EDITOR’S PICK

Access to Federally Qualified Health Centers and Emergency Department Use Among Uninsured and Medicaid-insured Adults: California, 2005 to 2013
Julia B. Nath, Shaughnessy Costigan, Feng Lin et al.

SYSTEMATIC REVIEWS (WITH OR WITHOUT META-ANALYSES)

Prognostic Accuracy of the HEART Score for Prediction of Major Adverse Cardiac Events in Patients Presenting With Chest Pain: A Systematic Review and Meta-analysis
Shannon M. Fernando, Alexandre Tran, Wei Cheng et al.

ORIGINAL CONTRIBUTIONS

A Case–Control Study of Sonographic Maximum Ovarian Diameter as a Predictor of Ovarian Torsion in Emergency Department Females With Pelvic Pain
Gavin Budhram, Tala Elia, Jeff Dan et al.

Adoption and Utilization of an Emergency Department Naloxone Distribution and Peer Recovery Coach Consultation Program
Elizabeth A. Samuels, Janette Baird, Eunice S. Yang et al.

Impact of Medical Scribes on Provider Efficiency in the Pediatric Emergency Department

Low-dose Magnesium Sulfate Versus High Dose in the Early Management of Rapid Atrial Fibrillation: Randomized Controlled Double-blind Study (LOMAGHI Study)
Wahid Bouida, Kaouthar Beltaiief, Mohamed Amine Msolli et al.

Injury-based Geographic Access to Trauma Centers
Ran Wei, N. Clay Mann, Mengtiao Dai et al.

Contents continued inside.
Referring Hospital Characteristics Associated With Potentially Avoidable Emergency Department Transfers
Kristin N. Ray, Jennifer R. Marin, Joyce Li et al. 205

Discriminatory Value of the Ascending Aorta Diameter in Suspected Acute Type A Aortic Dissection
Dustin G. Mark, Justin A. Davis, Yun-Yi Hung et al. 217

EVIDENCE-BASED DIAGNOSTICS
Accuracy of Dementia Screening Instruments in Emergency Medicine: A Diagnostic Meta-analysis
Christopher R. Carpenter, Jay Banerjee, Daniel Keyes et al. 226

HOT OFF THE PRESS
Hot Off the Press: Validation of the Pediatric NEXUS II Head Computed Tomography Decision Instrument for Selective Imaging of Pediatric Patients with Blunt Head Trauma
Corey Heitz, Justin Morgenstern, Christopher Bond et al. 246

THE BIROS SECTION ON RESEARCH ETHICS
Physician-assisted Death: Ethical Implications for Emergency Physicians
Arthur R. Derse, John C. Moskop, Norine A. McGrath et al. 250

THE BRASS TACKS: CONCISE REVIEWS OF PUBLISHED EVIDENCE
Epinephrine for Out-of-hospital Cardiac Arrest
Kyle Kelson, Ian S. deSouza 256

Video Laryngoscopy vs. Direct Laryngoscopy
Abdullah Bakhsh, Michael Ritchie 259

COMMENTARY–INVITED
ED Chest Pain Rules: Follow Your HEART?
Michael A. Kohn, Andrew Worster 261

COMMENTARIES - UNSOLICITED
Bedside End-tidal Carbon Dioxide in Evaluation for Pulmonary Embolism
Marianne C. Wallis, Matthew D. Wilson, Mihriye Mete et al. 263

High Diagnostic Uncertainty and Inaccuracy in Adult Emergency Department Patients With Dyspnea: A National Database Analysis
Katherine M. Hunold, Jeffrey M. Caterino 267

REFLECTIONS
“I Can't Imagine Doing What You Do”
Jillian Nickerson 272

ERRATUM
Re: SAEM Annual Meeting Abstracts 273
Access to Federally Qualified Health Centers and Emergency Department Use Among Uninsured and Medicaid-insured Adults: California, 2005 to 2013

Julia B. Nath, MD, Shaughnessy Costigan, MD, Feng Lin, MS, Eric Vittinghoff, PhD, and Renee Y. Hsia, MD, MSc

ABSTRACT

Background: While improved access to safety net primary care providers, like federally qualified health centers (FQHCs), is often cited as a route to alleviate potentially preventable emergency department (ED) visits, no studies have longitudinally established the impact of improving access to FQHCs on ED use among Medicaid-insured and uninsured adults. We aimed to determine whether improved access to FQHCs was associated with lower ED use by uninsured and Medicaid-insured adults.

Methods: Using data from the Uniform Data System, U.S. Census Bureau, and California Office of Statewide Health Planning & Development, we conducted a longitudinal analysis of 58 California counties from 2005 to 2013. For each county-year observation, we employed three measures of FQHC access: geographic density of FQHCs (delivery sites per 100 square miles), FQHCs per county resident (delivery sites per 100,000 county residents), and the proportion of Medicaid-insured or uninsured residents ages 19 to 64 years that utilized FQHCs. We then used a fixed-effects model to examine the impact of changes in the measures of FQHC access on ED visit rates by Medicaid-insured or uninsured adults in each county.

Results: Increasing geographic density of FQHCs was associated with a 26% to 35% decrease in ED use by uninsured but not Medicaid-insured patients. Increasing numbers of clinics per county resident and higher percentages of Medicaid-insured and uninsured adults seen at FQHCs were not associated with reduced rates of ED use among either uninsured or Medicaid-insured adults.

Conclusions: We were unable to detect a consistent association between our measures of FQHC access and ED use by Medicaid-insured and uninsured nonelderly California adults, underscoring the importance of investigating additional drivers to reduce ED use among these vulnerable patient populations.
Over a decade ago, the National Academy of Medicine described emergency departments (EDs) as the “safety net of the safety net,” catching patients who fall through the cracks of primary care and social services systems.1 Today, EDs still fill that gap in our health care system; emergency physicians provide the majority of visits for new health problems by uninsured and Medicaid-insured patients.2 Given the lack of easily accessible primary care or urgent care, EDs may also see patients who do not require care in an emergency setting, but seek care in EDs because they play an important role as safety net providers.3 While estimates differ, some portion of visits may represent potentially preventable ED use. In the context of ED overcrowding,4,5 which reduces quality of care,6 increases adverse outcomes, and impairs ED access, the inefficiency represented by patients who only go to the ED because they cannot obtain care elsewhere in a more appropriate setting can be seen as a socially undesirable outcome.7

As the nation’s primary care safety net providers, federally qualified health centers (FQHCs) are poised to help alleviate the growing demand for low-cost, high-quality primary care services and mitigate preventable use of the ED by vulnerable populations. FQHCs include community health clinics, public housing centers, school clinics, and other health care and social service delivery sites funded by enhanced Medicaid reimbursements and more than $9 billion in block grants in 2016 from the Health Resources and Services Administration (HRSA) within the U.S. Department of Health and Human Services.8 With over 9,000 delivery sites serving over 20 million patients in 2012,9 FQHCs provide essential access to high quality health care for medically underserved populations.10,11

Much of the current discussion regarding which policies to implement to reduce ED use is based on the assumption that increased access to primary care is the solution to overuse of the ED.4,12,13 However, the link between primary care availability and ED use remains largely unproven. While FQHCs comprise a minority of primary care providers in the United States, they are a crucial public health tool for filling market gaps in care for underserved populations and play a prominent role in this policy debate. Previous studies relating FQHC access and acute care use at the community level have either focused on Medicare–Medicaid dual-eligible patients,17 who represent a small proportion of the overall patients seen at FQHCs, or have been cross-sectional analyses of Medicaid-insured patients.18,19

As a result, the question remains, is the longitudinal growth and expansion of the FQHC program over the past decade reducing ED use? By employing a longitudinal fixed-effects model over 58 California counties from 2005 to 2013, we tested our hypothesis that FQHC access is inversely associated with ED use by Medicaid-insured and uninsured adults at the county level.

