et al. studied a P4P program implemented in Ontario since 2008 and found modest reduction in the average LOS in the participating EDs. Cheng and Sutherland studied a P4P program initiated in 2007 in British Columbia and reported mixed impacts of the P4P program on average LOS in different EDs. Other studies on similar P4P programs were either based on interviews or anecdotal evidence or focused on other possible consequences such as increased cost.

In this retrospective study, we took a granular approach to investigate the impact of an ED LOS P4P program. Instead of focusing on the average LOS, we looked into the distribution of the LOS, which better captured the nuances of patient dispositions decisions. In particular, we examined whether the LOS distribution had a continuous density near the LOS targets specified by the P4P program. A decreasingly discontinuous density means significantly more patients are discharged/admitted right before the LOS targets than after, suggesting that patient disposition may have been timed to meet the LOS targets. Although our analysis was limited to EDs in the Metro Vancouver area, the insights we gained on how ED performance changed in response to a LOS-based P4P program may be of great interest to policy makers in other regions/countries where P4P programs with similar structure are implemented.

We studied EDs in the Metro Vancouver area where the same P4P program as studied by Cheng and Sutherland was implemented. The pilot P4P program was initiated in 2007 by the British Columbia provincial government with the objective of reducing ED patient LOS. It was then implemented on a rolling basis in all EDs in British Columbia and terminated on March 31, 2014. EDs participating in the program received a $100 compensation for each discharged patient with a LOS less than 4 hours if the patient was of triage level 1, 2, or 3 and less than 2 hours if the patient was of triage level 4 or 5. EDs also received $600 for each admitted patient with LOS less than 10 hours regardless of triage level.

METHODS

Study Design
The study was approved by the Behavioural Research Ethics Board, Office of Research Services, the University of British Columbia (UBC BREB no. H16-00303).

Setting
There were 813,491 patient visits to the four major EDs in the metro Vancouver area recorded in the National Ambulatory Care Reporting System data in the period from April 1, 2013, to March 31, 2016. The study period can be divided into two phases: pretermination (from April 1, 2013, to March 31, 2014, when the P4P program was in effect in all the four study EDs) and posttermination (from April 1, 2014, to March 31, 2016, when the P4P program was terminated and no longer in effect). However, even after April 1, 2014, the regional health authority governing all four study EDs decided to internally fund the exact same admission incentive scheme, $600 per admitted patient with LOS less than 10 hours, and continue without interruption. Only the discharge incentive had completely disappeared postgovernment P4P policy termination.

For each patient visit, the data included the patient’s demographic information (age and sex), arrival mode (walk in or by ambulance), chief complaint system (CCS), chief complaint description, triage acuity code, disposition decision (discharged or admitted), time of arrival, time to see a doctor, and the LOS in the ED. We classified patients into discharged and admitted by the disposition decision.

In all four EDs, triage level 4 and 5 patients were treated in a separate area from triage level 1, 2, and 3 patients. We focused on the higher-acuity patients in triage levels 1, 2, and 3, with an average admission rate of 24.0% in the pretermination period and 21.8% in the posttermination period.

Regardless of triage level, the study population had a consistent P4P LOS target—4 hours for discharged patients and 10 hours for admitted patients. In April 2013 to March 2014, the policy effective period, two EDs (A and B) had patient volumes (for all triage levels) of less than 60,000 patients per year (lower volume), while the other two EDs (C and D) had annual patient volumes of greater than 75,000 patients (higher volume).

Table 1 summarizes the statistics for triage level 1, 2, and 3 patients in both the pretermination and
posttermination periods. In particular, it presents the average admission rate and number of visits in the time intervals near the 4-hours P4P discharge LOS target, which are relevant to the analysis.

Measurement

Our primary outcome was the LOS distribution as the P4P program directly monitored individual ED patient LOS. LOS refers to the duration between arrival and disposition from the ED, either admitted or discharged. In either case, the patient had to be physically out of the treatment area of the ED at the disposition time. For secondary outcome, we chose the 7-day return and admission (RA) rate—the probability that a patient, after being discharged home, returned to any ED within 7 days and was admitted to an inpatient unit—to measure the care quality impact of the P4P program. We chose the 7-day time window as it covers about half of the revisits that took place within 30 days of a prior discharge. We also calculated and compared the 30-day RA rate as a sensitivity test. Note that we only counted revisits that ended up being admitted to inpatient units, considering that those patients may have been discharged earlier than ideal and did not receive adequate treatment during the initial visit.

Data Analysis

First, we directly observed whether the P4P program had affected patients’ LOS by plotting the pre- and posttermination empirical LOS distribution of the discharged and admitted patients by individual EDs and aggregated data. Next, we rigorously tested whether the ED patient LOS distribution had discontinuous density at the P4P LOS targets, we used the hypothesis testing method suggested by McCrary. A detailed description of the discontinuity test method can be found in Data Supplement S1 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13635/full).

We tested the null hypothesis of continuous density for the discharged and admitted patient groups separately and tried both the 4- and the 10-hours P4P

<table>
<thead>
<tr>
<th>Phase</th>
<th>Statistic</th>
<th>LOS Interval (min)</th>
<th>ED A</th>
<th>ED B</th>
<th>ED C</th>
<th>ED D</th>
<th>All EDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretermination</td>
<td>Average admission rate</td>
<td>200–220</td>
<td>9.14%</td>
<td>6.00%</td>
<td>5.49%</td>
<td>11.85%</td>
<td>8.36%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>220–240</td>
<td>9.82%</td>
<td>7.13%</td>
<td>7.13%</td>
<td>14.79%</td>
<td>10.07%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240–260</td>
<td>20.24%</td>
<td>12.44%</td>
<td>9.02%</td>
<td>16.36%</td>
<td>14.47%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All</td>
<td>22.14%</td>
<td>20.70%</td>
<td>18.89%</td>
<td>30.84%</td>
<td>23.99%</td>
</tr>
<tr>
<td></td>
<td>Average number of daily</td>
<td>200–220</td>
<td>4.53</td>
<td>4.93</td>
<td>5.29</td>
<td>6.66</td>
<td>21.40</td>
</tr>
<tr>
<td></td>
<td>visits</td>
<td>220–240</td>
<td>5.64</td>
<td>4.65</td>
<td>4.96</td>
<td>6.32</td>
<td>21.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240–260</td>
<td>2.69</td>
<td>2.75</td>
<td>3.83</td>
<td>5.73</td>
<td>15.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All</td>
<td>87.03</td>
<td>82.88</td>
<td>114.92</td>
<td>148.62</td>
<td>433.45</td>
</tr>
<tr>
<td></td>
<td>Average LOS of discharged</td>
<td></td>
<td>259.88</td>
<td>251.30</td>
<td>244.71</td>
<td>272.69</td>
<td>257.87</td>
</tr>
<tr>
<td></td>
<td>patients (min)</td>
<td></td>
<td>757.74</td>
<td>806.86</td>
<td>729.58</td>
<td>681.34</td>
<td>726.28</td>
</tr>
<tr>
<td></td>
<td>Total number of visits</td>
<td></td>
<td>31,765</td>
<td>30,250</td>
<td>41,947</td>
<td>54,247</td>
<td>158,209</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase</th>
<th>Statistic</th>
<th>LOS Interval (min)</th>
<th>ED A</th>
<th>ED B</th>
<th>ED C</th>
<th>ED D</th>
<th>All EDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttermination</td>
<td>Average admission rate</td>
<td>200–220</td>
<td>7.16%</td>
<td>5.35%</td>
<td>4.71%</td>
<td>11.02%</td>
<td>7.36%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>220–240</td>
<td>8.79%</td>
<td>6.06%</td>
<td>5.97%</td>
<td>12.70%</td>
<td>8.67%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240–260</td>
<td>10.45%</td>
<td>9.62%</td>
<td>8.06%</td>
<td>14.17%</td>
<td>11.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All</td>
<td>19.01%</td>
<td>19.38%</td>
<td>17.41%</td>
<td>28.57%</td>
<td>21.75%</td>
</tr>
<tr>
<td></td>
<td>Average number of daily</td>
<td>200–220</td>
<td>6.08</td>
<td>5.12</td>
<td>6.45</td>
<td>7.81</td>
<td>25.45</td>
</tr>
<tr>
<td></td>
<td>visits</td>
<td>220–240</td>
<td>5.74</td>
<td>4.92</td>
<td>6.19</td>
<td>7.18</td>
<td>24.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240–260</td>
<td>4.65</td>
<td>3.45</td>
<td>4.81</td>
<td>6.81</td>
<td>19.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All</td>
<td>106.77</td>
<td>88.41</td>
<td>134.11</td>
<td>158.98</td>
<td>488.27</td>
</tr>
<tr>
<td></td>
<td>Average LOS of discharged</td>
<td></td>
<td>282.68</td>
<td>277.82</td>
<td>248.51</td>
<td>285.86</td>
<td>272.81</td>
</tr>
<tr>
<td></td>
<td>patients (min)</td>
<td></td>
<td>946.76</td>
<td>866.76</td>
<td>676.74</td>
<td>634.64</td>
<td>741.01</td>
</tr>
<tr>
<td></td>
<td>Total number of visits</td>
<td></td>
<td>78,052</td>
<td>64,626</td>
<td>98,034</td>
<td>116,217</td>
<td>356,929</td>
</tr>
</tbody>
</table>

LOS = length of stay.
LOS target for both patient groups as the discontinuous point. We applied the above test to all patients in the four study EDs as a whole. We also ran a separate test for each individual ED by anticipating that different EDs (even in the same geographic region) may have responded differently to the P4P program. In addition, we performed falsification tests at 2, 3, and 5 hours for discharged patients and 6, 8, and 12 hours for admitted patients, respectively, to validate that the discontinuities were indeed led by the LOS targets set by the P4P program.

Since it is possible that some patients were discharged prematurely to meet the discharge LOS target, we also evaluated the impact of those decisions on care outcomes. To that end, we followed the literature\textsuperscript{14,16} and measured patient outcomes by the 7-day RA rate. We compared the average RA rate of patients discharged between 220 and 240 minutes, just before the 4-hours LOS target, to those discharged during the 20-minute time windows before and after that interval—200 to 220 and 240 to 260 minutes, respectively. We focused on discharges during the interval of 220 to 240 minutes because that was the interval where the LOS density exhibited a spike in ED A, where the LOS discontinuity was most significant. We also compared the average 30-day RA rate in the three intervals and obtained similar results. So our conclusion is robust.

**RESULTS**

**Empirical LOS Distribution**

The empirical LOS distribution of the discharged and admitted patients showed evidence of the P4P program’s impact in the pretermination period (Figure 1, left and right columns, respectively). In the left column, for the lower-volume EDs, A and B, we observed sharp discontinuity in the LOS densities for discharged patients at the 4-hours P4P discharge target, whereas for the higher-volume EDs, C and D, such discontinuity was not evident. In the right column for admitted patients, we observed significant discontinuity in all four EDs at the 10-hours admission target. In EDs A, B, and D, the discontinuity was more prominent, with a spike just before the 10-hours target.

Figure 2 presents the empirical LOS distribution of discharged and admitted patients in the posttermination period. For all EDs, there was no discontinuity for discharged patients at 4 hours. This supports that the discontinuities at the discharge LOS target that we observed for ED A and B in the pretermination period were led by the P4P program. Nevertheless, the 10-hours admission target is still effective in most EDs (except ED C) after the termination of the P4P program.

**Discontinuous LOS Density at P4P Targets**

The detailed discontinuous density test results for patient LOS distribution are summarized in Table 2. For example, we found that the discontinuity estimate \( \hat{\rho} \) (a measure of change in density; see Data Supplement S1) for ED A, discharged patients at 4 hours in the pretermination period was \( -0.4111 \) and statistically significant at the 0.1% level (\( p < 0.001 \)), suggesting that the LOS density discontinuously changed at 4 hours. The negative sign of \( \hat{\rho} \) indicates a drop in density, which suggests that there were fewer patients discharged right after the 4-hours target compared to just before the target.

In the pretermination period, discharged patients had statistically significant discontinuous LOS density at their respective target of 4 hours in the two lower-volume EDs, A and B, while admitted patients had discontinuous LOS density at 10 hours in all four EDs. Each significant discontinuity corresponded to a downward jump, which was consistent with the empirical distribution (Figure 1). Note that for the pretermination period, we also tested discontinuity at 2, 3, 5, and 10 hours for discharged patients and at 4, 6, 8, and 12 hours for admitted patients (see Table 2 and Table S1 in Data Supplement S1) but did not find any significant discontinuity at those points. This supports the fact that the discontinuity was likely led by the P4P program rather than rounding behavior at the integer hours. In the posttermination period, we found that discharged patients in none of the four EDs had significantly discontinuous LOS density at 4 hours while admitted patients still had discontinuous LOS density at 10 hours in three of the four EDs, A, B, and D.

**LOS Density by CCS and Congestion Level**

Since the LOS distribution varied across different CCS codes, 19 categories recorded at the clinical department level, we further investigated whether the discontinuity in discharged patients from EDs A and B was led by a subgroup of patients or the entire patient population. We answered this question by looking into the LOS distribution for the eight most frequent CCS categories: cardiovascular, gastrointestinal, general and minor, genitourinary, neurologic, orthopedic, respiratory, and skin. Indeed, for all eight CCS categories, we found significant discontinuity at the 4-hours LOS
Figure 1. Length-of-stay (LOS) distribution for discharged and admitted patients by ED in pretermination period. Plots are truncated at LOS of 800 minutes. Plots on average cover 90.28% of the patient population. According to our conversation with the ED administrators, when the nurses entered a patient’s LOS into the electronic record system, sometimes they rounded it to full hour value for convenience, which likely led to the spikes at the end of each hour (most notable in ED C, discharged). Those spikes, however, would not affect the discontinuity test results as we compared the density function strictly before and after the LOS targets without using the frequency at the exact target times. [Color figure can be viewed at wileyonlinelibrary.com]
Figure 2. Length-of-stay (LOS) distribution for discharged and admitted patients by ED in posttermination period. Plots are truncated at LOS of 800 minutes. Plots on average cover 93.62% of the patient population. [Color figure can be viewed at wileyonlinelibrary.com]
Table 2
Discontinuous Density Test Results for ED Patient LOS Distribution

<table>
<thead>
<tr>
<th>Phase</th>
<th>ED</th>
<th>Disposition</th>
<th>Target Time (hr)</th>
<th>Discontinuity Estimator</th>
<th>p-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretermination</td>
<td>A</td>
<td>Discharged</td>
<td>4</td>
<td>-0.4111*</td>
<td>&lt;0.0001</td>
<td>(-0.5979 to -0.2243)</td>
</tr>
<tr>
<td>A</td>
<td>Admitted</td>
<td>4</td>
<td>0.3581</td>
<td></td>
<td>0.2874</td>
<td>(-0.3016 to 1.0177)</td>
</tr>
<tr>
<td>A</td>
<td>Discharged</td>
<td>10</td>
<td>0.1908</td>
<td></td>
<td>0.7189</td>
<td>(-0.8480 to 1.2295)</td>
</tr>
<tr>
<td>A</td>
<td>Admitted</td>
<td>10</td>
<td>-1.7668*</td>
<td></td>
<td>0.0001</td>
<td>(-2.6610 to -0.8725)</td>
</tr>
<tr>
<td>B</td>
<td>Discharged</td>
<td>4</td>
<td>-0.4578*</td>
<td></td>
<td>0.0005</td>
<td>(-0.7164 to -0.1991)</td>
</tr>
<tr>
<td>B</td>
<td>Admitted</td>
<td>4</td>
<td>-0.0715</td>
<td></td>
<td>0.8631</td>
<td>(-0.8835 to 0.7406)</td>
</tr>
<tr>
<td>B</td>
<td>Discharged</td>
<td>10</td>
<td>-0.0149</td>
<td></td>
<td>0.9777</td>
<td>(-1.0597 to 1.0298)</td>
</tr>
<tr>
<td>B</td>
<td>Admitted</td>
<td>10</td>
<td>-2.1861†</td>
<td></td>
<td>0.0028</td>
<td>(-3.6182 to -0.7539)</td>
</tr>
<tr>
<td>C</td>
<td>Discharged</td>
<td>4</td>
<td>-0.0595</td>
<td></td>
<td>0.5934</td>
<td>(-0.2780 to 0.1590)</td>
</tr>
<tr>
<td>C</td>
<td>Admitted</td>
<td>4</td>
<td>-0.0150</td>
<td></td>
<td>0.9684</td>
<td>(-0.7598 to 0.7297)</td>
</tr>
<tr>
<td>C</td>
<td>Discharged</td>
<td>10</td>
<td>0.5528</td>
<td></td>
<td>0.1926</td>
<td>(-0.2789 to 1.3844)</td>
</tr>
<tr>
<td>C</td>
<td>Admitted</td>
<td>10</td>
<td>-0.6859‡</td>
<td></td>
<td>0.0340</td>
<td>(-1.3198 to -0.0520)</td>
</tr>
<tr>
<td>D</td>
<td>Discharged</td>
<td>4</td>
<td>-0.1125</td>
<td></td>
<td>0.2771</td>
<td>(-0.3154 to 0.0904)</td>
</tr>
<tr>
<td>D</td>
<td>Admitted</td>
<td>4</td>
<td>-0.0421</td>
<td></td>
<td>0.8610</td>
<td>(-0.5128 to 0.4287)</td>
</tr>
<tr>
<td>D</td>
<td>Discharged</td>
<td>10</td>
<td>0.2542</td>
<td></td>
<td>0.4477</td>
<td>(-0.4021 to 0.9106)</td>
</tr>
<tr>
<td>D</td>
<td>Admitted</td>
<td>10</td>
<td>-1.1089*</td>
<td></td>
<td>&lt;0.0001</td>
<td>(-1.6117 to -0.6061)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Discharged</td>
<td>4</td>
<td>-0.1665‡</td>
<td></td>
<td>0.0021</td>
<td>(-0.2727 to -0.0604)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Admitted</td>
<td>4</td>
<td>-0.0693</td>
<td></td>
<td>0.6589</td>
<td>(-0.3772 to 0.2385)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Discharged</td>
<td>4</td>
<td>0.3023</td>
<td></td>
<td>0.1525</td>
<td>(-0.1118 to 0.7164)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Admitted</td>
<td>10</td>
<td>-0.6417*</td>
<td></td>
<td>&lt;0.0001</td>
<td>(-0.9485 to -0.3349)</td>
</tr>
<tr>
<td>Posttermination</td>
<td>A</td>
<td>Discharged</td>
<td>4</td>
<td>-0.0858</td>
<td>0.1040</td>
<td>(-0.1893 to 0.0177)</td>
</tr>
<tr>
<td>A</td>
<td>Admitted</td>
<td>4</td>
<td>0.0982</td>
<td></td>
<td>0.5465</td>
<td>(-0.2210 to 0.4174)</td>
</tr>
<tr>
<td>A</td>
<td>Discharged</td>
<td>10</td>
<td>-0.2845</td>
<td></td>
<td>0.2149</td>
<td>(-0.7342 to 0.1651)</td>
</tr>
<tr>
<td>A</td>
<td>Admitted</td>
<td>10</td>
<td>-1.9197*</td>
<td></td>
<td>&lt;0.0001</td>
<td>(-2.3861 to -1.4533)</td>
</tr>
<tr>
<td>B</td>
<td>Discharged</td>
<td>4</td>
<td>-0.0831</td>
<td></td>
<td>0.1694</td>
<td>(-0.2016 to 0.0354)</td>
</tr>
<tr>
<td>B</td>
<td>Admitted</td>
<td>4</td>
<td>-0.0025</td>
<td></td>
<td>0.9900</td>
<td>(-0.3971 to 0.3920)</td>
</tr>
<tr>
<td>B</td>
<td>Discharged</td>
<td>10</td>
<td>0.0061</td>
<td></td>
<td>0.9820</td>
<td>(-0.5257 to 0.5379)</td>
</tr>
<tr>
<td>B</td>
<td>Admitted</td>
<td>10</td>
<td>-1.2266‡</td>
<td></td>
<td>0.0131</td>
<td>(-2.1955 to -0.2576)</td>
</tr>
<tr>
<td>C</td>
<td>Discharged</td>
<td>4</td>
<td>-0.0351</td>
<td></td>
<td>0.4998</td>
<td>(-0.1372 to 0.0669)</td>
</tr>
<tr>
<td>C</td>
<td>Admitted</td>
<td>4</td>
<td>-0.2508</td>
<td></td>
<td>0.2107</td>
<td>(-0.6435 to 0.1419)</td>
</tr>
<tr>
<td>C</td>
<td>Discharged</td>
<td>10</td>
<td>0.0986</td>
<td></td>
<td>0.5729</td>
<td>(-0.2442 to 0.4413)</td>
</tr>
<tr>
<td>C</td>
<td>Admitted</td>
<td>10</td>
<td>0.1122</td>
<td></td>
<td>0.4697</td>
<td>(-0.1920 to 0.4165)</td>
</tr>
<tr>
<td>D</td>
<td>Discharged</td>
<td>4</td>
<td>0.0456</td>
<td></td>
<td>0.3310</td>
<td>(-0.0463 to 0.1375)</td>
</tr>
<tr>
<td>D</td>
<td>Admitted</td>
<td>4</td>
<td>0.0845</td>
<td></td>
<td>0.4843</td>
<td>(-0.1524 to 0.3215)</td>
</tr>
<tr>
<td>D</td>
<td>Discharged</td>
<td>10</td>
<td>0.1238</td>
<td></td>
<td>0.3820</td>
<td>(-0.1538 to 0.4014)</td>
</tr>
<tr>
<td>D</td>
<td>Admitted</td>
<td>10</td>
<td>-0.6311*</td>
<td></td>
<td>&lt;0.0001</td>
<td>(-0.8530 to -0.4092)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Discharged</td>
<td>4</td>
<td>-0.0314</td>
<td></td>
<td>0.2301</td>
<td>(-0.0828 to 0.0199)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Admitted</td>
<td>4</td>
<td>0.0200</td>
<td></td>
<td>0.8026</td>
<td>(-0.1368 to 0.1768)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Discharged</td>
<td>10</td>
<td>-0.0764</td>
<td></td>
<td>0.4144</td>
<td>(-0.2598 to 0.1071)</td>
</tr>
<tr>
<td>All EDs</td>
<td>Admitted</td>
<td>10</td>
<td>-0.6517*</td>
<td></td>
<td>&lt;0.0001</td>
<td>(-0.8099 to -0.4935)</td>
</tr>
</tbody>
</table>

LOS = length of stay.

* p < 0.001.
† p < 0.01.
‡ p < 0.05.

target for discharged patients in ED A and B and at the 10-hours LOS target for admitted patients in all four EDs (see Figures S1 and S2 in Data Supplement S1). We also plotted the LOS distributions for disposition decisions that were made in strain (when the total ED census was more than the median census) and nonstrain periods and did not find significant difference (see Figure S3 in Data Supplement S1).
Emergency care providers deliver health service by prioritizing ED patients based on their clinical needs, in addition to managing operational flow of the patients and resources throughout the ED. The decision to whether admit or discharge a patient is largely driven by the patient’s clinical conditions but is also affected by nonclinical factors. For example, disposition may depend on the patient’s insurance status\(^{17,18}\) or other administrative factors such as the congestion level of the ED.\(^{19–21}\) We studied how an exogenous factor, a P4P program that provided financial compensation for each patient visit that met a certain LOS target, affected the timing of patient disposition and resulting outcomes in the form of return to an ED and subsequent admission to an inpatient unit.

We showed that in all four study EDs, during the pretermination period, patients were more frequently admitted right before the 10-hours admission LOS target. This suggests that the P4P admission target provided enough incentive to speed up the patient admission process. However, empirical LOS distribution of admitted patients showed that EDs might have taken different means to achieve that. In three of the four study EDs, A, B, and D, just before the 10-hours target, a significant spike in the LOS distribution was observed, which suggests that the EDs might have rushed to reach the target for patients who without the target would have stayed longer than 10 hours, whereas in ED C, the LOS was much more evenly distributed before and up to the 10-hours target and then dropped off sharply at the target. This is more in line with a systematic change through overall process improvement rather than urgent admission decisions as in the three other EDs.

Nevertheless, we observed mixed results regarding the 4-hours LOS target for patients to be discharged. In the pretermination period, out of the four EDs, only the two lower-volume EDs seemed to have responded to the 4-hours target, and even the two lower-volume EDs operated in quite different manners, which led to contrasting outcomes. In ED A, patients were discharged in the last 20 minutes just before the 4-hours target at a higher-than-normal frequency (see the spike in LOS density during 220–240 minutes in Figure 1 and discontinuity estimate of \(-0.4111\) in Table 2), while the statistical analysis showed that such patients had a RA rate higher than patients with similar LOS (Table 3). In contrast, for

## Comparison of RA Rates Near the P4P Discharge Target

Table 3 summarizes the average 7-day RA and 30-day RA rate in the three LOS intervals near the 4-hours P4P discharge target, as well as the \(p\)-value of the two-sample \(z\)-test. With similar LOS in the ED, one would expect similar patient mix and thus similar average RA rates in these intervals. However, the test results showed that in ED A, patients who were discharged in the 20 minutes just before the P4P target, 220 to 240 minutes, had a statistically significantly (\(p < 0.05\)) larger 7-day RA rate, 0.0324, compared to patients in the two adjacent 20-minute intervals, 0.0220 and 0.0188, respectively. This suggests that some of these discharges right before the P4P LOS target may have been premature. However, such significant difference was observed only in ED A. We did not find significantly larger 7-day RA rates during the 220- to 240-minutes interval at the other ED (B) that also exhibited discontinuous LOS density at the 4-hours P4P discharge LOS target. The sensitivity test results indicated that patients who were discharged right before 4 hours in ED A had a statistically significantly (\(p < 0.08\)) larger 30-day RA rate, which was consistent with the result of 7-day RA rate test.

### Table 3

Two-sample Z-test for the RA Rate of 3 LOS Intervals

<table>
<thead>
<tr>
<th>ED</th>
<th>LOS (min)</th>
<th>7-day RA Rate (p-value)</th>
<th>30-day RA Rate (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>220–240</td>
<td>0.0324</td>
<td>0.0562</td>
</tr>
<tr>
<td></td>
<td>200–220</td>
<td>0.0220 (0.0300)*</td>
<td>0.0436 (0.0456)*</td>
</tr>
<tr>
<td></td>
<td>240–260</td>
<td>0.0188 (0.0172)*</td>
<td>0.0434 (0.0769)*</td>
</tr>
<tr>
<td>B</td>
<td>220–240</td>
<td>0.0266</td>
<td>0.0553</td>
</tr>
<tr>
<td></td>
<td>200–220</td>
<td>0.0236 (0.2854)</td>
<td>0.0449 (0.0881)*</td>
</tr>
<tr>
<td></td>
<td>240–260</td>
<td>0.0326 (0.2063)</td>
<td>0.0587 (0.3643)</td>
</tr>
<tr>
<td>C</td>
<td>220–240</td>
<td>0.0319</td>
<td>0.0750</td>
</tr>
<tr>
<td></td>
<td>200–220</td>
<td>0.0366 (0.2181)</td>
<td>0.0653 (0.1294)</td>
</tr>
<tr>
<td></td>
<td>240–260</td>
<td>0.0340 (0.3755)</td>
<td>0.0669 (0.2011)</td>
</tr>
<tr>
<td>D</td>
<td>220–240</td>
<td>0.0282</td>
<td>0.0612</td>
</tr>
<tr>
<td></td>
<td>200–220</td>
<td>0.0316 (0.2593)</td>
<td>0.0615 (0.4826)</td>
</tr>
<tr>
<td></td>
<td>240–260</td>
<td>0.0351 (0.1177)</td>
<td>0.0677 (0.2494)</td>
</tr>
<tr>
<td>All EDs</td>
<td>220–240</td>
<td>0.0329</td>
<td>0.0663</td>
</tr>
<tr>
<td></td>
<td>200–220</td>
<td>0.0329 (0.4979)</td>
<td>0.0617 (0.1295)</td>
</tr>
<tr>
<td></td>
<td>240–260</td>
<td>0.0350 (0.2753)</td>
<td>0.0700 (0.2136)</td>
</tr>
</tbody>
</table>

LOS = length of stay; RA = return-and-admission.
\(p\)-value refers to the \(z\)-test for the respective LOS interval compared to the 220- to 240-minutes interval.

\(\ast p < 0.05\).

\(^{\dagger}p < 0.1\)
ED B, the density curve below 4 hours was relatively flat without any spike (see Figure 1) albeit a statistically significant discontinuity (−0.4578 in Table 2), so a reasonable conjecture is that ED B had been consistently speeding up their discharges rather than rushing patients in the very last minute. Also, for ED B, we did not find that patients discharged right before the 4-hours target had a higher RA rate than those with similar LOS.

Overall, these results suggest that an ED discharge target can effectively reduce the LOS of discharged patients when the overall ED patient process is managed such as in ED B, but otherwise may lead to premature discharges due to the financial incentive to discharge patients within a limited time frame. It can also happen that the ED is not responsive to the discharge time target as we observed in the two higher-volume EDs. This can possibly be attributed to the much smaller financial reward for meeting the discharge time target ($100) compared to the reward for meeting the admission time target ($600) and the relatively more complex process and higher risk of discharging a patient out of the medical facility compared to transferring a patient into an inpatient unit.

We observed that lower-volume EDs were more responsive to the discharge LOS target. A similar phenomenon has been reported by Burgess et al., who found that a P4P program only impacted small teams in a major UK government agency. To sum up, the effectiveness of a P4P ED discharge time target in reducing LOS is not universal even within a single geographic region and its impact on care quality outcomes may depend on characteristic of the specific ED.

The significant discontinuity at the target hours in all of the eight most frequent CCS categories and suggest that the impact of the P4P program on the timing of patient disposition was systematic and did not depend on the patients’ specific condition. We also find that the LOS distribution does not critically depend on the congestion level of the ED.

After the P4P program was terminated, for discharged patients, we no longer observed significant discontinuity at the 4-hours target in ED A and B, suggesting that those EDs had stopped their attempts to discharge patients within 4 hours. For admitted patients, discontinuity at the 10-hours target disappeared only in ED C. This is due to the fact that while the discharge incentive had completely disappeared postgovernment P4P policy termination, the admission incentive was still provided by the regional health authority without interruption. One plausible explanation for ED C being the anomaly is that to begin with, even in the pretermination period, it was the least responsive ED out of the four having the smallest discontinuity and the response died down over time. Also, being an institution that is governed by an affiliated care provider of the regional health authority rather than under the direct control of the authority may have limited the reach of the incentive scheme resulting in less response post-P4P termination. Nevertheless, the comparison between pre- and posttermination analysis support that the discontinuities at the discharge LOS targets were led by the P4P program.

LIMITATIONS

When we measured clinical outcomes, we relied on the 7-day (and 30-day) RA rate rather than more direct measures such as mortality rate. The reason is that the incidences of death in EDs were too few, hence mortality rate did not provide a meaningful comparison.

Another limitation is the challenge in establishing causation between early discharge and the higher 7-day (and 30-day) RA rate in ED A due to the potential selection bias for patients in different LOS intervals. The Cochrane Effective Practice and Organization of Care recommends against inclusion of such results in their reviews, which makes the results difficult to influence the practice.

Finally, the study data are from four EDs in the greater Vancouver area only. The generality of the research results should be further verified.

CONCLUSION

The effectiveness of pay-for-performance programs has been well discussed in the literature, but the evidence of organizational responses to pay-for-performance incentives is unclear and mixed. Many studies indicated little effect of employing pay-for-performance incentives to improve the quality of health care. On the other hand, some showed that hospitals engaged in pay-for-performance achieved modestly greater improvements in quality than hospitals that did not. However, the positive effect might be attributed to side effects, such as public reporting and increased awareness of data recording. Our findings suggest that the pay-for-performance program implemented in
British Columbia affected the timing of patient disposition in the participating EDs. Such influence includes reduction in the ED length of stay for admitted patients, but possibly higher return-and-admission rate for patients who were discharged right before the target time in certain EDs. These observations suggest that we may only include the admission time target in a length of stay–based pay-for-performance program for now. While implementing the discharge target calls for further justification, one may consider combining it with other measurements preventing the potential risk of discharging patients prematurely.

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.13635/full

Data Supplement S1. Supplemental material.
Randomized Placebo-controlled Trial of Droperidol and Ondansetron for Adult Emergency Department Patients With Nausea

Robert Meek, MBBS, MClinEpi, Michaela J. Mee, MBBS, Diana Egerton-Warburton, MBBS, MPH, Andis Graudins, MBBS, PhD, Alastair Meyer, MBBS, Pourya Pouryahya, MD, Gabriel Blecher, MBBS, James Fahey, and Sallyanne Crow

ABSTRACT

Objective: The objective was to separately compare effectiveness of 1.25 mg of intravenous (IV) droperidol and 8 mg of IV ondansetron with 0.9% saline placebo for adult emergency department (ED) patients with nausea. A novel primary outcome measure, expected to aid clinical interpretation of reported results, was employed.

Methods: A randomized controlled trial was conducted at the three EDs of Monash Health, Melbourne, Australia. The design was to demonstrate superiority of the active drugs over placebo. The primary outcome measure of symptom improvement was defined as a visual analog scale (VAS) rating change of −8 mm or more from baseline at 30 minutes posttreatment. Mean VAS changes per group and percentages experiencing the desired treatment effect were also compared. The study was concluded after recruitment of 215 of the planned 378 patients, as interim analysis confirmed that continuation could not result in a finding of superiority.

Results: Of 215 patients, 73 (34%), 71 (33%), and 71 (33%) received droperidol, ondansetron, and placebo. Symptom improvement occurred in 75% (95% confidence interval [CI] = 64% to 85%), 80% (95% CI = 69% to 89%), and 76% (95% CI = 64% to 85%), respectively. Mean VAS changes were −29 mm (95% CI = −36 to −23 mm), −34 mm (95% CI = −41 to −28 mm), and −24 mm (95% CI = −29 to −19 mm), respectively. Desired treatment effects were experienced by 77% (95% CI = 65% to 86%), 73% (95% CI = 61% to 83%), and 59% (95% CI = 47% to 71%), respectively.

Conclusion: For adult ED patients with nausea, superiority was not demonstrated for droperidol or ondansetron over placebo.

From the Emergency Medicine Program, Monash Health (RM, MJM, DEW, AG, AM, PP, GB, SC), and the Department of Medicine, Monash University (RM, MJM, DEW, AG, AM, PP, GB, JF), Melbourne, Victoria, Australia.

Received May 22, 2018; revision received October 10, 2018; accepted October 22, 2018.

The authors received AUD$10,000.00 from the Morson Taylor Research Grant, administered by the Australasian College for Emergency Medicine, and AUD$25,000.00 from the Monash Health Emergency Research Fund. The majority of these funds were required for the preparation of the study medications by the Monash Clinical Trials Pharmacy. The residual covered various local administrative costs and payment for specialist statistical services.

The authors have no potential conflicts to disclose.

Author contributions: RM, DEW, and MJM conceived the study; all coauthors had significant input into study design. Study conduct was supervised at the Dandenong Hospital site by RM, AG, and SC; at Monash Medical Centre by MJM, DEW, GB, and JF; and at Casey Hospital by PP and AM. Study education at all sites was conducted by JF, with assistance at particular sites from RM, SC, MM, and PP. Data collection and data entry were managed by RM (recruited patients) and SC (nonrecruited patients). Statistical analyses were performed by RM, with advice from biostatisticians in the Department of Epidemiology and Preventive Medicine, Monash University. All authors contributed substantially to the manuscript; RM takes responsibility for the paper as a whole.

Supervising Editor: James Miner, MD.

Address for correspondence and reprints: A/Prof Robert Meek; e-mail: robertmeek66@hotmail.com.

ACADEMIC EMERGENCY MEDICINE 2019;26:867–877.
Emergency department (ED) patients commonly suffer nausea and vomiting as part of their presenting symptom complex. Effective treatment is desirable to alleviate patient distress and to reduce the potential for complications. Surveys report that ED patients with nausea expect to receive antiemetic drugs, and that ED doctors are willing to prescribe them. ED-based trials, however, have failed to demonstrate superiority for commonly used antiemetic drugs over placebo. Doubts have been expressed about these seemingly counterintuitive findings. Possible limitations of the currently used outcome measures for the detection of real differences and the difficulty of interpreting the main study results have been highlighted.