METHODS

Sample
We examined the impact of changes in FQHC access from 2005 to 2013 on ED use within 58 counties in California. We included all FQHCs that report data to HRSA in our sample and specifically examined the impact of FQHC access on rates of ED use per county of residents in two populations: Medicaid-insured and uninsured adults aged 19 to 64 years. We chose to study Medicaid-insured and uninsured patients specifically given previous work indicating that these groups are the most likely to use the ED for primary care–preventable conditions.20,21 In addition, they are target populations served by the FQHC system and are most likely to be affected by changes in access to FQHCs.

Data Sources
We obtained data on the number of patients served at a FQHC and number of sites per county from HRSA’s publicly available Uniform Data System (UDS). To calculate geographic density of FQHC sites in each county, we used 2010 county land mass information from the U.S. Census Bureau. To calculate clinics per county resident, we used county population estimates from the California Department of Finance. For our outcome measure, we obtained comprehensive records of all ED visits in California from the Office of Statewide Health Planning and Development (OSHPD).

We used Small Area Health Insurance Estimates files from the U.S. Census Bureau to estimate a denominator of uninsured adults by county and the California Department of Health Care Services Medi-Cal enrollment counts from the Medi-Cal Eligibility Data System to similarly establish a denominator of...
Medicaid-insured adults by county. These sources allowed us to establish a denominator both for our predictors of percentage of Medicaid-insured or uninsured adult county residents seen at FQHCs, as well as for individual use in our model to generate rates of ED use per population among Medicaid-insured or uninsured adults. We then used the Area Health Resource Files (AHRF) from HRSA to obtain potentially time-variant county characteristics to use as covariates in our model.

**Predictors**

We measured FQHC access in each county-year by geographic density of FQHC delivery sites, FQHC delivery sites per county resident, and percentage of Medicaid-insured or uninsured adult residents served at a FQHC. The process of creating these measures is outlined in Figure 1. We categorized our access measures into quartiles based on their overall distributions over the study period because all of their estimated effects violated the assumption of log-linearity. We assigned the lowest quartile of each measure as the reference group and compared each higher quartile to that group in our regressions.

**Geographical Clinic Access**

Health Resources and Services Administration funds FQHCs through block grants given to “grantees,” which operate multiple sites, for which they are required to list basic information, including location, in the UDS. We calculated the number of sites by zip code using these data; however, because most other data were presented at the grantee rather than site level, there were frequent inconsistencies in the raw listing of site information. To ensure validity, we manually verified each site’s years of operation using online documentation or phone calls to the clinics or grantees directly. Due to this manual verification, we were unable to include clinic measures from later-acquired 2013 data.

We included clinics for all years that services were provided, even if they were not listed in the UDS during the entire period, which is different from the unverified patient counts captured through the

---

**Figure 1.** Derivation of FQHC access measures. Notes: Census refers to the U.S. Census Bureau 2010 land mass estimates. The HUD crosswalk refers to the U.S. Department of Housing and Urban Development zip code to FIPS county code crosswalk files (used to collapse patient counts by county). Manual verification: using online documentation or phone calls to the clinics or grantees directly we manually verified the years of operation of different clinics listed in the UDS data given frequent inconsistencies in the raw data. Age and insurance weighting: we multiplied each grantee’s zip code patient counts by the proportion of their patients between ages 19 and 64 and by the proportion of Medicaid-insured or uninsured adult patients. HRSA = Health Resources and Services Administration; UDS = Uniform Data System; DHCS = California Department of Health Care Services; SAHIE = Small Area Health Insurance Estimates (U.S. Census Bureau); FQHC = federally qualified health center; HUD = Housing and Urban Development; FIPS = Federal Information Processing Standard Publication.
We included all types of California FQHC delivery sites listed in the UDS in our clinic count because in addition to medical care provided at traditional primary care centers, services provided at for instance, dental clinics, homeless health care centers, school, or social services centers, (e.g., dental care, case management) influence factors driving underserved patients to the ED. We merged the final counts of clinics by year and zip code into the OSHPD ED visit files, which were collapsed to the county level using an OSHPD county indicator. We then divided the delivery site count per county by the county’s land mass to get density of FQHCs per 100 square miles. This predictor provided a proxy measure for geographic access to care, indicating roughly how difficult it would be for patients to travel to receive care at a FQHC.

Clinics per County Population
We also included an additional measure of access to clinics, FQHC sites per 100,000 county residents, as a way to account for population density as well as geographic density of clinics. This measure was prepared similarly as above, but we divided clinic counts per county by the county’s population in each year to obtain our final predictor.

Percentage of Medicaid or Uninsured Residents Served
We first abstracted the number of California FQHC patients from the UDS by grantee, zip code, and year. A patient is defined by HRSA as anyone who has had at least one visit for face-to-face services—regardless of the volume or scope of services received. Each patient is counted once per grantee, regardless of how many visits that patient makes. If a patient received care at multiple grantees, he or she may be counted multiple times. To determine the number of adult patients served, we multiplied each grantee’s zip code patient counts by the proportion of their patients between ages 19 and 64. We then multiplied this total by the grantee’s proportion of Medicaid-insured or uninsured adult patients.

We totaled all the grantees’ adult patient counts by insurance type (Medicaid or uninsured) in a given zip code to get the total number of Medicaid-insured or uninsured adults served by any FQHC in that zip code. We collapsed these data to the county level using the Housing and Urban Development zip code to county crosswalk files for the year in question. We used the 2010 crosswalk files for years prior to 2010 given the lack of crosswalk files for those years. If zip codes crossed more than one county, the patient counts were divided between the counties by the resident ratio or the proportion of residential addresses in that zip code in each county. We then divided the patient counts by the number of uninsured or Medicaid-insured adult residents of the county-year in question to arrive at the proportion of uninsured or Medicaid-insured adults in that county seen at an FQHC. We used county codes to link this patient data to OSHPD ED use data. Because patient location data were not reliable prior to 2008, this predictor was limited to 2008 to 2013.

Outcome
We included all ED visits in each county by Medicaid-insured and uninsured patients ages 19 to 64. ED visits were assigned to a county based on each patient’s reported residential county. The county population of Medicaid-insured or uninsured adults was included as an offset in our negative binomial models, leading to a final outcome measure of ED visit rates per corresponding population.