From 2000 to the present time, ED-based antiemetic trials have all used the standard 100-mm visual analog scale (VAS). Patients rate severity at baseline and after a defined posttreatment period; the change is measured in millimeters. Use of the VAS to monitor nausea severity has a number of advantages. It reliably discriminates between severity subgroups, is sensitive for the detection of change, and is easy for patients to use and understand. Difference in mean VAS change between treatment groups has been the primary outcome measure for all three ED-based placebo-controlled trials conducted to date. Findings of superiority or equivalence have been based on the statistical significance of the between-group differences, but the clinical interpretation of these results is not straightforward. The “minimum clinically significant difference” (MCSD) for nausea on the VAS, defined as the mean VAS change reported by people who describe their symptoms as being “a little less,” was intended to address this difficulty. While the MCSD varies a little with baseline severity, it is generally accepted as being between −15 and −20 mm. Its usefulness as an aid in antiemetic research, however, has proved to be limited. The seven ED-based antiemetic studies published from 2000 to 2014 reported mean VAS changes of between −22 and −41 mm for 16 of the 19 different treatment groups. As these changes are greater than the MCSD, it can only be inferred that most patients in all groups were improved to some degree. Even if a between-group difference is statistically significant, the clinical significance is difficult to determine when the mean VAS change for both groups is in excess of the MCSD.

To reconsider the prerequisite for a primary outcome measure, it should provide the best evidence with regard to the primary objective. The primary objective of antiemetic treatment for ED patients with nausea is clinically significant symptom improvement. Mean group VAS change does not provide direct evidence for this objective. Recent research has demonstrated that ED patients with symptom improvement (“a little less” or “a lot less”) reliably report VAS reductions in excess of −5 mm. This is not surprising, since when symptoms remain “the same,” multiple studies have reported the upper 95% confidence limits of the reported VAS change for this group to be between −5 and −9 mm. It follows logically that if symptoms are no longer “the same,” they are almost certain to be improved. One study has also found that percentage VAS change from baseline also accurately predicts symptom improvement, reported best cut-off levels for percentage change of −20% and for measured change of −8 mm (R. Meek and A. Graudins, manuscript submitted for publication). From an analysis point of view, compared with treating VAS change as a continuous variable, its reduction to a binary outcome must lessen its sensitivity for detection of between-group differences. This may seem undesirable, but the advantage in this setting is that it will allow the primary objective of symptom improvement to be directly compared between groups. In conjunction with more standardly used outcome measures, this should aid the clinical interpretability of results. In turn, this may help clarify the issue of antiemetic drug effectiveness for ED patients, but the usefulness of VAS change cutoff levels is yet to be demonstrated in a prospective antiemetic trial.

The aim of this study was to separately compare droperidol and ondansetron to placebo for the treatment of adult ED patients with nausea. The primary outcome was symptom improvement, which was defined using a measured VAS change cutoff level of −8 mm. Mean measured VAS change, mean percentage VAS change, a percentage VAS change cutoff level of −20%, and patients’ experiencing of the desired treatment effect were included as secondary outcomes.

METHODS

Study Design, Setting, and Period
A triple-blind, randomized, controlled trial was designed to demonstrate the superiority of two antiemetic drugs, droperidol and ondansetron, over placebo. The study was conducted at the three EDs of Monash Health, Melbourne, Victoria, Australia. These are Monash Medical Centre (tertiary referral hospital,
ED annual census 79,000 patients), Dandenong Hospital (urban district hospital, ED annual census 72,000 patients), Casey Hospital (urban district hospital, ED annual census 67,000 patients). A convenience sample of eligible patients was recruited from April 1, 2017, to November 10, 2017. The trial was registered with the Australian New Zealand Clinical Trials Registry (ACTRN12617000224325). Study conduct was approved by the Monash Health Human Research Ethics Committee (HREC).

**Eligibility Criteria**

**Inclusion.** Patients aged 18 years or more, with nausea severity at recruitment of 4 or more from any underlying cause. The severity screening used an 11-point verbal rating scale, with 0 being described as no nausea and 10 as the worst nausea imaginable.

**Exclusion.** Exclusion criteria included the following: 1) allergy to ondansetron or droperidol; 2) prior use (previous 4 hours) of an antiemetic drug (including ondansetron, droperidol, metoclopramide, promethazine, chlorpromazine, prochlorperazine, and any steroid medication [this was done to prevent potential confounding from any of residual ongoing effects of the previously administered drug, possible receipt of varying doses of a study drug, or uncertain effect on outcome of receipt of multiple antiemetics from different drug groups]); 3) too unwell to participate for any reason (e.g., cardiovascular instability or altered mental state [this was subjectively at the discretion of the attending physician and was not further defined]); 4) contraindication to a normal saline infusion (e.g., fluid-restricted patients); 5) Parkinson’s disease or restless leg syndrome; 6) current use of a dopamine antagonist medication (including amisulpride, chlorpromazine, clopenthixol or flupenthixol, domperidone, haloperidol, paliperidone, quetiapine, risperidone, thioridazine); 7) cognitive impairment or language barrier compromising study understanding; 8) pregnant or breastfeeding women; or 9) chemotherapy- or radiotherapy-induced nausea.

**Objectives and Measures**

**Primary Objective.** The primary objective was the between-group comparisons of the number (percentage) of patients with a measured VAS change of −8 mm or more.

**VAS-related Secondary Objectives.** The VAS-related secondary objectives were the between-group comparisons of mean measured VAS change, the between-group comparisons of mean percentage VAS change, and the between-group comparisons of the number (percentage) of patients with a percentage VAS change of −20% or more.

**VAS-related Outcome Measures.** Nausea severity was rated at baseline (t0) and 30 minutes (t30) posttreatment on a VAS. The VAS was labeled as “no nausea” at the left and “worst nausea imaginable” at the right. Measures in millimeters were taken from the left end. Measured change was calculated as t30−t0, as per previous literature. Percentage change was calculated as [(t30−t0)/t0]. Measured and percentage VAS change cutoff levels of −8 mm and −20% were used to categorize patients as “improved” or “nonimproved.” At the time of trial registration, use of only a measured VAS change cutoff of −5 mm was planned. This was based on the one relevant report available at the time.11 The measured cutoff was altered to −8 mm and the −20% cutoff included due to the findings of a currently unpublished study. This analysis pooled data from three nausea measurement studies that linked VAS and described change in a similar way.3,11,17 Receiver operating characteristic curve analyses found that both measured and percentage VAS change had equally high accuracy for detection of symptom improvement, with best cut points of −8 mm and −20%.

**Other Secondary Objectives and Measures.** Other secondary objectives and measures included 1) between-group comparisons of number (percentage) of patients experiencing the desired treatment effect (this was elicited from direct questioning —“The drug I received had the desired effect for me: Yes or No”); 2) number (percentage) of patients requesting additional antiemetic drugs; and 3) adverse events are reported for each group. The most frequently expected events of agitation/sedation (droperidol), dizziness (droperidol), and headache (ondansetron) were specifically assessed. The presence and degree of agitation or sedation was rated on the Richmond Agitation-Sedation Scale (RASS) at the time of the 30-minute nausea severity rating by the attending physician (+4 = combative, +3 = very agitated, +2 = agitated, +1 = restless, 0 = alert and calm, −1 = drowsy, −2 = light sedation, −3 = moderate sedation, −4 = deep sedation, −5 = unrousable). Presence/severity of headache and dizziness were rated on...
an adjectival scale as none, mild, moderate, or severe. Any other adverse events of any type were to be noted as free text.

**Randomization, Blinding, and Study Drug Preparation**

A simple (nonblock, nonstratified) randomization list was generated in the Monash Clinical Trials Pharmacy, where the study drugs were then prepared. Study drugs appeared identical as 4 mL of clear fluid in a 5-mL syringe. Each syringe was labeled with a unique study identification number, the HREC study reference number, and an expiry date. These were kept refrigerated and had a shelf life of 7 days. Prepared study drugs were delivered to each ED as required. While randomization was nonblock, initial deliveries to the Dandenong Hospital, Monash Medical Centre, and Casey Hospital sites commenced, respectively, at number one, number 200, and number 300 on the randomization list. Subsequent deliveries to each site followed in numerical order from those starting numbers.

**Study Drugs**

**Droperidol** (Droleptan, Phebra Pty Ltd): 0.5 mL from the 2.5 mg/mL ampoule was diluted with 3.5 mL of 0.9% saline to make a total of 1.25 mg in 4 mL.

**Ondansetron** (Ondansetron MYX, Mayne Pharma International Pty Ltd): Two of the 4 mg/mL two mL ampoules remained undiluted to make a total of 8 mg in 4 mL.

**Placebo**: The syringe contained 4 mL of 0.9% sodium chloride.

**Study Drug Choice**. Droperidol 1.25 mg intravenous (IV) is the only antiemetic drug to have shown a statistically significant greater reduction in mean VAS rating in comparison with placebo. Ondansetron is the most commonly used antiemetic in the ED setting; the 8 mg IV dose was chosen as studies have reported 4 mg IV ondansetron to be equivalent with placebo.

**Recruitment and Study Procedure**

Study education took place prior to and throughout the study period. Sixteen final-year medical students volunteered to assist with recruitment and underwent training with regard to conduct of the study. They were present at a range of times between 08:00 and 24:00, on a variety of days at any of the three study sites. When present, the student assistants monitored the electronic ED tracking system to identify patients with nausea. On these occasions the student would check eligibility by completing an enrollment form which detailed the inclusion and exclusion criteria. If eligible, the student obtained consent and engaged with the clinical staff for the required study drug and IV fluid prescriptions. Attending emergency physicians were also asked to consider recruitment of any patient for whom they intended prescribing an IV antiemetic drug for nausea from any underlying cause. If a student assistant was present, he or she was notified and assisted with the study requirements. If no student was present, the attending physician was asked to complete the enrollment form and obtain informed consent. This enabled patients to be recruited at any time of any day, regardless of whether a student assistant was present or not. It was not feasible to have student assistants present at all sites on every shift of every day. Following enrollment, an IV infusion of 0.9% saline at a rate of 1,000 mL over 4 hours was commenced and the study drug was obtained from the ED medication room refrigerator. After the baseline VAS rating was recorded, the study drug was administered as a hand-delivered, 2-minute IV infusion. At 30 minutes post-treatment, the second VAS rating was taken. At this time, the baseline rating was overleaf and not readily visible, although the patient was not prevented from viewing it if he or she wished. At this time, the patient-centered efficacy question was asked by either the student assistant or the attending clinician (nurse or physician), and information on the specified adverse events was completed. Student assistants confirmed the RASS rating with the attending physician. Other undefined adverse events of any type could be added by either the student or an attending clinician at any time during the ED episode of care. Regardless of other recorded responses, the patient was offered further antiemetic medication. Ondansetron 8 mg IV was recommended, but final choice was at physician discretion. When the inclusion criterion was met but the patient was not recruited, student assistants and recruiting ED clinicians were asked to record reasons for this (e.g., exclusion criteria, patient declined) on an enrollment form.

**Data Analysis**

Participant flow is reported using the Consort methodology; the analysis is intention to treat. Baseline information of age, sex, initial severity, and underlying condition are reported for each study site. Patients
who were improved and those experiencing the desired effect are reported as number (%) and compared using the chi-square test. As distribution approximates normal, VAS rating change is reported as mean millimeters with 95% confidence intervals (CIs); mean VAS change was compared using an independent-samples t-test. Use of additional medication and occurrence of adverse events are described.

Data were entered by one investigator (RM) into a secure database (Microsoft Excel 2007) at which time it was deidentified. A random sample of 10% was checked for accuracy by another investigator (SC). Data were analyzed using Stata Version 12.0 statistical software.

**Sample Size**

This was informed by reanalysis of the raw data from one previous ED-based study, which compared ondansetron (4 mg IV) with placebo.\(^6\) VAS reduction of \(\geq 8\) mm or more was reported by 79 and 57% of patients, respectively.\(^{10}\) Replication of this result required a sample of 111 per group to demonstrate superiority for ondansetron over placebo \((\alpha = 0.05, \beta = 0.90)\). This would give a potential between-group difference of 22% (95% CI = 10%–34%) and number needed to treat (NNT) of 5 (95% CI = 3–10). While dependent on the clinical circumstances, a single-digit NNT with an upper 95% confidence limit of 10 or less would generally be considered clinically worthwhile.\(^{20}\) For this reason, this level of difference was accepted as being clinically significant for this study. No corresponding information was available for droperidol. To allow for a dropout rate of up to 10%, the aim was to recruit 126 patients per group, for a total of 378. The secondary outcomes were not considered relevant for sample size calculation.

| Table 1 Baseline Variables: Total Population and Comparison Between Treatment Groups |
|-----------------|-----------------|-----------------|-----------------|
| Variable        | Total \((n = 215)\) | Droperidol \((n = 73)\) | Ondansetron \((n = 71)\) | Placebo \((n = 71)\) |
| Study site      |                 |                 |                 |                 |
| DH              | 145 (68%)       | 49 (67%)        | 49 (69%)        | 47 (66%)        |
|                 | [55–78]         | [57–79]         | [57–79]         | [54–77]         |
| MMC             | 50 (23%)        | 18 (25%)        | 15 (25%)        | 17 (24%)        |
|                 | [15–36]         | [12–32]         | [12–32]         | [15–36]         |
| CH              | 20 (9%)         | 6 (8%)          | 7 (10%)         | 7 (10%)         |
|                 | [3–17]          | [4–19]          | [4–19]          | [4–19]          |
| Age (years), median (IQR) | 44 (32–60) | 42 (31–61) | 47 (36–63) | 44 (26–58) |
| Male sex, \(n\) (%) \([95\% CI]\) | 87 (40%) \([34–67]\) | 30 (41%) \([30–53]\) | 26 (37%) \([25–49]\) | 31 (44%) \([32–56]\) |
| Baseline VAS (mm), median (IQR) | 60 (47–75) | 60 (47–80) | 59 (47–75) | 62 (46–75) |
| Major diagnostic groups \((n > 10)\) |                 |                 |                 |                 |
| Gastroenteritis | 42 (20%)        | 11 (15%)        | 12 (17%)        | 19 (27%)        |
|                 | [8–25]          | [9–28]          | [9–28]          | [17–39]         |
| Infective illness | 40 (19%) | 17 (23%) | 9 (13%) | 14 (20%) |
|                 | [14–35] | [6–23] | [6–23] | [11–31] |
| AP–U            | 29 (13%)        | 7 (10%)         | 12 (17%)        | 10 (14%)        |
|                 | [4–19]          | [9–28]          | [9–28]          | [7–24]          |
| AP–S            | 28 (13%)        | 10 (14%)        | 10 (14%)        | 8 (11%)         |
|                 | [7–24]          | [7–24]          | [7–24]          | [5–21]          |
| Opioid-related  | 17 (8%)         | 3 (4%)          | 11 (15%)        | 3 (4%)          |
|                 | [1–12]          | [8–26]          | [8–26]          | [1–12]          |
| Gastritis (type unspecified) | 13 (6%) | 9 (12%) | 0 (0%) | 4 (6%) |
|                 | [6–22] | [0–5] | [0–5] | [2–14] |
| Drug/alcohol (excluding opioids) | 12 (6%) | 4 (5%) | 4 (6%) | 4 (6%) |
|                 | [2–13] | [2–14] | [2–14] | [2–14] |
| Other           | 34 (16%)        | 12 (16%)        | 13 (18%)        | 9 (13%)         |
|                 | [9–27] | [10–29] | [10–29] | [6–23] |

AP–S = abdominal pain, associated with specified condition (e.g., appendicitis, pancreatitis); AP–U = abdominal pain, underlying condition unspecified/unknown; CH = Casey Hospital; DH = Dandenong Hospital; IQR = interquartile range; MMC = Monash Medical Centre.
Due to ongoing concerns about the limited support for the calculated sample size, an interim analysis was performed after recruitment of 215 patients. Specialist statistical advice confirmed that there was no realistic prospect of demonstrating superiority for the active drugs over placebo by continuing recruitment to the planned sample size. A sensitivity analysis was

**Interim Analysis and Sensitivity Analysis**

Due to ongoing concerns about the limited support for the calculated sample size, an interim analysis was performed after recruitment of 215 patients. Specialist statistical advice confirmed that there was no realistic prospect of demonstrating superiority for the active drugs over placebo by continuing recruitment to the planned sample size. A sensitivity analysis was
conducted: additional potential treatment successes (“best imaginable” for the active drugs and “lowest imaginable” for placebo) were calculated as follows: (remaining number per group to reach \( n = 111 \)) × (upper 95% confidence limit for active drugs or lower 95% confidence limit for placebo). This number was added to the actual number of improved patients in each group at the time of the analysis. “Best imaginable” between-group differences were calculated from these theoretical treatment success rates.

RESULTS

Characteristics of the Study Subjects
A total of 215 patients were recruited, 145 (68%) at Dandenong Hospital, 50 (23%) at Monash Medical Centre, and 20 (9%) at Casey Hospital. The median age of all participants was 44 years (range = 18–91 years), 40% were male, and the mean baseline VAS rating was 61 mm (95% CI = 58–65 mm). There were no significant differences in baseline characteristics between sites (Table 1). Patient flow is detailed in Figure 1.

Of the 215 patients, 73 (34%), 71 (33%), and 71 (33%) received droperidol, ondansetron, and placebo, respectively. Similar proportions were recruited to each group at each study site; between-group differences for age, sex, baseline severity and diagnostic groupings were not significant (Table 1). The median time between study drug administration and the second VAS rating was 30 minutes (interquartile range [IQR] = 30–35 minutes).

Main Results

Primary Outcome. Numbers with VAS change of –8 mm or more for droperidol, ondansetron, and placebo were similar, being 55 of 73 (75%, 95% CI = 64%–85%), 57 of 71 (80%, 95% CI = 69%–89%), and 54 of 71 (76%, 95% CI = 64%–85%), respectively (\( p = 0.75 \), Pearson chi-square). The between-group differences and NNT are shown in Table 2.

Secondary VAS-related Outcomes. The mean measured VAS changes for the droperidol, ondansetron, and placebo groups were –29 mm (95% CI = –36 to –23 mm), –34 mm (95% CI = –41 to –28 mm), and –24 mm (95% CI = –29 to –19 mm); the mean percentage VAS changes were –50% (95% CI = –59% to –40%), –55% (95% CI = –64% to –46%), and –41% (95% CI = –49% to –33%), respectively. Percentage VAS change of –20% or more was reported by 74% (95% CI = 62% to 84%), 75% (95% CI = 63% to 84%), and 73% (95% CI = 61% to 83%), respectively. Treatment having the desired effect was reported for droperidol, ondansetron, and placebo by 77% (95% CI = 65%–86%), 73% (95% CI = 66%–89%), and 72% (95% CI = 60%–83%), respectively.
CI = 61%–83%), and 59% (95% CI = 47%–71%), respectively. Full values, between-group differences and NNT (where applicable) are shown in Table 2. Measured and percentage VAS reductions were significantly greater when the desired effect was experienced versus not (both p < 0.001, independent-samples t-test). For measured VAS change this was −39 mm (95% CI = −43 to −36 mm) versus −7 mm (95% CI = −11 to −2 mm); for percentage VAS change this was −66% (95% CI = −70% to −62%) versus −8% (95% CI = −15% to −1%). Individual patient percentage VAS changes for those experiencing the desired treatment effect versus not are illustrated for each treatment group in Figure 2.

Other Secondary Outcomes. Additional antiemetic medication was requested by 11 of 73 (15%, 95% CI = 8%–25%), 16 of 71 (23%, 95% CI = 13%–34%), and 21 of 71 (30%, 95% CI = 19%–42%), respectively. Of the 48 who requested extra medication, 43 (90%) had not experienced the desired treatment effect.

A reduced level of alertness (moderate sedation, light sedation or drowsiness) was noted significantly more often in the droperidol group, compared with the ondansetron and placebo groups (27/73 [37%, 95% CI = 26%–49%] vs. 9/71 [13%, 95% CI = 6%–23%] and 12/71 [17%, 95% CI = 9%–28%], respectively, p = 0.001 [Pearson chi-square]). Restlessness or agitation was noted for four of 73 (5%, 95% CI = 2%–13%), two of 71 (3%, 95% CI = 0%–10%), and two of 71 (3%, 95% CI = 0%–10%), respectively. Headache was reported by 12 of 73 (16%), 13 of 71 (18%), and 20 of 71 (28%), respectively. Dizziness was reported by 11 of 73 (15%), five of 71 (7%), and 11 of 71 (15%), respectively.

Sensitivity Analysis and Quality Control

Calculations for the sensitivity analysis, as defined, found that the greatest imaginable treatment success rates for droperidol and ondansetron, and lowest imaginable for placebo, were 87 of 111 (79%, 95% CI = 71%–87%) and 93 of 111 (84%, 95% CI = 77%–91%) versus 80/111 (72%, 95% CI = 64–80), respectively. The differences between the three groups was not statistically significant (p = 0.12, Pearson chi-square). VAS change was remeasured from 22 randomly selected case report forms. Of these, the measured VAS change differed by 0 to 1 mm for 19 (87%) and by 2 to 3 mm for three (13%).

Nonenrolled Patients

Data were collected on 159 nonenrolled patients who met inclusion criteria. Median age was 49 years (IQR = 32–67 years) and 124 (78%) were female. Of the 159, a total of 106 (67%) had exclusion criteria. The most frequent were as follows: 43 (41%) received an antiemetic drug prior to ED arrival, 21 (20%) had cognitive impairment, and 16 (15%) were pregnant. Of the 53 without exclusion criteria, 39 (74%) declined participation; the remaining 14 were not recruited for a variety of reasons including ED activity at the time and lack of an available study drug syringe.

DISCUSSION

For a population of adult ED patients with nausea from any underlying cause, this study did not demonstrate superiority for either droperidol or ondansetron in comparison with placebo. VAS reductions of −8 mm or more were reported by 75, 80, and 76%, respectively. While the between-group comparison favored ondansetron over placebo, the 4% difference was not statistically significant; the NNT of 25 is not clinically worthwhile. This is not to say that all treatments are equally effective. This was not designed as an equivalence trial, which would be unusual for a placebo-controlled study. It should also be noted that in this setting, placebo does not equate with “no treatment.” Patients are still being actively managed for the primary condition to which their nausea relates. As expected, the symptom improvement rates predicted by the percentage VAS change cutoff level of −20% were almost identical to those detected by the measured VAS change cutoff.

Although the primary outcome measure of symptom improvement differs from that used in the previous research on the topic, the finding remains generally consistent.4–7 Ever since the first ED-based, placebo-controlled antiemetic trial was published in 2006,4 there has been a consistent lack of support for the effectiveness of antiemetic drugs in the ED setting.4–7 In the past, a number of reasons have been proposed in the literature as to why the multiple study findings might be erroneous.8,9 The difficulty in accepting that antiemetic drugs may offer little for ED patients might stem from the decades of apparent support for their effectiveness in the postoperative and oncology settings. In those fields, studies have consistently demonstrated that the prophylactic administration of antiemetic drugs reduces the incidence of poststimulus
(anesthetic or chemotherapy) nausea and vomiting.\textsuperscript{21–25} Interestingly, however, when nausea does develop after delivery of chemotherapy or in postoperative patients, difference in its severity has not been demonstrated between treatment groups.\textsuperscript{24,26} Perhaps antiemetic drugs are more effective for prevention than they are for cure.

A number of the secondary outcomes were of interest. While the percentages reporting symptom improvement were similar between all groups, experiencing of the desired treatment effect was not. This was reported by 77 and 73\% for droperidol and ondansetron, but only by 59\% of the placebo group. This may seem inconsistent with the symptom improvement results, but these outcomes reflect different amounts of symptom change. The VAS change cutoff levels identify patients whose symptoms are either “a little less” or “a lot less,” while experiencing the desired effect requires symptoms to be “a lot less.”\textsuperscript{3} Given this, the difference in the desired effect findings is most likely explained by the relative difference in the mean VAS changes between groups. For measured change, these were –29, –34, and –24 mm for the droperidol, ondansetron, and placebo groups. It is conceivable that the somewhat lesser mean VAS change for the placebo group resulted from fewer patients having the greater VAS reductions, which are reported when symptoms become “a lot less.” This inference is supported by relatively more in the placebo group requesting additional medication (30\%) in comparison with the droperidol and ondansetron groups (15 and 23\%). It has previously been reported that most patients wanting additional antiemetic medication are those who have improved but not by the desired amount.\textsuperscript{3} The NNT for experiencing the desired effect of 5 and 7 for droperidol and ondansetron, in comparison with placebo, may be clinically worthwhile but these point estimates are fairly imprecise. Despite this, it may still be useful to balance this information against other factors such as drug costs and side effects when making individual treatment recommendations. For the drugs used in this study, the costs are low and the reasonably minor adverse effects did not require any treatment.

The purpose of primarily comparing symptom improvement rates between groups was to enable findings that directly related to the primary treatment objective to be presented in a format that is easy to clinically interpret.\textsuperscript{10} It was hoped that this would aid understanding of relative treatment effectiveness in a way which might be beneficial for both treating doctors and patients. It is not useful, for example, to inform a patient that without antiemetic drug treatment their nausea severity is likely to improve by about –24 mm on the VAS, but that on average, ondansetron might reduce it by –34 mm. The following seems far more helpful: “Whether or not you have an antinausea drug, there is a 75 to 80\% chance that your nausea will ease as your underlying condition is treated. Ondansetron might give a little extra benefit to about one-in-seven of those who do improve. There are some people, however, whose nausea will not quickly settle no matter what we do. Ondansetron may have some side-effects, but these are usually fairly mild.”

With regard to future ED-based antiemetic research, it should be remembered that the studies to date have only examined the response to a single administration of one drug at 30 or 60 minutes posttreatment. The response to higher drug doses, repeated dosages over a longer time period, or the concurrent delivery of antiemetic drugs from different groups may be quite different. Characterizing treatment responders versus non-responders could also be of value and the need for condition-specific research has never been entirely discounted.\textsuperscript{27}

**LIMITATIONS**

Consideration of the outcome measures used in ED-based antiemetic studies remains important. Between-group comparisons of mean VAS change have previously failed to demonstrate superiority for antiemetic drugs over placebo. The VAS change cutoff level included as an outcome measure in this study may have aided clinical interpretation of the results, but it also failed to demonstrate superiority for the active drugs. For ED patients, antiemetic drugs may truly provide little additional benefit to that derived from treatment of their underlying condition. It may also be that outcome measures and methods of analysis capable of detecting a real difference are yet to be determined and successfully trialed.

The original sample size calculation for the study was based on “anticipated” symptom improvement rates for ondansetron and placebo of 79 and 57\%. As this was drawn from a post hoc analysis of only one study,\textsuperscript{6,10} doubts about the accuracy of the estimate persisted. For this reason, conduct of an interim analysis was deemed prudent, and in retrospect, not
preplanning this was an error. At that time, it was found that the “actual” and “anticipated” symptom improvement rates for the placebo group were markedly different (76% vs. 57%). Also, the mean VAS change of $-29$ mm for droperidol was much lower than the $-55$ mm previously reported.\(^4\) The degree of these differences is probably not surprising given the known variation in mean VAS changes for the same treatment regimens in different ED-based studies. For example, two different studies reported posttreatment mean VAS changes for ondansetron (4 mg IV) of $-34$ and $-22$ mm;\(^5,13\) two other studies reported mean VAS changes for placebo of $-39$ and $-16$ mm.\(^4,5\) This is despite patient populations appearing otherwise similar.

Other potential limitations include that the convenience sample may not be representative of all ED patients with nausea. Recruitment was probably more frequent when student research assistants were present, but the number of patients enrolled by students versus duty clinical staff was not recorded. Although incomplete, monitoring of reasons for nonrecruitment was attempted. Prehospital antiemetic administration was the most frequent reason for exclusion. This may have led to recruitment of fewer patients with severe nausea; the potential impact of this on results is unknown. Although the study instructions dictated that all patients have 1000 mL of IV 0.9% saline running at a 4-hourly rate, exact amounts of fluid received during the 30-minute study period may have varied.

**CONCLUSIONS**

For adult ED patients with nausea, this study did not demonstrate superior symptom improvement rates for 1.25 mg of intravenous droperidol or 8 mg of intravenous ondansetron in comparison with placebo. The marginally greater mean visual analog scale reductions and rates of experiencing the desired treatment effect in the active drug groups may aid treatment decision making in individual cases.

Specialist statistical advice regarding the interim and sensitivity analyses was obtained from Professor Rory Wolfe of the Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia. This was invaluable. The assistance in recruitment from the following Monash University medical students is also acknowledged: Jeremy Lee, Sarah Monagle, Amos Liew, Helen Huang, Geoffrey Liu, Annabel Jones, Lucia Hadinata, Luke Fletcher, Tom de Vries, Liyin Yip, Stephanie See, Sean Ng Ying Kin, Novia Tan, Monique Kowitt, James Jiang, and Sandy Chu.

**References**


Depression in Emergency Department Patients and Association With Health Care Utilization

David G. Beiser, MD1, Charlotte E. Ward, MPH2, Milkie Vu, MA3, Neda Laiteerapong, MD4, and Robert D. Gibbons, PhD5

ABSTRACT

Background: Depression is one of the most common illnesses in the United States, with increased prevalence among people with lower socioeconomic status and chronic mental illness who often seek care in the emergency department (ED). We sought to estimate the rate and severity of major depressive disorder (MDD) in a nonpsychiatric ED population and its association with subsequent ED visits and hospitalizations.

Methods: This prospective cohort study enrolled a convenience sample of English-speaking adults presenting to an urban academic medical center ED without psychiatric complaints between January 1, 2015, and September 21, 2015. Patients completed a computerized adaptive depression diagnostic screen (CAD-MDD) and dimensional depression severity measurement test (CAT-DI) via tablet computer. Primary outcomes included number of ED visits and hospitalizations assessed from index visit until January 1, 2016. Negative binomial regression modeling was performed to assess associations between depression, depression severity, clinical covariates, and utilization outcomes.

Results: Of 999 enrolled patients, 27% screened positive for MDD. The presence of MDD conveyed a 61% increase in the rate of ED visits (incidence rate ratio [IRR] = 1.61, 95% confidence interval [CI] = 1.27 to 2.03) and a 49% increase in the rate of hospitalizations (IRR = 1.49, 95% CI = 1.06–2.09). For each 10% increase in MDD severity, there was a 10% increase in the relative rate of subsequent ED visits (IRR = 1.10, 95% CI = 1.04 to 1.16) and hospitalizations (IRR = 1.10, 95% CI = 1.02 to 1.18). Across the range of the severity scale there was over a 2.5-fold increase in the rate of ED visits and hospitalization rates.

Conclusions: Rates of depression were high among a convenience sample of English-speaking adult ED patients presenting with nonpsychiatric complaints and independently associated with increased risk of subsequent ED utilization and hospitalization. Standardized assessment tools that provide rapid, accurate, and precise classification of MDD severity have the potential to play an important role in identifying ED patients in need of urgent psychiatric resource referral.
Major depressive disorder (MDD) is a significant public health problem, affecting 16.2 million adults in the United States (6.7%) per year and is the leading cause of disability among adults in high-income countries. MDD can have significant negative impact on physical, mental, and social well-being and is major multiplier of health care costs. The burden of MDD is disproportionately experienced by patients with low socioeconomic status, Medicaid recipients, the elderly, and those with chronic medical conditions. This same population is also less likely to receive primary care and more likely to access the emergency department (ED) for both urgent and ambulatory care sensitive conditions.

Importance
Early detection and appropriate treatment of MDD increases the likelihood of achieving remission, preventing relapse, and decreasing overall health care costs. Yet MDD remains underdetected and undertreated with barely half of Americans and only 40% of African Americans with MDD receiving treatment. Recognizing this gap in care, the United States Preventive Services Task Force (USPSTF) recently recommended routine depression screening in adult primary care settings. Given the ED’s role as a primary safety net provider and the risk profiles of its patients, several have asked whether or not MDD screening should be extended to the general adult ED population. Yet more research is needed regarding the scope of the problem, implementation of screening, and potential impacts before such secondary prevention efforts are broadly adopted in the ED.

Goals of This Investigation
This study represents a first attempt at estimating the scope of the problem of depression in the ED and the needed capacity for mental health referral resources for future ED-based screening programs. Specifically, the goals of this study were to 1) estimate the rate of MDD and spectrum of severity of symptoms in a nonpsychiatric adult ED population, 2) examine the health service implications of depression (diagnosis and severity) through the metric of ED and hospital utilization in this sample, and 3) cross-validate a computerized adaptive test for depression severity (CAT-DI) in this population.

METHODS
Study Design and Setting
This was a prospective observational study conducted between January 1, 2015, to September 21, 2015, of patients over the age of 18 years old presenting with nonpsychiatric chief complaints to the ED of an urban hospital.
academic medical center with a triage patient volume of 41,373 during that period. The institutional review board of the University of Chicago approved this study.

Selection of Participants
Adult ED patients were recruited during weekdays and weekends between the hours of 8:00 AM and 12:00 AM based on research assistant availability. To avoid potential issues of incapacity around informed consent, patients with acute psychiatric complaints at triage were not approached. In addition, patients triaged with an Emergency Severity Index\(^35\) of 1 (requiring immediate lifesaving intervention) or 2 (high risk of deterioration) were deemed ineligible to avoid the potential for interference with acute patient care delivery. Patients were also excluded if they declined participation, were unable to consent, or did not speak English. Patients were recruited by a group of nine volunteer research assistants and a paid research coordinator, who were all trained to screen patients, enroll participants, and operate a tablet computer used to deliver the depression instruments. The majority of recruitment sessions occurred during the hours of 1 PM and 11 PM, which corresponds to the peak census of our ED. Recruitment sessions were uniformly distributed across the days of the week.

While the overall design of the study was nonrandom, efforts were made to reduce sampling bias by randomizing the screening process. Patients were randomly selected for screening during each recruitment session using the following strategy. At the start of each session, the current ED census was printed. A reduced list of patients was developed for potential screening by matching the last digit of a patient’s age with a randomly selected number between 0 and 9. Patients on the reduced list were then screened for eligibility and approached using a standardized script if eligible. Once all eligible patients were either approached or removed from the ED tracking board, the recruitment session ended or an updated census was printed and the process repeated. Patients could only be approached once for the study. Written informed consent was obtained for all study participants.

MDD Screening and Severity Scoring
The self-administered Computerized Adaptive Diagnost-
ic Test for Major Depressive Disorder (CAD-MDD)\(^32\) and Computerized Adaptive Test-Depression Inventory (CAT-DI)\(^36\) were used to obtain rapid diagnostic depression screens and severity estimates. Once enrolled, patients were handed a tablet computer that they used to complete the CAD-MDD and CAT-DI. Patients were given the option of text or text plus audio administration. Test administration process measures, including time for completion in seconds and number of administered questions, were recorded to track test administration burden.

The CAD-MDD\(^32\) is a computerized adaptive depression screening tool based on a random forest\(^37\) machine learning algorithm that adapts to patient responses to questions about depression by asking the most diagnostically informative question from a bank of 88 items. The CAD-MDD item bank was created based on a review of over 500 items from 73 commonly used depression tools. Items were then filtered by an expert panel to include only those that closely aligned with nine DSM-IV criteria for MDD diagnosis: depressed mood, loss of interest or pleasure in activities, loss or gain of weight, insomnia or hypersomnia, agitation or slowed behavior, fatigue, thoughts of worthlessness or guilt, inability to think or concentrate, and suicidality. The final item bank included only those items in the public domain.\(^32\)

A prior study showed that the CAD-MDD was on average shorter than the PHQ-9 (an average of four items versus nine items) and that overall sensitivity and specificity for the CAD-MDD was 0.95 and 0.87, respectively, compared to 0.70 and 0.91 for the PHQ-9, compared to the Structured Clinical Interview for DSM-IV (SCID) criterion standard.\(^32\) An independent validation study including patients presenting to an outpatient mental health clinic and healthy controls showed similar test performance.\(^38\) We have also recently validated the CAD-MDD within our institution’s primary care population with comparable results.\(^39\)

The CAT-DI\(^36\) is a computerized adaptive dimensional severity measure for depression that utilizes a bank of 389 depression items whose response patterns are fitted to a multidimensional item response theory model.\(^40,41\) The CAT-DI produces a continuous depressive severity estimate on a 0 to 100 point scale with 5 points of precision.\(^36\) Prior work has shown that an average of 12 items and a median administration time of 137 seconds had a correlation of \(r = 0.95\) with the 389 total item bank score.\(^36\) In terms of diagnostic validity, using the continuous CAT-DI depressive severity scale scores as a linear predictor of DSM-IV MDD diagnoses, there was a 24-fold increase in the
probability of MDD across the range of the CAT-DI scale (odds ratio = 24.19, 95% CI = 10.51 to 55.67). Gibbons et al. have also shown that using an empirically derived threshold based on a normal mixture distribution, the CAT-DI has a sensitivity of 0.92 and specificity of 0.88 for predicting of MDD based on a SCID. We cross-validated the CAT-DI by comparing responses of our adult ED sample to those of the original psychiatric outpatient base sample to test for the presence of differential item functioning (DIF). DIF occurs when people from different subgroups with the same underlying level of a latent trait, in this case depression, have different likelihoods of endorsing certain survey items about depression. A description of these cross-validation methods and results can be found in Data Supplement S1 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13726/full).

Chart Review

Descriptive statistics were used to summarize CAD-MDD depression screening and CAT-DI depression severity results. Sociodemographic, health care–related, and utilization covariates and outcome variables were abstracted from the hospital’s electronic medical record (EMR; Epic Systems) following procedures outlined in a coding manual created by study team members. Sociodemographic covariates included patient sex, age, insurance status, and race/ethnicity. Comorbid diagnoses were obtained from EMR data within the history, problem list, and ED clinical impression fields and recorded using 9th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-9). Additional diagnoses recorded in the free text of all ED provider EMR notes were also scanned. To reduce variability, all records were reviewed by one research assistant, who was blinded to depression testing results. Inconsistencies were adjudicated by the principal investigator (DGB) who was also blinded to depression testing results.

Outcomes

Utilization outcomes included the number of subsequent ED visits and hospital admissions at the study institution for 1 year following the index ED visit.

Statistical Analysis

Summary statistics were presented as frequencies (with percentages), means (with standard deviations [SDs]), or medians (with interquartile ranges [IQRs]), as appropriate. We performed multivariable analyses using four models that assessed associations between depression and utilization outcomes, one for each unique pair of depression and utilization measures. Use of a Poisson model for count outcomes requires the variance of the dependent variable to be equal to the mean (no overdispersion). In the case of both outcomes, ED visits and admissions, overdispersion was detected through a deviance or Pearson chi-squared value substantially exceeding 1.0, and therefore, negative binomial models were estimated controlling for demographics, comorbidity burden, having a primary care provider; and current use of illicit drugs, alcohol, and tobacco. Likelihood ratio chi-square tests were used to assess final model fit by comparing the likelihoods of the intercept-only and full (all covariates included) model. The full model fit the data better than the null model (p < 0.0001 for both outcome measures).

Primary covariates of interest included depression screen status (positive/negative) and symptom severity (severity percentile decile on a 0–100 scale) as measured by the CAD-MDD and CAT-DI, respectively. These measures were included in separate models. The same covariates were used to control for confounding in modeling both count outcomes of hospital admissions and ED visits. Demographic variables included patient sex, age, insurance status, and race/ethnicity. Age was modeled as a continuous covariate, while sex (male/female), insurance status (commercial/Medicaid/Medicare/uninsured/miscellaneous), race (white/black/other), and ethnicity (Hispanic/not Hispanic) were included in models as categorical covariates. Healthcare-related dichotomous covariates included having a primary care provider and current use of illicit drugs, alcohol, and tobacco. Comorbidity burden was quantified using the enhanced version of the validated Charlson–Deyo comorbidity index (CCI) for administrative data, which weights 17 selected comorbidities, where higher scores are associated with greater burden of comorbid disease. The CCI was included in models as a continuous covariate.

The incidence rate ratio (IRR), a ratio of two incidence rates, was used as a relative measure of the effect of depression on utilization. We defined the incidence rate as the number of events (i.e., ED visits or hospitalizations) divided by the person-time at risk in person-years. All IRRs are presented with 95% CIs. All
analyses were performed using SAS software version 9.4.

An enrollment goal of 1,000 patients was established based on an a priori power analysis to provide a margin of error in rate of MDD of 2% based on a conservative estimated base rate of 15%,22–25,27,29,46,47 a 95% confidence, and a population of 10,000 ED patients during the recruitment period.

RESULTS

Descriptive Statistics

During the study period, a total of 1,000 patients were enrolled (Figure 1). Of those enrolled, one patient had a missing CAD-MDD test score leaving 999 patients for analysis. The median time to complete the CAD-MDD was 62 seconds (IQR = 41 seconds), with a median number of four items (IQR = 1 item). Median time for completion of the CAT-DI was 101 seconds (IQR = 62 seconds), with a median number of nine items (IQR = 4 items). Among enrollees, 26.5% screened positive for MDD by the CAD-MDD. Patients who screened positive for MDD (Table 1) were predominantly female (65.4% vs. 57.7%), used illicit drugs (14.7% vs. 9.4%), and currently smoked (25.2% vs. 15.4%).

Utilization

Patients who screened positive for MDD were more likely to revisit the ED and to be admitted to the hospital during follow-up compared to those with a negative screen (ED, 3.51 vs. 2.15 events per person-year; hospitalization, 1.50 vs. 1.10 events per person-year). In adjusted analyses, a positive MDD screen was associated with a 61% increase in subsequent ED utilization and a 49% increase in subsequent hospital admissions (see Table 2). In terms of depressive severity (deciles), the relative rate of subsequent ED visits and hospitalizations increased by 10% for every 10% increase in MDD severity (10 points on the 100-point scale). Across the entire range of the scale (from lowest to highest severity) the rate of ED visits increased, and the rate of hospitalizations increased more than 2.5-fold. Full model results can be found in Data Supplement S1, Tables S1 and S2.

DISCUSSION

This study presents one of the largest prospective reports of major depression screening in a general adult ED population. We describe for the first time the distribution of depression severity in the adult ED and its association with future health care utilization.
Our ED sample had a rate of MDD of 26.5% among patients without psychiatric complaints who present during daytime hours. This rate is consistent with the published ED literature. The rate of MDD in our ED sample is also more than double of the 12.5% estimated prevalence of MDD for adult primary care patients in the United States. Consistent with suggestions by Booth et al., high rates of depression may reflect the impact of low socioeconomic status on MDD as approximately 29% of the population within the University of Chicago Medical Center catchment area live below the Federal Poverty Level and the 40% of rate of Medicaid coverage in our sample. Also, it may represent the under treatment and under diagnosis in our catchment area due to a documented local shortage of primary care and mental health professionals. Similar shortages have been documented nationally and are projected to worsen over the next decade.

We also found that a positive MDD screen was associated with increased subsequent health care utilization during the 1-year follow-up period. While associations between depression and health care utilization have been previously reported in the primary care clinic, specialty clinic, elderly ED patient populations, and adult ED patients with abdominal pain, our study represents one of the first reports specific to the general adult ED population.

Severity of depression had an even stronger association with utilization: 2.55-fold increase across the continuum from the least to the most severe for ED visits and 2.53-fold increase for the rate of hospitalizations. To our knowledge, we are among the first to report the correlation between depression severity and ED utilization. In one small study of homebound elderly patients, self-reported ED visit frequency during the preceding 6 months was positively associated with scores from the 24-item Hamilton Rating Scale for Depression (HAMD) during baseline interviews. In addition, ED visit frequency at 12- and 24-week

### Table 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CAD-MDD Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative (n = 733)</td>
</tr>
<tr>
<td><strong>Sociodemographic</strong></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (±SD)</td>
<td>47.0 (±18.2)</td>
</tr>
<tr>
<td>Female</td>
<td>423 (57.7)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>108 (14.7)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>599 (81.7)</td>
</tr>
<tr>
<td>Other</td>
<td>26 (3.6)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>25 (3.4)</td>
</tr>
<tr>
<td><strong>Insurance type</strong></td>
<td></td>
</tr>
<tr>
<td>Commercial</td>
<td>204 (27.8)</td>
</tr>
<tr>
<td>Medicare</td>
<td>218 (29.7)</td>
</tr>
<tr>
<td>Medicaid</td>
<td>267 (36.4)</td>
</tr>
<tr>
<td>Uninsured</td>
<td>40 (5.5)</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>4 (0.6)</td>
</tr>
<tr>
<td><strong>Charlson Index Score, median (IQR)</strong></td>
<td>0 (0–1)</td>
</tr>
<tr>
<td><strong>Health care-related</strong></td>
<td></td>
</tr>
<tr>
<td>Has primary care provider</td>
<td>426 (58.1)</td>
</tr>
<tr>
<td>Any illicit drug use</td>
<td>69 (9.4)</td>
</tr>
<tr>
<td>Any alcohol use</td>
<td>222 (30.3)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>113 (15.4)</td>
</tr>
<tr>
<td><strong>ED utilization</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion with ≥ 1 ED visit</td>
<td>353 (48.2)</td>
</tr>
<tr>
<td>Number of ED visits among those with ≥ 1 ED visit, median (IQR)</td>
<td>2 (1–3)</td>
</tr>
<tr>
<td>Time to first revisit to the ED (days), mean (±SD)</td>
<td>101.2 (-100.1)</td>
</tr>
<tr>
<td><strong>Inpatient utilization</strong></td>
<td></td>
</tr>
<tr>
<td>Proportion with ≥ 1 hospitalization</td>
<td>158 (21.6)</td>
</tr>
<tr>
<td>Number of hospitalizations among those with ≥ 1 hospitalization, median (IQR)</td>
<td>1 (1–3)</td>
</tr>
<tr>
<td>Time to first readmission (days), mean (±SD)</td>
<td>98.0 (+97.9)</td>
</tr>
<tr>
<td><strong>Depression severity (CAT-DI), mean (±SD)</strong></td>
<td>23.6 (+13.4)</td>
</tr>
</tbody>
</table>

**CAD-MDD = Computerized Adaptive Diagnostic-Major Depressive Disorder; CAT-DI = Computerized Adaptive Testing-Depressive Disorder Inventory (severity classifier); IRR = incident rate ratio.**

### Table 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Number of ED Visits (95% CI)</th>
<th>p-value</th>
<th>Number of Admissions (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CAD-MDD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Ref Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>1.61 (1.27–2.03)</td>
<td>&lt;0.0001</td>
<td>1.49 (1.08–2.09)</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>CAT-DI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per increase in level of depression severity</td>
<td>1.10 (1.04–1.17)</td>
<td>&lt;0.001</td>
<td>1.10 (1.02–1.18)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

**CAT-MDD = Computerized Adaptive Diagnostic Test for Major Depressive Disorder; CAT-DI = Computerized Adaptive Testing-Depression Inventory (severity classifier); IRR = incident rate ratio.**

*Adjusted for sociodemographic and health care-related covariates.
follow-up intervals were associated with changes in HAMD scores from baseline. In one prior report in the general adult ED population, patients with moderate or severe depression reported a median of two ED visits in the past 6 months while those who screened negative for depression reported one visit. Notably, that study also documented strong correlations between mental health scores for depression and anxiety and perceived barriers to care.

To integrate MDD screening into the busy ED clinical workflow, it is essential that screening and diagnostic instruments be convenient, brief, and accurate. Our results demonstrate that computer adaptive CAD-MDD and CAT-DI assessments can be delivered to ED patients using a minimal amount of patient burden, although the feasibility of integrating this approach into the clinical workflow without the use of research assistants has yet to be demonstrated. Others have reported on the feasibility, advantages, and challenges of deploying technology-based behavioral health intervention, screening, and referral programs in the ED (see Choo et al. for review). There is also evidence to suggest that self-administered screening has high patient acceptability and may increase disclosure rates by at-risk individuals. In many examples, computers enable the delivery of kiosk-based self-administered assessments with automated scoring and little to no staff intervention. We envision integrating a variety of patient-facing adaptive screening instruments such as the CAD-MDD and CAT-DI into the ED workflow via tablet computers or in-room television-based patient response system in a manner that minimizes clinical provider time and interfaces with the EMR.

An optimal ED-based screening program must include provisions for those who screen positive to provide appropriate diagnosis, initial treatment, and evidence-based care or referral to a proper care setting. ED implementation of traditional screening and referral programs based on standard MDD short-form screening instruments with binary, i.e., positive or negative, outcomes would quickly overwhelm the existing psychiatric resources of most health systems. Our study, by documenting the full spectrum of MDD disease severity in an ED population, provides a refined estimate of the demand for psychiatric referral services. Specifically, we can apply empirically derived thresholds established for the CAT-DI in an outpatient psychiatric population to estimate that approximately 7% of our ED sample were experiencing moderate to severe depression symptoms and thus might require urgent referral for psychiatric services. Assuming a representative sample of our entire ED census, this result implies the need for approximately 11 urgent psychiatric referrals from the ED per day. This number of additional urgent referrals would quickly saturate available psychiatry consultation and clinic capacity at our institution.

In the past, we have addressed the needed for additional follow-up clinic capacity, e.g., primary care, by developing referral networks with unaffiliated community physician providers. However, given the aforementioned shortages of mental health providers in our city and nationally, it is vital that we consider more innovative strategies for addressing this gap in care. Several care models from primary care may provide useful guidance in this regard. For example, Project ECHO utilizes expert-led video conferences to provide advanced training of primary care providers in mental health treatment. Integrated behavioral health care models embed mental health providers of various levels directly into the primary care clinic environment. Alternatively, stepped-care models align referral resource intensity (e.g., primary care, social work, or psychiatry) with depression symptom severity (mild, moderate, or severe). In addition, at our institution, we have adopted a primary care–behavioral health model, have educated our primary care providers to provide primary psychiatric care, and have community mental health partnerships. Finally, digital technology, including telepsychiatry, Web- and mobile-based applications may help extend the capacity of the existing mental health workforce.

LIMITATIONS

The use of a convenience sample in this study introduces the potential for bias from several important sources. By excluding patients with primary psychiatric complaints, acute life-threatening illness, cognitive impairment, and those who decline to participate, we may have excluded patients at higher risk for depression or with greater depression severity and, thus, underestimated the rate of MDD in a general adult
ED population. Patients presenting overnight between 12 AM and 8 AM were also not included in the sample. While patients presenting overnight represented only approximately 10% of our patient volume, there is the potential that patients with MDD might not present uniformly throughout the day and thus impact our estimate of MDD rate in this population. For example, one report suggests that patients with psychiatric issues are more likely to present to a psychiatric emergency service facility during day, rather than night, shifts.72 If this were true in our population, it could lead to an overestimate of the rate of MDD.

Another limitation of our study is that patient follow-up data were only available for a single hospital site. Since the primary service area of our hospital overlaps with those of other institutions with EDs, it is likely that we have underestimated recidivism rates in both the MDD-positive and the MDD-negative groups. In addition, we did not account for patient mortality. It is difficult to predict how incomplete utilization and mortality data might bias the observed association between depression severity and utilization.

The interpretation of this study’s utilization findings is also limited by our reliance on chart review methods for capturing clinical and demographic data.73–75 Also, our study did not control for a variety of social determinants of health, which have both been shown to have associations with depression and health care utilization. Not accounting for these additional covariates in our model may have biased our estimates, leading to overestimation of the effects. Finally, while the CAD-MDD has been previously validated against the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) in an outpatient psychiatric population,32 this validation step has yet to be performed in the ED and will be the subject of future work.

CONCLUSIONS

This work documents the high rate of depression and its range of severity within a population of adult patients presenting to an urban ED with nonpsychiatric complaints. It also describes the positive association between depression and depression severity and future ED visits and hospitalizations in this selected population. Given the potential for increased ED utilization under Medicaid expansion,76,77 the ED may be ideally positioned to interrupt the vicious cycle of depression, poor medical outcomes, and high utilization. Together these results lend support to the idea of a future trial testing the impact of depression screening and treatment on ED utilization and hospital readmissions. Additional research is necessary to demonstrate the feasibility, effectiveness, and cost savings associated with MDD screening, diagnosis, and severity classification tools as part of an ED-based tiered psychiatric referral and follow-up pathway.

The authors acknowledge the volunteer data collection efforts of Jenifer Goldberg, BA, as well as the following individuals affiliated with the University of Chicago: Maha Ahmad, BA; Alexandra Berthiaume; Cody Davis; Ali Fadhil, MBChB; Annie Hao; Ellen Harris; Davina Moossazadeh; Ivy Sandquist; and Anna Shin, BA.

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.13726/full

Data Supplement S1. Supplemental material.
Psychiatric Outcomes of Patients With Severe Agitation Following Administration of Prehospital Ketamine

Jacob A. Lebin, MD1, Arvin R. Akhavan, MD1, Daniel S. Hippe, MS2, Melissa H. Gittinger, DO3, Jagoda Pasic, MD4, Andrew M. McCoy, MD, MS1, and Marie C. Vrablik, MD, MCR1

ABSTRACT

Background: Ketamine is an emerging drug used in the management of undifferentiated, severe agitation in the prehospital setting. However, prior work has indicated that ketamine may exacerbate psychotic symptoms in patients with schizophrenia. The objective of this study was to describe psychiatric outcomes in patients who receive prehospital ketamine for severe agitation.

Methods: This is a retrospective cohort study, conducted at two tertiary academic medical centers, utilizing chart review of patients requiring prehospital sedation for severe agitation from January 1, 2014, to June 30, 2016. Patients received either intramuscular (IM) versus intravenous (IV) ketamine or IM versus IV benzodiazepine. The primary outcome was psychiatric inpatient admission with secondary outcomes including ED psychiatric evaluation and nonpsychiatric inpatient admission. Generalized estimating equations and Fisher’s exact tests were used to compare cohorts.

Results: During the study period, 141 patient encounters met inclusion with 59 (42%) receiving prehospital ketamine. There were no statistically significant differences between the ketamine and benzodiazepine cohorts for psychiatric inpatient admission (6.8% vs. 2.4%, difference = 4.3%, 95% CI = –2% to 12%, p = 0.23) or ED psychiatric evaluation (8.6% vs. 15%, difference = –6.8%, 95% CI = –18% to 5%, p = 0.23). Patients with schizophrenia who received ketamine did not require psychiatric inpatient admission (17% vs. 10%, difference = 6.7%, 95% CI = –46% to 79%, p = 0.63) or ED psychiatric evaluation (17% vs. 50%, difference = –33%, 95% CI = –100% to 33%, p = 0.55) significantly more than those who received benzodiazepines, although the subgroup was small (n = 16). While there was no significant difference in the nonpsychiatric admission rate between the ketamine and benzodiazepine cohorts (35% vs. 51%, p = 0.082), nonpsychiatric admissions in the benzodiazepine cohort were largely driven by intubation (63% vs. 3.8%, difference = 59%, 95% CI = 38% to 79%, p < 0.001).

Conclusions: Administration of prehospital ketamine for severe agitation was not associated with an increase in the rate of psychiatric evaluation in the emergency department or psychiatric inpatient admission when compared with benzodiazepine treatment, regardless of the patient’s psychiatric history.

A cute, undifferentiated agitation is a commonly encountered problem in the prehospital and emergency department (ED) setting.1,2 When verbal deescalation is ineffective, pharmacologic management may be required to ensure safety for both patients and providers. Traditional pharmacologic treatments for
severe agitation, such as benzodiazepines or haloperidol, have relatively slow onset of action and often require redosing to achieve adequate sedation.\textsuperscript{3} Ketamine, a dissociative agent with rapid onset and wide therapeutic index, is an emerging alternative in the management of undifferentiated, severe agitation in the prehospital and ED setting.\textsuperscript{3–6}

In addition to rapid sedation, ketamine is well known to cause emergence reactions, often described as vivid dreams or hallucinations, which has tempered its use in certain populations.\textsuperscript{7} Currently, practice guidelines recommend against the use of ketamine for procedural sedation in patients with schizophrenia given concern for exacerbation of underlying psychotic disorders.\textsuperscript{8} Several studies suggest that the glutamatergic N-methyl-D-aspartic acid (NMDA) receptor is involved in the pathophysiology of schizophrenia and ketamine, through NMDA receptor antagonism, can trigger recrudescence of psychotic symptoms.\textsuperscript{9,10} However, there are little data evaluating whether patients who receive ketamine in the prehospital or ED setting have clinically significant psychiatric exacerbations. This study compares psychiatric outcomes in patients administered prehospital ketamine to patients administered prehospital benzodiazepines in the setting of severe agitation. Specifically, our primary objective was to compare rates of clinically significant psychiatric exacerbation, in patients with and without schizophrenia, using admission to an inpatient psychiatry service as a surrogate marker. Our secondary objective was to assess other markers of psychiatric exacerbation, such as need for psychiatric evaluation in the ED and ED length of stay (LOS). We hypothesized that ketamine would not result in worsened psychiatric outcomes when compared to benzodiazepines in patients with acute, undifferentiated prehospital severe agitation.

**METHODS**

**Study Design and Setting**

This was a retrospective cohort study performed at two urban, academic centers, in conjunction with the fire department–based emergency medical services (EMS) system. The EDs have a combined annual census of 100,000 visits and are served by the same EMS system. Consecutive adult patients who received prehospital sedatives for acute, undifferentiated severe agitation were evaluated from January 1, 2014, to June 30, 2016. During the first 17 months of the study period, prehospital providers administered benzodiazepines, specifically midazolam or diazepam, for sedation of severe agitation. Ketamine was added to the prehospital pharmacy for the management of severe agitation on June 1, 2015. During the last 13 months of the study, prehospital providers could administer benzodiazepines, ketamine, or both agents for sedation at their discretion. Institutional review board approval was obtained for all study procedures.

**Selection of Participants**

In this EMS system, all advanced life support charts are reviewed by two EMS medical directors. The presence of severe agitation, defined as profound agitation with imminent risk of injury to patient or provider, was determined by EMS medical director review of each chart. Patients who were determined to have severe agitation, required sedation, and presented to the ED were eligible for the study. Exclusion criteria included patients who were administered sedating medications for indications other than severe agitation (such as a seizure or other medical indication), patients who were taken to other hospitals so their medical records were unavailable, and patients missing other critical data such as hospital record number or administered medications.

**Variables and Outcomes**

We defined two cohorts: patients who received prehospital ketamine, with or without benzodiazepines to manage severe agitation, and patients who received prehospital benzodiazepines only. Prehospital medications were administered at the discretion of prehospital providers with approval of hospital-based physician medical control. Prehospital guidelines in this system suggest the following doses: ketamine 3 to 5 mg/kg intramuscular (IM) or 1 to 2 mg/kg intravenous (IV); midazolam 5 to 10 mg IM, 1 to 10 mg IV, or 2.5 to 10 mg intranasal (IN); and diazepam 2.5 to 10 mg IV. The dose calculation was made by prehospital providers based on estimated weight in the field. Due to the retrospective nature of the study, medication dosages were not uniform across cohorts. Patient weights were not available for all patients and weight-based dosing could not be reported.

The primary outcome of the study was admission to an inpatient psychiatry service as a surrogate marker for clinically relevant exacerbation of psychiatric disease. We did not find a prior consensus definition for psychiatric exacerbation in the literature. Secondary
outcomes included need for psychiatric evaluation in the ED, which occurred after medical clearance from ED providers, and ED LOS, defined as time of presentation to final disposition (discharge or admission). Patients requiring psychiatric evaluation were usually seen in a separate psychiatric unit within the ED but may have been seen in their ED bed depending on bed availability. Disposition was categorized as discharge, admission to inpatient psychiatry, or admission for any other nonpsychiatric indication. Nonpsychiatric hospital admission diagnoses were based on the primary hospital discharge diagnosis and broadly categorized as altered mental status, drug-related, medical indication, trauma related, or due to intubation. Notably, intubated patients were excluded from the outcomes of psychiatric admission and psychiatric evaluation in the ED as their psychiatric illness could not be assessed.

Data Collection and Processing

Data were abstracted from EMS reports and our hospital-based electronic medical record (EMR). The date of service and prehospital medications were abstracted from EMS reports by two reviewers on a structured data abstraction form and reviewed for consensus. EMS reports were then matched to hospital-based EMRs. The covariates of ED visits in the prior year and history of substance use, identified by ICD code, were abstracted from the hospital EMR. Additional covariates, including prior psychiatric diagnoses, ED LOS, ED disposition, and hospital LOS were manually abstracted from the hospital EMR by two reviewers on a structured data abstraction form. Prior psychiatric diagnosis was defined as any of the following: depression, bipolar disorder, personality disorder, substance-induced psychosis, schizophrenia, posttraumatic stress disorder, anxiety, psychosis not otherwise specified, and personality disorder. All data abstracted by manual review were confirmed by both reviewers. Full consensus between authors was required for inclusion. If consensus was not reached, the case was adjudicated by another reviewer. A written definition of each variable guided the data abstraction (Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13725/full).

Data Analysis

Continuous variables were summarized as median (range) and categorical variables as count (percentage). Demographics and outcomes were compared between benzodiazepines and ketamine cohorts using generalized estimating equations (GEE)-based logistic regression to account for any nonindependence of multiple visits by the same patient. Confidence intervals (CIs) for absolute differences in percentages and medians were calculated using GEE-based regression with the identity link or the nonparametric bootstrap, with resampling performed by patient and stratified by prehospital medication group. Fisher’s exact test was used to compare outcomes between groups in the subgroup with schizophrenia. All statistical calculations were conducted with the statistical computing language R (version 3.1.1, R Foundation for Statistical Computing). Throughout, two-sided tests were used, with statistical significance defined as p < 0.05.

RESULTS

During the study period, 173 patients received sedation for prehospital severe agitation. There were 141 visits by 134 patients that met inclusion and exclusion criteria and were included in the analysis (Figure 1). Benzodiazepines were administered in 82 visits (58%) and ketamine was administered in 59 visits (42%). During the time period when ketamine was available, patients were exposed to ketamine in all but nine of the 68 visits (87%). Demographic data are presented in Table 1. Age (median = 32 years vs. 33 years, p = 0.81), site (p = 0.59), history of substance use (87% vs. 88%, p = 0.79), ED visits in the prior year (p = 0.30), and history of mental illness (p = 0.32) were not statistically significantly different between the benzodiazepines and ketamine treatment groups, respectively. The benzodiazepine group was more often male (92% vs. 80%, p = 0.033) than the ketamine group, and there were also differences in the racial distribution between the two groups (p < 0.001), although race was categorized as “other or not reported” in 36% of visits (Table 1). Dosages of prehospital medications are summarized in Table 2.

The rate of inpatient psychiatric admission was low in both in both cohorts at 2.4% for the benzodiazepine cohort and 6.8% for the ketamine cohort (Table 3). The difference of 4.3% (95% CI = −2.4% to 12.1%, p = 0.23) in these two rates was relatively small, although the upper bound of the CI included a 12% higher inpatient admission rate in the ketamine cohort. In contrast, the ketamine cohort was observed to have a somewhat lower rate of psychiatric
evaluation in the ED compared to the benzodiazepine cohort (15% vs. 8.6%, difference = −6.8%, 95% CI = −18% to 4.7%, p = 0.23). In this case, the upper bound of the CI was a 5% higher rate in the ketamine cohort.

The estimated difference between groups in the rate of psychiatric evaluation in the ED became somewhat larger (greater in the benzodiazepine group than the ketamine group) after adjusting for history of any mental illness (difference = −10.3%, 95% CI = −21% to 0.8%, p = 0.069 for the difference between the two groups), but was little changed by separate adjustments for male sex (difference = −5.9%), history of schizophrenia (difference = −5.4%), history of substance use (difference = −7.2%), or number of prior ED visits in the past year (difference = −4.3%). There was no significant difference in median ED LOS between the ketamine and the benzodiazepine cohort (median = 394 minutes vs. 353 minutes, difference = −40 minutes, 95% CI = −152 to 68 minutes; Table 3).

Out of the 124 visits analyzed, there were 16 visits by 15 patients with schizophrenia, with 10 in the benzodiazepine cohort and 6 in the ketamine cohort (Table 4). Only one patient in each cohort had an inpatient psychiatric admission (10% vs. 17%, difference = 6.7%, 95% CI = −35% to 49%). Three patients in the benzodiazepine cohort underwent psychiatric evaluation in the ED compared to one in the ketamine cohort (50% vs. 17%, difference = −33%, 95% CI = −100% to 33%).

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Visit Group*</th>
<th>p-value†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Benzodiazepine</td>
<td>Ketamine</td>
</tr>
<tr>
<td></td>
<td>(n = 82)</td>
<td>(n = 59)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>32 (17–58)</td>
<td>33 (19–59)</td>
</tr>
<tr>
<td>Male sex</td>
<td>76 (92.7)</td>
<td>47 (79.7)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>45 (54.9)</td>
<td>29 (49.2)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0 (0.0)</td>
<td>10 (16.9)</td>
</tr>
<tr>
<td>Asian</td>
<td>5 (6.1)</td>
<td>1 (1.7)</td>
</tr>
<tr>
<td>Other/not reported</td>
<td>32 (39.0)</td>
<td>19 (32.2)</td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic medical center</td>
<td>5 (6.1)</td>
<td>5 (8.5)</td>
</tr>
<tr>
<td>Level I trauma center</td>
<td>77 (93.9)</td>
<td>54 (91.5)</td>
</tr>
<tr>
<td>History of substance use</td>
<td>71 (86.6)</td>
<td>52 (88.1)</td>
</tr>
<tr>
<td>ED visits in the prior year</td>
<td>1 (0–30)</td>
<td>1 (0–21)</td>
</tr>
<tr>
<td>0</td>
<td>39 (47.6)</td>
<td>27 (45.8)</td>
</tr>
<tr>
<td>1–2</td>
<td>17 (20.7)</td>
<td>10 (16.9)</td>
</tr>
<tr>
<td>3–10</td>
<td>22 (26.8)</td>
<td>13 (22.0)</td>
</tr>
<tr>
<td>≥10</td>
<td>4 (4.9)</td>
<td>9 (15.3)</td>
</tr>
<tr>
<td>History of mental illness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>10 (12.2)</td>
<td>6 (10.2)</td>
</tr>
<tr>
<td>Other</td>
<td>28 (34.1)</td>
<td>28 (47.5)</td>
</tr>
<tr>
<td>None</td>
<td>44 (53.7)</td>
<td>25 (42.4)</td>
</tr>
</tbody>
</table>

*Values are reported as median (range) or n (%).
†Generalized estimating equations based logistic regression (Wald test).
‡Fisher's exact test was used due to the cell with 0 observations (Black or African American).
Overall, there was no significant difference in the rate of hospital admission between the benzodiazepine and ketamine cohorts (35% vs. 51%, difference = 16%, 95% CI = –2% to 33%, p = 0.082; Table 3). The majority of hospital admissions were for nonpsychiatric indications for both the benzodiazepine (93%) and the ketamine (87%) cohort. As with the overall analysis, in the schizophrenia subgroup, there was no significant difference in the rate of hospital admission between the ketamine and benzodiazepine cohort (50% vs. 67%, difference = 17%, 95% CI = –46% to 79%, p = 0.63), but these were mainly for nonpsychiatric indications (Table 4).

Of the nonpsychiatric hospital admission diagnoses, drug-related (15% vs. 62%) and trauma-related indications (0% vs. 12%) tended to be more common in the ketamine cohort while intubations (63% vs. 4%) tended to be less common in the ketamine cohort, relative to the benzodiazepine cohort (Table 3). In patients with schizophrenia, there were three

### Table 3
Comparison of Outcomes Between Benzodiazepine and Ketamine Cohorts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Benzodiazepine (n = 82)</th>
<th>Ketamine (n = 59)</th>
<th>Difference (Ketamine Minus Benzodiazepine)</th>
<th>Value† (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED psychiatric evaluation (excludes intubated patients)</td>
<td>10 (15.4)</td>
<td>5 (8.6)</td>
<td>–6.8% (-18.0% to 4.7%)</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>ED LOS (min)</td>
<td>394 (115–2,352)</td>
<td>353 (133–2,412)</td>
<td>–40 (-152 to 68)</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Hospital admission</td>
<td>29 (35.4)</td>
<td>30 (50.8)</td>
<td>15.50% (-1.9% to 32.7%)</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>Psychiatric hospital admission</td>
<td>2 (2.4)</td>
<td>4 (6.8)</td>
<td>4.30% (-2.4% to 12.1%)</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Nonpsychiatric hospital admission</td>
<td>27 (32.9)</td>
<td>26 (44.1)</td>
<td>11.10% (-5.8% to 28.4%)</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Nonpsychiatric hospital admission diagnoses†</td>
<td>2 (7.4)</td>
<td>1 (3.8)</td>
<td>–3.6% (-16.7% to 8.7%)</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Drug-related</td>
<td>4 (14.8)</td>
<td>16 (61.5)</td>
<td>46.70% (21.4% to 69.3%)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>4 (14.8)</td>
<td>5 (19.2)</td>
<td>4.40% (-16.9% to 27.0%)</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>0 (0.0)</td>
<td>3 (11.5)</td>
<td>11.50% (-1.05% to 24.1%)</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>Intubated</td>
<td>17 (63.0)</td>
<td>1 (3.8)</td>
<td>–59.1% (-79.35% to –37.9%)</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Hospital LOS (days)§</td>
<td>2 (1–89)</td>
<td>2 (1–14)</td>
<td>0 (-0.5 to 1.0)</td>
<td>&gt;0.99</td>
<td></td>
</tr>
</tbody>
</table>

†The estimated difference in percentages (binary variables) or medians (continuous variables).

‡Of visits with an inpatient nonpsychiatric admission; there were n = 27 visits in the benzodiazepine group and n = 26 in the ketamine group.

§Of visits with an admission, there were n = 28 visits in the benzodiazepine group and n = 27 in the ketamine group after excluding patients who were transferred to outside facilities after admission.
nonpsychiatric admissions in the ketamine group, of which two (67%) were drug-related and one (33%) had a medical indication (Table 4). The benzodiazepine group had four nonpsychiatric admissions, all of which were due to intubation. Of those admitted, the median hospital LOS was 2 days in both groups (Tables 3 and 4). Outcomes comparing patient cohorts with and without history of any mental illness were completed (see Data Supplement S1).

**DISCUSSION**

In this retrospective evaluation of prehospital ketamine for the management of severe, undifferentiated agitation, we did not find significantly increased rates of psychiatric evaluation when compared to treatment with benzodiazepines alone. However, patients in both cohorts were likely to require inpatient admission, although primarily for nonpsychiatric indications. Despite the small cohort size, this is one of the largest studies examining psychiatric outcomes of patients with schizophrenia who received high-dose ketamine.

Even with the widespread use of ketamine for procedural sedation, rapid and delayed sequence intubation, and more recently behavioral control of acute agitation, few studies have examined ketamine associated psychiatric outcomes in the ED or prehospital setting. In a single-center observational study by Riddell et al.,6 patients who received ketamine for undifferentiated agitation in the ED did not require psychiatric admission more than patients who received benzodiazepines, haloperidol, or a combination of benzodiazepines or haloperidol. Our study adds important subgroup analyses for patients with a history of mental illness and, specifically, schizophrenia. While the small sample size limits our ability to reach statistically significant conclusions in this population, our study further supports the hypothesis that in undifferentiated, highly agitated patients, receiving ketamine rather than benzodiazepines for sedation has a minimal impact on meaningful psychiatric outcomes, notably psychiatric admission.