Covariates
We included nine variables from the AHRF that we anticipated might impact ED use in a given county based on previous literature, including percentage of population living in poverty, median household income, unemployment rate, primary care doctors per population, short-term hospital beds per population, percentage of population with a college degree, and health professional shortage area status.

Data Analysis
We used negative binomial models to estimate the dependence of the number of ED visits in each county and year on our three measures of FQHC access. The models included county as a fixed effect and controlled for prespecified factors that potentially changed over the study period within counties. Accordingly, access effect estimates depend solely on within-county contrasts across years, avoiding bias stemming from fixed but unmeasured county-level confounders. We included year as a series of indicator variables to flexibly control for secular trends and used robust standard errors to account for overdispersion. In addition to estimating ED visit rate ratios by quartile of the access measure, we also...
tested for the overall trend across quartiles using orthogonal linear combinations in the coefficients for the quartiles. To examine a potential mechanism for the association between geographic density of FQHCs and ED use, we conducted an additional analysis in which we determined whether geographic clinic density was associated with the proportion of Medicaid or uninsured county residents seen at FQHCs in the following year. We also conducted the same analysis examining the association between clinics per county resident and the proportion of Medicaid or uninsured county residents seen at FQHCs in the following year for completeness. We performed analyses using SAS 13.1 and STATA 14.0. Two-tailed significance was assessed at \( p < 0.05 \). The UCSF Committee on Human Research approved this study.

**RESULTS**

From 2005 to 2012, the number of FQHC sites in California increased from 910 to 1,197, with the average number of sites per county rising from 18 to 24. FQHC density by county also rose, with most counties rising in number of FQHC sites per 100 square miles (Figure 2) and sites per 100,000 county residents. The number of patients ages 19 to 64 served by a FQHC in California grew from 1.4 million in 2008 to over 2 million in 2013, rising from approximately 6% to 8% of the California adult population.

Using geographic density of FQHC clinics as a measure of access, we found that ED use by Medicaid-insured adults did not change significantly as counties added clinics and crossed into higher quartiles of FQHCs per 100 square miles (\( p = 0.45 \) for overall

---

**Figure 2.** Quartile of clinics per 100 square miles, by county. Quartile 1 = <0.12 sites/100 square miles; Quartile 2 = 0.12-0.45 sites/100 square miles; Quartile 3 = 0.46-1.38 sites/100 square miles; Quartile 4 = >1.38 sites/100 square miles. N/A refers to the one county with insufficient data to calculate this metric.
linear trend; Table 1). In contrast, we found that an increase from less than 0.12 clinics per 100 square miles to greater than 0.46 clinics per 100 square miles was associated with a 35% decrease in ED use by uninsured adults in that county from 2005 to 2012 (p < 0.01, Table 1). In additional analyses, we found that geographic FQHC density was independently associated with increases in the percentage of uninsured patients using FQHCs in the following year (p = 0.88; 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.13494/full).

When using FQHC delivery sites per county resident as a measure of FQHC access, we found no statistically significant association between increasing numbers of FQHCs per 100,000 county residents and ED visit rates among Medicaid-insured or uninsured adults (p = 0.10 and p = 0.37 for overall trend, respectively; Table 2). In our additional analyses, we found that FQHC delivery sites per 100,000 county residents were not statistically significantly associated with changes in the proportion of Medicaid-insured or uninsured adults seen at FQHCs the following year (p = 0.18 and p = 0.07; Data Supplement S1).

Increases in our third measure of access—percentages of Medicaid-insured or uninsured adults served at a FQHC—were not associated with a significant change in ED use by Medicaid-insured or uninsured patients between 2008 and 2013 (p = 0.30 and p = 0.26 for overall linear trend; Table 3).

Table 1
The Association of Geographic FQHC Delivery Site Density With ED Visit Rates Among Adults Ages 19 to 64 Without Insurance or Insured by Medicaid, 2005 to 2012

<table>
<thead>
<tr>
<th>Density of FQHC Sites (Quartile, Sites/100 Square Miles)</th>
<th>Medicaid</th>
<th></th>
<th></th>
<th>Uninsured</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
<td>95% CI</td>
<td>p-value</td>
<td>IRR</td>
<td>95% CI</td>
<td>p-value</td>
</tr>
<tr>
<td>&lt;0.12</td>
<td>Ref</td>
<td></td>
<td>0.45*</td>
<td>Ref</td>
<td></td>
<td>0.01*</td>
</tr>
<tr>
<td>0.12–0.45</td>
<td>1.04</td>
<td>0.96–1.13</td>
<td>0.30</td>
<td>0.74</td>
<td>0.63–0.87</td>
<td>0.01</td>
</tr>
<tr>
<td>0.46–1.38</td>
<td>0.89</td>
<td>0.76–1.05</td>
<td>0.17</td>
<td>0.65</td>
<td>0.54–0.78</td>
<td>0.01</td>
</tr>
<tr>
<td>&gt;1.38</td>
<td>0.97</td>
<td>0.82–1.15</td>
<td>0.73</td>
<td>0.65</td>
<td>0.52–0.81</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Results of a negative binomial model regressing ED visits on geographic density of FQHC delivery sites in each county each year. Estimates adjusted for year and county as fixed effects as well as time-varying county-level covariates including percentage of population living in poverty, median household income, unemployment rate, primary care doctors per population, short-term hospital beds per population, percentage of the population with a college degree, and health professional shortage area status. IRRs are interpreted as the incidence of ED visits in quartile listed/incidence of ED visits in reference quartile, similarly to an OR or hazard ratio for their respective outcomes. Additional results are available in Data Supplement S2.

FQHC = federally qualified health center; IRR = incidence rate ratio.

* p-value for trend across four quartiles.

Table 2
The Association of FQHC Delivery Sites per 100,000 County Residents With ED Visit Rates Among Adults Ages 19 to 64 Without Insurance or Insured by Medicaid, 2005 to 2012

<table>
<thead>
<tr>
<th>FQHCs per 100,000 Population (Quartile)</th>
<th>Medicaid</th>
<th></th>
<th></th>
<th>Uninsured</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
<td>95% CI</td>
<td>p-value</td>
<td>IRR</td>
<td>95% CI</td>
<td>p-value</td>
</tr>
<tr>
<td>&lt;2.12</td>
<td>Ref</td>
<td></td>
<td>0.10*</td>
<td>Ref</td>
<td></td>
<td>0.37*</td>
</tr>
<tr>
<td>&gt;2.12–4.59</td>
<td>1.02</td>
<td>0.96–1.08</td>
<td>0.62</td>
<td>0.97</td>
<td>0.91–1.04</td>
<td>0.46</td>
</tr>
<tr>
<td>&gt;4.59–7.27</td>
<td>0.93</td>
<td>0.85–1.01</td>
<td>0.09</td>
<td>0.98</td>
<td>0.89–1.08</td>
<td>0.74</td>
</tr>
<tr>
<td>&gt;7.27</td>
<td>0.92</td>
<td>0.82–1.04</td>
<td>0.20</td>
<td>1.08</td>
<td>0.87–1.45</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Results of a negative binomial model regressing ED visits on FQHC delivery sites per 100,000 county residents in each county each year. Estimates adjusted for year and county as fixed effects as well as time-varying county-level covariates including percent of population living in poverty, median household income, unemployment rate, primary care doctors per population, short-term hospital beds per population, percentage of the population with a college degree, and health professional shortage area status. IRRs are interpreted as the incidence of ED visits in quartile listed/incidence of ED visits in reference quartile, similarly to an OR or hazard ratio for their respective outcomes. Additional results are available in Data Supplement S2.