While ketamine has seen a resurgence in recent years, clinical practice guidelines recommend against using ketamine for procedural sedation in patients with schizophrenia due to the potential for psychiatric exacerbation and decompensation.8 However, the evidence for this precaution must be applied in the context of the literature. In the oft-cited paper by Lahti et al.,11 nine patients with schizophrenia were administered subanesthetic doses of ketamine (0.1–0.5 mg/kg IV) and demonstrated positive symptoms

<table>
<thead>
<tr>
<th>Variable</th>
<th>Benzodiazepine (n = 10)</th>
<th>Ketamine (n = 6)</th>
<th>Value† (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED psychiatric evaluation</td>
<td>3 (50.0)</td>
<td>1 (16.7)</td>
<td>-33.3% (-99.9% to 33.2%)</td>
<td>0.55</td>
</tr>
<tr>
<td>ED LOS (minutes)</td>
<td>406 (148–2,532)</td>
<td>301 (167–1,365)</td>
<td>-104 (-898 to 689)</td>
<td>0.64</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>5 (60.0)</td>
<td>4 (66.7)</td>
<td>16.70% (-45.5% to 78.8%)</td>
<td>0.63</td>
</tr>
<tr>
<td>Psychiatric hospital admission</td>
<td>1 (10.0)</td>
<td>1 (16.7)</td>
<td>6.70% (-35.1% to 48.5%)</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>Nonpsychiatric hospital admission</td>
<td>4 (40.0)</td>
<td>3 (50.0)</td>
<td>10.00% (-50.2% to 70.2%)</td>
<td>&gt;0.99</td>
</tr>
</tbody>
</table>

Nonpsychiatric hospital admission diagnoses‡

<table>
<thead>
<tr>
<th>Diagnoses</th>
<th>Benzodiazepine (n = 10)</th>
<th>Ketamine (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altered mental status</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Drug-related</td>
<td>0 (0.0)</td>
<td>2 (66.7)</td>
</tr>
<tr>
<td>Medical</td>
<td>0 (0.0)</td>
<td>1 (33.3)</td>
</tr>
<tr>
<td>Trauma</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Intubated</td>
<td>4 (100.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

| Hospital LOS (days)§     | 2 (1–27)                | 2 (1–2)         |

LOS = length of stay.
*Values are reported as n (%) or median (range).
†The estimated difference in percentages (binary variables) or medians (continuous variables).
‡Of visits with an inpatient psychiatric admission, there were n = 4 visits in the benzodiazepine group and n = 3 in the ketamine group.
§Of visits with an admission; there were n = 5 visits in the benzodiazepine group and n = 3 in the ketamine group after excluding patients who were transferred to outside facilities after admission.
(hallucinations, delusions, thought disorder), similar to the patient’s psychotic episodes over the course of 30 minutes. However, a subsequent study by the same group demonstrated that subanesthetic doses of ketamine in both healthy volunteers and schizophrenic patients induced a similar pattern of symptoms. We hypothesize that patients with schizophrenia are at no greater risk for developing psychotic symptoms than the general population following ketamine administration and that the risk for psychiatric symptom exacerbation may lie with subanesthetic ketamine dosing. Additional studies assessing healthy volunteers and patients with schizophrenia have demonstrated similar symptom patterns when using subanesthetic doses of ketamine.13,14

Consistent with prior work, patients in this study frequently required inpatient admission following pharmacologic management of prehospital severe agitation. The majority of these admissions, as in our study and others, are for nonpsychiatric indications.3,6 While not measured as a primary outcome in this study, patients who received ketamine were less likely to be intubated when compared to patients who received benzodiazepines (3.8% vs. 63%, p < 0.001). This is notable as other studies have demonstrated high rates of intubation following ketamine for severe agitation.3,15,16 However, we cannot rule out the presence of confounding factors, such as presentation severity, which may have biased the results toward a particular agent for a given presentation. In this study, the majority of intubations (76%) in both cohorts occurred prior to the introduction of ketamine to the prehospital pharmacy, further supporting the presence of temporal confounders that may have contributed to the differences observed.

There is little discussion in the literature regarding the indications for these nonpsychiatric admissions, but one study refers to the need for “ongoing workup of altered mental status.”5,6,15 In this study, drug-related admissions comprised the majority of admissions, especially in the ketamine cohort. For nonpsychiatric admissions, the hospital LOS was usually brief, suggesting readily reversible etiologies. While no study to date has evaluated the long-term psychiatric outcomes in patients who receive ketamine for acute agitation, long-term outcomes have been documented in schizophrenic volunteers who receive subanesthetic ketamine. In an observational study of schizophrenic volunteers who received multiple doses of subanesthetic ketamine over 2 weeks, there were no differences in adverse events, PRN medication administration, or amount of discharge neuroleptic medication when compared to schizophrenic patients who did not receive ketamine at multiple times points over a mean follow-up period of 8 months. Further studies are needed to evaluate the long-term psychiatric outcomes in patient who receive high-dose ketamine for acute agitation.

LIMITATIONS

The findings of this study must be applied in the context of its limitations. First, this is a nonrandomized, retrospective study conducted at two sites, both affiliates under the same academic institution. Given the retrospective nature, we were unable to control the selected sedative for any given patient. Therefore, we cannot establish causal relationships between our outcomes and the received sedative. Additionally, the data used in this study were collected through chart review, which can introduce additional biases, although we utilized a structured data abstraction form and two trained abstractors who achieved consensus to reduce some of these potential biases.17 Second, the observed differences in racial distribution are significantly limited by missing data as race was categorized as “other or not reported” in over one-third of visits. Third, the overall sample size was not large, although it was larger than most prior studies. Nevertheless, some small differences between cohorts may not have been statistically detectable due to sample size and the subgroup analysis of schizophrenic patients should be considered exploratory due to the limited number of such patients available. The CIs reported throughout provide a useful measure of the margin of error of each analysis and ranges of potential differences between groups that are still consistent with the observed data. However, CIs do not account for any bias due to retrospective bias or confounders. Finally, we did not account for further interventions taken during the ED course. It is possible that patients received additional diagnostics or interventions that contributed to their ultimate disposition.

Notably, intubated patients were excluded from both the primary and ED psychiatric evaluation outcome, which may bias the results towards more mild presentations of severe agitation. Intubated patients were excluded from this part of the analysis as psychiatric outcomes could not be reliably evaluated in the
ED in these patients. Potential confounders, such as ICU delirium and other administered medications, would be difficult to control for in evaluating psychiatric outcomes during or after intubation. Further, time to initial psychiatric evaluation may be significantly delayed following intubation. It is unlikely that the observed psychiatric outcomes following intubation would be related to the ketamine exposure, which may have occurred days earlier.

**CONCLUSION**

In summary, administration of prehospital ketamine for severe agitation was not associated with an increase in rate of psychiatric evaluation or psychiatric inpatient admission when compared with benzodiazepine treatment, regardless of the patient’s psychiatric history. However, patients who received a sedative for prehospital agitation were likely to require inpatient hospitalization for nonpsychiatric indications.

**References**

9. Newcomer JW, Farber NB, Jevtovic-Todorovic V, et al. Ketamine-induced NMDA receptor hypofunction as a model of memory impairment and psychosis. Neuropsychopharmacology 1999;20:106–18.

**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.13725/full

**Data Supplement S1.** Supplemental material.
Drinking and Intimate Partner Violence Severity Levels Among U.S. Ethnic Groups in an Urban Emergency Department

Raul Caetano, MD, PhD, Carol B. Cunradi, MPH, PhD, Harrison J. Alter, MD, MS, Christina Mair, PhD, and Rebecca K. Yau, PhD

ABSTRACT

Background: Emergency departments (EDs) provide care to ethnically diverse populations with multiple health-related risk factors, many of which are associated with intimate partner violence (IPV). This paper examines ethnic-specific 12-month rates of physical IPV by severity and their association with drinking and other sociodemographic and personality correlates in an urban ED sample.

Methods: Research assistants surveyed patients at an urban ED regarding IPV exposure as well as patterns of alcohol and drug use, psychological distress, adverse childhood experiences (ACEs), and other sociodemographic features.

Results: The survey (N = 1,037) achieved an 87.5% participation rate. About 23% of the sample reported an IPV event in the past 12 months. Rates were higher (p < 0.001) among blacks (34%), whites (31%), and multiethnic (46%) respondents than those among Asians (13%) and Hispanics (15%). Modeled results showed that black respondents were more likely than Hispanics (reference) to report IPV (adjusted odds ratio [AOR] = 1.69, 95% confidence interval [CI] = 1.98–2.66, p < 0.05) and that respondents’ partner drinking was associated with IPV (AOR = 1.85, 95% CI = 1.25–2.73, p < 0.01) but respondents’ drinking was not. Use of illicit drugs, younger age, impulsivity, depression, partner problem drinking, ACEs, and food insufficiency were all positively associated with IPV.

Conclusions: There was considerable variation in IPV rates across ethnic groups in the sample. The null results for the association between respondents’ drinking and IPV was surprising and may stem from the relatively moderate levels of drinking in the sample. Results for ethnicity, showing blacks as more likely than Hispanics to report IPV, support prior literature.

From the Prevention Research Center, Pacific Institute for Research and Evaluation (RC, CBC, RKY), Berkeley, CA; the Department of Emergency Medicine, Highland Hospital–Alameda Health System (HJA), Oakland, CA; and the Department of Behavioral and Community Health Sciences, University of Pittsburgh Graduate School of Public Health (CM), Pittsburgh, PA.

Received August 29, 2018; revision received November 19, 2018; accepted November 22, 2018.

A summary of results has been presented at the following meetings: the Research Society on Alcoholism Annual Meeting (poster), San Diego, CA, June 17–20, 2018; the Society for Academic Emergency Medicine Western Regional Meeting, Albuquerque, NM, February 3, 2018; and the Society for Academic Emergency Medicine Annual Meeting, Indianapolis, IN, May 18, 2018.

Research reported in this publication was supported by R01-AA022990 and P60-AA006282 from the National Institute on Alcohol Abuse and Alcoholism to the Pacific Institute for Research and Evaluation.

RC has received grant funding to the University of Texas Houston Health Science Center and the Pacific Institute for Research and Evaluation for investigator-initiated research from the National Institute on Alcohol Abuse and Alcoholism as well as paid participation in grant review meetings from the National Institute of Health; CBC has received grant funding from the Pacific Institute for Research and Evaluation for investigator-initiated research from the National Institute on Alcohol Abuse and Alcoholism as well as paid participation in grant review meetings from the National Institute of Health; HJA has received grant funding to the Andrew Levitt Center for Social Emergency Medicine for investigator-initiated research from the National Institute on Alcohol Abuse and Alcoholism, the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and the Robert Wood Johnson Foundation; CM has received grant funding to the University of Pittsburgh for investigator-initiated research from the National Institute on Alcohol Abuse and Alcoholism, as well as paid participation in grant review meetings from the National Institutes of Health. She has also received payment for consultation from the Pacific Institute of Research and Evaluation and payment for review services from Rockman et al., an independent research and evaluation firm. She has received grant funding from the University of Pittsburgh’s Clinical & Translational Science Institute; RKY is presently supported by a T32 award from the National Institute on Alcohol Abuse and Alcoholism to the University of California at Berkeley. The authors have no potential conflicts to disclose.

© 2019 by the Society for Academic Emergency Medicine
doi: 10.1111/acem.13706
Intimate partner violence (IPV) is the physical violence, sexual violence, stalking, and psychological aggression (including coercive tactics) that is perpetrated against a partner in a romantic relationship. This paper focuses on physical IPV, which as other types of IPV, remains a major public health problem in the United States. Surveys with face-to-face interviews of heterosexual couples in the community over the past 30 years have shown that about one in five couples in the United States reports at least one episode of physical IPV in the past 12 months. These rates also vary considerably by ethnicity. For instance, in one survey of couples 17% of Hispanic couples, 23% of black couples, and 11% of white couples reported an incident of male-to-female physical IPV in the past 12 months. Rates of female-to-male physical IPV were 21% among Hispanics, 30% among blacks, and 15% among whites. More recent cross-sectional data from the 2010–2012 National Intimate Partner and Sexual Violence Survey show 12-month rates for physical IPV of 3.9% among women and 4.7% among men. These rates are lower than those above likely due to differences in survey methods, especially telephone interviewing versus face-to-face and interviews with one person only and not with both persons in the couple.

Emergency departments (EDs) remain the entry point and sometimes the only setting for clinical care for physical and behavioral problems for a large part of the U.S. population, especially the 8.8% (28.3 million) without health insurance. EDs have higher prevalence rates of IPV than other health care settings, second only to substance use treatment units. IPV screening of patients presenting at EDs show rates ranging from 9% to 37% for a 12-month time frame and as high as 46% for lifetime. EDs are therefore an excellent setting for screening and identification of IPV and subsequent referral for treatment. Identification of ED patients involved in IPV also helps ED personnel to arrive at a better understanding of patients’ reasons for seeking care given that these patients may present with problems associated with IPV (e.g., bone fractures, posttraumatic stress disorder [PTSD], depression).

Unfortunately, the association between IPV and ethnicity has not been well examined in ED samples, even though EDs serve a high proportion of U.S. ethnic minority patients (e.g., Bazargan-Hejazi et al., Houry et al., Lipsky et al.). Also, when ED data on ethnicity are analyzed, this is done in a limited way by using the variable as a correlate in multivariable analyses. The estimated rates usually show that black women compared to whites, Hispanics, or all others combined have higher rates of IPV. Black women reporting IPV are more likely to use tobacco, to be depressed, and to abuse alcohol and drugs. Both Black and Hispanic women reporting IPV are also more likely to binge drink, use drugs, and have a higher level of impulsivity personality traits than women who do not report IPV. Men who perpetrate IPV are more likely than nonperpetrators to be young, not married, black, or Hispanic, with lower education and income, being unemployed, with heavier alcohol use, and higher rates of both alcohol and illicit drug abuse/dependence.

Drinking has been consistently associated with IPV both in the general population and in ED samples. Among U.S. couples in the community, about one-third of those who report an IPV event in the past 12 months also report that one or both partners were drinking during the event. Alcohol has also been associated with more severe violence, with more severe injuries from violence, and with greater chronicity of violence. In ED samples, binge drinking, heavier drinking, drinking problems, hazardous drinking as measured by higher Alcohol Use Disorder Identification Test (AUDIT) scores, mental health problems, a positive diagnosis of alcohol and or illicit drug abuse, and dependence have all been associated with IPV.

This paper examines ethnic-specific 12-month rates of victimization, perpetration, and mutual physical IPV and their association with drinking and other sociodemographic and personality correlates in an urban ED sample. It provides two distinct contributions to the existing literature: 1) a detailed focus on the association between drinking, ethnicity, and IPV and 2) analyses that recognize two levels of IPV severity, moderate and
severe. Following results in the existing literature reviewed above, the analyses will: 1) describe ethnic specific sociodemographic characteristics, drinking, drug use, psychological risk factors and IPV rates (perpetration, victimization, mutual violence, moderate and severe IPV) in the sample; 2) test the expectation that blacks and those who are younger will have a higher rate of IPV in general and also a higher rate of severe IPV; 3) test the expectation that respondents weekly mean drinking volume, the mean monthly frequency of alcohol intoxication, and illicit drug use will be independently and positively associated with IPV, as will their partners’ problem drinking; and 4) test the expectation that adverse childhood experiences (ACEs), PTSD, depression, impulsivity, and anxiety will also be positively associated with IPV, with severity of IPV increasing as these problems become more severe.

**METHODS**

Sample and Data Collection

Trained, bilingual (English and Spanish) research assistants (RAs) recruited nonemergent patients in the ED of a hospital Level I trauma center and county safety-net hospital. After receiving training about the study’s conceptual framework, data collection techniques, and protection of human subjects, the RAs pilot tested the survey with 41 patients (39% African American, 41% Hispanic, 44% male) who met the study’s eligibility criteria. The pilot testing process helped the research team identify obstacles to study recruitment and refine data collection procedures and provided the team with baseline information as to average survey interview length. After the RAs were debriefed with project staff and making minor adjustments to the questionnaire, the finalized survey data collection effort was launched.

The initial sample size estimate called for the enrollment of 800 married, cohabiting, or dating adults aged 18 to 50 years. This was based on calculations that using linear regression analyses, power would be 80% to detect a small overall effect ($R^2 = 0.02$) with 20 predictors, $\alpha = 0.05$, and $n = 800$. Power would be 85% to detect small incremental changes of adding single variables to the regression equations ($\Delta R^2 = 0.01$) with 19 prior predictors, a prior $R^2$ value of 0.10, and $\alpha = 0.05$. The analytical objectives that guided these power analyses were: 1) identify aspects of alcohol consumption and drinking context (e.g., type and frequency of venue utilization, amount of alcohol consumed per venue) that are associated with the occurrence and frequency of IPV and 2) determine the extent that known IPV risk factors (e.g., ACEs, impulsivity, depression) are associated with drinking contexts. Once the survey was being conducted, costs were lower than anticipated, and the sample N was increased to increase power for planned analyses and additional analyses.

Participants eligibility criteria included 18 to 50 years old; English or Spanish speaker; residence in the county where the study was conducted; and married, cohabiting, or in a romantic (dating) relationship for the past 12 months. Patients who were intoxicated, experiencing acute psychosis or suicidal or homicidal ideation, were cognitively and/or psychologically impaired and unable to provide informed consent, in custody by law enforcement, or in need of immediate medical attention were excluded. Thirty-four patients who could not speak English or Spanish were excluded. Two interviewers per shift staffed the ED during weekday peak volume hours (9AM–9PM) to recruit eligible patients to the study.

Data were collected from February through December 2017. Patients could opt to be interviewed in English or Spanish. In this latter case, a Spanish version of the questionnaire, which had been validated through translation into Spanish and retranslation into English followed by verification, was then used. Figure 1 shows in sequence from top to bottom the number of patients identified in electronic health records, screened, found eligible, and interviewed. The side arrows and boxes show the number of patients from the preceding inclusion step that were not considered further. Initially, 1,066 patients (90% of those eligible) agreed to participate. Of these, 1,037 were interviewed (87.5% participation rate, 53% female), of which 376 were interviewed in Spanish (36% of the full sample and 72% of respondents who self-identified as Hispanic/Latino).

The RAs obtained informed consent in a private area adjacent to the ED waiting room or in the patient’s room without others present. Twenty-nine patients terminated the survey interview before completion, mostly due to interruption for medical services. Ten interviewed patients identified themselves as American Indian/Alaskan Native, seven as Native Hawaiian/Pacific islanders, and 26 did not report their race/ethnicity. These respondents are not part of the analytical sample herein ($N = 994$). Patient survey data were collected by the RAs using computer-assisted
personal interview with computer tablets running the Qualtrics platform. Questionnaire development was guided by the study’s conceptual model as described in Cunradi et al. and formative qualitative research conducted at the outset of the study. Regarding the latter, project ethnographers conducted semistructured interviews with 30 nonacute ED patients in which they asked open-ended questions as to respondents’ use of alcohol and the venues where they drank. Missing data were negligible; none of the variables analyzed in this paper had more than 4% information missing. These data therefore were left as missing. The project was approved by the Committee for the Protection of Human Subjects of the hospital where the study was conducted.

Measurements

Ethnicity. This was based on self-identification. Respondents were asked: What racial or ethnic group (s) best describes you? (More than one category may be checked): Asian; black, African American; Latino, Hispanic; white, Caucasian; Native American Indian/Alaskan Native; Native Hawaiian/other Pacific Islander; and some other race (specify). Respondents who selected more than one category were identified as multiethnic.

IPV. Physical IPV was measured with the 12 items on physical assault in the Revised Conflict Tactics Scale (CTS2), which have been used in prior ED-based IPV studies. Two levels of IPV severity, moderate and severe, were operationalized based on previously published reports. Moderate violence consisted of at least one of the following acts: threw something at partner that could hurt, pushed or shoved, grabbed, slapped, or twisted partner’s arm or hair. Severe violence consisted of kicked, punched or hit with something that could hurt, beat up, choked, burned or scalded on purpose, slammed against a wall, or used a knife or gun. Cronbach’s alpha for the scale in the data set under analysis was 0.85.

Quantity and Frequency of Drinking. Respondents were asked the frequencies with which they had

![Figure 1. Study sample recruitment.](image-url)
one or more, two or more, three or more, six or more, and nine or more drinks in the past 4 weeks. A “drink” was defined as a 12-ounce can of beer, a 5-ounce glass of wine, or a 1-ounce shot of liquor. Respondents who did not use alcohol in the past 4 weeks were asked the same questions over the past year. Using a mathematical model described in Grue-newald et al., a measure of the weekly mean drinking volume was calculated. This measure was log-transformed due to skewness for inclusion in the ordered logistic regression. Test–retest reliabilities of these measures vary from $r = 0.65$ for drinking quantities to $r = 0.85$ for drinking frequencies.

**Alcohol Intoxication.** This was assessed with the following question: “During the past 12 months, about how many times did you drink enough to feel intoxicated or drunk, that is, when your speech was slurred, you felt unsteady on your feet, or you had blurred vision?” This measure was log-transformed due to skewness for inclusion in the ordered logistic regression.

**Illicit Drug Use.** This measure covered drug use in the 12 months preceding the interview. Respondents were asked how many days they used the following drugs: marijuana or hashish (without a doctor’s prescription), amphetamines, cocaine, heroin, and pain relievers not prescribed for you. In the current study, illicit drug used was operationalized as a dichotomous variable of either any or no drug use.

**Partner Problem Drinking.** The three-item AUDIT-C was used to measure the respondent’s assessment of his/her spouse/partner’s drinking. Male partners with a score above 4 and female partners with a score above 3 in the test 0–12 scale were considered hazardous drinkers. Internal consistency reliability for this scale in the data set under analysis as measured by Cronbach’s alpha was 0.81.

**ACEs.** The modified ACEs measures exposure to six adverse experiences the respondent may have had “during their first 18 years of life”: 1) exposure to a mentally ill person in the home, 2) parent/caregiver alcoholism, 3) sexual abuse, 4) physical abuse, 5) psychological abuse; and 6) violence directed against the respondent’s mother. These six exposures are summed to create the ACE variable (range = 0–6). Internal consistency reliability (Cronbach’s alpha) in the data set under analysis was 0.74.

**Impulsivity.** This was measured with three items that assessed respondents’ agreement with the following statements: “I often act on the spur-of-the-moment without stopping to think”; “You might say I act impulsively”; and “Many of my actions seem to be hasty.” Four response categories could be chosen ranging from “not at all” to “quite a lot.” Alpha reliability for the scale in the data set under analysis was 0.79.

**Anxiety and Depression.** This was measured with Hospital Anxiety and Depression Scale, which has been successfully used in previous ED studies (see Bokma et al., Perruche et al.36). Both anxiety and depression were measured with seven items each on a 4-point Likert-type scale (e.g., 1 = not at all, 4 = very often). Alpha reliability in the dataset under analysis was 0.69 for the depression scale and 0.81 for the anxiety scale. Following Brennan et al.37 a cutoff point equal to or higher than eight identified positives in both scales. This cutoff gives sensitivities of 0.82 and 0.78 and specificity of 0.74 and 0.78 for depression and anxiety, respectively.

**PTSD.** This measure is from the primary care screener for PTSD, and it too has been successfully used in ED studies (see Hanley et al., Mills et al.40). It asks subjects about past-month symptoms they may have felt in response to a “frightening, horrible, or upsetting” experience. Answers were coded yes or no, and a score of 3 or more is considered positive. Internal consistency of this scale in the data set under analysis was $\alpha = 0.83$.

**Perceived Neighborhood Disorder.** This was measured with Hill and Angel’s 10-item measure of neighborhood disorder. Items cover the extent to which assaults, muggings, drug dealing, gangs, unsafe streets, thefts, teenage pregnancy, abandoned houses, police not available, unsupervised children, and high unemployment are neighborhood problems. Respondents could select one of the following three categories to answer each item: not a problem, somewhat of a problem, or a big problem. Cronbach’s alpha was 0.88 in the data set under analysis.

**Other Sociodemographic Variables.**

- Gender. Gender was a dichotomous variable coded as male and female (reference).
• **Age.** The age of respondents was used as a categorical variable: 18–29, 30–39, and 40 years and older (reference).

• **Level of Education.** Respondents were categorized into four education categories: 1) less than high school (reference), 2) completed high school or GED, 3) some college or technical or vocational school, or 4) completed 4-year college or higher.

• **Importance of Religion.** This variable had four categories: very important (reference), somewhat important, not very important, or no important at all.

• **Marital Status.** This is a four-category variable: 1) married (reference), 2) living with partner, 3) separated or divorced, or 4) never married. Widowers (n = 33) were dropped from the analyses because 23 had no alcohol use disorder, which created estimation problems in the multivariable analysis.

• **Food Insufficiency.** Respondents were asked their level of agreement with the statement, “In the past 12 months, the food we bought ran out and we didn’t have money to get more.” Response categories were never (reference), sometimes true, or often true.

### Data Analyses

All analyses were conducted with Stata 15.0.42 Associations in bivariate analyses (Tables 1–3) were tested with chi-square tests with level of statistical significance adjusted using a Bonferroni correction when necessary. This indicated that significance levels for Tables 1 and 2 should be 0.05/10 = p < 0.005. Denominators for estimating drinking indicators in Table 2 were drinkers only. All other rates use the full sample as the denominator.

Multivariable analysis of IPV severity in Table 4 was conducted with Stata’s “ologit” procedure, which implements an ordered logistic regression under a proportional odds assumption. Test results indicated that the model tested fit the proportional odds assumption: $\chi^2 = 36.01$ with df = 27 and $p = 0.11$. Therefore, adjusted odds ratios (AORs) in Table 4 represent both the odds of moderate plus severe IPV contrasted with no IPV, and the odds of severe IPV contrasted with no IPV plus moderate IPV. Independent variables were entered in the model in one step. These variables were selected for inclusion in the model based on previous results in the literature indicating they had a statistically significant association with IPV in community samples (e.g., neighborhood disorder), ED samples, or both (e.g., gender, age, marital status, ethnicity, religion, drinking volume and intoxication, illicit drug use, impulsivity, ACEs, depression).4,9,15,25,26,43,44

### RESULTS

#### Sociodemographic Indicators

There was no variation in the mean age of participants nor in the proportion of women across the ethnic groups (Table 1). However, the proportion of high school graduates, of participants with an annual income up to $40,000/year, of married participants, and of unemployed participants varied significantly across ethnic groups. The proportion of respondents reporting not being able to “make ends meet” and the proportion reporting food insufficiency “at least sometimes” did not vary significantly across ethnic groups.

#### Drinking, Drug Use, and Psychological Status Indicators

Pairwise comparisons indicated that there were significant differences in the monthly mean number of

### Table 1: Sociodemographic Indicators by Ethnicity in an Urban ED Sample

<table>
<thead>
<tr>
<th>Sociodemographic Characteristics</th>
<th>Hispanics (n = 520)</th>
<th>Asians (n = 51)</th>
<th>Blacks (n = 299)</th>
<th>Whites (n = 68)</th>
<th>Multiethnic (n = 56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SE)*NS</td>
<td>35.3 (0.35)</td>
<td>36.9 (1.11)</td>
<td>34.9 (0.53)</td>
<td>35.7 (0.99)</td>
<td>34.0 (1.22)</td>
</tr>
<tr>
<td>% Women NS</td>
<td>50</td>
<td>51</td>
<td>58</td>
<td>53</td>
<td>63</td>
</tr>
<tr>
<td>% High School Graduate†</td>
<td>48</td>
<td>90</td>
<td>86</td>
<td>85</td>
<td>93</td>
</tr>
<tr>
<td>% Annual Income up to $40,000†</td>
<td>61</td>
<td>67</td>
<td>62</td>
<td>53</td>
<td>68</td>
</tr>
<tr>
<td>% Married†</td>
<td>51</td>
<td>69</td>
<td>24</td>
<td>22</td>
<td>14</td>
</tr>
<tr>
<td>% Unemployed†</td>
<td>22</td>
<td>26</td>
<td>39</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>% “Cannot make ends meet”NS</td>
<td>22</td>
<td>12</td>
<td>17</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>% Reports food insufficiency “at least sometimes”NS</td>
<td>49</td>
<td>37</td>
<td>49</td>
<td>47</td>
<td>59</td>
</tr>
</tbody>
</table>

*Significance level for differences between age means after Bonferroni correction: 0.05/10, all pairwise t tests = NS; $\chi^2$ NS = not significant. †p < 0.001
alcohol intoxication events between Hispanics and blacks, Hispanics and Whites, and Hispanic and multiethnic participants (Table 2). The proportion of participants that used drugs in the past 12 months was higher among blacks, whites, and multiethnic respondents, with over half of these respondents reporting such use. About a third of blacks, whites, and multiethnic respondents had a partner with a positive AUDIT-C, while among Hispanics and Asian Americans this was true of less than one-fifth of respondents.

The analyses of drinking indicators in Table 2 was repeated with each ethnic group split between those who reported and those who did not report IPV (data not shown). There were significant differences for the proportion of Hispanics and Blacks with an AUDIT-C–positive partner.

The proportion of respondents reporting two or more ACEs, were PTSD positive, and were anxiety positive was significantly different across ethnic groups. The proportion with two or more ACEs and were PTSD positive was higher among whites and those with a multiethnic background. Blacks, whites, and multiethnic respondents were more likely to score positively in the anxiety scale.

### IPV

About 23% of the sample reported at least one incident of IPV in the past 12 months (data not shown). Any IPV as well as IPV perpetration, victimization, and mutual violence were significantly higher among

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinking, Drug Use, and Psychological Indicators by Ethnicity in an ED Sample</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Drinking indicators*</td>
</tr>
<tr>
<td>Weekly mean number of drinks (SE)†</td>
</tr>
<tr>
<td>Monthly mean number of intoxication events (SE)‡</td>
</tr>
<tr>
<td>Respondents drug use and partner drinking indicator</td>
</tr>
<tr>
<td>% Used any illicit drug past 12 months***</td>
</tr>
<tr>
<td>% Partner AUDIT-C positive§</td>
</tr>
<tr>
<td>Psychological indicators</td>
</tr>
<tr>
<td>% Two or more ACEs§</td>
</tr>
<tr>
<td>% PTSD positive§</td>
</tr>
<tr>
<td>% Anxiety positive§</td>
</tr>
</tbody>
</table>

ACEs = adverse childhood experiences; PTSD = posttraumatic stress disorder.
*Denominators for drinking indicators are drinkers only.
†Pairwise t-test Hispanic × black p < 0.01; Hispanic × white p < 0.01; Hispanic × multiethnic p < 0.01; all others p = NS.
‡All pairwise t-tests NS; χ² NS = not significant.
§p < 0.001.

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPV Rates (Proportions) by Ethnicity in an ED Sample</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>% Any IPV*</td>
</tr>
<tr>
<td>% Perpetration*</td>
</tr>
<tr>
<td>% Victimization</td>
</tr>
<tr>
<td>% Mutual violence</td>
</tr>
<tr>
<td>IPV severity*</td>
</tr>
<tr>
<td>% Moderate IPV</td>
</tr>
<tr>
<td>% Severe IPV</td>
</tr>
</tbody>
</table>

IPV = intimate partner violence.
*χ² p < 0.001.
multietnic respondents, followed by blacks and whites (Table 3). Mutual violence was significantly higher among blacks and multietnic respondents. Multietnic respondents, blacks, and whites reported higher levels of both moderate and severe IPV than Hispanics and Asian Americans.

Correlates of IPV

Correlates with statistically significant associations with IPV severity are shown first in Table 4. Blacks were a little over 1.5 times more likely than Hispanics to report IPV. Respondents who reported illegal drug use in the past 12 months were over two times more likely than those who did not report drug use to report IPV. Younger respondents, those who have higher scores in the impulsivity scale, those who had a partner with a positive AUDIT-C score, those with a positive PTSD screener, and those with a higher ACE were also more likely to report IPV. In addition, respondents who reported that they experienced food insufficiency “sometimes” or “often” were about 1.76 and 1.95 times more likely, respectively, than those who never experienced food insufficiency to report IPV. Finally, respondents who scored higher in the neighborhood social disorder scale as well as those with higher scores in the depression scale were also more likely to report IPV.

DISCUSSION

The only hypothesis put forward in the introduction that was not confirmed was that about a positive association between anxiety and IPV. Results in Tables 1, 2, and 3 and the multivariable model in Table 4 show that the sample analyzed herein has a profile of IPV and its correlates that closely resembles that of other ED samples. This includes an overall rate of IPV (23%) that is in the midrange of the rates of 9% to 37% for a 12-month timeframe that have been described in ED samples.7–9,11

The model in Table 4 indicates that, with controls for several confounders, black respondents compared to Hispanics, those who used illicit drug use, and those ages 18 to 29 compared to age 40+ have increased odds of IPV. Regarding ethnicity, it is possible that the relatively small number of Asian, white, and multietnic respondents lead to lack of power to assess some of the associations in these three groups. As in other community and ED samples, blacks were significantly more likely to report IPV.8,9,26,27,45

Other variables that increased significantly the odds of IPV were impulsivity, partner with a positive score in the AUDIT-C, PTSD, ACEs, food insufficiency, neighborhood social disorder, and depression. Impulsive behavior is associated with lower behavior control and thus a higher chance of acting out a violent behavior.12,46 A partner with a positive score in the AUDIT-C underscores that although drinking by

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Multivariate Ordered Logistic Regression of IPV Severity on Sociodemographic, Psychological and Drinking-Related Variables (n = 917)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnicity (Ref: Hispanics)</td>
<td>AOR 95% CI</td>
</tr>
<tr>
<td>Asian American</td>
<td>1.63 0.61–4.35</td>
</tr>
<tr>
<td>Black*</td>
<td>1.69 1.98–2.66</td>
</tr>
<tr>
<td>White</td>
<td>1.17 0.55–2.45</td>
</tr>
<tr>
<td>Multietnic</td>
<td>1.61 0.80–3.23</td>
</tr>
<tr>
<td>Any illicit drug past 12 months†</td>
<td>2.44 1.65–3.62</td>
</tr>
<tr>
<td>Age (Ref: 40+ years)</td>
<td></td>
</tr>
<tr>
<td>18–29 years‡</td>
<td>2.08 1.31–3.33</td>
</tr>
<tr>
<td>30–39 years</td>
<td>1.24 0.79–1.94</td>
</tr>
<tr>
<td>Impulsivity scale score†</td>
<td>1.15 1.07–1.24</td>
</tr>
<tr>
<td>Partner AUDIT-C positive†</td>
<td>1.85 1.25–2.73</td>
</tr>
<tr>
<td>PTSD*</td>
<td>1.57 1.06–2.34</td>
</tr>
<tr>
<td>ACEs*</td>
<td>1.14 1.02–1.29</td>
</tr>
<tr>
<td>Food insufficiency (Ref: never true)</td>
<td></td>
</tr>
<tr>
<td>Sometimes true‡</td>
<td>1.76 1.19–2.61</td>
</tr>
<tr>
<td>Often true*</td>
<td>1.95 1.14–3.35</td>
</tr>
<tr>
<td>Neighborhood social disorder*</td>
<td>1.04 1.004–1.07</td>
</tr>
<tr>
<td>Depression scale†</td>
<td>1.07 1.01–1.13</td>
</tr>
<tr>
<td>Male (Ref: female)</td>
<td>1.04 0.72–1.50</td>
</tr>
<tr>
<td>Anxiety scale</td>
<td>0.96 0.91–1.01</td>
</tr>
<tr>
<td>Log weekly mean drinking volume</td>
<td>1.15 0.93–1.42</td>
</tr>
<tr>
<td>Log monthly mean frequency of intoxication</td>
<td>1.11 0.80–1.55</td>
</tr>
<tr>
<td>Marital status (Ref: married)</td>
<td></td>
</tr>
<tr>
<td>Living with partner</td>
<td>1.14 0.72–1.81</td>
</tr>
<tr>
<td>Separated/divorced</td>
<td>1.62 0.62–4.20</td>
</tr>
<tr>
<td>Never married</td>
<td>1.02 0.63–1.66</td>
</tr>
<tr>
<td>Education (Ref: no high school)</td>
<td></td>
</tr>
<tr>
<td>High school completed</td>
<td>1.36 0.86–2.14</td>
</tr>
<tr>
<td>Some college/technical</td>
<td>1.36 0.80–2.30</td>
</tr>
<tr>
<td>College degree or more</td>
<td>0.68 0.29–1.60</td>
</tr>
<tr>
<td>Importance of religion (Ref: very important)</td>
<td></td>
</tr>
<tr>
<td>Somewhat important</td>
<td>0.67 0.43–1.04</td>
</tr>
<tr>
<td>Not very important</td>
<td>0.78 0.39–1.55</td>
</tr>
<tr>
<td>Not important at all</td>
<td>0.79 0.32–1.96</td>
</tr>
</tbody>
</table>

AOR = adjusted odds ratio; ACEs = adverse childhood experiences; IPV = intimate partner violence; PTSD = posttraumatic stress disorder.
†p < 0.05.
‡p < 0.001.
§p < 0.01.
one partner in the dyad may be enough to increase the odds of IPV, frequently (in one-third of the cases), IPV occurs when both partners are drinking.4,26

Posttraumatic stress disorder is associated with increased levels of anxiety and depression, which can be a result of victimization from IPV8 or increase the odds of aggressive behavior and thus IPV. ACEs, which often include neglect, maltreatment, sexual, emotional and physical abuse, as well as observing IPV between parents or other household members, has been consistently identified as a factor of risk for involvement in IPV during adult life.44,47 Food insufficiency is an indicator of socioeconomic difficulties for a couple and other household members, but it perhaps triggers more anxiety and a stronger sense of insecurity than, say, less education or lower income, which then increases the odds of IPV. Finally, depression and perceived neighborhood social disorder are also positively associated with IPV. Recent studies have reported a relatively important irritability component in depression, which when present could lead to increased conflicts in a couple and then to IPV.48–50 Depression can also be a consequence of perpetration or victimization by IPV.26,46,51 Perceived neighborhood social disorder can lead to conditions where behavioral norms are more lax and more accepting of violence and informal social controls that minimize violence (e.g., neighbors who call the police or intervene) are not present.43

Altogether, the results show that IPV is associated with an array of risk factors and that it is comorbid with other behavioral and psychological disorders. Health professionals working in ED settings must be aware of this comorbidity and look for common factors of risk across behavioral problems, apply screens to detect these problems, and be ready to intervene and refer to specialized services if necessary. Given that IPV is often associated with physical trauma (e.g., lacerations, fractures), the presence of IPV factors of risk identified in this paper in patients with such presentations, especially if the patient is a woman, should guide ED personnel to investigate history of IPV. ED personnel can also use the knowledge about IPV correlates and risk factors in this paper to develop more personalized and culturally appropriate assessments of patients who have screened positive for IPV, building rapport and thus maximizing the chances that these patients will follow up with suggested referrals.

LIMITATIONS

The major limitation in the analyses is the small number of respondents in some of the ethnic groups interviewed. This reflects the ethnic composition of ED samples in urban areas and in the region the study took place. This limitation leads to lack of power for some of the analyses presented herein. Because the study was conducted in a single ED, results may not generalize to other EDs and other health settings. Exclusion of subjects who were alcohol intoxicated from the study may have led to an underestimation of the association between drinking and IPV. In addition, recall bias may have affected subjects’ information about events that reached back over 12 months. Finally, if under-reporting of IPV varies across ethnic groups, those groups that are more willing to report will appear to have higher prevalence rates.

CONCLUSIONS

First, as seen here and in other ED studies reviewed above, the association between intimate partner violence and ethnicity matters; blacks and perhaps multiethnic groups are more at risk for involvement in this type of problem than other groups. In the long term, this identification of population groups with higher prevalence of intimate partner violence will further understanding of potential social and cultural correlates of this public health problem, which may help develop more effective approaches to address intimate partner violence in and outside ED settings. More immediately, this informs decisions about what groups should be the focus of specialized personnel and specific intimate partner violence–related actions such as brief intervention or referral to treatment.

References

INTIMATE PARTNER VIOLENCE AND ETHNICITY


31. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions


42. Stata. Stata Statistical Software. College Station, TX: Stata Corp LP, 2015


Variation in Emergency Department Adherence to Treatment Guidelines for Inpatient Pneumonia and Sepsis: A Retrospective Cohort Study

Stacy A. Trent, MD, MPH, Zachary J. Jarou, MD, Edward P. Havranek, MD, Adit A. Ginde, MD, MPH, and Jason S. Haukoos, MD, MSc

ABSTRACT

Objectives: Evidence-based clinical practice guidelines (CPGs) for the treatment of pneumonia and sepsis have existed for many years with multiple studies suggesting improved patient outcomes. Despite their importance, little is known about variation in emergency department (ED) adherence to these CPGs. Our objectives were to estimate variation in ED adherence across CPGs for pneumonia and sepsis and identify patient, provider, and environmental factors associated with adherence.

Methods: This was a multicenter retrospective study using standard medical record review methods. The population consisted of consecutive adults hospitalized for pneumonia or sepsis (identified by discharge ICD-9 codes) at five Colorado hospitals (two academic, three community) who were admitted to the hospital from the ED and for whom the ED diagnosed or initiated treatment. The outcome measured was ED adherence to the CPG (primary) and in-hospital mortality (secondary). Hierarchical generalized linear models were used for analysis.

Results: Among 827 patients, ED care was 57% adherence to CPGs with significant variation in adherence across CPGs (sepsis 50%, pneumonia 64%, p < 0.001). Patients were less likely to receive adherent care if they presented with chief complaints that were associated but not typical of the diagnosis (odds ratio [OR] = 0.6, 95% confidence interval [CI] = 0.4–0.8), received an ED diagnosis that was not specific to the CPG (associated diagnosis OR = 0.3 [95% CI = 0.2–0.5]; unrelated diagnosis OR = 0.4 [95% CI = 0.2–0.6]) or presented to a community hospital (OR = 0.6, 95% CI = 0.4–0.9). ED CPG nonadherence was associated with higher in-hospital mortality (OR = 2.4, 95% CI = 1.2–4.8).

Conclusion: Adherence to ED infectious CPGs for pneumonia and sepsis varies significantly across diseases and types of institutions with significant room for improvement, especially in light of a significant association with in-hospital mortality.
Pneumonia and sepsis are two of the most common reasons for hospital admission and death in the United States, accounting for 2.4 million hospitalizations, 200,000 in-hospital deaths, and $35.8 billion in aggregate hospital costs annually.1 Emergency departments (EDs) play a vital role in providing evidence-based care for the management of pneumonia and sepsis as the initial evaluation and treatment is most often initiated in the ED, and both conditions have clinical practice guidelines (CPGs) relevant to ED management that have been shown to improve mortality, hospital length of stay (LOS), and costs.2–26

While ED guidelines for the treatment of pneumonia and sepsis have existed for more than a decade, both have undergone recent updates. For pneumonia, the recommendations to obtain blood cultures on all patients and administered antibiotics within 6 hours of ED arrival have been retired from the Center for Medicare and Medicaid Services (CMS) Pneumonia Core Measure. The Infectious Disease Society of America (IDSA) and the American Thoracic Society (ATS), however, continue to recommend that guideline concordant antibiotics be given in accordance with a patient’s risk for atypical organisms.27,28 For sepsis, the early management bundles advocated by the Surviving Sepsis Campaign (SSC) have broken down the original 6-hour early resuscitation bundle to two distinct resuscitation bundles with 3- and 6-hour goals, with the 3-hour bundle specifically targeted toward ED management.4 Recognizing the importance and need to improve evidence-based care for patients hospitalized with sepsis, CMS introduced a sepsis core measure (SEP-1) in 2016 that parallels the SSC’s 3-hour bundle and mandates that hospitals publicly report their adherence to the SEP-1 guideline.29

Previous literature on CPG adherence for pneumonia and sepsis does not reflect current guidelines. In addition, previous literature on sepsis CPG adherence using the SSC registries has largely mixed ED and inpatient care, making it difficult to assess guideline adherence specifically initiated in the ED.15,16 Thus, the primary objective of this study was to estimate ED adherence to CPGs for inpatient community acquired pneumonia and sepsis treatment. Secondary objectives were to identify patient, physician, and environmental factors associated with ED adherence and estimate the association between adherence and in-hospital patient outcomes including mortality and hospital LOS.

METHODS

Study Design
We performed a retrospective study using standardized medical record review to identify a large, consecutive patient population to determine variation in ED adherence to CPGs for inpatient community acquired pneumonia treatment and early identification and management of sepsis and septic shock. The institutional review boards at each participating hospital approved the study with a waiver of consent.

Study Setting and Population
This study was performed at five hospitals in Colorado with heterogeneous and diverse practice environments that represent the main types of EDs including: 1) urban academic safety-net hospital, 2) suburban academic tertiary care hospital, and 3) urban and rural community hospitals (Table 1). Each ED was staffed by emergency medicine board-certified or board-eligible physicians at all times.

Consecutive patients were identified retrospectively by any hospital discharge ICD-9 codes for pneumonia (481–486.xx) or severe sepsis/septic shock (785.52, 995.92).30,31 Starting on January 1, 2013, investigators initially obtained a list of consecutive patients with these ICD-9 codes from the safety-net and tertiary care hospitals. Sufficient sample sizes were obtained from the safety net hospital after reviewing 4 months of consecutive patient charts (i.e., September 2012 to January 1, 2013) and after reviewing 5 months of consecutive charts at the tertiary care hospital (i.e., August 2012 to

Table 1
Characteristics of Study Sites in 2014–2015

<table>
<thead>
<tr>
<th></th>
<th>Urban Safety Net</th>
<th>Tertiary Care</th>
<th>Rural Community</th>
<th>Suburban Community</th>
<th>Urban Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual adult ED census</td>
<td>80,000</td>
<td>97,000</td>
<td>25,000</td>
<td>48,000</td>
<td>81,000</td>
</tr>
<tr>
<td>% ED patients admitted</td>
<td>15</td>
<td>13</td>
<td>19</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Beds in ED</td>
<td>72</td>
<td>76</td>
<td>20</td>
<td>30</td>
<td>77</td>
</tr>
<tr>
<td>% Patients seen by residents</td>
<td>70</td>
<td>55</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>0</td>
</tr>
<tr>
<td>% Patients seen by PA/NP</td>
<td>10</td>
<td>30</td>
<td>33</td>
<td>33</td>
<td>11</td>
</tr>
</tbody>
</table>

NP = nurse practitioner; PA = physician assistant.
January 1, 2013). The study was then expanded to the three community hospitals to increase generalizability. Investigators, similarly, obtained a list of consecutive patients from the three community hospitals starting on January 1, 2015. Sufficient sample sizes were obtained from each of the three community hospitals after reviewing 12 months of consecutive patient charts at each hospital from January 2014 to January 1, 2015. From the initial cohort, each chart was screened by a physician abstractor for inclusion using the following criteria: 1) a discharge diagnosis in the medical record of pneumonia, severe sepsis, or septic shock; 2) admission to the hospital from the ED; and 3) diagnosis or initiated treatment of the disease process in the ED. Pneumonia was present in the ED if definitively identified on imaging by a radiologist or treated in the ED based on documentation of clinical suspicion. Pneumonia was not considered to be present if azithromycin was given for the treatment of a chronic obstructive pulmonary disease (COPD) exacerbation alone. Severe sepsis or septic shock was present in the ED if the patient met all criteria for severe sepsis or septic shock as defined by the SSC while in the ED. Pneumonia was present in the ED if definitively identified on imaging by a radiologist or treated in the ED based on documentation of clinical suspicion. Pneumonia was not considered to be present if azithromycin was given for the treatment of a chronic obstructive pulmonary disease (COPD) exacerbation alone. Severe sepsis or septic shock was present in the ED if the patient met all criteria for severe sepsis or septic shock as defined by the SSC while in the ED. Exclusion criteria were age < 18 years, repeat visits by the same patients, and patients transferred from another facility as the initial management would not have occurred in the included EDs. Additionally, patients were not included in both the pneumonia and the sepsis cohorts. If the patient met criteria for the sepsis cohort due to pneumonia, the patient was included in the sepsis cohort rather than the pneumonia cohort.

Study Protocol

Once the study cohort was obtained, structured medical record abstraction was performed using established, standard methodology. To maximize validity and reliability of the medical record abstraction process, we used the following established methodologies: 1) physician abstractors, blinded to the purpose of the study, to ensure expert familiarity with medical records and documentation; 2) abstractors trained by the lead author using a set of test cases to standardize approaches; 3) use of a previously developed and refined closed-response data collection instrument (Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.111/ace m.13639/full); 4) performance of 10 pilot reviews, using actual cases sampled from each hospital but not included for analysis to gain familiarity with each hospital’s medical record system; 5) reabstraction of 15% of randomly selected included cases to estimate interrater reliability of the primary outcome, with the intention of performing reabstraction with adjudication of 100% of the cases if agreement of the 15% is less than \( \kappa < 0.8 \); and 6) routine oversight of the abstractor team by the lead author, who was also available throughout the data collection process to address questions and problems that occurred. Using a structured data abstraction form, abstractors documented the presence of all prespecified variables necessary to assess adherence with each CPG. Using the same data abstraction form, data were collected related to patient, physician, and environmental characteristics that had been shown to be associated with CPG adherence in previous studies on other emergency conditions.

Patient factors included patient demographics, primary health insurance, primary language, infectious disease–related comorbidities, and chief complaint. Patient demographics, insurance, and language were obtained directly from each hospital’s administrative database. Missing data were abstracted directly from the patient’s medical record when available and when unavailable were recorded as missing. All remaining characteristics were obtained directly from the medical record. Infectious disease–related comorbidities included diabetes, acquired immune deficiency syndrome, and iatrogenic immunosuppression (e.g., chemotherapy or other immunosuppressive medication). Patient chief complaints were stratified into three groups based on how typical the complaint was for the diagnosis. Stratification of chief complaints into three groups was defined by the lead and senior author based on frequency and specificity of the chief complaint for the diagnosis. Typical chief complaints for pneumonia included cough, shortness of breath, and fever. The only typical chief complaint for sepsis was fever. Associated chief complaints for pneumonia included chest pain, abdominal pain, flu, upper respiratory infection, congestion, hemoptysis, chills, myalgias, altered mentation, hypoxia, hypotension, tachycardia, and weakness. Associated chief complaints for sepsis included cough, dysuria, abdominal pain, flank pain, back pain, cellulitis, abscess, wound infection, blood infection, vomiting, diarrhea, altered mentation, chills, myalgias, shortness of breath, hypotension, and tachycardia. All other chief complaints were grouped into an “other” category.

Physician factors included the individual ED physician, ED physician’s experience, type of medical
degree, and ED diagnosis as well as the admitting hospital unit (i.e., floor vs. intensive care). Patients who were admitted under observation status or admitted to intermediate care units were considered floor admissions. ED physician’s experience was determined as the number of years of independent practice at the time the patient was seen (i.e., years following completion of residency training). ED physician’s medical degree was categorized into MD or DO. Physician’s ED diagnosis was categorized into three groups based on its association with pneumonia or sepsis. If the physician documented pneumonia, sepsis, severe sepsis, or septic shock as the primary ED diagnosis, then the ED diagnosis was designated as “primary.” For patients with pneumonia, if the physician documented COPD, hypoxia, pleural effusion, respiratory failure, or sepsis as the primary diagnosis, then the ED diagnosis of pneumonia was designated as “associated” with the primary diagnosis. Similarly, for patients with severe sepsis, if the physician documented a specific type of infection (e.g., pneumonia, cellulitis, and pyelonephritis) as the primary diagnosis, then the ED diagnosis of sepsis was designated as “associated.” All other primary ED diagnoses were categorized as “other.”

Environmental factors included time of day, day of week, ED occupancy, and hospital. Time of day was categorized into four groups: day (6 AM–11:59 AM), afternoon (12 PM–5:59 PM), evening (6 PM–11:59 PM), and night (12 AM–5:59 AM). Day of week was categorized into two groups: weekday (Monday 7 AM–Friday 4:59 PM) and weekend (Friday 5 PM–Monday 6:59 AM).

Outcome Measures
The primary outcome was ED adherence to the respective CPG for community-acquired pneumonia and severe sepsis/septic shock as written or endorsed by the IDSA/ATS and the SSC.4,27,32 Table 2 describes how adherence was determined for each CPG. Secondary outcomes included hospital LOS and all-cause in-hospital mortality. Hospital LOS was measured in days from time of hospital admission order to time of hospital discharge order.

Data Management and Statistical Analyses
All data management and statistical analysis were performed using SAS version 9.4. Descriptive statistics were calculated for all variables. Continuous data were reported as medians with interquartile ranges (IQRs) and categorical variables as percentages with 95% confidence intervals (CIs). Prevalence estimates with 95% CIs were used to report adherence with CPGs, and a chi-square test was used to test the a priori hypothesis that a statistically significant difference in adherence existed between the two CPGs. A p-value of <0.05 was considered statistically significant.

Unadjusted logistic regression was used to estimate the association of each patient, physician, and environmental variable with ED adherence to CPGs within the combined cohort and each disease subgroup. Hierarchical generalized linear models were used to estimate adjusted associations between patient, physician, and environmental factors and ED adherence with CPGs within the combined cohort. Adherence for all CPGs was initially modeled as a composite outcome to evaluate for factors associated with ED adherence to pneumonia and sepsis CPGs. Secondary models for each individual CPG were also developed, incorporating additional disease-specific patient factors. Models were developed by first creating a full model followed by dropping variables found to be collinear. Hospital was included as a random effect. Effect modification, using interaction terms, was assessed for sex, primary language, and race/ethnicity by complaint category and included if they contribute significantly to the model (p < 0.05).

Sample Size Estimation
In an effort to report estimates with reasonable precision, we chose a priori to include numbers of patients based on an upper 95% confidence limit of 5% (10%
This degree of precision allowed for appropriate statistical separation between estimates across institutions with relatively high and relatively low adherence and allowed for separation of all prevalence estimates from our a priori defined 95% adherence threshold. Thus, using estimates from our preliminary data, we estimated needing a minimum sample size of 350 total patients for each disease process (700 total patients) to achieve the above stated degree of precision. To provide a more balanced sample between the academic and community hospitals, we increased the sample size needed from the community hospitals as the data from the academic hospitals had already been collected prior to adding the community hospitals to the study.

## RESULTS

### Patient Characteristics

Overall, 827 patients were included in the study including 414 patients with pneumonia and 413 patients with severe sepsis or septic shock. Inter-rater reliability of abstraction of the primary outcome exceeded our predefined threshold ($\kappa > 0.8$). Table 3 describes the characteristics of the patients included in the study. The median age was 60 years (IQR = 49–74 years), and 53% were male. Patients were primarily non-Hispanic white (66%), spoke English primarily (91%), and were insured by Medicare (46%). While 60% of pneumonia patients presented with complaints typical of pneumonia, only 14% of sepsis patients presented with complaints typical of sepsis.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Combined ($n = 827$)</th>
<th>Pneumonia ($n = 414$)</th>
<th>Sepsis ($n = 413$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td>60 (49–74)</td>
<td>60 (48–76)</td>
<td>61 (49–73)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>435 (53)</td>
<td>211 (51)</td>
<td>224 (54)</td>
</tr>
<tr>
<td>Female</td>
<td>392 (47)</td>
<td>203 (49)</td>
<td>189 (46)</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>543 (66)</td>
<td>281 (68)</td>
<td>262 (63)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>161 (20)</td>
<td>65 (16)</td>
<td>96 (23)</td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>82 (10)</td>
<td>46 (11)</td>
<td>36 (9)</td>
</tr>
<tr>
<td>Other</td>
<td>41 (5)</td>
<td>22 (5)</td>
<td>19 (5)</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>748 (91)</td>
<td>382 (92)</td>
<td>366 (89)</td>
</tr>
<tr>
<td>Spanish</td>
<td>64 (8)</td>
<td>23 (6)</td>
<td>41 (10)</td>
</tr>
<tr>
<td>Other</td>
<td>15 (2)</td>
<td>9 (2)</td>
<td>6 (2)</td>
</tr>
<tr>
<td><strong>Primary health insurance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicare</td>
<td>383 (46)</td>
<td>197 (48)</td>
<td>186 (45)</td>
</tr>
<tr>
<td>Medicaid</td>
<td>160 (19)</td>
<td>71 (17)</td>
<td>89 (22)</td>
</tr>
<tr>
<td>Uninsured</td>
<td>158 (19)</td>
<td>96 (23)</td>
<td>62 (15)</td>
</tr>
<tr>
<td>Private</td>
<td>110 (13)</td>
<td>42 (10)</td>
<td>68 (17)</td>
</tr>
<tr>
<td>Other source</td>
<td>16 (2)</td>
<td>8 (2)</td>
<td>8 (2)</td>
</tr>
<tr>
<td><strong>Comorbidities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>20 (2)</td>
<td>16 (4)</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>101 (12)</td>
<td>47 (11)</td>
<td>54 (13)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>198 (24)</td>
<td>85 (21)</td>
<td>113 (27)</td>
</tr>
<tr>
<td>End-stage liver disease</td>
<td>20 (2)</td>
<td>3 (1)</td>
<td>17 (4)</td>
</tr>
<tr>
<td>End-stage renal disease</td>
<td>20 (2)</td>
<td>1 (0)</td>
<td>19 (5)</td>
</tr>
<tr>
<td>Immunosuppression</td>
<td>69 (8)</td>
<td>34 (8)</td>
<td>35 (9)</td>
</tr>
<tr>
<td><strong>Chief complaint</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical for disease</td>
<td>305 (37)</td>
<td>249 (60)</td>
<td>56 (14)</td>
</tr>
<tr>
<td>Associated with disease</td>
<td>419 (51)</td>
<td>125 (30)</td>
<td>294 (71)</td>
</tr>
<tr>
<td>Other</td>
<td>103 (13)</td>
<td>40 (10)</td>
<td>63 (15)</td>
</tr>
</tbody>
</table>

Data are reported as median (IQR) or $n$ (%). IQR = interquartile range.
Prevalence of Adherence

Overall, the prevalence of adherence to ED infectious CPGs was 57% (95% CI = 54%–61%) (Table 4). Physicians were more adherent to prescribing IDSA-concordant antibiotics to patients with pneumonia (64%, 95% CI = 59%–69%) than completing the SSC’s 3-hour bundle (50%, 95% CI = 45%–55%; \(p < 0.001\)). Overall adherence to the SSC’s 3-hour bundle was no different between patients with severe sepsis and septic shock (\(p = 0.9\); Figure 1). However, while the composite adherence to the SSC’s 3-hour bundle was only 50%, completion of individual components of the bundle were markedly better with 92% obtaining a screening lactate, 82% obtaining blood cultures before antibiotics, 72% receiving antibiotics within 3 hours of ED arrival, and 69% of septic shock patients receiving 30 mL/kg IV fluids within 3 hours of ED arrival.

Patient, Physician, and Environmental Variables Associated With Adherence

Table 5 shows the results of our adjusted multivariable analysis for the combined cohort. Patients were more likely to receive adherent care in the ED if they presented with chief complaints that were typical for the diagnoses and if the primary diagnosis in the ED was specific to the CPG. When patients presented with symptoms that were associated but not typical for cultures before antibiotics, 72% receiving antibiotics within 3 hours of ED arrival, and 69% of septic shock patients receiving 30 mL/kg IV fluids within 3 hours of ED arrival.

Table 4

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Pneumonia*</th>
<th>Sepsis and Septic Shock†</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adherence to CPG</td>
<td>57.2 (53.7–60.6)</td>
<td>64.0 (59.2–68.6)</td>
<td>50.4 (45.4–55.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hospital type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tertiary care</td>
<td>65.4 (58.9–71.5)</td>
<td>68.4 (59.1–76.7)</td>
<td>62.4 (53.0–71.2)</td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>52.9 (47.6–58.2)</td>
<td>56.7 (49.1–64.0)</td>
<td>49.2 (41.6–56.7)</td>
<td></td>
</tr>
<tr>
<td>Safety net</td>
<td>55.6 (48.9–62.0)</td>
<td>70.9 (61.8–79.0)</td>
<td>40.2 (31.2–49.6)</td>
<td></td>
</tr>
</tbody>
</table>

Data are reported as % (95% CI).

CPGs = clinical practice guidelines.

*Infectious Disease Society of America guideline-concordant antibiotics.

†Surviving Sepsis Campaign 3-hour bundle.
the disease, the odds of receiving adherent care in the ED were 0.6 (95% CI = 0.4–0.8). When the primary ED diagnosis was associated but not specific to the CPG, the odds of receiving adherent care were 0.3 (95% CI = 0.2–0.5) and 0.4 (95% CI = 0.2–0.6) for other primary diagnoses. Finally, patients who presented to a community hospital, were less likely to receive adherent care than patients who presented to an academic tertiary care hospital (adjusted odds ratio [AOR] = 0.6, 95% CI = 0.4–0.9).

Table 6 shows the results of our adjusted multivariable analysis for the sepsis cohort. Patients were
significantly more likely to receive all components of the SSC’s 3-hour bundle in the ED if they presented at night (AOR = 2.5, 95% CI = 1.2–5.3) compared to morning hours. Patients were significantly less likely to receive all components of the SSC’s 3-hour bundle in the ED if their infectious source was abdominal (AOR = 0.3, 95% CI = 0.1–0.8) or soft tissue (AOR= 0.4, 95% CI = 0.2–0.9) compared to a respiratory source; if they had fluid-sensitive comorbidities such as end-stage renal disease on hemodialysis, congestive heart failure, or end-stage liver disease (AOR = 0.6, 95% CI = 0.3–0.99); were diagnosed by the ED physicians with a specific infection rather than severe sepsis or septic shock (AOR = 0.5, 95% CI = 0.3–

<table>
<thead>
<tr>
<th>Table 6 Multivariable Model of ED Adherence to SSC’s 3-Hour Bundle, Adjusted for Clustering by Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sepsis Cohort (n = 413)</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Race/ethnicity</td>
</tr>
<tr>
<td>Non-Hispanic white</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>Non-Hispanic black</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Comorbidities</td>
</tr>
<tr>
<td>Fluid sensitive*</td>
</tr>
<tr>
<td>SIRS</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>ED source</td>
</tr>
<tr>
<td>Respiratory</td>
</tr>
<tr>
<td>Urinary</td>
</tr>
<tr>
<td>Skin/soft tissue</td>
</tr>
<tr>
<td>Abdominal</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>Primary ED diagnosis</td>
</tr>
<tr>
<td>Sepsis/severe/shock</td>
</tr>
<tr>
<td>Specific infection</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Admitting hospital unit</td>
</tr>
<tr>
<td>ICU</td>
</tr>
<tr>
<td>Floor</td>
</tr>
<tr>
<td>Hospital type</td>
</tr>
<tr>
<td>Tertiary care</td>
</tr>
<tr>
<td>Community</td>
</tr>
<tr>
<td>Safety net</td>
</tr>
<tr>
<td>Time of day</td>
</tr>
<tr>
<td>Morning (6 AM–11:59 AM)</td>
</tr>
<tr>
<td>Afternoon (noon–5:59 PM)</td>
</tr>
<tr>
<td>Evening (6 PM–11:59 PM)</td>
</tr>
<tr>
<td>Night (midnight–5:59 AM)</td>
</tr>
</tbody>
</table>

Data are reported as n (%) or OR (95% CI).

ICU = intensive care unit; IQR = interquartile range; SIRS = systemic inflammatory response syndrome; SSC = Surviving Sepsis Campaign.

*Congestive heart failure, end-stage renal disease on hemodialysis, and end-stage liver disease.
were admitted to a general medical or surgical floor (AOR = 0.5, 95% CI = 0.3–0.9) rather than an intensive care unit; or presented to a community or safety-net hospital (AOR = 0.4, 95% CI = 0.2–0.7, respectively) rather than a quaternary care hospital.

Unadjusted associations between ED adherence to infectious disease CPGs and all patient, provider, and environmental variables for the combined cohort and each disease subgroup are provided in Data Supplement S1 (Tables S1–S3). Adjusted multivariable analysis for the pneumonia cohort is also presented in Data Supplement S1 (Table S4).

### Secondary Outcomes

In the combined cohort, 40 (4.8%) patients died during the index hospitalization, 95% of whom were patients in the sepsis cohort. Adjusted for patient age, sex, admitting disease, admitting hospital unit, and ED CPG adherence, the odds of in-hospital mortality were significantly increased in patients who did not receive adherent CPG care in the ED (AOR = 2.4, 95% CI = 1.2–4.8; Table 7). The median hospital LOS for pneumonia patients receiving guideline adherent care in the ED was 1 day shorter than pneumonia patients receiving nonadherent care in the ED. In contrast, the median hospital LOS was 1 day longer for sepsis patient receiving guideline adherent ED care than sepsis patients receiving nonadherent care in the ED (Table 8).

### DISCUSSION

Our results suggest considerable variation in guideline adherence for two of the most prevalent and deadly infectious diseases encountered in the ED, with only 64% of patients with community-acquired pneumonia and 50% with sepsis receiving recommended therapy in the ED. To our knowledge, this is the only study to examine differences in adherence to contemporaneous guidelines for pneumonia and sepsis treatment in multiple, diverse ED settings in the United States. Similar to our findings among cardiovascular and cerebrovascular ED guideline adherence, chief complaint and primary ED diagnosis were significantly associated with ED adherence, such that the more straightforward the complaint and diagnosis, the more likely ED care was to be adherent to the relevant guideline.40 While the
random effect of hospital was small, accounting for only 1% of the variability in ED adherence, hospital type was significantly associated with adherence with community EDs less likely to adhere to infectious disease guidelines compared to an academic, tertiary care hospital. Finally, and perhaps most importantly, our results showed a significant association between guideline-adherent care in the ED and in-hospital mortality. After patient age, sex, admitting disease, and acuity of illness were adjusted for, patients who did not receive guideline-adherent care in the ED were 2.4 times more likely to die in the hospital compared to patients who did receive guideline-adherent care.

Since CMS retired its pneumonia core measure in 2014, little has been written on ED adherence to IDSA/ATS recommended antibiotic administration for patients admitted for community-acquired pneumonia. While the CMS pneumonia core measure was criticized for the lack of evidence related to blood cultures and timing of antibiotics, the recommendation for appropriate antibiotic therapy is supported by the literature, which suggests decreased mortality and hospital LOS when patients are administered guideline-recommended antibiotics. Additionally, it is important that patients receive appropriate therapy without being exposed to unnecessarily broad therapy, which can result in increased resistance and other adverse effects. This is particularly relevant now that the updated IDSA/ATS guidelines for hospital-acquired and ventilator-associated pneumonia recently removed the concepts of health care–associated pneumonia from the guideline given new evidence that contact with the health care system alone is less important than underlying patient characteristics for predicting risk of multidrug-resistant organisms.

Whether health care–associated pneumonia is differentiated from community acquired pneumonia by the IDSA/ATS guidelines for community acquired pneumonia remains unclear at this time. The IDSA and ATS are actively updating their guideline on community-acquired pneumonia with a projected release in the fall 2018. If the concept of health care-associated pneumonia is removed from the guideline, antimicrobial treatment of immunocompetent patients with pneumonia who present to the ED from the community will be greatly simplified, likely leading to improved adherence and antimicrobial stewardship.

For sepsis, most previous literature on adherence to the SSC’s resuscitation bundle did not differentiate adherence to components at 3 and 6 hours. While the 3-hour bundle is frequently initiated and completed in the ED, the 6-hour components are more likely completed in the inpatient setting, making it difficult to assess composite guideline adherence from larger studies. However, a handful of studies in the past few years have begun to report adherence to the SSC 3-hour bundle and the CMS SEP-1 guideline. The IMPReSS study by Rhodes et al. showed that adherence to the SSC 3-hour bundle was poor with overall adherence within their multicenter, international cohort being only 19% and rising to only 29% among the subset of North American participating hospitals. Our data are similar to Venkatesh et al., who recently showed that among U.S. EDs participating in the American College of Emergency Physician’s Emergency Quality Network Sepsis Initiative, overall adherence with the CMS SEP-1 was 54%. Both of these studies as well as ours suggest significant room for improvement in early ED sepsis care. Importantly, the SSC released a new “hour-1” bundle in May 2018, which replaced both the 3- and 6-hour bundles. The impact of the new SSC recommendation on ED sepsis care is likely to be limited in the United States given its current discordance with the CMS SEP-1 guideline. Moreover, the SSC hour-1 bundle has been highly criticized, particularly within the emergency medicine community, given the low-quality evidence to support these recommendations as well as the risk of over treating some sepsis patients and diverting attention away from nonsepsis patients.

**LIMITATIONS**

The use of discharge ICD-9 codes to identify ED patients is limited because discharge diagnoses may not be relevant to the reasons for admission from the ED. Consequently, using discharge ICD-9 codes was coupled with direct chart review to ensure that the sample only represented patients with the diagnoses of interest, who were admitted to the hospital from the ED specifically for these diagnoses. However, we may have missed some sepsis patients who were not coded as such. Additionally, including hospital admission as an inclusion criterion may have excluded patients who died in the ED, had an unknown disposition, or were discharged from the ED. Limiting patients to those who were admitted helped limited chart reviews to patient who were most likely to truly have the disease and in whom the guideline recommended care could actually have been enacted. Missing documentation within the medical chart is a known limitation to
medical record abstraction. Missing documentation could have affected our estimates of adherence especially in the pneumonia subgroup where details related to a patient’s immune competence could have been missing. Although we abstracted a comprehensive list of potential patient, physician, and environmental factors that have been shown to be associated with CPGs in other studies, additional variables may have been left out of the model, a known limitation of retrospective analyses. We used admitting hospital unit as a proxy for illness severity. The use of a more robust illness severity score may have resulted in a more specific variable for illness severity.

CONCLUSIONS

Adherence to ED infectious clinical practice guidelines for community-acquired pneumonia and sepsis varies significantly across diseases and institutions with significant room for improvement, especially in light of a significant association with in-hospital mortality.

The authors acknowledge Erica Ashley Morse, MD, and Michael Susalla, MD, who both contributed to data acquisition.

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.13639/full

Data Supplement S1. Supplemental material.
A Prospective, Multicenter Evaluation of Point-of-care Ultrasound for Small-bowel Obstruction in the Emergency Department

Brent A. Becker, MD1, Shadi Lahham, MD2, Mark A. Gonzales, DO1, Jason T. Nomura, MD3, Michelle K. Bui2, Taylor A. Truong2, Barbara A. Stahlman, MS1, John C. Fox, MD2, and Thompson Kehrl, MD1

ABSTRACT

Objective: The main objective of this study was to evaluate the accuracy of emergency physician-performed point-of-care ultrasound (POCUS) for the diagnosis of small-bowel obstruction (SBO) compared to computed tomography (CT).

Methods: We performed a prospective, multicenter, observational study examining a convenience sample of adult patients with potential SBO presenting to the emergency department (ED) between July 2014 and May 2017. Each POCUS was interpreted at the bedside by the performing emergency physician and retrospectively by an expert reviewer. Test characteristics were calculated for POCUS, blinded expert interpretation, and specific POCUS parameters.

Results: A total of 217 subjects were included in the primary analysis with an overall SBO prevalence of 42.9%. For the diagnosis of SBO, POCUS demonstrated an overall sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio of 0.88 (95% confidence interval [CI] = 0.80 to 0.94), 0.54 (95% CI = 0.45 to 0.63), 1.92 (95% CI = 1.56 to 2.35), and 0.22 (95% CI = 0.12 to 0.39), respectively. Expert review yielded a similar sensitivity (0.89 [95% CI = 0.81 to 0.95]) with a significantly higher specificity (0.82 [95% CI = 0.74 to 0.88]). The more sensitive sonographic parameters for both POC sonographers and expert reviewers were small-bowel dilation ≥ 25 mm (0.87 [95% CI = 0.79 to 0.93], 0.87 [95% CI = 0.79 to 0.93]) and abnormal peristalsis (0.82 [95% CI = 0.72 to 0.89], 0.85 [95% CI = 0.76 to 0.87]). The more specific parameters for both groups were transition point (0.82 [95% CI = 0.74 to 0.89], 0.98 [95% CI = 0.94 to 1.00]), intraperitoneal free fluid (0.82 [95% CI = 0.74 to 0.89], 0.93 [95% CI = 0.87 to 0.97]), and bowel wall edema (0.76 [95% CI = 0.67 to 0.83], 0.93 [95% CI = 0.87 to 0.97]).

Conclusion: POCUS is moderately sensitive for SBO, although less specific, when performed by a diverse group of emergency physicians across multiple EDs. Interpretation of acquired POCUS images is significantly more accurate when performed by physicians with prior emergency ultrasound fellowship training and familiarity with the sonographic appearance of SBO.
Background
Small-bowel obstruction (SBO) represents a clinical entity frequently encountered by emergency physicians. In the United States, published data suggest approximately 300,000 adults are hospitalized annually for SBO despite increasing use of laparoscopic surgery. Imaging plays a significant role in making the diagnosis of SBO, as history and physical examination are unreliable. Due to poor accuracy and frequently inconclusive results, traditional plain radiographs have been increasing eschewed in favor of computed tomography (CT). CT is significantly more sensitive and specific for SBO, particularly with 64-slice scanners; however, its use confers significant expense, potential delays, and ionizing radiation. Multiple studies have shown that ultrasound outperforms plain radiography for the detection of SBO. The use of point-of-care ultrasound (POCUS) for the evaluation of SBO has grown in recent years and ultrasound is increasingly being touted as a first-line imaging modality for SBO.