FQHC = federally qualified health center; IRR = incidence rate ratio.

* p-value for linear trend across four quartiles.
DISCUSSION

We could not identify a significant association between increasing access to and utilization of FQHCs and ED use among underinsured adults in California counties between 2005 and 2013 in five of our six analyses. The lack of consistent statistically significant associations with reduced ED use across all access measures challenges the assumption that expansion of FQHCs will necessarily reduce ED use by uninsured and Medicaid-insured patients. While these findings do not address other potential benefits of increased access to primary care and social services via FQHCs, which increasingly provide an important source of care to racial and ethnic minorities, low-income, and uninsured or Medicaid-insured individuals, they serve as a sobering reminder that reduced ED use may not necessarily be among them.

However, our findings do suggest that improved geographic access to FQHCs is associated with lower rates of ED use among uninsured adults. We found that increasing from the lowest quartile (<0.12 clinics per 100 square miles) to either of the top two quartiles of geographic access to FQHCs (>0.46 clinics per 100 square miles) was associated with a 35% decrease in rates of ED use by uninsured adults in those counties. These findings concur with previous cross-sectional work showing that the presence of a FQHC in a geographic area was related to lower rates of preventable hospitalizations or ED use among vulnerable populations.14,15

Our additional analysis on determining whether or not this association is mediated through additional uninsured patients seen at these newly opened clinics showed that the proportion of uninsured adults seen at FQHCs increased the year following an increase in FQHC delivery sites per 100 square miles, but the proportion of Medicaid-insured adults seen at FQHCs did not. This supports the theory that utilization by uninsured adults is the mechanism by which geographic FQHC density impacts ED utilization rates. However, our later analysis was unable to detect any correlation between increasing proportions of uninsured adults seen at FQHCs and ED use, which does not support this proposed mechanism. This disconnect could be secondary to our relatively small sample size of 58 counties, which limits our ability to detect such a relationship if the effect size is small. However, the mechanism by which geographic FQHC density reduced ED use by the uninsured could also be through other measures not captured in our analysis. For instance, more FQHCs could increase focus on and services for community health and reduce ED use as uninsured patients receive these additional public health benefits rather than more direct services at FQHCs.

The rest of our analyses were unable to detect a significant association between FQHC access and ED visit rates in uninsured and Medicaid-insured adults. When measuring access by geography, we found that ED visit rates among Medicaid-insured adults did not respond to increases in the number of clinics per 100 square miles. Previous evidence has shown that in addition to distance, factors such as lack of transportation, up-front costs, limited clinic hours, limited services, long wait times, difficulty getting timely appointments, patient education, and health literacy

<table>
<thead>
<tr>
<th>Percentage of Medicaid-insured Adults Seen at FQHCs (Quartile)</th>
<th>ED Visits by Adults With Medicaid IRR 95% CI p-value</th>
<th>Percentage of Uninsured Adults Seen at FQHCs (Quartile)</th>
<th>ED Visits by Uninsured Adults IRR 95% CI p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10.8</td>
<td>Ref Ref 0.30*</td>
<td>&gt;7.2–17.9 1.01 0.91–1.12 0.83</td>
<td></td>
</tr>
<tr>
<td>&gt;10.8–20.8</td>
<td>1.04 1.00–1.08 0.04</td>
<td>&gt;17.9–32.8 1.08 0.94–1.25 0.29</td>
<td></td>
</tr>
<tr>
<td>&gt;20.8–39.3</td>
<td>1.07 1.02–1.12 0.01</td>
<td>&gt;33.1–71.8 1.08 0.93–1.26 0.31</td>
<td></td>
</tr>
<tr>
<td>&gt;39.3–95.4</td>
<td>1.03 0.96–1.10 0.41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results of a negative binomial model regressing ED visit rate among adults ages 19 to 64 covered by Medicaid or uninsured adults by county on percentage of Medicaid-insured adults or percentage of uninsured adults ages 19 to 64 who are FQHC patients. Estimates adjusted for year and county as fixed effects as well as time-varying county-level covariates including percentage of population living in poverty, median household income, unemployment rate, primary care doctors per population, short-term hospital beds per population, and percentage of the population with a college degree, and health professional shortage area status. IRRs are interpreted as the incidence of ED visits in quartile listed/incidence of ED visits in reference quartile, similarly to an OR or hazard ratio for their respective outcomes. Additional results are available in Data Supplement S2.

FQHC = federally qualified health center; IRR = incidence rate ratio.
*p-value for trend across four quartiles.
drive patients to use the ED instead of local FQHC clinics for nonemergent concerns.\textsuperscript{1,27} These additional factors likely contributed to the lack of impact of increased geographic access to FQHCs on eventual ED use among Medicaid patients and may be less pronounced among uninsured patients who are more likely to seek alternative, less-expensive sources of care when available, because they are not as shielded from medical costs as Medicaid enrollees.\textsuperscript{28}

However, we were also unable to detect an association between the proportion of Medicaid-insured adults seen at FQHCs and their rate of ED visits. This measure of access captures patients who overcame nongeographic barriers and received some care from an FQHC, yet we still could not find an association with ED use. We also did not find an association between FQHC sites per 100,000 county residents and ED use among Medicaid-insured adults. This points to alternative factors beyond FQHC access that drive Medicaid-insured patients to the ED.

For example, Medicaid-insured adults have significantly worse health than either uninsured or privately insured adults.\textsuperscript{28–30} Health care use, including ED use, in medically complex populations is less elastic and, subsequently, would likely be less affected by changes in access to primary care.\textsuperscript{31} Our previous work with pediatric populations,\textsuperscript{32} which found decreased ED use with increased geographic access for both uninsured and Medicaid-insured children, provides some support for this possibility, as children in general (whether Medicaid or uninsured) are healthier than adults. Health status or other Medicaid-specific concerns are unlikely to be the primary explanation behind our findings, however, as we also did not find an association between the proportion of uninsured adults or clinics per 100,000 uninsured residents and ED use among uninsured adults.