Importance
Ultrasound for SBO was first described in case series during the 1970s and has been shown to be both sensitive and specific in subsequent studies. The ability of POCUS to accurately diagnose SBO could potentially improve patient care by decreasing time to diagnosis and expediting consultation, as seen with other POCUS applications. A recent meta-analysis reflects a strong performance of ultrasound for SBO in the emergency department (ED) setting; however, only two trials examined POCUS performed by emergency physicians and analysis on this specific subgroup was not possible. Moreover, the available published research on POCUS for SBO largely consists of small, single-center studies with variable reference standards.

Goals of This Investigation
The main objective of this study was to conduct a multicenter evaluation of the accuracy of emergency physician–performed POCUS for the diagnosis of SBO compared to CT. We also aimed to compare POCUS interpretation to that of emergency ultrasound fellowship–trained experts and assess the role of specific sonographic parameters in confirming the diagnosis of SBO by POCUS.

METHODS
Study Design and Setting
We performed a prospective, multicenter, observational study examining the diagnostic accuracy of POCUS for SBO. A convenience sample of adult ED patients presenting between July 2014 and May 2017 with suspicion for SBO underwent goal-directed POCUS of the abdomen for the evaluation of SBO. POCUS findings were interpreted at bedside by a physician sonographer blinded to laboratory and imaging results, including CT, and retrospectively by an expert reviewer after deidentification of the images. Each POCUS interpretation was compared to abdominal CT as the reference standard.

Subjects were enrolled at three separate facilities, including two suburban, academic community hospitals and an urban, university-based tertiary referral center. The combined annual ED census of the three centers is approximately 250,000 visits. All three facilities support independent three-year (PGY-1 to -3) emergency medicine (EM) residency training programs and emergency ultrasound fellowships. The institutional review boards at each site provided approval, and written informed consent was obtained from all subjects. The study was preregistered with ClinicalTrials.gov (NCT0219081) and conducted in accordance with Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines.

Selection of Participants
Potential subjects were identified either by direct clinical contact by a member of the study team or referral by other ED providers not directly participating in the study. Patients were eligible for enrollment only when a physician participating in the study was available. The study team member obtained patient consent prior to performing POCUS for SBO. Patients were eligible for enrollment if they were at least 18 years of age, able to provide consent in English, not pregnant, had not yet undergone radiology imaging, and presented with symptoms concerning for possible SBO. The latter criterion was not explicitly defined, but was based on the clinical assessment by the treating physician. Patients were excluded if they did not receive CT imaging.
Data Collection and Measurements
Data were collected using universal standardized data collection forms common to all participating sites. Sonographic data and bedside interpretations were recorded at the time of POCUS. Clinical features were also recorded at the time of POCUS, including diarrhea within 24 hours, vomiting, duration of symptoms, timing of last bowel movement, and presence of diffuse abdominal pain or tenderness. Follow-up data were collected by review of the electronic medical record and included patient demographics, discharge diagnosis, operative reports, abdominal x-ray imaging, and CT results. CT scanners used included: Phillips Ingenuity 128 slice (hospital 1), Phillips Brilliance iCT 256 slice (hospital 1 and 2), and Siemens SOMATOM Definition AS 64 slice (hospital 3). Board-certified attending radiologists rendered all final CT interpretations.

Ultrasound Technique and Interpretation
Point-of-care ultrasound for SBO was performed in supine patients using a standardized protocol: a curvilinear probe between 1 and 5 MHz was placed on the patient’s anterior abdomen. Sonographers conducted a systematic evaluation of the entire abdomen, including dedicated views of the right upper quadrant, left upper quadrant, right lower quadrant, and left lower quadrant. Sonographers were instructed to adjust the probe orientation with respect to the body (transverse, sagittal, or coronal) to optimize longitudinal images of the small bowel. All POCUS examinations required at least one video clip of the small bowel in each quadrant. Maximum bowel diameter was measured and recorded as a still image. Any suspected transition point, defined as the region between the proximal segment of dilated small bowel and the distal segment of decompressed small bowel, was recorded via a video clip. Each patient was also specifically assessed for four other sonographic parameters previously associated with SBO: small-bowel dilation, abnormal peristalsis, small-bowel wall edema, and intraperitoneal free fluid (Table 1).

Intraperitoneal free fluid was considered present if anechoic, extraluminal collections were visualized between bowel loops. Sonographers utilized a checklist on the bedside data collection form to ensure all aforementioned views and measurements were obtained. Still images and video clips were saved, exported, and deidentified for independent interpretation by a blinded expert reviewer.

Each POCUS was performed by an attending ED physician, emergency ultrasound fellow, or upper-level EM resident (PGY-2 or PGY-3). The attending group included physicians both with and without emergency ultrasound fellowship training. All participating fellows and residents received a 30-minute lecture on the POCUS technique for SBO, followed by brief hands-on scanning practice on normal individuals without SBO. Residents had completed at least one emergency ultrasound rotation and performed at least 50 POCUS examinations; however, they had no prior exposure to or training in POCUS for SBO prior to joining the study. Standardized didactic materials were distributed among the three enrolling centers to ensure consistency of POCUS training, but there was no prerequisite number of POCUS for SBO before participants were cleared to enroll. Sonographers were blinded to the results of other diagnostic tests, including laboratory values and subsequent imaging results. The treating physicians and radiologists were similarly blinded to the results of the POCUS. POCUS findings were not used in clinical decision making, and each patient otherwise received normal ED standard of care.

The specific ultrasound devices used depended on the site and included Mindray TE7 (Mindray North America), Sonosite M-Turbo (FUJIFILM Sonosite), Sonosite X-porte (FUJIFILM Sonosite), Ultrasonix SonixTOUCH (BK Ultrasound), and Zonare ZS3 (Mindray North America). All images were acquired using a low-frequency, convex array transducer (1–5 MHz) and archived using site-specific software, namely, SonixHUB (Analogic Corporation), AGFA Healthcare

Table 1
Specific Ultrasound Parameters Assessed During POCUS for SBO

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Small bowel dilation ≥ 25 mm</td>
<td>Measures outer wall to outer wall maximum bowel diameter</td>
</tr>
<tr>
<td>2. Abnormal peristalsis</td>
<td>Defined as “to-and-fro,” shuttling or swirling movements of intraluminal bowel contents</td>
</tr>
<tr>
<td>3. Intraperitoneal free fluid</td>
<td>Visualized extraluminal anechoic collections</td>
</tr>
<tr>
<td>4. Small-bowel wall edema</td>
<td>Present if thickened bowel wall results in a “keyboard sign”</td>
</tr>
<tr>
<td>5. Transition point</td>
<td>Defined as the region between proximal dilated and distal decompressed bowel</td>
</tr>
</tbody>
</table>

POCUS = point-of-care ultrasound; SBO = small-bowel obstruction.
Enterprise Imaging (AGFA-Gevaert Group), and QPath (Telexy Healthcare).

The performing sonographer completed a closed-response data collection form at the time of the examination. Each POCUS was classified as positive, negative, or indeterminate for SBO, based on the presence of small bowel dilation \( \geq 25 \text{ mm} \) or abnormal peristalsis. The additional three ultrasound parameters (small-bowel wall edema, intraperitoneal free fluid, and transition point) served to augment the sonographers’ overall impression based on small-bowel diameter and peristalsis.

Blinded interpretations of the deidentified POCUS images were conducted in an analogous fashion, utilizing the same standardized closed-response form as the POC sonographer. Expert reviews were conducted by the primary author (BB) or the emergency ultrasound director at two of the sites (TK, SL). All reviewers had completed an emergency ultrasound fellowship and two were ARDMS certified. If these individuals were involved in the index POCUS or familiar with the clinical details of a given case, the blinded overread was delegated to another emergency ultrasound fellowship–trained member of the ED faculty.

**Outcomes**
The primary outcome of the study was POCUS-mediated diagnosis of SBO, confirmed by CT. Secondary outcomes included the diagnosis of SBO by blinded expert interpretation of POCUS images and diagnostic accuracy of each of the five specific sonographic parameters.

Any POCUS classified as indeterminate was considered “positive” for SBO. This approach was chosen to align with typical ED practice, namely, the tendency to pursue equivocal results with further workup or consultation. Similarly, noncommittal CT interpretations (i.e., “ileus versus SBO”) were treated as SBO. Unequivocal CT diagnosis of ileus was considered negative for SBO. Obstructive processes seen on CT that did not specifically involve the small bowel were also classified as negative for SBO, including large-bowel obstruction, volvulus, and pseudo-obstruction.

**Data Analysis**
Analysis of patient demographics was descriptive with continuous and categorical variables reported as medians with interquartile ranges (IQRs) and percentages, respectively. Standard 2 × 2 tables were used to calculate sensitivity, specificity, and positive/negative likelihood ratios (LR+/LR−) with 95% confidence intervals (CIs) for POCUS and blinded expert interpretation. Subgroup analysis included analogous calculations based on sonographer level of training. Inter-rater reliability between POCUS and blinded expert review was assessed using Cohen’s kappa coefficient (κ).

Given the unilateral treatment of indeterminate imaging results in the primary analysis, sensitivity analysis was performed. Primary analysis was repeated with 1) reclassification of all indeterminate POCUS as negative for SBO and 2) removal of patients with indeterminate POCUS or CT interpretations. Additionally, we assessed the reliability of CT as the reference standard by recalculating POCUS test characteristics based on final discharge diagnosis. This incorporated additional clinical information beyond the initial CT, including subsequent diagnostics, operative findings, and hospital course.

Previously published literature examining ultrasound for SBO predicted an average sensitivity and specificity of approximately 90 and 91%, respectively, with an expected SBO incidence of 40%. With these assumptions, it was calculated a priori that 96 patients would be required to yield precision of 0.10 and confidence level of 95%. We initially planned a single-site enrollment of 106 patients anticipating an approximate 10% exclusion rate; however, prior prospective studies had enrolled similar numbers and our study subsequently expanded to include additional sites. Thus, a total enrollment of 212 patients was planned across all participating centers.

Data were compiled using Remark Office OMR 7 (Gravic) and Microsoft Excel 2010 (Microsoft). Data analysis was performed using IBM SPSS Statistics for Windows version 24.0 (IBM Corp, Armonk, NY) and MedCalc statistical software 18.10.2 (MedCalc Software bvba; https://www.medcalc.org; 2018).

**RESULTS**

**Characteristics of Study Subjects**
A total of 232 subjects were initially enrolled across the three hospitals. Fifteen (6.5%) patients did not receive CT imaging and were excluded, leaving 217 subjects for the primary analysis consisting of 111 (51.2%), 103 (47.5%), and three (1.4%) patients at the three study sites, respectively. Figure 1 depicts the overall study flow chart.
The median (IQR) age of the included cohort was 55 (45–67) years and 114 (52.5%) patients were female. The overall prevalence of SBO was 42.9%. Patient characteristics were generally similar between study sites and are presented in Table 2.

Point-of-care ultrasound was performed by an attending, fellow, or resident physician in 77 (35.5%), 72 (33.2%), and 68 (31.3%) subjects, respectively. A total of 41 unique physicians performed a median (IQR) of 4 (2.5–15.5) POCUS examinations each with individual counts ranging from 1 to 35.

**Main Results**

The performance of POCUS for the diagnosis of CT-confirmed SBO is displayed in Table 3, illustrating the 11 false-negative and 57 false-positive POCUS examinations. The corresponding CT findings and discharge diagnoses for these divergent cases are reported in Data Supplement S1, Tables S1 and S2 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13713/full).

Point-of-care ultrasound performed by trainees (fellows/residents) demonstrated a sensitivity of 0.91 (95% CI = 0.80 to 0.97) and specificity of 0.51 (95% CI = 0.40 to 0.62), compared to a sensitivity of 0.85 (95% CI = 0.70 to 0.94), and specificity of 0.61 (95% CI = 0.43 to 0.76) for attending-performed POCUS. Subgroup analysis of trainee-performed POCUS by level of training is reported in Table 4.

**Secondary Results**

Blinded expert interpretation of deidentified POCUS images was performed for 216 subjects, due to missing images for a single patient (Table 3). The sensitivity of expert review for SBO was similar to that of POC sonographers (0.89 [95% CI = 0.81 to 0.95]), while specificity was significantly greater (0.82 [95% CI = 0.74 to 0.88]; Table 4). Accuracy of expert interpretation was significantly higher than that of POC sonographers (0.85 [95% CI = 0.80 to 0.90] vs. 0.69 [95% CI = 0.62 to 0.75]). Comparison of original POC and expert interpretation of POCUS for SBO (positive, indeterminate, or negative) yielded a percentage agreement of 68.1% with associated $\kappa = 0.468$ (95% CI = 0.378 to 0.558), $p < 0.001$. After reclassification of all indeterminate cases as positive, percentage agreement increased to 78.2% with associated $\kappa = 0.568$ (95% CI = 0.464 to 0.672), $p < 0.001$.  

---

**Figure 1.** Study flow diagram. IND = indeterminate; OOF = other obstructive finding; POCUS = point-of-care ultrasound; SBO = small-bowel obstruction.
The relative incidence, sensitivities, specificities, and LRs of the specific POCUS parameters for the diagnosis of CT-confirmed SBO are reported in Table 5. The more sensitive sonographic parameters for both POC sonographers and expert reviewers were small-bowel dilation ≥ 25 mm (0.87 [95% CI = 0.79 to 0.93], 0.87 [95% CI = 0.79 to 0.93]) and abnormal peristalsis (0.82 [95% CI = 0.72 to 0.89], 0.85 [95% CI = 0.76 to 0.87]). The more specific parameters for both groups were transition point (0.82 [95% CI = 0.74 to 0.89], 0.98 [95% CI = 0.94 to 1.00]), intraperitoneal free fluid (0.82 [95% CI = 0.74 to 0.89], 0.93 [95% CI = 0.87 to 0.97]), and bowel wall edema (0.76 [95% CI = 0.67 to 0.83], 0.93 [95% CI = 0.87 to 0.97]).

In the sonographer group, there was significant association between all POCUS parameters (p ≤ 0.025). There were similar associations observed in the expert reviewer group (p < 0.001) with the exception of transition point, which was not demonstrably associated with bowel wall edema (p = 0.822) and intraperitoneal free fluid (p = 0.330).

**Sensitivity Analysis**
Reclassification of indeterminate POCUS interpretations as “negative” for SBO decreased sensitivity to 0.77

---

**Table 2**
Patient Demographics and Clinical Characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All Sites</th>
<th>Hospital 1</th>
<th>Hospital 2</th>
<th>Hospital 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number (%) of patients</td>
<td>217 (100.0)</td>
<td>111 (51.2)</td>
<td>103 (47.5)</td>
<td>3 (1.4)</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>55 (45–67)</td>
<td>61 (48–72)</td>
<td>52 (41–62)</td>
<td>73 (-)</td>
</tr>
<tr>
<td>Female</td>
<td>114 (52.5)</td>
<td>52 (46.8)</td>
<td>60 (58.3)</td>
<td>2 (66.7)</td>
</tr>
<tr>
<td>BMI</td>
<td>25 (22–30)</td>
<td>27 (22–32)</td>
<td>24 (21–28)</td>
<td>33 (-)</td>
</tr>
<tr>
<td><strong>Clinical characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diarrhea ≤ 24 hours</td>
<td>51 (23.5)</td>
<td>25 (22.5)</td>
<td>25 (24.3)</td>
<td>1 (33.3)</td>
</tr>
<tr>
<td>Diffuse abd pain</td>
<td>108 (49.8)</td>
<td>55 (49.5)</td>
<td>51 (49.5)</td>
<td>2 (66.7)</td>
</tr>
<tr>
<td>Diffuse abd tenderness</td>
<td>92 (42.4)</td>
<td>45 (40.5)</td>
<td>45 (43.7)</td>
<td>2 (66.7)</td>
</tr>
<tr>
<td>Last BM (days)</td>
<td>1 (0-3)</td>
<td>1 (0-2)</td>
<td>1 (0-3)</td>
<td>0 (-)</td>
</tr>
<tr>
<td>Sx duration (days)</td>
<td>2 (1-4)</td>
<td>2 (1-4)</td>
<td>2 (1-4)</td>
<td>2 (-)</td>
</tr>
<tr>
<td>Vomiting</td>
<td>129 (59.4)</td>
<td>65 (58.6)</td>
<td>61 (59.2)</td>
<td>3 (100.0)</td>
</tr>
<tr>
<td><strong>Sonographer level of training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attending</td>
<td>77 (35.5)</td>
<td>51 (45.9)</td>
<td>23 (22.3)</td>
<td>3 (100.0)</td>
</tr>
<tr>
<td>Fellow</td>
<td>72 (33.2)</td>
<td>29 (26.1)</td>
<td>43 (41.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Resident (PGY-2/PGY-3)</td>
<td>68 (31.3)</td>
<td>31 (27.9)</td>
<td>37 (35.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td><strong>Imaging</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXR performed</td>
<td>64 (29.5)</td>
<td>48 (43.2)</td>
<td>15 (14.6)</td>
<td>1 (33.3)</td>
</tr>
<tr>
<td>CT contrast type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>34 (15.7)</td>
<td>25 (22.5)</td>
<td>8 (7.8)</td>
<td>1 (33.3)</td>
</tr>
<tr>
<td>IV</td>
<td>150 (69.1)</td>
<td>81 (73.0)</td>
<td>67 (65.0)</td>
<td>2 (66.7)</td>
</tr>
<tr>
<td>Oral</td>
<td>1 (0.5)</td>
<td>0 (0.0)</td>
<td>1 (1.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>IV + oral</td>
<td>32 (14.7)</td>
<td>5 (4.5)</td>
<td>27 (26.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td><strong>POCUS impression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indeterminate</td>
<td>33 (15.2)</td>
<td>16 (14.4)</td>
<td>16 (15.5)</td>
<td>1 (33.3)</td>
</tr>
<tr>
<td>No SBO</td>
<td>78 (35.9)</td>
<td>37 (33.3)</td>
<td>41 (39.8)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>SBO</td>
<td>106 (48.8)</td>
<td>58 (52.3)</td>
<td>46 (44.7)</td>
<td>2 (66.7)</td>
</tr>
<tr>
<td><strong>Discharge diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ileus</td>
<td>7 (3.2)</td>
<td>6 (5.4)</td>
<td>1 (0.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Other obstructive process</td>
<td>8 (3.7)</td>
<td>6 (5.4)</td>
<td>2 (1.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>No SBO</td>
<td>115 (53.0)</td>
<td>57 (51.4)</td>
<td>58 (56.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>SBO</td>
<td>87 (40.1)</td>
<td>42 (37.8)</td>
<td>42 (40.8)</td>
<td>3 (100.0)</td>
</tr>
</tbody>
</table>

Data are reported as n (%) or median (IQR).
Abd = abdominal; AXR = abdominal x-ray; BM = bowel movement; BMI = body mass index; CT = computed tomography; IQR = interquartile range; IV = intravenous; POCUS = point-of-care ultrasound; SBO = small-bowel obstruction; Sx = symptoms.
and specificity of expert reviewers were 0.95 (95% CI = 0.87 to 0.99) and 0.82 (95% CI = 0.74 to 0.88), respectively.

Exclusion of all cases with either an indeterminate POCUS or CT interpretation (n = 43) yielded a slightly lower sensitivity of 0.85 (95% CI = 0.75 to 0.92) and improved specificity to 0.66 (95% CI = 0.56 to 0.75) for POC sonographers. Expert review showed slightly improved sensitivity and specificity of 0.94 (95% CI = 0.86 to 0.98) and 0.86 (95% CI = 0.79 to 0.92).

Recalculation of POCUS test characteristics based on discharge diagnosis in lieu of CT interpretation resulted in findings similar to the primary analysis. For POC sonographers, sensitivity was 0.91 (95% CI = 0.82 to 0.96) and specificity was 0.53 (95% CI = 0.44 to 0.62), while for expert reviewers, sensitivity was 0.92 (95% CI = 0.83 to 0.97) and specificity was 0.79 (95% CI = 0.71 to 0.85). There were six false-negative and 13 false-positive CT interpretations compared to discharge diagnosis, representing a sensitivity of 0.93 (95% CI = 0.85 to 0.97) and specificity of 0.90 (95% CI = 0.84 to 0.95) for SBO. Percentage agreement between CT and discharge diagnosis was 91.2% with associated \( \kappa = 0.820 \) (95% CI = 0.744 to 0.896), \( p < 0.001 \).

### DISCUSSION

Overall, we found emergency physician–performed POCUS to be relatively sensitive (0.88) for SBO, but considerably less specific (0.54), with attending and trainee physicians performing similarly. Blinded over-read of POCUS images by fellowship-trained faculty

---

**Table 3**

<table>
<thead>
<tr>
<th></th>
<th>SBO</th>
<th>No SBO</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CT Interpretation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POCUS interpretation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBO</td>
<td>82</td>
<td>57</td>
<td>139</td>
</tr>
<tr>
<td>No SBO</td>
<td>11</td>
<td>67</td>
<td>78</td>
</tr>
<tr>
<td>Totals</td>
<td>93</td>
<td>124</td>
<td>217</td>
</tr>
<tr>
<td>Expert reviewer interpretation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBO</td>
<td>83</td>
<td>22</td>
<td>105</td>
</tr>
<tr>
<td>No SBO</td>
<td>10</td>
<td>101</td>
<td>111</td>
</tr>
<tr>
<td>Totals</td>
<td>93</td>
<td>123</td>
<td>216</td>
</tr>
</tbody>
</table>

CT = computed tomography; POCUS = point-of-care ultrasound; SBO = small-bowel obstruction.

---

**Table 4**

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POCUS interpretation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>217 (100.0)</td>
<td>0.88 (0.80–0.94)</td>
<td>0.54 (0.45–0.63)</td>
<td>1.92 (1.56–2.35)</td>
<td>0.22 (0.12–0.39)</td>
</tr>
<tr>
<td>Attending</td>
<td>77 (35.5)</td>
<td>0.85 (0.70–0.94)</td>
<td>0.61 (0.43–0.76)</td>
<td>2.14 (1.41–3.25)</td>
</tr>
<tr>
<td>All trainees</td>
<td>140 (64.5)</td>
<td>0.91 (0.80–0.97)</td>
<td>0.51 (0.40–0.62)</td>
<td>1.86 (1.47–2.34)</td>
</tr>
<tr>
<td>Fellow (PGY-4)</td>
<td>72 (33.2)</td>
<td>0.93 (0.77–0.99)</td>
<td>0.40 (0.25–0.56)</td>
<td>1.54 (1.19–2.00)</td>
</tr>
<tr>
<td>PGY-3</td>
<td>41 (18.9)</td>
<td>0.92 (0.64–1.00)</td>
<td>0.57 (0.37–0.76)</td>
<td>2.15 (1.37–3.40)</td>
</tr>
<tr>
<td>PGY-2</td>
<td>27 (12.4)</td>
<td>0.83 (0.52–0.98)</td>
<td>0.73 (0.45–0.92)</td>
<td>3.12 (1.30–7.51)</td>
</tr>
<tr>
<td>Expert reviewer interpretation</td>
<td>216 (100.0)</td>
<td>0.89 (0.81–0.95)</td>
<td>0.82 (0.74–0.88)</td>
<td>4.99 (3.39–7.33)</td>
</tr>
</tbody>
</table>

CI = confidence interval; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PGY = postgraduate year; POCUS = point-of-care ultrasound; SBO = small-bowel obstruction.
was notably more accurate than POCUS interpretation for the diagnosis of SBO, demonstrating both greater sensitivity (89%) and greater specificity (82%). Small-bowel dilation ≥ 25 mm (0.87) and abnormal peristalsis (0.82 to 0.85) proved more sensitive for SBO, while transition point (0.82 to 0.98), intraperitoneal free fluid (0.82 to 0.93), and bowel wall edema (0.76 to 0.93) were more specific, however, almost all POCUS parameters were significantly correlated.

Multiple studies have evaluated the diagnostic accuracy of ultrasound for SBO, but it is most appropriate to compare our findings to the two prospective trials examining emergency physician–performed POCUS for SBO published by Jang et al.⁶ and Unluer et al.¹³ These studies reported greater sensitivity (0.91, 0.98) and specificity (0.84, 0.93), respectively, with similar SBO prevalence (43%, 52%) and reference standards (CT and a combination of CT, operative reports, and phone follow-up).⁶,¹³ These divergent results could be due to differences in the POCUS training of study participants: Jang et al. required each sonographer to amass five SBO-positive POCUS examinations prior to enrolling and Unluer et al. provided 6 hours of training. Physician sonographers in our study received a significantly shorter didactic session and did not have to demonstrate competency prior to enrolling patients. Moreover, the group of sonographers in our study likely represents a broader cross-section of emergency physicians with significantly less experience with POCUS for SBO, including resident physicians.

The overall impression of expert reviewers proved the most accurate method of diagnosing SBO and the heightened performance we observed with fellowship-trained experts aligns more with that seen in previously published literature. Sonographic diagnosis of SBO likely hinges more on recognizing a characteristic appearance or constellation of patterns and our findings suggest that successful application of POCUS for SBO is dependent on prior experience and exposure to normal/abnormal cases. This highlights the widely recognized operator-dependence of POCUS and reiterates the need for adequate education, via both didactics and hands-on experience. We propose that proficiency in POCUS for SBO likely necessitates dedicated training, as is the case for other more established POCUS applications, such as focused assessment with sonography in trauma and POC echocardiography. Future research should investigate methods of educating learners and assessing competency for this application, as well as the economic and operational effects of POCUS as a first-line test for SBO.

### Table 5
Test Characteristics of Specific Ultrasound Parameters for SBO

<table>
<thead>
<tr>
<th>Parameter</th>
<th>n (%)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR− (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POCUS interpretation (n = 217)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowel ≥ 25 mm</td>
<td>141 (65.0)</td>
<td>0.87 (0.79–0.93)</td>
<td>0.60 (0.51–0.69)</td>
<td>2.20 (1.75–2.78)</td>
<td>0.21 (0.12–0.37)</td>
</tr>
<tr>
<td>Abnormal peristalsis</td>
<td>147 (67.7)</td>
<td>0.82 (0.72–0.89)</td>
<td>0.51 (0.42–0.60)</td>
<td>1.66 (1.36–2.04)</td>
<td>0.36 (0.23–0.57)</td>
</tr>
<tr>
<td>Bowel edema</td>
<td>75 (34.6)</td>
<td>0.43 (0.33–0.54)</td>
<td>0.76 (0.67–0.83)</td>
<td>1.78 (1.20–2.62)</td>
<td>0.75 (0.61–0.92)</td>
</tr>
<tr>
<td>Free fluid</td>
<td>57 (26.3)</td>
<td>0.34 (0.25–0.45)</td>
<td>0.82 (0.74–0.89)</td>
<td>1.94 (1.21–3.11)</td>
<td>0.8 (0.67–0.94)</td>
</tr>
<tr>
<td>Transition point</td>
<td>48 (22.1)</td>
<td>0.25 (0.16–0.35)</td>
<td>0.82 (0.74–0.89)</td>
<td>1.39 (0.83–2.34)</td>
<td>0.92 (0.79–1.05)</td>
</tr>
<tr>
<td><strong>Expert reviewer interpretation (n = 216)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowel ≥ 25 mm</td>
<td>104 (48.1)</td>
<td>0.87 (0.79–0.93)</td>
<td>0.81 (0.73–0.88)</td>
<td>4.66 (3.02–6.79)</td>
<td>0.16 (0.09–0.27)</td>
</tr>
<tr>
<td>Abnormal peristalsis</td>
<td>103 (47.7)</td>
<td>0.85 (0.76–0.87)</td>
<td>0.80 (0.72–0.87)</td>
<td>4.35 (3.01–6.30)</td>
<td>0.19 (0.11–0.31)</td>
</tr>
<tr>
<td>Bowel edema</td>
<td>34 (15.7)</td>
<td>0.27 (0.18–0.37)</td>
<td>0.93 (0.87–0.97)</td>
<td>3.67 (1.80–7.49)</td>
<td>0.79 (0.69–0.90)</td>
</tr>
<tr>
<td>Free fluid</td>
<td>23 (10.6)</td>
<td>0.15 (0.08–0.24)</td>
<td>0.93 (0.87–0.97)</td>
<td>2.06 (0.93–4.55)</td>
<td>0.92 (0.83–1.01)</td>
</tr>
<tr>
<td>Transition point</td>
<td>17 (7.8)</td>
<td>0.16 (0.09–0.25)</td>
<td>0.98 (0.94–1.00)</td>
<td>9.92 (2.33–42.3)</td>
<td>0.85 (0.78–0.93)</td>
</tr>
</tbody>
</table>

CI = confidence interval; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; POCUS = point-of-care ultrasound; SBO = small-bowel obstruction.
LIMITATIONS

There are several limitations to the study that warrant further address. The observational design and convenience sampling introduce an inherent potential for bias. Enrollment at the third site was relatively low due to changes in faculty. Participating physicians possessed variable ultrasound experience, received relatively limited training in POCUS for SBO, and did not receive additional retraining over the span of study enrollment. The study was conducted at centers that previously were not routinely using POCUS for the evaluation of SBO; thus the results may not generalize to more experienced providers. Multiple different ultrasound machines were used during the course of the study and the effect of any given machine on the accuracy of POCUS was not assessed. While efforts were made to standardize the ultrasound technique, there was not an explicit, “stepwise” algorithm that was universally employed.

Sonographers in our study had the unique option of interpreting the POCUS as “indeterminate” for SBO, and this may have resulted in the lower test specificity observed. A prospective study of radiologist-performed ultrasound for SBO by Schmutz et al.15 included a subset of “gassy” patients, for which the ultrasound was essentially deemed indeterminate and were excluded from analysis. The study reported a specificity of 0.84, but recalculation of this value including gassy patients yields a lower specificity of 0.72. Similar findings were observed in the reanalysis of our data after the exclusion of indeterminate examinations, although the specificity remained low (0.66) relative to antecedent studies.

CONCLUSION

Point-of-care ultrasound is moderately sensitive for small-bowel obstruction, although notably less specific, as performed by a diverse group of emergency physicians across multiple EDs. Interpretation of acquired point-of-care ultrasound images is significantly more accurate when performed by physicians with prior emergency ultrasound fellowship training and familiarity with the typical sonographic appearance of small-bowel obstruction.

We thank Theodore Bell, MS, Rodney Grim, PhD, and Soheil Saadat, MD, PhD, for their assistance with the study.

References


**Supporting Information**

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13713/full

**Supplemental Table S1.** CT interpretations and discharge diagnoses for false negative cases in the primary POCUS cohort (n = 11).

**Supplemental Table S2.** CT interpretations and discharge diagnoses for false positive cases in the primary POCUS cohort (n = 57).
ABSTRACT

Background: Ocular complaints are common presentations to the emergency department (ED). Among these, retinal detachment can cause significant vision loss if not rapidly diagnosed and referred for appropriate treatment. Point-of-care ultrasound has been suggested to identify the diagnosis rapidly when the ocular examination is limited or the ophthalmology service is not readily available. However, prior studies were limited by small sample sizes, resulting in wide ranges of potential accuracy. The primary outcome for this review was to determine the test characteristics of point-of-care ocular ultrasound for the diagnosis of retinal detachment.

Methods: PubMed, CINAHL, Scopus, LiLACS, the Cochrane Database of Systematic Reviews, the Cochrane Central Register of Controlled Trials, and bibliographies of selected articles were assessed for all prospective and randomized controlled trials assessing the accuracy of point-of-care ultrasound for identifying retinal detachment. Data were dual extracted into a predefined worksheet and quality analysis was performed using the QUADAS-2 tool. Data were summarized and a meta-analysis was performed with planned subgroup analyses by location and provider specialty. This review was registered with PROSPERO CRD42018097288. There was no funding for this review.

Results: Eleven studies (n = 844 patients) were identified. Overall, ultrasound was 94.2% (95% confidence interval [CI] = 78.4% to 98.6%) sensitive and 96.3% (95% CI = 89.2% to 98.8%) specific for the diagnosis of retinal detachment with a positive likelihood ratio of 25.2 (95% CI = 8.1 to 78.0) and a negative likelihood ratio of 0.06 (95% CI = 0.01 to 0.25). Subgroup analysis found that ultrasound was more accurate among ED patients, but was not significantly different when performed by ED or non-ED providers.

Conclusions: Point-of-care ocular ultrasound is sensitive and specific for the diagnosis of retinal detachment. Future studies should determine the ideal training protocol and the influence of color Doppler and contrast-enhanced ultrasound on diagnostic accuracy.

Ocular complaints are common emergency department (ED) presentations, accounting for approximately 3.4% of all ED visits over a 5-year period.¹ The differential diagnosis of flashes and floaters includes posterior vitreous detachment, retinal detachment, posterior uveitis, vitreous hemorrhage, oculodigital stimulation, rapid eye movements, neovascular age-related macular degeneration, migraine with aura, occipital lobe disorders, and postural hypotension.² While most patients will have a benign etiology, up to 10% to 26% of patients presenting with acute onset of flashes or floaters will be diagnosed with a
retinal detachment or tear.\(^2\),\(^3\) One study found that the incidence of retinal detachment ranged from 6.3 to 17.9 per 100,000 people.\(^4\) Rapid diagnosis is critical in these patients, as delays in diagnosis can be associated with irreversible loss of vision.\(^5\)

Unfortunately, historical features are poorly predictive of the diagnosis of retinal detachment.\(^2\) Floaters and flashes have a positive likelihood ratio (LR+) of 1.2 and a negative likelihood ratio (LR–) of 0.9, while vision loss with floaters or flashes has a LR+ of 5.0 and LR– of 0.6.\(^2\) Additionally, direct ophthalmoscopy is limited in the ED setting, with one study demonstrating that the majority of fundoscopic examinations were either skipped or performed inadequately.\(^6\) This may be due to limited fundoscopic training, inadequate skill maintenance, or limitations in the technique (e.g., examining nondilated eyes, performance of the examination in a bright room). When performed by ophthalmologists, indirect ophthalmoscopy has good test characteristics, with the presence of a vitreous hemorrhage having a LR+ of 10 and LR– of 0.49, while visualization of vitreous pigment has a LR+ of 44 and LR– of 0.23. Therefore, indirect ophthalmoscopy performed by an ophthalmologist is generally considered to be the criterion standard. However, rapid ophthalmologist evaluation is not universally available, especially during nonbusiness hours. Consequently, point-of-care ultrasound has been suggested to be a potential alternate modality. Ultrasound machines are widely available in most EDs and the American College of Emergency Physicians has included ocular ultrasound as a core application within emergency medicine.\(^7\) However, prior studies were limited by small sample sizes, resulting in wide ranges of potential accuracy.\(^8\)–\(^11\) Since then, a number of additional studies have been published, significantly increasing the available data.\(^12\)–\(^19\)

The primary outcome for this systematic review and meta-analysis was to determine the test performance characteristics of point-of-care ocular ultrasound for the diagnosis of retinal detachment. A priori subgroup analyses of the primary outcome were planned for location (ED vs. non-ED location) and provider specialty (emergency medicine vs non–emergency medicine).

**METHODS**

Our study conforms to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Diagnostic Test Accuracy (PRISMA-DTA) guidelines for systematic reviews and was performed in accordance with best practice guidelines (see Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13682/full).\(^20\) This review was registered with PROSPERO CRD42018097288. There was no funding for this review. In conjunction with a medical librarian, we conducted a search of PubMed, CINAHL, LILACS, Scopus, the Cochrane Database of Systematic Reviews, and the Cochrane Central Register of Controlled Trials to include citations from inception to June 12, 2018. Details of the search strategy are included in Data Supplement S1, Appendix S1. We reviewed the bibliographies of identified studies and review articles for potential missed articles. We also consulted with topic experts to help identify any further relevant studies.

**Inclusion and Exclusion Criteria**

Inclusion criteria consisted of all prospective or randomized controlled trials assessing point-of-care ocular ultrasound for retinal detachment without age, language, or date restrictions. All studies must have had a confirmatory test, defined as formal ophthalmologic examination, surgical findings, computed tomography (CT), magnetic resonance imaging, or clinical follow-up. Point-of-care ultrasound was defined as ocular ultrasound performed with a portable ultrasound machine. Studies using static ophthalmologic ultrasound devices (i.e., where the eye is stationary and an automated ultrasound machine performs a complete ocular ultrasound examination) were excluded. We also excluded case reports, case series, retrospective studies, cadaver studies, and conference abstracts. Conference abstracts were excluded because they often contain only preliminary data, have limited methodologic descriptions, and have not undergone formal peer review. A priori subgroup analyses of the primary outcome were planned for location (ED vs. non-ED location) and provider specialty (emergency medicine vs. non–emergency medicine).