Numerous other factors beyond primary care use could be contributing to the lack of overall association between FQHC access and ED use among vulnerable populations. For example, we were unable to account for clinic and care-specific factors that influence the likelihood of patients using the ED, including physician attitudes, referral practices, continuity of care, quality of care, and patient-centeredness.\textsuperscript{5,33–35} The specific interventions individual clinics can implement to reduce ED use is an important area for continued research. In addition, previously cited factors in qualitative studies that drive patients to visit the ED for potentially preventable or nonurgent concerns such as time constraints, flexibility, provider referrals, perception of severity of illness, and perception of quality of care could be further investigated as potential intervention points to reduce preventable ED use.\textsuperscript{5,27,36}

While our findings contradict previous cross-sectional studies that show Medicaid-insured patients seen at FQHCs having fewer ED visits,\textsuperscript{18,19} other earlier works have similarly suggested that increased access to health services may not reduce ED use. In the 2008 Oregon Medicaid expansion, subjects randomized to receive Medicaid coverage actually had increased rates of ED use compared to those not given insurance.\textsuperscript{37} In addition, projections have shown that expanding health care access alone is likely to increase ED use.\textsuperscript{38} While these studies focus on expanding access to health insurance and our analysis attempted to capture expanded access to primary care, they do point out that expanded healthcare access can lead to increased use of all health care services.

Recent policy shifts may impact the results we found over our study period. In California, extensive Medicaid insurance expansion following the first Affordable Care Act enrollment period in 2013 reduced the uninsured population by 68%, half of which gained insurance through Medicaid.\textsuperscript{39} The California FQHC Program, which receives the majority of its funding from enhanced Medicaid reimbursements,\textsuperscript{40} is poised to benefit from this shift. Furthermore, California FQHCs were already growing at a rate that outpaced the national average before 2013\textsuperscript{41} and have taken the initiative to begin comprehensive payment and delivery reform to help reduce spending, raise the quality of care and improve access to primary and preventive care.\textsuperscript{42} Because these changes in the environment of FQHC care in California have primarily occurred after our study period, it is unclear how they will affect the relationships we find between FQHC access and ED use. However, if increasing geographic FQHC density reduces ED visits for uninsured, but not for Medicaid-insured, adults, as our findings suggest, then a growing Medicaid population and dwindling uninsured population may reduce any net benefit realized by increasing FQHC density in a given county. Further research could help elucidate the impact of these policy changes.

**LIMITATIONS**

While our fixed-effects model allowed us to remove all between-county omitted variable bias, there is still the
possibility of confounding in the relationship between FQHC access and ED use by factors that changed within a county over time. We attempted to control for this by using potentially time-variant county characteristics, which did not meaningfully change the results. However, we could not control for all factors, such as measures of health status of the population over time, which could have shifted within counties over our study period and affected ED visits and FQHC utilization. Our analysis is also subject to the limitations of an ecological study; we cannot determine if the patients served at a FQHC were the same people making fewer visits to the ED, which would require longitudinal patient data that are not available, particularly for uninsured patients.

While counties as proxies for market areas approximate geographic access, the measures are imperfect. Boundary bias may exist if a clinic in a neighboring county is actually closer for some residents than clinics in their own county. However, no other geographic boundary allows us to combine the necessary sources of data for this analysis, and county boundaries serve as a common measure used in studies evaluating access to care.

Finally, our measures of access to FQHCs are imperfect. Geographic access, as we defined it, does not take into account transportation time to the nearest clinic and thus is only a proxy for geographic access to care. In addition, the number of FQHC delivery sites per county resident is measured uniformly across each county, rather than within specific population centers, and may not capture the meaningful number of clinics per person. Further, while our measures of the proportion of Medicaid-insured or uninsured adults seen at an FQHC attempt to capture the penetrance of FQHC access into the target patient population, it does not account for a number of competing factors, such as the large number of Medicaid-insured patients that receive primary care at non-FQHC clinics. While we attempted to control for the underlying supply of primary care in a given county using our covariates of primary care doctors per population, as well as hospital access using hospital beds per population, we did not have data on alternative care sources used by Medicaid-insured and uninsured patients in that county. Therefore, we do not know whether changes in the proportion of patients served at a FQHC represent true penetrance of services into an area of unmet need (i.e., a true increase in access) or the transferring of patients between alternative care sources.

CONCLUSIONS

Our analysis could not detect a consistent association between multiple measures of federally qualified health center access and ED use among uninsured and Medicaid-insured adults in California counties between 2005 and 2013. These results suggest the need for the investigation of intervention points beyond nonspecific primary care access in the effort to reduce ED overuse among vulnerable populations.

We thank the California Office of Statewide Health Planning and Development for their assistance in preparing the data sets used in this project and Sarah Sabbagh, MPH, and Joanna Guo, BA, for their administrative and editorial support.

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.13494/full

Data Supplement S1. Association between geographic FQHC density and FQHC sites per county resident and percentage of Medicaid-insured or uninsured adults seen at FQHCs in the following year.

Data Supplement S2. Supplemental material: full regression results.
Prognostic Accuracy of the HEART Score for Prediction of Major Adverse Cardiac Events in Patients Presenting With Chest Pain: A Systematic Review and Meta-analysis

Shannon M. Fernando MD, MSc, Alexandre Tran MD, MSc, Wei Cheng PhD, Bram Rochwerg MD, MSc, Monica Taljaard PhD, Venkatesh Thiruganasambandamoorthy MBBS, MSc, Kwadwo Kyeremanteng MD, MHA, and Jeffrey J. Perry MD, MSc

ABSTRACT

Objective: The HEART score has been proposed for emergency department (ED) prediction of major adverse cardiac events (MACE). We sought to summarize all studies assessing the prognostic accuracy of the HEART score for prediction of MACE in adult ED patients presenting with chest pain.

Methods: We searched MEDLINE, PubMed, EMBASE, Scopus, Web of Science, and the Cochrane Database of Systematic Reviews from inception through May 2018 and included studies using the HEART score for the prediction of short-term MACE in adult patients presenting to the ED with chest pain. The main outcome was short-term (i.e., 30-day or 6-week) incidence of MACE. We secondarily evaluated the prognostic accuracy of the HEART score for prediction of mortality and myocardial infarction (MI). Where available, accuracy of the Thrombolysis in Myocardial Infarction (TIMI) score was determined.

Results: We included 30 studies (n = 44,202) in analysis. A HEART score above the low-risk threshold (≥4) had a sensitivity of 95.9% (95% confidence interval [CI] = 93.3%–97.5%) and specificity of 44.6% (95% CI = 38.8%–50.5%) for MACE. A high-risk HEART score (≥7) had a sensitivity of 39.5% (95% CI = 31.6%–48.1%) and specificity of 95.0% (95% CI = 92.6%–96.6%) for MACE, whereas a TIMI score above the low-risk threshold (≥2) had a sensitivity of 87.8% (95% CI = 80.2%–92.8%) and specificity of 48.1% (95% CI = 38.9%–57.5%) for MACE. A high-risk TIMI score (≥6) was 2.8% sensitive (95% CI = 0.8%–9.6%), but 99.6% (95% CI = 98.5%–99.9%) specific for MACE. A HEART score ≥ 4 had a sensitivity of 95.0% (95% CI = 87.2%–98.2%) for prediction of mortality and 97.5% (95% CI = 93.7%–99.0%) for prediction of MI.
Conclusions: The HEART score has excellent performance for prediction of MACE (particularly mortality and MI) in chest pain patients and should be the primary clinical decision instrument used for the risk stratification of this patient population.