Two investigators (MG, DH) independently assessed all studies for eligibility based upon the above criteria. All abstracts meeting initial criteria were reviewed as full manuscripts. Studies determined to meet the eligibility criteria on full text review by both extractors (MG, DH) were included in the final data analysis. Any discrepancies were resolved by consensus with the addition of a third reviewer (GDP).
Data Collection and Processing
Two investigators (MG, DH) independently dual extracted data from the included studies. The investigators underwent initial training and extracted data into a predesigned data collection form. The following information was abstracted: last name of the first author, publication year, study country, study population size, type of study (e.g., prospective or randomized controlled trial), study location (e.g., ED, outpatient clinic), mean age of the study patients, gender of the study patients, number of retinal detachments, sonographer training, operator specialty, operator experience (i.e., attending or resident physician), true positives, false positives, true negatives, and false negatives.

Studies were independently assessed for quality by two separate investigators (MG, DH) utilizing the QUADAS-2 tool. Any discrepancies were resolved by consensus with the addition of a third reviewer (GDP). The authors used several a priori conditions to evaluate an individual study’s risk of bias and degree of applicability.

- In the assessment of “patient selection,” studies that included all consecutive patients or a random selection of patients were considered “low risk” of bias. Convenience sampling was considered “unclear risk” of bias. As the study question was intended to assess point-of-care ultrasound, we considered any ultrasound performed by an emergency physician at the point of patient care with a portable ultrasound machine to be low risk for applicability. Studies performed by non–emergency physicians were deemed at unclear risk for applicability.

- In the assessment of “index test,” studies were determined to be low risk of bias if the sonographer was blinded to the results of the reference standard. With regard to applicability, studies were deemed low risk if they used a linear transducer and imaged the eye in both the transverse and the sagittal planes. Studies that utilized a nonlinear probe or did not clearly describe the use of both imaging planes in their protocol were deemed unclear risk for applicability.

- In the assessment of the “reference standard,” studies were determined to be low risk of bias if they were interpreted without knowledge of the index test (i.e., ocular ultrasound examination). Studies that used an accepted criterion standard for diagnosis (i.e., as formal ophthalmologic examination, surgical findings, CT, magnetic resonance imaging, or clinical follow-up) were deemed low risk.

- In the assessment of “flow and timing,” we defined low-risk studies as those where all participants were followed up and had the same reference standard. Studies with different reference standards for some patients were deemed unclear risk of bias. Because retinal detachments are not expected to resolve without an intervention and the likelihood of a new spontaneous retinal detachment occurring in the short interval between studies is unlikely, we did not require a specific time interval between when the index test and reference standard were performed.

Data Analysis
The primary analysis was completed using a bivariate random effects model to calculate sensitivity, specificity, LR+, and LR− with 95% confidence intervals (CIs). Chi-square and I² statistics were utilized to assess heterogeneity of included studies with a p-value of <0.1 or I² > 50% considered significant for heterogeneity. A kappa value was calculated for study selection and the QUADAS-2 assessments. A summary receiver operating characteristic (ROC) curve with observed study data and a 95% CI region was constructed. We used a linear regression test of funnel plot asymmetry with a p-value of <0.1 for the slope coefficient. Bayes’ theorem was used to determine posttest probability of a positive and negative ultrasound using the cumulative prevalence from all included studies, as well as the lowest and highest prevalence of studies examined individually. For the subgroup analyses where fewer than four studies were included and precluded use of a bivariate model, then a univariate model was utilized.

Statistical analysis was completed with StataMP, version 13 (StataCorp LP). The MIDAS module was used to perform analyses, including bivariate random-effects analyses, summary ROC curve analysis, and assessment of publication bias, and to construct the graphs. The DIAGT module was utilized to perform univariate analyses. Forest plots were constructed using RevMan, version 5.3 (The Nordic Cochrane Centre, Copenhagen, Denmark), with the studies listed by year.

RESULTS
A total of 2,621 studies were identified. PubMed yielded 1,011 studies, Scopus identified 1,427 studies, CINAHL found 71 studies, LILACS discovered 112,
the Cochrane Database of Systematic Reviews yielded no studies, and the Cochrane Central Register of Controlled Trials identified no studies. After removing duplicates, 1,967 abstracts were reviewed with 39 selected for full-text review (Figure 1). No additional papers were identified by the topic experts or through bibliographic review.

Eleven studies, comprising 844 patients, were selected for the final analysis (Table 1). The kappa value for study selection was 1.0. All eleven studies were prospective, observational trials. Studies were published between 1995 and 2018 with population sizes ranging from 32 to 139. Three studies were conducted in the United States,9–11 two were performed in Canada,12,13 one took place in China,14 one was conducted in India,15 one was performed in Ireland,16 one was conducted in Korea,17 one took place in Nigeria,17 and one was performed in Thailand.18 Five studies were conducted in the ED setting,9–11,13,14 two were performed in radiology departments,17,18 one was performed in an ophthalmology clinic,12 and three did not describe the study setting.15,16,19 Emergency medicine providers performed the examination in six studies,9–14 radiologists performed the examination in four studies,16–19 and one study did not describe the provider specialty.15 All studies used a linear probe except for Ukponmwan and Marchien, who used a curvilinear probe.18 Among included studies, the mean age of patients was 52.9 years and 48.4% of patients were male. Overall, retinal detachments were present in 21.4% of the study participants. There were no indeterminate ultrasound examinations.

Overall, ultrasound had a 94.2% (95% CI = 78.4% to 98.6%) sensitivity and 96.3% (95% CI = 89.2% to 98.8%) specificity for the diagnosis of retinal detachment with a LR+ of 25.2 (95% CI = 8.1 to 78.0) and a LR– of 0.06 (95% CI = 0.01 to 0.25; Figure 2). The area under the ROC curve indicated high accuracy (0.988; 95% CI = 0.974 to 0.994; Figure 3). Statistical heterogeneity was moderate with an $I^2$ of 0.59. Funnel plot analysis demonstrated no evidence of publication bias (Figure 4).

Among ED providers, ultrasound was 92.0% (95% CI = 67.2% to 98.5%) sensitive and 91.4% (95% CI = 84.9% to 95.3%) specific (Data Supplement S1, Figure S1). Among non-ED providers, ultrasound was 91.1% (95% CI = 67.5% to 98.0%) sensitive and 98.6% (95% CI = 81.7% to 99.9%) specific (Data Supplement S1, Figure S2). Note that the overall pooled sensitivity is higher than the pooled sensitivities
of either subgroup due to the inclusion of Nagaraju et al.\textsuperscript{15} (with 100% sensitivity and 100% specificity) among the 11 studies, but exclusion of that same study from either subgroup due to uncertainty about whether ED providers performed the retinal ultrasound.

Among ED patients, ultrasound was 93.9% (95% CI = 78.7% to 98.5%) sensitive and 92.4% (95% CI = 85.6% to 96.1%) specific (Data Supplement S1, Figure S3). Among non-ED patients, ultrasound was 74.1% (95% CI = 61.0% to 84.7%) sensitive and 85.3% (95% CI = 75.3% to 92.4%) specific (Data Supplement S1, Figure S4).

Studies were at overall low risk of bias for most parameters, except patient selection and reference standard (Table 2). Five studies were at unclear risk of bias for patient selection because they utilized a convenience sample.\textsuperscript{10,12,15,17,19} Five studies were at unclear risk of applicability for patient selection because they were performed by non-EM physicians.\textsuperscript{15–19} One study was at unclear risk for reference standard because ophthalmologists were not blinded to the index test.\textsuperscript{11} Four studies were at unclear risk of flow and timing due to different confirmatory testing.\textsuperscript{9,15–17} Four studies were at risk of applicability for the index test because they used a curvilinear probe\textsuperscript{18} or did not explicitly state that both transverse and sagittal views were obtained.\textsuperscript{11,16,19} No studies were at high risk of bias for any of the parameters. The kappa value for the QUADAS-2 selection was 0.84 with 97.4% inter-observer agreement.

**DISCUSSION**

To the best of our knowledge, this is the largest review to date and the first meta-analysis to demonstrate the diagnostic accuracy of point-of-care ultrasound for retinal detachment. This systematic review and meta-analysis demonstrated that point-of-care ultrasound is both sensitive and specific for the diagnosis of retinal detachment. Additionally, there were no significant differences in the accuracy of point-of-care ultrasound among ED and non-ED providers. Accuracy was also maintained when only ED patients were included.

Importantly, the use of point-of-care ultrasound for retinal detachment depends on the pretest probability of the disease. In the included studies, the overall prevalence ranged from 7.0% to 53.8%, Using the lowest reported prevalence (7.0%), the posttest probability of a positive ultrasound was 65.4% and the
Therefore, an ultrasound examination without evidence of a retinal detachment could reliably exclude the disease, while a positive examination would be insufficient to confirm the diagnosis. Using the highest reported prevalence of retinal detachment (53.8%), the posttest probability of a positive ultrasound was 96.7% while a negative ultrasound was 6.5%. In this case, a positive examination is strongly supportive of the diagnosis; however, a negative examination is insufficient to exclude the disease. One particularly valuable potential use of point-of-care ocular ultrasound would be in resource-limited settings where telemedicine could be used to help identify the diagnosis where access to ophthalmology may be more limited. This has been previously found to be successful with cardiac and lung ultrasound. Further studies should assess the utility of ocular ultrasound in this area.

One prior systematic review by Vrablik and colleagues evaluated the use of point-of-care ultrasound for retinal detachment and found that it was 97% to 100% sensitive and 83% to 100% specific. However, that review was performed in 2012 and was limited by small sample sizes and inability to perform meta-analysis of the data. The current study utilized a more targeted search strategy combined with the inclusion of three additional databases and an updated search, resulting in the identification of eight additional studies with a greater than fourfold increase in the number of patients. As a result, our study was able to perform a meta-analysis of the data, allowing for more accurate test characteristics of this modality. Additionally, our study was the first to perform subgroup analyses by provider type and patient location.

While this modality can be beneficial for the diagnostic assessment of retinal detachment, it is also

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>McNicholas 1995</td>
<td>17</td>
<td>0</td>
<td>4</td>
<td>40</td>
<td>0.81 [0.58, 0.95]</td>
<td>1.00 [0.91, 1.00]</td>
<td>1.00 [0.91, 1.00]</td>
<td>1.00 [0.91, 1.00]</td>
</tr>
<tr>
<td>Han 2001</td>
<td>11</td>
<td>4</td>
<td>3</td>
<td>14</td>
<td>0.79 [0.49, 0.95]</td>
<td>0.78 [0.52, 0.94]</td>
<td>0.78 [0.52, 0.94]</td>
<td>0.78 [0.52, 0.94]</td>
</tr>
<tr>
<td>Ukponmwan 2001</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>1.00 [0.84, 1.00]</td>
<td>1.00 [0.81, 1.00]</td>
<td>1.00 [0.81, 1.00]</td>
<td>1.00 [0.81, 1.00]</td>
</tr>
<tr>
<td>Blavas 2002</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>52</td>
<td>1.00 [0.66, 1.00]</td>
<td>1.00 [0.93, 1.00]</td>
<td>1.00 [0.93, 1.00]</td>
<td>1.00 [0.93, 1.00]</td>
</tr>
<tr>
<td>Younessi 2010</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>1.00 [0.81, 1.00]</td>
<td>0.83 [0.65, 0.94]</td>
<td>0.83 [0.65, 0.94]</td>
<td>0.83 [0.65, 0.94]</td>
</tr>
<tr>
<td>Kongspap 2011</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>87</td>
<td>1.00 [0.59, 1.00]</td>
<td>0.99 [0.94, 1.00]</td>
<td>0.99 [0.94, 1.00]</td>
<td>0.99 [0.94, 1.00]</td>
</tr>
<tr>
<td>Shinar 2011</td>
<td>28</td>
<td>5</td>
<td>1</td>
<td>58</td>
<td>0.97 [0.82, 1.00]</td>
<td>0.92 [0.82, 0.97]</td>
<td>0.92 [0.82, 0.97]</td>
<td>0.92 [0.82, 0.97]</td>
</tr>
<tr>
<td>Nagaraju 2015</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>93</td>
<td>1.00 [0.59, 1.00]</td>
<td>1.00 [0.96, 1.00]</td>
<td>1.00 [0.96, 1.00]</td>
<td>1.00 [0.96, 1.00]</td>
</tr>
<tr>
<td>Woo 2016</td>
<td>11</td>
<td>7</td>
<td>12</td>
<td>32</td>
<td>0.48 [0.27, 0.69]</td>
<td>0.82 [0.66, 0.92]</td>
<td>0.82 [0.66, 0.92]</td>
<td>0.82 [0.66, 0.92]</td>
</tr>
<tr>
<td>Chu 2017</td>
<td>14</td>
<td>16</td>
<td>2</td>
<td>107</td>
<td>0.88 [0.62, 0.98]</td>
<td>0.87 [0.80, 0.92]</td>
<td>0.87 [0.80, 0.92]</td>
<td>0.87 [0.80, 0.92]</td>
</tr>
<tr>
<td>Kim 2018</td>
<td>12</td>
<td>6</td>
<td>4</td>
<td>93</td>
<td>0.75 [0.48, 0.93]</td>
<td>0.94 [0.87, 0.98]</td>
<td>0.94 [0.87, 0.98]</td>
<td>0.94 [0.87, 0.98]</td>
</tr>
</tbody>
</table>

Figure 2. Forest diagram of the overall sensitivity and specific of ultrasound for retinal detachment.

Figure 3. Summary receiver operating characteristics curve of ultrasound for retinal detachment.

Figure 4. Funnel plot of ultrasound for retinal detachment.
important to consider several limitations with respect to the use of ocular ultrasound. First, as with many ultrasound applications, it is provider-dependent. Therefore, it is important to ensure that providers obtain adequate training and practice with this technique. Based on an average of 3.4% of patients presenting with ocular complaints and because ocular ultrasound is not limited to the assessment of retinal detachments, the average provider would have the opportunity to utilize this examination once every 29 patients (or approximately once per shift).1 With an annual incidence of retinal detachment ranging from 6.3 to 17.9 per 100,000 people, providers would encounter a retinal detachment approximately once every 5,000 to 15,000 patients.4 Therefore, it is important for providers to refresh their experience with identification of retinal detachment on ultrasound. An excellent review of the technique and findings for ocular ultrasound is available for free in the electronic book Introduction to Bedside Ultrasound: Volume 2 (available at: https://itunes.apple.com/us/book/introduction-to-bedside-ultrasound-volume-2/id647356692?mt=11), as well as the following online resource: https://www.youtube.com/watch?v=gNLTZipajwM.

Additionally, researchers should assess the potential value of simulation and how many abnormal examinations are needed for adequate training, as well as the feasibility of this in daily practice.7,26,27 Importantly, ultrasound has the potential to increase intraocular pressure, so it is important to ensure that the hand is properly balanced on the nasal bridge or maxilla and that adequate gel is applied. Finally, it can be challenging to differentiate retinal detachment from posterior vitreous detachment and providers should have a low threshold to refer to ophthalmology if the diagnosis is unclear.28 Future studies should assess the inter-rater reliability of point-of-care ocular ultrasound for retinal detachment among providers and between providers of different specialties (e.g., emergency medicine, radiology, ophthalmology). Studies should also assess whether high-frequency linear probes, color Doppler, or contrast-enhanced ultrasound will improve the diagnostic accuracy.

**LIMITATIONS**

It is important to consider several limitations with respect to the current study. First, studies involved a variety of provider experience levels and specialties, with only half involving ED providers. This may explain some of the heterogeneity identified in this meta-analysis. However, all studies utilized similar ultrasound machines and scanning protocols. Additionally, subgroup analyses did not demonstrate a significant difference in accuracy with respect to providers. Unfortunately, we were unable to comment on the provider accuracy with respect to provider experience, as most
studies included a combination of resident and attending physician sonographers. The training protocols also varied between studies and, while the majority of training sessions ranged from 30 to 120 minutes, it is unclear what the ideal training protocol is. Further studies are needed to determine the ideal training protocol and learning curve for this modality. Additionally, the included studies only assessed B-mode ultrasound. Some studies have suggested that color or focused Doppler ultrasound may be more accurate than B-mode, but the current data are limited. While most studies had similar test characteristics, Woo and colleagues had a notable decrease in sensitivity compared with the other publications. This may be secondary to inadequate training, with most studies being performed by a medical student.

As discussed, four studies used two different confirmatory tests, which creates a risk of verification bias. Additionally, one study used a curvilinear probe and three did not explicitly state that they included both transverse and sagittal views, which may limit the applicability. Moreover, only one of the included studies performed an a priori sample size calculation and none of the studies adhered to the Standards for Reporting of Diagnostic Accuracy Studies (STARD) criteria, which may suggest lower-quality studies. Future studies should adhere to the STARD criteria to ensure high-quality research. Given the relatively wider CIs, it would be valuable to include additional studies of point-of-care ocular ultrasound using more homogenous groups of providers and training. Additionally, randomized, controlled trials may be beneficial to better assess the influence on patient-centered outcomes, such as the influence on referral rates and procedural interventions. Moreover, the cost-effectiveness of this approach should also be assessed. This should take into account initial skill acquisition, skill maintenance, patient care time spent on the examination, referral and consultation rates, and potential billing revenue.

It is possible that new studies may have been published between completion of our literature review and when this article was completed. However, we are unaware of any new publications since that time and a repeated search from the time of the initial search and when this manuscript was submitted did not identify any new articles. Our exclusion of retrospective studies and conference abstracts may have limited our findings. However, this was determined a priori to ensure that only higher-quality studies were included. Finally, while there was no evidence of publication bias on our assessment, the evaluation of publication bias has been suggested to be more limited in systematic reviews of diagnostic test accuracy.

**CONCLUSION**

Point-of-care ocular ultrasound is both sensitive and specific for the diagnosis of retinal detachment. This may be a valuable adjunct when ophthalmology is not immediately available.

The authors thank Jennifer C. Westrick, MSLIS, for her assistance with the literature search.

**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.13682/full
Data Supplement S1. Supplemental material.
Underdosing of Benzodiazepines in Patients With Status Epilepticus Enrolled in Established Status Epilepticus Treatment Trial

Abhishek G. Sathe, MS1,2, Holly Tillman, MS2, Lisa D. Coles, MS, PhD1, Jordan J. Elm, PhD2, Robert Silbergleit, MD3, James Chamberlain, MD4, Jaideep Kapur, MD, PhD5,6, Hannah R. Cock, MBBS, MD7, Nathan B. Fountain, MD5, Shlomo Shinnar, MD, PhD8, Daniel H. Lowenstein, MD9, Robin A. Conwit, MD10, Thomas P. Bleck, MD11, and James C. Cloyd, PharmD1

Benzodiazepines, including diazepam (DZP), lorazepam (LZP), and midazolam (MDZ), are considered the initial drugs of choice for status epilepticus (SE) treatment. A number of trials have demonstrated their safety and efficacy; however, the failure rate ranges from 10% to 55%.1,2 This may be attributable, in part, to suboptimal benzodiazepine dosing and timing of administration.

The Neurocritical Care Society (NCS) and American Epilepsy Society (AES) have published evidence-based guidelines for benzodiazepine use in SE that specify drugs, doses, and routes of administration.1,2 Initial benzodiazepine treatment should consist of either a 10-mg dose of intramuscular (IM) MDZ for patients weighing > 40 kg or 5 mg for those 13 to 40

From the1Department of Experimental and Clinical Pharmacology, College of Pharmacy and Center for Orphan Drug Research, University of Minnesota, Minneapolis, MN; the2Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC; the3Department of Emergency Medicine, University of Michigan, Ann Arbor, MI; the4Division of Emergency Medicine, Children’s National Health System, and the Department of Pediatrics and Emergency Medicine, School of Medicine and Health Sciences, George Washington University Washington, DC; the5Department of Neurology; and the6Department of Neuroscience, Brain Institute, University of Virginia, Charlottesville, VA; St. George’s University of London and St. George’s University Hospitals NHS Foundation Trust, London, UK;7Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY; the8Department of Neuroscience, University of California, San Francisco, CA;9National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD; and10Feinberg School of Medicine, Northwestern University and Rush Medical College, Chicago, IL.

Received April 1, 2019; revision received May 13, 2019; accepted May 25, 2019.

This work has been presented as an abstract at the 2017 London Innsbruck Colloquium on Status Epilepticus and Acute Seizures, Salzburg, Austria, April 6–8, 2017.

Research reported in this publication was supported by National Institutes of Health, National Institutes of Neurological Disorders and Stroke under Awards U01NS088034, U01NS088023, U01NS056975, U01NS059041, and R01NS099653 (Clinical trials.gov identifier NCT01960075).

Author contributions: AGS, LDC, HT, JJE, and JCC interpreted the data and prepared the initial draft of the manuscript; HT and JJE performed the statistical analysis; JK, RS, JC, HRC, TPB, and JJE conceived and designed the ESETT study and analysis; all authors significantly contributed to article revisions; and AGS takes responsibility for the paper as a whole.

All authors were supported by the ESETT study grant from NIH/NINDS (U01NS088034). Dr. Coles reports grants from NIH/NINDS, during the conduct of the study, and personal fees from Neurelis Pharmaceuticals, grants from Sollevio, outside the submitted work. Dr. Shinnar reports grants from NINDS, during the conduct of the study, personal fees from Eisai, and personal fees from Ipsys, outside the submitted work. Prof. Cock reports grants from NINDS, during the conduct of the study, and personal fees from Sage Pharmaceuticals Ltd, personal fees from Eisai Europe Ltd, personal fees from UCB Pharma Ltd, personal fees from UK Epilepsy Nurse Specialist Association, nonfinancial support from Special Products Ltd, nonfinancial support from International League Against Epilepsy, Epilepsy Certification (education) Task Force, nonfinancial support from European Academy of Neurology, and personal fees from Bial and Eisai, outside the submitted work. Dr. Fountain reports grants from NINDS, during the conduct of the study, and grants from SK Lifesciences, grants from Neurelis, grants from Takeda, grants from GW Pharma, grants from Biogen, and grants from UCB, outside the submitted work. Dr. Cloyd reports a grant from National Institute of Neurological Disorders and Stroke during the conduct of the study, personal fees from Neurelis Pharmaceuticals, and grants from Sollevio, outside the submitted work.

Supervising Editor: Peter Panagos, MD.

Address for correspondence and reprints: Abhishek G. Sathe, MS; e-mail: sathe134@umn.edu.


© 2019 by the Society for Academic Emergency Medicine
doi: 10.1111/acem.13811
kg or intravenous (IV) LZP 0.1 mg/kg/dose (maximum 4 mg/dose) or IV DZP 0.15 to 0.2 mg/kg/dose (maximum 10 mg/dose). The LZP and DZP doses can be repeated if the initial dose fails to stop the seizure. Although not included in the guidelines, based on pharmacokinetics, 10 mg IV MDZ dose can be considered adequate therapy.

Reports have documented underdosing of benzodiazepines used in SE; however, comprehensive information regarding patient age, setting, drugs, doses, timing of doses, and routes is limited. This report describes patterns of benzodiazepine use in SE in a geographically diverse population.

The Established Status Epilepticus Treatment Trial (ESETT) provided an opportunity to systematically observe benzodiazepine administration in patients subsequently determined to have SE unresponsive to benzodiazepines. Using preenrollment data from ESETT subjects, we describe benzodiazepine treatment with respect to: 1) drug choice, dose, and route of administration; 2) timing and setting in which the drugs were administered; and 3) patient weight (< 40 or ≥ 40 kg for LZP, ≤ 40 or > 40 kg for MDZ, and < 66.7 or ≥ 66.7 kg for DZP). NCS and AES guidelines were used to define underdosing for our analyses. These weight-based cutoffs were per published guidelines.

Because patients could receive more than one benzodiazepine, the cumulative dose was determined using LZP equivalents to account for differences in drug potencies. Transmucosal benzodiazepines, e.g., rectal DZP or intranasal/buccal MDZ, given prior to emergency medical services (EMS) arrival are included in the calculation of cumulative benzodiazepine dose. For patients weighing ≥ 32 kg, 10 mg MDZ or DZP was considered equal to 4 mg LZP. For patients weighing < 32 kg, 0.3 mg/kg DZP IV or 0.2 mg/kg MDZ IV or 0.3 mg/kg MDZ IM were considered equal to 0.1 mg/kg LZP IV. There was no upper limit for the benzodiazepine dose required to qualify for ESETT enrollment. While the ESETT protocol stipulated a minimum cumulative adequate dose for enrollment (Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/ace.m.13811/full), instructions on the rate and frequency of dosing were not provided. ESETT sites were expected to dose benzodiazepines as per their local standards of care. The settings in which benzodiazepines were administered were categorized as: 1) prior to EMS, 2) EMS, and 3) emergency department (ED).

Of these, 86.7% of DZP, 14.5% of MDZ, and 23.2% of LZP doses met minimum recommendations per guidelines. Among all subjects, 102 received their first dose of any benzodiazepine in the ED. Overall, 29.8% of first doses met minimum recommendations per guidelines. Of these, 86.7% of DZP, 14.5% of MDZ, and 23.2% of LZP administrations met the minimum dose recommendations. Figure 1 shows that for subjects < 40 kg the guideline recommended LZP (≥ 0.1 mg/kg) or MDZ (≥ 5 mg) dose was administered as a first dose in 41.9 and 12.5% of the cases, respectively. In contrast, for those weighing ≥ 40 kg the recommended LZP (≥ 4 mg) or MDZ (≥ 10 mg) dose was administered in 14.7 and 15.4% of the subjects, respectively. A DZP dose ≥ 10 mg was administered in 60% of the subjects ≥ 66.7 kg, while 96% of DZP administrations were ≥ 0.15 mg/kg in those < 66.7 kg.

**FIRST DOSE OF FIRST BENZODIAZEPINE**

Among all subjects, 102 received their first dose of any benzodiazepine in the ED. Overall, 29.8% of first doses met minimum recommendations per guidelines. Of these, 86.7% of DZP, 14.5% of MDZ, and 23.2% of LZP administrations met the minimum dose recommendations. Figure 1 shows that for subjects < 40 kg the guideline recommended LZP (≥ 0.1 mg/kg) or MDZ (≥ 5 mg) dose was administered as a first dose in 41.9 and 12.5% of the cases, respectively. In contrast, for those weighing ≥ 40 kg the recommended LZP (≥ 4 mg) or MDZ (≥ 10 mg) dose was administered in 14.7 and 15.4% of the subjects, respectively. A DZP dose ≥ 10 mg was administered in 60% of the subjects ≥ 66.7 kg, while 96% of DZP administrations were ≥ 0.15 mg/kg in those < 66.7 kg.

**DOSE PER ADMINISTRATION**

Seventy-seven percent of DZP, 10.7% of MDZ, and 21.8% of LZP doses administered were at or above the recommendations (Data Supplement S1). Prior to EMS, most administrations were DZP (25/37) given at or above the minimum recommended doses, whereas in both the EMS and the ED settings, most
of the administered benzodiazepine doses were below recommendations.

**CUMULATIVE BENZODIAZEPINE DOSES**

Cumulative dosing patterns were examined using LZP equivalents (Data Supplement S1). Among 138 adults and older children weighing ≥ 32 kg, the cumulative dose in LZP equivalents was < 4 mg in 9%, 4 mg in 42%, 5 to 6 mg in 25%, and > 7 mg in 24%. In 68 children weighing < 32 kg, the cumulative dose was < 0.1 mg/kg in 18%, 0.1 to < 0.2 mg/kg in 44%, 0.2 to < 0.3 mg/kg in 28%, and >0.3 mg/kg in 10% of subjects.

The results of this study suggest that many patients with SE who fail benzodiazepine treatment are not receiving recommended initial doses of benzodiazepines. The observed practice was not consistent with published evidence-based guidelines, which stipulate that the initial treatment of SE begin with a benzodiazepine administered as early as possible, as a single full dose, and by an appropriate route. In contrast, we found a pattern of administering multiple, small doses with approximately 70% of patients receiving a lower than guideline recommended first dose of the first drug. If, however, rectal DZP is excluded, the first doses of MDZ and LZP, mostly administered by EMS and/or ED personnel, were below guideline recommendations 80% of the time. Administration of subsequent doses continued the pattern of underdosing. Regardless of the number of administrations, approximately 12% of patients never received the required cumulative dose needed to meet ESETT eligibility criteria. This potentially reduced response to benzodiazepines as delay in administering appropriate therapy is thought to place patients at risk for longer seizures and poor outcomes.
Our results extend the findings from earlier reports on initial management of SE. In a multicenter study of adults, the investigators found that > 80% of patients with SE received a lower than recommended LZP dose. Langer and Fountain, in a retrospective study of generalized convulsive SE in 170 children and adults, found that only 11% of the patients, all children, received an adequate initial benzodiazepine dose. The problem of benzodiazepine underdosing in SE may be attributable to the perceived risk of cardiorespiratory compromise associated with benzodiazepines. However, Alldredge et al. showed that the rate of respiratory or circulatory complications was nearly doubled (p = 0.08) in untreated SE patients versus those treated with benzodiazepines. We also noted that on 17 occasions LZP was administered by IM, IN, or buccal routes. These routes do not support rapid LZP absorption and are inappropriate for SE therapy.

Our analysis is limited to SE patients who continued to have seizures despite benzodiazepine treatment. Since initial benzodiazepine underdosing is likely associated with treatment failure, our population may overestimate the rate of underdosing among patients treated for SE. While this limits the generalizability of our findings, benzodiazepine underdosing is particularly important in this subgroup in whom seizures continue and may progress to refractory SE with attendant high rates of morbidity and mortality. Conversely, this analysis may under estimate the rate of underdosing because only those given an adequate cumulative benzodiazepine dose were eligible for ESETT enrollment. It is possible that eagerness to enroll subjects could bias toward lower cumulative benzodiazepine doses. However, in this scenario, EDs would be more likely to administer larger individual doses to meet the minimum adequate dose sooner and should not affect EMS practice. Lastly, our sample size precluded the analysis of specific factors such as regional effects on dosing patterns.

Benzodiazepine underdosing for the treatment of SE was common in this geographically diverse set of EDs. This phenomenon may contribute to decreased efficacy. Further, the low doses used per administration in both ED and EMS settings suggests this represents practice culture rather than an artifact in practice driven by study enrollment. Hence, greater educational efforts and overcoming systematic and structural barriers are needed to change clinical practice.

We acknowledge the ESETT Data and Safety Monitoring Board. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke, National Institutes of Health, or the U.S. Government.

References

Supporting Information
The following supporting information is available in the onine version of this paper available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13811/full  
Data Supplement S1. Supplemental material.  
Figure S1. Total number of administrations that met (blue) and did not meet (red) guideline recommendations for DZP, MDZ and LZP (N=511) (Numbers on top of the bars represent % administrations for each drug)  
Figure S2. Distribution of the cumulative benzodiazepine dose in lorazepam equivalents for subjects weighing ≥ 32 kg (top panel) and < 32 kg (bottom panel)
Emergency Department Risk Stratification After Opiate Overdose Is Just the Beginning

Brian M. Clemency, DO¹, Joshua J. Lynch, DO¹, Terrance Creighton, MD², and Heather A. Lindstrom, PhD¹

The January 2019 issue of Academic Emergency Medicine reported on the Hospital Observation Upon Reversal (HOUR) with Naloxone Study. This study examined the ability of emergency medicine providers to use clinical judgment, a decision rule, or both to evaluate patients for safe discharge from the emergency department (ED) 1 hour after naloxone administration for suspected opiate overdose.¹ The topic of safe and timely discharge is important for the practice of emergency medicine, but not nearly as important as what comes next for our patients.

When a patient arrives at an ED following an out-of-hospital naloxone administration, clinicians must first focus on emergency evaluation and stabilization. If the patient survives the initial event, attention must next turn to strategies to prevent future occurrences. In this way, opioid use disorder must be treated like any other disease. The importance of addressing a particular disease is a function of its likelihood to result in morbidity and mortality for a patient. Weiner et al.² demonstrated that among patients treated with naloxone by EMS who did not die the same day, the 1-year mortality rate was approximately 10%. However, the exact short-term mortality rate for patients following an opioid overdose is unclear. Better understanding this rate would aid in the design of public health interventions and inform shared decision making between emergency medicine providers and patients.

We sought to determine the risk of 60-day mortality among males ages 18 to 54, enrolled in the HOUR study, who were discharged home and survived the initial 48-hour study period. Mortality was determined by a review of local medical examiner records. This protocol was approved by the University at Buffalo Institutional Review Board.

There were 350 patients included in this follow-up analysis. Although all survived to ED discharge and survived the initial 48-hour study period, tragically, eight (2.3%) died within 60 days of their original overdose. Half of those who died were less than 35 years of age.

The follow-up analysis relied on county medical examiner records and may have systematically underestimated the true mortality rate since deaths that were attributed to natural causes or that occurred outside of the jurisdiction would not be represented in medical examiner records. Many more patients are likely to have had nonfatal adverse overdose in the 60 days following ED discharge.

To put the observed 60-day mortality rate of 2.3% in perspective, a similar short-term mortality rate would likely be viewed as unacceptable if the same young patients presented to the ED for chest pain. In
the context of ED evaluation of chest pain, a 1% to 2% acceptable miss rate is often discussed, but this typically applies to an older population and a composite endpoint that includes outcomes beyond just mortality, yet the current standard of care following opioid overdose frequently results in patients simply being discharged home after a period of observation. Given the short- and long-term mortality risks for these patients, ED visits can, and should, be used as an opportunity to intervene.

Distributing naloxone kits to patients with opioid use disorder and their families is one such intervention strategy. Widespread community distribution of naloxone, when coupled with opioid overdose recognition training, has been shown to decrease mortality rates. An ED encounter following an overdose is an opportunity to provide the patient and family with a brief training and an emergency naloxone kit.

Another promising harm reduction technique is ED-initiated buprenorphine/naloxone treatment for opioid use disorder. Multiple randomized controlled trials have demonstrated that patients in medication-assisted treatment programs are more likely to remain in treatment compared to patients who are given a placebo or no medications at all. D’Onofrio et al. demonstrated that patients who were started on buprenorphine/naloxone (Suboxone) and referred to outpatient treatment within 72 hours of ED discharge were more likely to be engaged in treatment at the 2-month mark. In the case of patients who have recently overdosed, a home induction may be considered in which the patient is discharged from the ED and takes the first dose following onset of withdrawal symptoms. In patients who present to the ED in opioid withdrawal, induction on buprenorphine/naloxone is also safe and appropriate. Both groups of patients should receive urgent follow-up with a qualified treatment program.

Opioid overdose patients depend on emergency medicine providers to provide more than just immediate lifesaving measures. Many emergency physicians understand and are willing to participate in opioid harm reduction programs, yet few actually do. Further study is needed to better understand barriers to implementation of ED-based harm reduction strategies.

Patients who receive naloxone for opioid overdose and survive are at risk for death in the next 60 days. ED visits following opioid overdoses may be an underutilized opportunity to implement harm reduction strategies and reduce patients’ short-term risk of death.

References
Hot Off the Press: A Systematic Review And Meta-analysis of Ketamine as an Alternative to Opioids for Acute Pain in the Emergency Department

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

ABSTRACT
Ketamine has been studied as an alternative to opioids for acute pain in the emergency department setting. This review compares the effectiveness of intravenous ketamine at a dose of <0.5mg/kg to opioids for acute pain in adult patients. Measurements were taken within 60 minutes of administration. Ketamine was found to have similar effectiveness to opioids. Increased, but short-lived, side effects were seen with ketamine.

BACKGROUND
Opiates are commonly used for acute pain in the emergency department (ED) setting. However, in recent years, an increased desire for alternatives has been prompted in an attempt to reduce opiate usage. Several studies have been performed evaluating low-dose ketamine (LDK) for acute pain, with a variety of methodologic designs, time endpoints, and doses. This review attempts to determine the utility of LDK, alone, in comparison to opiates for acute pain in the ED.

ARTICLE SUMMARY
This is a systematic review and meta-analysis with strict inclusion criteria, which limited the studies to English-language randomized controlled trials of adult patients using doses of <0.5mg/kg of ketamine compared to IV opioids for acute pain in the ED, with second pain scale measurements performed within 60 minutes of initial measurement. Three studies were included, for a total of 261 patients. Ketamine was found to be statistically noninferior to opioids (in morphine equivalents).

QUALITY ASSESSMENT
This was a high-quality systematic review. An exhaustive search of the literature was performed, and due to the strict inclusion criteria, only three studies were included, but this allowed for close comparison as the trials were similar in ketamine and opioid dosing and did not include pediatric patients, patients who had coadministration of other medications that could confound the results, and alternate routes of administration. The authors performed two preplanned sensitivity analyses, the first of which could not be performed because of incomplete data and the second of which supported the results of the review. Adverse

From LewisGale Medical Center (CH), Salem, VA; Markham Stouffville Hospital (JM), Markham, Ontario, Canada; the University of Calgary (CB), Calgary, Alberta, Canada; and the University of Western Ontario (WKM), Goderich, Ontario, Canada.
Received October 30, 2018; accepted October 31, 2018.
The authors have no relevant financial information or potential conflicts to disclose.
Supervising Editor: Esther K. Choo, MD, MPH.
Address for correspondence and reprints: Corey Heitz, MD; e-mail: coreyheitzmd@gmail.com.
ACADEMIC EMERGENCY MEDICINE 2019;26:946-948.
reactions were summarized but not analyzed due to insufficient data.

KEY RESULTS

There were 261 patients included from three studies. All studies used similar dosing of ketamine, and all use the same dosing of morphine. Ketamine was found to be noninferior to morphine for acute pain.

- **Primary outcome**—Numerical rating scale (NRS) or visual analog scale (VAS) within 60 minutes:
  - Pooled estimate of difference between ketamine and morphine equivalents was 0.42 (95% CI = –0.70 to 1.54). The lower limit of –0.70 is less than the lower end of inferiority established at –1.4.

- **Secondary outcome**—Adverse events:
  - No severe adverse events were reported.
  - Higher rates of nonsevere adverse events were seen with ketamine.

AUTHORS’ COMMENTS

Low-dose ketamine, given intravenously, is similarly effective as opioids dosed at 0.1 mg/kg at treating acute pain in the ED. With longer measurement periods and increased patient numbers, differences in duration of effectiveness and more precise outcome estimates may be possible.

TOP SOCIAL MEDIA COMMENTARY

**Josh Farkas (@PulmCrit):**
We need to seriously reconsider where opioids should be on the analgesic ladder. In critically ill patients, might be time to demote them to third place (behind acetaminophen and ketamine). https://emcrit.org/pulmcrit/analgesic-ladder/

**Sergey Motov (@painfreeED):** Wholeheartedly agree that it’s time to modify analgesic ladder with respect to role of opioids for any painful conditions but especially for critically ill patients. Non-opioid analgesic modalities first and opioids to rescue...

**Evan Schwarz (@TheSchwarziee):** I’d agree that in the critically ill depending on the physiology, ketamine likely has some benefit over at least certain opioids. If we weren’t on shortage & could do a drip, I’d use it more in them.

**Michael Reindl (@MichaelReindl):**
In Germany iv ketamine is in preclinical #ems widely used an accepted but rarely seen in the #ED. In this setting ketamine is more often seen for sedation and anesthesia

TWITTER POLL

Ketamine as an opioid alternative in acute pain
Systematic review & meta-analysis: RCTs 1946-2017; adult ED patients

Paper-in-a-pic from Kirsty Challen, @EMOttawa

TAKE-TO-WORK POINTS

Ketamine is a reasonable alternative to opioids for acute pain in the ED. Dosing of <0.5 mg/kg appears to be similarly effective to 0.1 mg/kg morphine equivalents. Increased, short-lived side effects have been reported. Longer-term effectiveness, or analgesic options at discharge, were not assessed.
References


Aspirin for Preventing a First Heart Attack or Stroke

Kristopher Roach, MD¹, Michael Ritchie, MD¹, and Shahriar Zehtabchi, MD²

Summary heading No overall benefit for primary prevention
Benefits in NNT No deaths were prevented
1 in 333 avoided a nonfatal heart attack
Unclear if ischemic strokes avoided
Benefits in percentages No deaths were prevented
0.30% lower risk of heart attack
Unclear if ischemic strokes avoided
Harms in NNT (NNH) 1 in 250 suffered a major bleeding event
Harms in percentages 0.40% higher risk of major bleeding
Efficacy endpoints Death, heart attack, stroke, measured over 5–7 years
Harm endpoints Major bleeding events, hemorrhagic strokes
Who was in the studies Approximately 164,000 subjects at varying risk for cardiovascular disease

Cardiovascular disease (CVD) is a major cause of death worldwide. Aspirin inhibits platelet aggregation, which reduces clot formation. Thus, aspirin can help prevent cardiovascular problems caused by blood clots, but it can also increase risk of bleeding. For persons with known CVD, the beneficial effect of aspirin use for preventing cardiovascular events outweighs the harmful side effects (e.g., bleeding).

Mortality
For the outcome of greatest interest, mortality, the older USPSTF analysis found a statistically significant overall benefit for aspirin at all doses, but no statistically significant benefit for aspirin at low doses.
The two updated reviews by Mahmoud et al.3 and Zheng et al.4 found no benefit in mortality regardless of the dose of aspirin (Table 1).

### Nonfatal Heart Attacks

All three reviews found statistically significant relative reductions of 15% to 20% in nonfatal heart attacks from the use of aspirin, with NNT values of 250,2,3 and 361.4

### Stroke

The older USPSTF analysis and the newer review by Mahmoud et al. did not find that aspirin prevented ischemic stroke, combined fatal and nonfatal.2,3 Only the systematic review by Zheng et al.4 showed a small reduction in risk of ischemic stroke in patients allocated to the aspirin group, with an NNT value of 625 (Table 1).

### Major Bleeding

All three reviews found statistically significant relative increases of 30% to 50% in major bleeding events with NNH values of 142 to 357 (stratified by baseline risk),2,3 and 213.4 Major bleeding was defined differently within each trial and could have included intracranial hemorrhage, major gastrointestinal bleeding, ocular bleeding, major epistaxis, or any extracranial bleeding requiring transfusion or hospitalization.

### Subgroups

Harms outweighed benefits in all three reviews when analyzing all patients, with no mortality reduction. The USPSTF,2 however, projected one subgroup that may have benefits outweighing harms. USPSTF examined different age groups and divided each age group into different CVD risk groups using the AHA calculator to predict the 10-year risk of a cardiovascular event. In a computer model by Health Partners Institute,8 50- to 59-year-old patients with >10% risk over 10 years saw a projected benefit from reduction in nonfatal heart attacks that outpaced the increase risk of major bleeding. No other subgroup realized a net overall benefit.

### Dosing

Optimal dosing of aspirin is unknown and some data suggest weight-based dosing.9 In the USPSTF analysis, low-dose aspirin (100 mg or less) was not associated with lower mortality but higher doses were, while the opposite association was true for nonfatal stroke. Aspirin dosing was not addressed in the Mahmoud et al.3 and Zheng et al.4 reviews. The conflicting results for dose finding in the older report and lack of newer analysis in the updated reviews makes dosing risk-benefit analysis unclear at this time.

### CAVEATS

The older USPSTF report has limitations. The finding of overall benefit for 50- to 59-year-old patients is a computer projection based on a statistical model.10 The model uses data from subgroups across several trials and applies the benefits found with aspirin to a hypothetical person—in this case, a 50- to 59-year-old American male—with a baseline cardiovascular risk estimated using the AHA risk calculator. Unfortunately, that calculator substantially overestimates risk (by anywhere from 20% to 100% or more).11-13 Given the razor-thin benefit margins found, any overestimate of baseline risk would convert the finding of overall benefit to a finding of overall harm. Moreover, the model is out of date as three new large randomized controlled trials have been published since its release.

### Table 1

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Absolute Risk Difference</th>
<th>Relative Risk (95% CI)</th>
<th>NNT/NNH</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-cause mortality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahmoud et al.3b</td>
<td>0.10</td>
<td>0.98 (0.93–1.02)</td>
<td>No benefit</td>
</tr>
<tr>
<td>Zheng et al.4</td>
<td>0.13</td>
<td>0.94 (0.88–1.01)</td>
<td>No benefit</td>
</tr>
<tr>
<td>Death from cardiovascular causes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahmoud et al.3b</td>
<td>0.08</td>
<td>0.92 (0.83, 1.01)</td>
<td>No benefit</td>
</tr>
<tr>
<td>Zheng et al.4</td>
<td>0.07</td>
<td>0.94 (0.83–1.05)</td>
<td>No benefit</td>
</tr>
<tr>
<td>Heart attack</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahmoud et al.3b</td>
<td>0.3</td>
<td>0.82 (0.71–0.94)</td>
<td>NNT: 333</td>
</tr>
<tr>
<td>Zheng et al.4</td>
<td>0.28</td>
<td>0.85 (0.73–0.99)</td>
<td>NNT: 361</td>
</tr>
<tr>
<td>Ischemic stroke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahmoud et al.3b</td>
<td>0.10</td>
<td>0.94 (0.86–1.02)</td>
<td>No benefit</td>
</tr>
<tr>
<td>Zheng et al.4</td>
<td>0.16</td>
<td>0.81 (0.76–0.87)</td>
<td>NNT: 625</td>
</tr>
<tr>
<td>Major bleeding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahmoud et al.3b</td>
<td>0.40</td>
<td>1.47 (1.31–1.65)</td>
<td>NNH: 250</td>
</tr>
<tr>
<td>Zheng et al.4</td>
<td>0.47</td>
<td>1.43 (1.30–1.56)</td>
<td>NNH: 213</td>
</tr>
<tr>
<td>Intracranial bleeding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahmoud et al.3b</td>
<td>0.10</td>
<td>1.33 (1.13–1.58)</td>
<td>NNH: 1,000</td>
</tr>
<tr>
<td>Zheng et al.4</td>
<td>0.11</td>
<td>1.34 (1.14–1.57)</td>
<td>NNH: 909</td>
</tr>
</tbody>
</table>

NNH = number needed to harm; NNT = number needed to treat. The data from USPSTF analysis2 is not reported here because the most recent trials were not included. As of January 2019, the USPSTF analysis was not updated. NNTs in the review by Mahmoud et al. are based on 6.6 years of follow-up; we adjusted these to 5 years for comparison across reviews.

(100 mg or less). The two updated reviews by Mahmoud et al.3 and Zheng et al.4 found no benefit in mortality regardless of the dose of aspirin (Table 1).
There is debate across reviews about the definition of “primary prevention.” In the ETDRS study,14 half of the patients had known CVD, and all patients in the POPADAD study5 and the AAA study had arterial disease. The USPSTF and older meta-analyses included these studies, Mahmoud et al. excluded them, and Zheng et al. included ETDRS only. These patients constitute fewer than 9,000 subjects (5%) of the total patients analyzed. The varying inclusion of the three studies who enrolled patients with apparent CVD resulted in 96% overlap between the review by Mahmoud et al. (157,000 subjects) and the review by Zheng et al. (164,000 subjects). The inclusion of some patients with CVD in the study by Zheng et al. may explain why Zheng et al. found a statistically significant small stroke prevention benefit, while Mahmoud et al. did not (Table 1).

Less clear are differences in the two new meta-analyses regarding the heterogeneity of the trials. Zheng et al. found no heterogeneity (I² = 0%) for heart attack reduction, while Mahmoud et al. concluded “a high degree” of heterogeneity (I² = 67%). These differences in the results could be from existing heterogeneity among different trials that were included (11 trials in the analysis by Mahmoud et al. and 13 trials in the analysis by Zheng et al.), but more likely reflects differences in the methodology of the meta-analyses.

Regardless of these minor differences, both updated reviews found no consensus finding of benefits outweighing harms in patients regardless of CVD risk, contradicting the statistical model projection from the USPSTF report that high-risk subgroups may benefit specifically. While overall benefit may be true in secondary prevention in high-risk patients, results from these primary prevention reviews are uniform in the benefits not outweighing the harms. In fact, the recent ASPREE trial found that high-risk patients had increased harm compared to low-risk patients.7 The proportion of high-risk patients was highest among the newer studies (25%–30%), and no statistically significant benefit was found in any outcome—only harms from mortality and bleeding.5–7

Further concern regarding the USPSTF report reliance on the AHA calculator to project a subgroup benefit is that coronary events occurred at less than one-third the predicted rate in the ASPREE trial7 and less than half predicted in the ARRIVE5 and ASCEND6 trials. In practice, clinicians often apply the AHA calculator used in trials and estimate risk conservatively. If faulty calculators and conservative gestalt lead to overestimation of risk, and clinicians wrongly believe that higher risk means greater benefit from aspirin, overall harm due to aspirin prescribing for primary prevention is probably widespread.

By the same token, overestimation means true ultra-high-risk patients (10-year risk of CVD > 30%) were potentially misdiagnosed as having preexisting CVD and were not enrolled in primary prevention trials and may potentially benefit. Future studies should tackle this question.

Last but not least, patient preference is an important factor for making the decision regarding aspirin use for primary prevention of CVD. Some patients may value avoiding nonfatal heart attacks or possibly avoiding ischemic strokes as being worth the increased risk of major bleeding.

We chose a red color recommendation (no benefit) because of consistent findings of harm outweighing benefit. We considered black (harmful) but recognize that there may be subgroups studies that will identify patients who can benefit.

References

Early Endovascular Thrombectomy for Large-vessel Ischemic Stroke Reduces Disability at 90 Days

Mona Al Banna, MB, BCh, BAO, MSc(Res) and Christopher D. Streib, MD, MS

NARRATIVE

Anterior circulation large-vessel occlusion (LVO) of the internal carotid or middle cerebral artery is one of the most devastating ischemic stroke subtypes. Prior to 2015, evidence supporting endovascular thrombectomy for acute ischemic stroke was limited.1,2 This was due to multiple factors, including low recanalization rates with previous-generation thrombectomy devices, inadequate neuroimaging inclusion criteria (patients enrolled in trials lacked target LVO and/or had large preexisting core infarcts), and selection bias—patients considered most likely to benefit from thrombectomy underwent the procedure outside of clinical trials.1,2 However, five randomized controlled trials3–7 published in 2015 established that endovascular thrombectomy significantly reduces disability in acute ischemic stroke when performed with stent retrievers in the setting of anterior circulation LVO and minimal core infarct burden. Four of these five trials were terminated early due to overwhelming treatment benefit, limiting analysis of secondary outcomes. A patient-level meta-analysis of all five trials was published in The Lancet in 2016.8 All subjects included in the meta-analysis had an anterior circulation LVO stroke. The median (interquartile range [IQR]) age was 68 (57–77) years with median (IQR) National Institutes of Health Stroke Scale (NIHSS) 17 (14–20) and Alberta Stroke Program Early CT Score (ASPECTS) 9 (7–10). Median (IQR) time to recanalization of the
ocluded vessel with endovascular thrombectomy was 285 (210–362) minutes.

The aggregate clinical trial data strongly favor endovascular thrombectomy over medical management for anterior circulation LVO with a number needed to treat (NNT) of 2.6 for one additional patient to achieve improved functional outcome, defined as improvement by at least one level on the modified Rankin Scale (mRS) at 90 days (adjusted odds ratio \( \text{aOR} = 2.49 \), 95% confidence interval \( \text{CI} = 1.79–3.53 \); please refer to Figure 1A for the mRS). Additionally, for every five patients treated with endovascular thrombectomy, one additional patient achieved functional independence (mRS 0–2) at 90 days (46% vs. 26.5%, absolute risk difference = 19.5%, \( \text{aOR} = 2.71 \), 95% CI = 2.07–3.55; Figures 1B and 1C). These findings were consistent across all subgroups (age, sex, NIHSS, ASPECTS, site of intracranial occlusion, intravenous tPA eligible or ineligible, and time from onset to randomization), suggesting that the benefit of thrombectomy is generalizable to a broad spectrum of patients with anterior circulation LVO.\(^8\)

Importantly, this strengthens the evidence that endovascular therapy should not be withheld on the basis of factors such as age or tPA eligibility.\(^8\)

**CAVEATS**

All five clinical trials were conducted at experienced, high-volume tertiary stroke centers. Therefore, these results may not be replicated at clinical sites without similar levels of stroke and neurointerventional expertise. In addition, these results cannot be extrapolated to all stroke patients. The benefit of endovascular treatment for stroke patients with large infarcts (representing irreversible ischemic injury), but substantial viable ischemic tissue remains to be determined. Additional stroke subgroups for which thrombectomy is yet to be definitively

![Figure 1. (A) The modified Rankin scale; scores on the modified Rankin Scale at 90 days in the intervention and control groups in (B) ordinal shift analysis and (C) dichotomized to dependent and independent functional status.](image-url)
studied include posterior circulation stroke, distal vessel occlusion, patients with minor deficits, and those with premorbid disability. Notably, two recent clinical trials have demonstrated a benefit for delayed thrombectomy (6–24 hours from stroke onset) in patients selected using more stringent clinical and advanced neuroimaging criteria.9,10 These highly selected patients represent a more homogenous stroke population distinct from the trials in the meta-analysis of Goyal et al.8

There was no significant difference in symptomatic intracerebral hemorrhage (aOR = 1.07, 95% CI = 0.62–1.84), parenchymal hematoma type 2 (aOR = 1.04, 95% CI = 0.63–1.72), and mortality (aOR = 0.73, 95% CI = 0.47–1.13) between the thrombectomy group and the medical group.8 One plausible explanation for this finding is that, despite the associated procedural risk, thrombectomy patients ultimately have smaller ischemic stroke volumes than nonthrombectomy patients leading to lower rates of postprocedural spontaneous hemorrhagic transformation.6

The mRS, which is commonly used in stroke clinical trials, is a discrete functional outcome score with high inter-rater reliability and little room for subjective interpretation.11 Although participants and their families were not blinded to the treatment arm, to minimize bias, each individual trial employed a PROBE (prospective randomized open blinded endpoint) design, in which all outcome assessors remain blinded. Additionally, all outcomes were adjudicated by study personnel certified in scoring the mRS.

In acute stroke trials, functional outcome can be analyzed by binary analysis (dichotomized ordinal outcome scales, i.e., “good” vs. “poor”) or by shift analysis (which evaluate outcomes over the entire scale range). One limitation of the mRS shift analysis employed in these trials is that it values each mRS strata equally. However, studies have shown, for example, that most people attach a higher value to an improvement from mRS 3 to 2 than to an improvement from mRS 1 to 0. This was accounted for in the DAWN trial,9 which employed a weighted shift analysis as one of the co-primary endpoints. Despite this limitation, the shift analysis has become convention in stroke trials because it provides information across a wider range of outcomes and, hence, is considerably more clinically relevant than traditional dichotomized outcomes.

In conclusion, we assigned a color recommendation of green (benefit > harm) to this intervention because of evidence of patient-centered benefit for early endovascular thrombectomy in anterior circulation LVO stroke and absence of significant harm. This benefit was evident across a wide range of ages and was present irrespective of tPA eligibility.

References

The Folly of the R: A Case Study

INTRODUCTION

The “You should apply for this” message from a colleague popped up on my work e-mail seven months ago. Reading through the funding announcement, I thought “Yeah, this is pretty much exactly in line with what I’m doing. But ugh!! Do I want to put myself through this again?”

For the sake of clarity, I have entered the autumn of my career, secure as a Professor of Emergency Medicine (step 4), with no major push to get funded. I have always been proud of my .400 grant funding batting average. I opened up hot with a solid first-round NIH-funded grant, followed by a similarly scored and funded second proposal. But my last three submissions resulted in the dreaded You seem like a nice guy and we don’t want to hurt your feelings 40–45 score. Another few failures would plant me firmly on the Mendoza line.

After an agonizing month of “Do I really want to do this?” I decided to take the plunge. What follows is a time record of my activities during the 6 months that it took me to complete the R series grant proposal. My objective in creating this record was to inform my next grant application decision by logging my hours in real time.

METHODS

My data collection methods were simple. I set up an Excel spreadsheet with days 1 to 175 as rows and columns consisting of grant proposal activities. I logged my hours—rounded to the nearest 15-minute interval—daily under the most appropriate column. I included any time that was expressly devoted to work on the grant and excluded travel times to meetings and the many other hours in which I contemplated my research strategy during other activities (runs and other exercise). In terms of statistical analysis, I contemplated using Stata v27 to perform a semiparametric bootstrap logistic regression but ultimately settled on simple addition, subtraction, multiplication, and division on my iPhone 7 calculator.

RESULTS

Over the 175-day (25-week) time period, I spent 332.75 hours or 13.3 hours per week on my grant proposal (Table 1). During the first month, I averaged approximately 4 hours per week but this ramped up to nearly 20 hours per week over the last 3 months. The majority (78%) of those hours were spent writing

<table>
<thead>
<tr>
<th>Days</th>
<th>Reading, Background</th>
<th>Assembling team</th>
<th>Writing first draft</th>
<th>Letters of support</th>
<th>Budgeting</th>
<th>Editing and Revising</th>
<th>All Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-30</td>
<td>7.5</td>
<td>2.5</td>
<td>11.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21.5 (6.5%)</td>
</tr>
<tr>
<td>31-60</td>
<td>4.25</td>
<td>7.75</td>
<td>16.5</td>
<td>2.25</td>
<td>0</td>
<td>0</td>
<td>30.75 (9.2%)</td>
</tr>
<tr>
<td>61-90</td>
<td>3</td>
<td>6.25</td>
<td>34.25</td>
<td>4.5</td>
<td>0.75</td>
<td>0</td>
<td>48.75 (14.7%)</td>
</tr>
<tr>
<td>91-120</td>
<td>1.75</td>
<td>3</td>
<td>59</td>
<td>14.75</td>
<td>2.5</td>
<td>2.5</td>
<td>83.5 (25.1%)</td>
</tr>
<tr>
<td>121-150</td>
<td>2.25</td>
<td>0</td>
<td>50.5</td>
<td>3</td>
<td>3.5</td>
<td>18</td>
<td>77.25 (23.2%)</td>
</tr>
<tr>
<td>151-175</td>
<td>3.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>67.25</td>
<td>71 (21.3%)</td>
</tr>
<tr>
<td>1-175</td>
<td>22.5 (6.8%)</td>
<td>19.5 (5.9%)</td>
<td>171.75 (51.6%)</td>
<td>24.5 (7.4%)</td>
<td>6.75 (2.0%)</td>
<td>87.75 (26.4%)</td>
<td>332.75</td>
</tr>
</tbody>
</table>

The authors have no relevant financial information or potential conflicts of interest to disclose.
and revising, followed by gathering letters of support (7%), background reading (7%), and assembling the research team (6%).

**DISCUSSION**

Although other investigators have reported theoretical models assessing cost/benefit ratios for grant writing,1 this is the first real-world time study of grant proposal preparation. As an investigator whose research is centered on determining ways to deliver efficient, high-yield health care, my results were disturbing—to put it mildly. From a straight economics vantage point, my efforts toward this grant made absolutely no sense. To monetize my total of 333 hours, I reviewed eight of those pesky e-mails for outside emergency medicine moonlighting and calculated a mean pay per hour of $231.25. If I had moonlighted this amount of time, I would have earned nearly $77,000 in gross income—the equivalent of the same FTE of salary I was requesting for year 1 in my proposal. Considering the 9% 2018 NIH payline for established investigators, moonlighting for guaranteed money makes this 1 of 11 prospects in my proposal. Considering the 9% of NIH funding have not similarly evolved, remaining virtually unchanged for the past half-century. Instead of fostering collaboration toward common research aims between institutions, the current funding system promotes secrecy with work in isolated silos, which in turn leads to inefficient, redundant research practice. I was once recruited to be a site investigator for grant proposals by two separate, experienced research teams at sister institutions located 300 miles from each other. As I heard the second group’s pitch that was nearly identical to the first, I felt extremely conflicted, knowing that I could not reveal the overlapping research plans to either group. I was happy but disconcerted when both proposals were granted on the same funding round. Although both culminated in publications in a high-impact journal, the two groups came to nearly the same conclusions with only slight, nuanced differences. It is likely that a combined team proposal and approach would have produced the same results at nearly half the effort and cost.

The primary limitation of this narrative is that it is essentially a case report—a sample size of one that may not reflect the time and effort put forth by other, less plodding investigators toward grant proposals. However, I conducted an informal poll of six other seasoned investigators at other institutions and they estimated between 175 and 300 hours of time for an R grant. I suspect that their response was subject to considerable recall bias favoring underestimation, but even the lowest number represents a daunting amount of effort.

Like our American health care funding system, our primary medical research funding mechanism is broken—investigators are encumbered by an archaic grant application process that stifles collaboration and precludes informed choice about investing their time. Noting that investigators spend as much time writing proposals as they do on their actual projects, other investigators have similarly noted the gross inefficiencies of our current mechanism.1,2 While many private industry groups have readily adopted digital sharing and collaboration models that have sparked a renaissance of biotechnological advances, the central concepts underlying NIH funding have not similarly evolved, remaining virtually unchanged for the past half-century. Instead of fostering collaboration toward common research aims between institutions, the current funding system produces the same results at nearly half the effort and cost.

Provision of information to allow patients to weigh risks and benefits is a cornerstone of medical practice. Yet, when embarking on laborious grant proposals, prospective investigators are armed with too little information to adequately render informed decisions. At early points of project development (before diving into hundreds of hours of work), investigators should be provided with the answers to at least three questions: 1) Am I on the right track with my proposal aims? 2) What are my chances of getting funded if I develop a full proposal? and 3) Is there another group with whom I should collaborate on this project? Out of fear of introducing bias or providing unfair advantage, grant program officers offer only very cursory guidance toward question 1; the only readily information available toward question 2 is the extremely vague prior year’s NIH payline; and there’s really nothing for question 3—if anything, the current system rewards secrecy instead of idea sharing. For limited institutional submissions, many universities have developed succinct, preproposal applications that answer all three of these questions at early stages, thereby eliminating redundancies, strengthening collaborative proposals, and saving countless hours of futile effort. Why can’t NIH enact a similar practice? If, after day 10 of preparation, I was told that my proposal had a less than 5% chance of being funded, I definitely would not have continued on to day 11. If afforded such information, I imagine that many other investigators would make similar choices, saving them hundreds of grinding hours that could be applied to other research—or even better, to make time for their outside lives.

In the end, national research funding should seek not only to fund the most promising grant proposals...
but also to preserve the well-being of investigators. A trial of a more transparent, informative early-action grant review system would be one laudable step toward that goal.

CONCLUSIONS

This real-time case study has informed my decision to pass on the funding announcement that arrived in my inbox today. After all, many landmark papers have been published without NIH funding.3–7

Robert M. Rodriguez, MD
(Robert.rodriguez@ucsf.edu)
Department of Emergency Medicine,
The University of California San Francisco School of Medicine, Zuckerberg San Francisco General Hospital, San Francisco, CA

Supervising Editor: John H. Burton, MD

References

Love, Sanity, or Medical School: A Memoir. 
By Stephanie Benjamin, MD. 
Seattle: Amazon, 2019; 355 pp; $14.99 (softcover).

Love, Sanity, or Medical School: A Memoir is the personal journal of a third-year medical student juggling the core clinical rotations, while trying to maintain a social life and sanity. The author became curious about the stress, exhaustion, and confusion described by recent third-year medical student veterans and decided to journal her experiences to find how these things occurred and to what extent they would happen to her. The author succeeded at her intent, to give a first-hand account of third year of medical school in an uncensored point of view.

The intended audience is the general public. Readers may find some of the activities discussed surprising or relatable—completely dependent on their role whether learner, administrator, patient, or educator. The approach is through a self-identified gonzo-style type of journalism which is a highly personalized style, written without claims of objectivity, and tells the author’s story in a first-person narrative.

This softcover book is organized into 19 chapters with three intersessions and an epilogue. Chapters are written in the order of each rotation as the author experiences them. The chapter begins with a suggested drink and song pairing that sets the tone for the rotation ahead. The music selections added a fun element to reading the book. The songs containing words were a little distracting, but the solely instrumental songs created a nice back drop to the chapters. Although the book was written as a diary of sorts, the time stamps before each activity sometimes broke the flow of reading.

This novel gives a personal account of a medical student bound for emergency medicine (EM) early in her training. It’s an interesting look at what may influence a medical student to pursue the specialty and how they may feel and act in other rotations before committing to EM. While this is a memoir intended to give a raw look into the life of a medical student, her emotions and thoughts are sometimes abrasive and could easily rub a non-EM reader the wrong way.

After one author initially read this story, it was apparent that this review would be much more meaningful if we included several different opinions in this review to provide a fuller perspective of this book. Therefore, we have put together three authors for this review—an EM resident from a similar medical school, a senior academic emergency physician, and an administrator who has worked with medical learners but has no formal medical school training.

From a current EM physician, not far removed from medical school, I believe the author provides a rather accurate depiction of the third year of medical school. The thoughts and emotions are brutally honest, and she doesn’t try to temper down any of the more unpleasant aspects. Her frustration with patients, her superiors, and herself are all real emotions that medical students experience but don’t often verbally express. The third year is tough and arguably the most awkward, uncomfortable, and difficult transition in a doctor-in-training’s career. The author does a nice job illustrating this by not holding anything back and allowing her true emotions and feelings be seen. This is the year that medical students, once sheltered in the comfort of their books and lectures, are caring for patients for the first time and it’s not a pretty transition.

The personal life of the author is not perfect. She struggles with moving on from a failed relationship and dating new people. Medical school is tough not just academically but on one’s personal life as well, with divorce, breakups, and heartbreaks rampant. The author’s failed long-term relationship is an important and unique aspect of the story and something not often discussed.

While the author tries to be open minded toward other specialties, it is clear very early on that she has one specialty in mind: emergency medicine. One weakness of this memoir is her attitude toward other specialties is sometimes harsh and can be a turnoff for readers, especially since in EM, we are involved with an aspect of every specialty.
From the view of an administrator with no clinical background, I found this book enlightening in several aspects. The descriptions and explanation of medical terminology and procedures were helpful for someone without medical experience. As a patient, I now appreciate the wide range of training that every physician receives. I will, however, be checking the name badges of anyone treating me since the author (and other students) felt comfortable performing a procedure on a patient despite direct instructions not to do so!

The author introduces a new way of thinking to an administrator. For example, one would hope a third-year medical student would take the stance of utilizing every opportunity to learn—but the system or personal experience may affect this in unexpected ways. Overwhelmed learners may have a point in a day or rotation where they believe they have learned enough and are backing away to focus on other parts of their lives. Another example is found in the frustration of the author when she felt her diagnoses and treatment options were ignored or disregarded. This seemed to decrease her respect toward her supervisors and attending physicians and may have been avoided by faculty awareness and a different style of precepting. Yet another example was when the author and many other students did not complete an assignment because it was “buried in a 30-page syllabus.” The message seems that many learners today will not read a syllabus because it appears to be without explicit focus. This may be to the dismay of educators who write and distribute syllabi, so reading this book may help guide administrators to create what students feel may be more relatable content.

The author initially displays a lack of self-awareness when she tends to negatively discuss the pace, patients, precepting, and procedures of every other specialty, but becomes defensive and offended when she encounters someone that does not share her views regarding EM. Her candid summary of post–third-year life displays a maturity and self-awareness of her thoughts and actions during her third year, which was refreshing and encouraging to see unfold.

Decades have passed since my third year of medical school and, as a physician, I still found this book fascinating. The fear of not knowing anything and providing no value to patients is a universal, timeless feeling of medical students caring for patients the first time. The thrill of making a difference by caring enough about a patient during their illness to being present and listening is an equally communal reaction. Yet the descriptions of her experiences were sometimes new for me since much in the training has changed from “the good old days.” So as an academic leader in a medical school, reading this text was a helpful reminder of the raw, unprocessed feelings of a junior learner who is just being exposed to the intimate bustle that create our careers.

When I look at this as a faculty member teaching and administrating educational activities for students and residents, it was very interesting to read about the perception of the process from the author’s experience. Although initially tempted to describe this as generational differences, I realized the change in complexity of the patients, the laws that have decreased the medical students’ importance in documentation, and the patients’ more sophisticated understanding of the process has changed to create a new set of challenges for students.

One area of disappointment in the story was that there seemed to be less discussion about how she felt regarding patients than how she was feeling about her personal life. Given that her only information was the narrative she recorded at the time, it might have been nice for her to have tried to provide some more perspective at the end of each rotation. Her retrospective discussion in the final pages provide some of this but were a long time coming while reading the book.

This book could be helpful for readers in a leadership role of the upcoming third-year and fourth-year medical student, as well as today’s resident physicians. Seeing the world through this view may help in understanding the thought processes of the current genre of learner. The different work ethic, and the focus on wellness may be in contrast to the thinking of an administrator or faculty member. Addressing changes in curriculum development, faculty will be more conscious of the impact of the joys, frustrations, and stresses after reading these chapters.

The family of a third-year medical student would find this book highly enlightening and terrifying at the same time. The coping mechanisms of this student and her friends seemed to tend toward the self-destructive between the relationships and the substance use—all while commenting on how her patients should be making better decisions about their own health. The support of family and friends seems crucial to the success of this or other third-year medical student. Any relative of a student will understand the importance of keeping close communication so support can be readily available to address concerns including stress, burnout, and poor personal decisions.
An easy-to-read and fascinating book, *Love, Sanity, or Medical School: A Memoir* is highly recommended reading to a wide demographic of readers. There is much to be learned in these pages, and each reader can take away something different. This book provides the means of having a perspective on an important time during the training to become a physician, which will be an enjoyable read for anyone interested.
He was just another old man in a busy emergency department during my clinical attending shift. An ambulance brought him in after he fell in the grocery store. My senior resident Carlo and I were called to see him because he was struggling with a C-collar that he didn’t need. He clearly broke his wrist even though he said it didn’t hurt. He proved his point to us by moving his broken wrist back and forth while we winced and splinted his hand. There were bruises on his face. It looked like he broke his jaw as well even though he spoke clearly. We put him in a room and ordered his imaging.

His wrist was broken. Orthopedics reduced the fracture and put him in a cast. His jaw was broken. I paged plastics surgery for his facial trauma. The old man didn’t ask for anything. I offered him pain medicine, water, a phone call, and he declined politely. He was just another uneventful patient who fell.

... Until he started to get himself dressed. Carlo and I saw him from afar. We wondered where he was going; he still needed care of his badly broken jaw. “Where are you going?” we asked. “I have to get home to my wife,” he answered in immaculate Spanish. “We live together; I can’t leave her alone. She’s waiting for me to come home,” he continued, as he pulled his pants up and buckled his belt.

“Where do you live?” Carlo asked. The old man could not give us an address.

“How will you get home?” Carlo asked.

“I’ll take the bus!” he replied.

“Which bus?” I asked.

“Any bus,” he answered.

“Do you have any family?” I pursued, knowing that we could not let the old man leave alone.

“My granddaughter!”

“What’s her phone number?” I asked. He didn’t know.

“Do you know where you are?” I asked. He didn’t know.

“I have to get home to my wife,” he repeated. “She’s waiting for me to come home.”

We put the old man back into his hospital gown and placed him on close observation. Fortunately, the old man had been to our hospital before and there was a phone number and an address. I called the number. Someone answered immediately. It was the old man’s granddaughter. I told her that he was with us in the ED and informed her of his injuries. She confirmed his living situation. He lived with his wife and he took care of her.

“Umm, he seems to have some dementia,” I said over the phone.

“Yeah, we think so,” answered the granddaughter.
We spoke about the old man’s disposition. Would you prefer to have him go home or keep him in the hospital? Who’s going to take care of him if he goes home? The granddaughter said they had already brought the patient’s wife to her home when they heard about the accident. They could take care of him at her home as well, she said.

“Besides, if you keep him in the hospital, he’ll keep trying to go home to his wife,” she said, echoing the old man’s very own words.

Wow, I thought. The old man’s love and devotion to his wife was so entrenched even the onset of dementia could not erase her from his mind. She’s waiting for me to come home, I heard again. We all have notions of enduring love—finding the one meant-to-be despite life’s obstacles, faith in our love in a world of hurt, togetherness to the bitter end. There was something incredibly loving in the old man’s dedication to his wife. It wouldn’t be so bad, I thought, finding someone you loved so deeply that you would remember her even when you didn’t know your way home. You still knew in your heart that you loved her and you had to go to her because she was there waiting for you to come home. We should all be so lucky.

Siu Fai Li, MD
(litfreak73@gmail.com)

Department of Emergency Medicine, Jacobi Medical Center, Bronx, NY

Supervising Editor: Brian Zink, MD