Chest pain is a commonly encountered presentation in the emergency department (ED), with approximately 8 million visits per year in the United States, and many of these patients are ultimately admitted to hospital for observation or intervention. Approximately 10% to 20% of these patients are diagnosed with an acute coronary syndrome (ACS), characterized by myocardial ischemia, and benefit from early identification and initiation of treatment, including revascularization. The goal of the clinician is to differentiate between patients presenting with ACS and those with other (typically more benign) conditions. While various historical features and laboratory values may help to identify patients with true ACS, none are sufficiently accurate to be used independently. As a result, 2% to 5% of patients with true ACS are inappropriately discharged from the ED annually, and missed cases of ACS represent a significant proportion of malpractice claims in the United States. Therefore, there is a tendency for clinicians to overinvestigate chest pain patients with further, often more invasive testing, even in low-risk patients. This practice results in increased resource utilization without improved outcomes.

Several decision instruments to identify low-risk chest pain patients who may be suitable for discharge without further testing are currently in use. One of the most well-recognized risk scores is the TIMI (Thrombolysis in Myocardial Infarction) score, which was originally derived and validated in a population of inpatients with unstable angina and non-ST-elevation myocardial infarction (MI), to determine their 14-day risk of major adverse cardiac events (MACE). MACE is a composite outcome that includes death, MI, and revascularization (either percutaneous coronary intervention [PCI] or coronary artery bypass grafting [CABG]). While the American Heart Association (AHA) and the American College of Cardiology (ACC) have recommended the TIMI score for the initial evaluation of a patient with chest pain, the use of this tool for the identification of low-risk chest pain patients has been associated with conflicting results.

To better risk stratify ED patients with chest pain, the HEART score was derived through a process involving expert opinion and review of the existing medical literature. The HEART score was created specifically to identify ED patients presenting with chest pain who were at a low risk of short-term MACE, who could then be discharged from the ED with appropriate follow-up, as well as patients with high-risk of MACE, who may require immediate intervention. The five predictors included in the HEART score are: history (H), electrocardiogram (ECG, E), age (A), risk factors (R), and troponin (T; Data Supplement S1, Table S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13649/full). A HEART score of 0 to 3 identifies a patient at “low risk” of MACE and suggests consideration of discharge. Patients with a score of 4 to 6 are considered “intermediate risk,” and those with a score of 7 to 10 are considered “high risk.” Since its initial validation, the HEART score has been independently validated in a number of studies worldwide. Given its potential role in the risk stratification of patients with chest pain, a better understanding of the overall prognostic accuracy of this tool is needed. We conducted a systematic review and meta-analysis to summarize the prognostic accuracy of the HEART score for prediction of short-term MACE in adult patients presenting with chest pain.

METHODS

We structured this systematic review according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, and existing guidelines for reviews of diagnostic accuracy. We chose to perform a meta-analysis of diagnostic test accuracy rather than predictive ability, for the specific purpose of informing accuracy of screening decisions by clinicians, which is how the HEART score is primarily utilized (i.e., to rule out MACE in low-risk patients and avoid unnecessary downstream testing). When evaluating a decision instrument in the context of screening, the most important test characteristics are the sensitivity, specificity, and positive and negative likelihood ratios. Indeed, these are the characteristics that are provided...
in the large majority of included studies in this meta-
analysis. The study protocol was registered with PROS-
PERO (CRD42018087034).

Data Sources and Searches
We searched MEDLINE, PubMed, EMBASE, Scopus, Web of Science, and the Cochrane Database of Systematic Reviews from inception until May 1, 2018. An experienced health sciences librarian assisted in the development of the search strategy. The search was conducted using the terms “HEART score,” “heart-score,” “HEART tool,” “HEART pathway,” and “HEART pathway score” (search strategy is depicted in Data Supplement S1, Figure S1). We used Science Citation Index to retrieve reports citing the relevant articles identified from our search and then entered them into PubMed. We also conducted further surveillance searches, utilizing the “related articles” feature.

Study Selection
We included all English-language full-text articles describing retrospective and prospective observational studies, as well as randomized controlled trials and quasi-randomized controlled trials. We included studies meeting the following criteria: 1) enrolled adult patients (≥16 years) with suspected ACS; 2) conducted in the ED; and 3) applied the HEART score for prediction of short-term MACE (in-hospital, 28-day, 30-day, 6-week, or 3-month). We excluded studies that evaluated MACE over longer or unspecified time periods. We similarly excluded case reports, case series, and studies only evaluating the prognostic accuracy of a modified version of the HEART score. We excluded conference abstracts, as the data often change between abstracts and full-text publications. Furthermore, the data presented in these abstracts had not been verified through peer review. To be eligible for meta-analysis, each study was required to have a 2 × 2 table of true-positive, true-negative, false-positive, and false-negative counts for at least the low-risk threshold (the primary threshold of interest), either extracted from the original article or calculated from other reported information such as declared sensitivity and specificity. We contacted authors in instances where these values could not be obtained from the reported data. We excluded the study if the corresponding author did not respond after three attempts.

We screened studies using Covidence software (Melbourne, Australia). Titles were imported into Covidence directly from the search databases, and duplicates were removed. In the first phase, two reviewers (SMF, AT) independently screened the titles and abstracts of all identified citations. Disagreement was resolved by consensus; no third-party adjudication was necessary. In phase two, the same two reviewers independently assessed full texts of the selected articles from phase one. Disagreements were resolved by consensus.

Data Extraction and Quality Assessment
We used a predesigned data extraction sheet (Data Supplement S1, Table S2) to minimize the risk of transcriptional errors. Two investigators independently collected the true-positive, false-positive, false-negative, and true-negative counts of the HEART score and TIMI score (in studies where this score was also reported), total number of MACE, and stated sensitivity and specificity of HEART and TIMI (only when included) from all studies. Disagreements were resolved through consensus. All extracted data were independently verified by a third investigator.

Two reviewers (SMF, AT) independently assessed the risk of bias for the included studies, using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool. Disagreements were resolved through consensus. The QUADAS-2 assesses four potential areas for bias and applicability of the research question: patient selection, index test, reference standard, and flow and timing.

Data Synthesis and Analysis
We presented individual study results graphically by plotting sensitivity and specificity estimates on onedimensional forest plots (ordered by sensitivity) as well as on the receiver operating characteristic (ROC) space, to visually assess for heterogeneity. To pool the results, we applied the hierarchical summary ROC (HSROC) model and obtained summary point estimates of the pairs of sensitivity and specificity, as well as diagnostic odds ratios (ORs) and likelihood ratios, with their 95% confidence intervals (CIs). The HSROC model incorporates both within-study and between-study variability. The summary point of test accuracy estimates was plotted in the ROC space together with the summary ROC curve. Whenever the number of studies was too few (seven studies or fewer), we fit a HSROC model that assumes a symmetric SROC curve (by restricting the shape parameter to be 0). The analyses were conducted using MetaDAS (Version 1.3), Univariate tests for heterogeneity in
sensitivity and specificity are not recommended by the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, as they do not account for heterogeneity explained by phenomena such as positive threshold effects. Instead, it is preferable to demonstrate heterogeneity graphically through the ROC curve and forest plots and through the use of multiple subgroup and sensitivity analyses, as done previously.

We conducted subgroup analyses of studies utilizing 6-week incidence of MACE versus 30-day incidence of MACE, ED physician-interpreted ECG versus cardiologist-interpreted ECG, and patients presenting with chest pain versus patients presenting with “suspected ACS.” We conducted sensitivity analyses excluding studies with high-sensitivity troponin and those with high risk of bias.

We assessed the overall confidence in pooled diagnostic effect estimates using the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach (performed by BR). The overall confidence in effect estimates were categorized into one of four levels which included high, moderate, low, or very low. A GRADE evidence profile was created using the guideline development tool (gradepro.org).

RESULTS

Search Results

Our search identified 778 citations (Figure 1) and following removal of duplicates, we screened 557 studies, from which 62 studies underwent full-text review. We included 29 distinct cohorts from 30 studies in the meta-analysis. All included studies evaluated the prognostic accuracy of the HEART score using a low-risk threshold (score between 0 and 3). Twenty-one studies also evaluated the prognostic accuracy of the HEART score using a high-risk threshold (score between 7 and 10). Eight studies additionally evaluated the prognostic accuracy of a low-risk TIMI (score of either 0 or 1) for prediction of MACE, while three evaluated the prognostic accuracy of a high-risk TIMI (score of either 6 or 7). Table 2, with corresponding curves depicted in Figure 3. All summary estimates described are pooled values. The sensitivity of a HEART score of ≥4 for prediction of short-term MACE was 95.9% (95% confidence interval [CI] = 93.3%–97.5%), and the specificity was 44.6% (95% CI = 38.8%–50.5%). In comparison, the pooled sensitivity of a TIMI score of ≥2 for prediction of MACE was 87.8% (95% CI = 80.2%–92.8%), and the specificity was 48.1% (95% CI = 38.9%–50.5%). A high-risk HEART score (7–10) was associated with a sensitivity of 39.5% (95% CI = 31.6%–41.8%) and specificity of 95.0% (95% CI = 92.6%–96.6%) for MACE. In comparison, a high-risk TIMI score (6–7) was associated with a sensitivity of 2.8% (95% CI = 0.8%–9.6%) and specificity

Quality Assessment

Quality assessments using QUADAS-2 criteria are summarized in Data Supplement S1, Figure S3. Twenty studies (66.7%) had unclear risk of bias in either/both the index test (HEART/TIMI) and/or the reference standard (MACE), as it was not explicitly stated whether these values were interpreted by blinded assessors. Seventeen studies were judged to be at high risk of bias in patient selection or application of the reference standard.

Primary Analyses: Overall Accuracy. Figure 2 depicts the forest plots of the sensitivity and specificity of HEART score ≥ 4 and TIMI score ≥ 2 (above the low-risk threshold). Summary estimates of all diagnostic accuracy measures from the HSROC model are tabulated in Table 2, with corresponding curves depicted in Figure 3. All summary estimates described are pooled values. The sensitivity of a HEART score of ≥4 for prediction of short-term MACE was 95.9% (95% confidence interval [CI] = 93.3%–97.5%), and the specificity was 44.6% (95% CI = 38.8%–50.5%). In comparison, the pooled sensitivity of a TIMI score of ≥2 for prediction of MACE was 87.8% (95% CI = 80.2%–92.8%), and the specificity was 48.1% (95% CI = 38.9%–50.5%). A high-risk HEART score (7–10) was associated with a sensitivity of 39.5% (95% CI = 31.6%–41.8%) and specificity of 95.0% (95% CI = 92.6%–96.6%) for MACE. In comparison, a high-risk TIMI score (6–7) was associated with a sensitivity of 2.8% (95% CI = 0.8%–9.6%) and specificity
of 99.6% (95% CI = 98.5%–99.9%). GRADE evidence profiles are included in the supplemental data (Data Supplement S1, Tables S4 and S5).

The prognostic accuracy of the HEART score for prediction of individual components of the MACE composite was also evaluated (Table 2 and Data Supplement S1, Figures S6–S11). For prediction of mortality, a HEART score above the low-risk threshold (≥4) had a sensitivity of 95.0% (95% CI = 87.2%–98.2%) and specificity of 34.2% (95% CI = 28.7%–40.2%). A high-risk HEART score (≥7) had a sensitivity of 48.4% (95% CI = 31.7%–65.4%) and specificity of 91.9% (95% CI = 88.4%–94.3%). For prediction of acute MI, a HEART score of ≥4 had a sensitivity of 97.5% (95% CI = 93.7%–99.0%) and specificity of 40.5% (95% CI = 33.6%–47.9%). A high-risk HEART score (≥7) had a sensitivity of 42.5% (95% CI = 28.9%–57.3%) and specificity of 96.9% (95% CI = 94.5%–98.3%). Finally, for prediction of coronary revascularization (i.e., PCI or CABG), a HEART score of ≥4 had a sensitivity of 89.7% (95% CI = 87.2%–91.8%) and specificity of 41.8% (95% CI = 39.4%–44.2%). A high-risk HEART score (≥7) had a sensitivity of 30.0% (95% CI = 20.2%–42.1%) and specificity of 94.5% (95% CI = 91.2%–96.6%). Prognostic accuracy of the TIMI score for prediction of individual components of the MACE composite outcome could not be evaluated due to a lack of sufficient studies.
Subgroup and Sensitivity Analyses. The results of the subgroup and sensitivity analyses to examine the prognostic accuracy of the HEART score in selected populations are depicted in Data Supplement S1, Table S6. Forest plots and HSROC curves for these analyses are displayed in Data Supplement S1, Figures S12–S27. There was no difference in pooled sensitivity among studies with ECGs interpreted by any physician compared to cardiologists specifically. Among studies evaluating only patients presenting with chest pain (as in the original HEART score study), the sensitivity of a HEART score of ≥4 for MACE was 96.1% (95% CI = 93.2%–97.7%). However, there was no difference when compared to studies evaluating any patient with suspected ACS regardless of presence of chest pain. For studies only utilizing conventional troponin assays (i.e., excluding high-sensitivity troponin assays), the sensitivity of a HEART score of ≥4 for MACE was 94.9% (95% CI = 91.8%–96.9%). There were insufficient studies utilizing high-sensitivity troponin to generate pooled estimates from the HSROC model. A sensitivity analysis excluding studies with high risk of bias did not significantly alter the findings.

DISCUSSION

We performed a systematic review and meta-analysis to summarize the prognostic accuracy of the HEART score for prediction of short-term MACE among patients presenting with chest pain. A previous meta-analysis also evaluated the prognostic accuracy of HEART, but only included nine studies and did not investigate the accuracy of HEART for individual prediction of death and MI.60 We condensed the findings from many external validation studies to provide a single estimate of the true prognostic performance of the score. We found a HEART score above the low-risk threshold (≥4) had high sensitivity (95.9%) for short-term MACE and was superior to the sensitivity of a TIMI score above the low-risk threshold (≥2; 87.8%). In particular, a HEART score of ≥4 had high sensitivity for both short-term mortality (95.0%) and MI (97.5%). Finally, a high-risk HEART score (≥7) had high specificity (95.0%) for short-term MACE, which was slightly lower than the specificity of a high-risk TIMI score (99.6%). Taken together, this work supports the utilization of the HEART score over the TIMI score for risk stratification of patients presenting with chest pain.

Given the difficulties associated with accurate risk stratification of patients presenting with chest pain, and the potential consequences associated with inappropriate discharge, clinicians often elect to admit patients that they believe to be at low risk of MACE.6 As a result, the AHA/ACC guidelines have recommended that risk stratification scores should be used to aid in clinical decision making.11 Specifically, these guidelines reference the TIMI score and the Global Registry of Acute Coronary Events (GRACE) score.61 However, neither TIMI nor GRACE was designed for ED chest pain risk stratification, but rather for prognostication among inpatients with confirmed ACS.

In a previous meta-analysis by Hess et al., the sensitivity of TIMI above the low-risk threshold (≥2) for MACE among ED patients was found to be 90.6%, which was similar to what was seen in our population (87.8%). This was despite the fact that our meta-analysis did not include any of the same citations, as none of the included studies by Hess et al. evaluated the HEART score. This suggests that the utilization of the
TIMI score among ED patients with chest pain will miss approximately one out of every 10 patients with short-term MACE. Our work demonstrates the HEART score offers superior prognostic accuracy to TIMI. A previous study has shown that the most predictive components of the TIMI score are those common to the HEART score (age, ECG changes, and troponin). Therefore, after using their clinical judgment to determine a pretest probability of MACE, physicians should utilize the HEART score as the decision instrument of choice in determining a final probability of MACE in their patients presenting with chest pain (Data Supplement S1, Tables S7 and S8). For example, a patient with a pretest probability of MACE of 25% and a HEART score below the low-risk threshold (≤3) would have a posttest probability of 3.0%. That same patient with a TIMI score below the low-risk threshold (≤1) would have a posttest probability of 7.8%. It was not possible to assess the prognostic accuracy of the GRACE score, as the included studies that analyzed this score utilized different thresholds for evaluation. These demonstrated findings have important implications for relevant guidelines and clinical policies. As mentioned, the AHA/ACC guidelines...
the dilemma of evaluating the diagnostic value of troponins for identifying “clinically significant” cardiac ischemia with need for revascularization. This is a subjective diagnosis without a criterion standard and any potential outcome assessor would be likely influenced by the presence of elevated troponins. In addition, utilization of composite outcomes implicitly suggests that each component is equivalent in importance. Previous work has demonstrated that varying definitions of composite endpoints (namely MACE) in cardiovascular research studies have been associated with substantially different results and conclusions due to sample sizes being driven largely by more common, less important outcomes (i.e., PCI). 63,64 We therefore individually evaluated the prognostic accuracy of HEART for mortality, MI, and coronary revascularization. We found the sensitivity of a HEART score above the low-risk threshold (≥4) for mortality (95.0%) and MI (97.6%) was substantially better than the sensitivity for coronary revascularization (89.7%). These findings demonstrate that the HEART score offers excellent ability to identify the most initially better than the sensitivity for coronary revascularization. This is a pragmatic approach to defining clinically significant ischemia based on the occurrence of adverse outcomes or need for major intervention. In this diagnostic test accuracy review, we characterized the target outcome as clinically significant cardiac ischemia, utilizing MACE as the reference standard and the HEART score as the index test.

While MACE is the most commonly utilized outcome in cardiovascular research due to its clinical importance, there is notably the potential for incorporation or verification bias—where the diagnostic test contributes to the definition of the disease. For example, consider the dilemma of evaluating the diagnostic value of troponins for identifying “clinically significant” cardiac ischemia with need for revascularization. This is a subjective diagnosis without a criterion standard and any potential outcome assessor would be likely influenced by the presence of elevated troponins. In addition, utilization of composite outcomes implicitly suggests that each component is equivalent in importance. Previous work has demonstrated that varying definitions of composite endpoints (namely MACE) in cardiovascular research studies have been associated with substantially different results and conclusions due to sample sizes being driven largely by more common, less important outcomes (i.e., PCI). 63,64 We therefore individually evaluated the prognostic accuracy of HEART for mortality, MI, and coronary revascularization. We found the sensitivity of a HEART score above the low-risk threshold (≥4) for mortality (95.0%) and MI (97.6%) was substantially better than the sensitivity for coronary revascularization (89.7%). These findings demonstrate that the HEART score offers excellent ability to identify the most initially better than the sensitivity for coronary revascularization. This is a pragmatic approach to defining clinically significant ischemia based on the occurrence of adverse outcomes or need for major intervention. In this diagnostic test accuracy review, we characterized the target outcome as clinically significant cardiac ischemia, utilizing MACE as the reference standard and the HEART score as the index test.

When evaluating patients with chest pain, the emergency physician’s priority is effectively to diagnose “clinically significant” cardiac ischemia. However, as discussed extensively in the cardiovascular literature, there is no objective criterion standard to establish this diagnosis. As a result, MACE is most commonly utilized as the reference standard—a pragmatic approach to defining clinically significant ischemia based on the occurrence of adverse outcomes or need for major intervention. In this diagnostic test accuracy review, we characterized the target outcome as clinically significant cardiac ischemia, utilizing MACE as the reference standard and the HEART score as the index test.

currently recommend that clinicians utilize a clinical decision instrument in the risk stratification of patients with chest pain.11 Our results suggest that the HEART score should be the preferred tool for these purposes, particularly when interested in identifying a low-risk population suitable for immediate discharge.

Table 2
Summary Estimates of the Performance of the HEART Score and TIMI Score

<table>
<thead>
<tr>
<th></th>
<th>No. of Cohorts</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Diagnostic OR (95% CI)</th>
<th>Positive Likelihood Ratio (95% CI)</th>
<th>Negative Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MACE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HEART score ≥ 4</strong></td>
<td>29 (44,202)</td>
<td>95.9 (93.3 to 97.5)</td>
<td>44.6 (38.8 to 50.5)</td>
<td>18.68 (12.44 to 28.06)</td>
<td>1.73 (1.57 to 1.90)</td>
<td>0.09 (0.06 to 0.14)</td>
</tr>
<tr>
<td><strong>HEART score ≥ 7</strong></td>
<td>21 (38,475)</td>
<td>39.5 (31.6 to 48.1)</td>
<td>95.0 (92.6 to 96.6)</td>
<td>12.40 (9.28 to 16.56)</td>
<td>7.89 (5.95 to 10.47)</td>
<td>0.64 (0.56 to 0.72)</td>
</tr>
<tr>
<td><strong>TIMI score ≥ 2</strong></td>
<td>8 (26,397)</td>
<td>87.8 (80.2 to 92.8)</td>
<td>48.1 (38.9 to 57.5)</td>
<td>6.68 (4.50 to 9.90)</td>
<td>1.69 (1.47 to 1.94)</td>
<td>0.25 (0.17 to 0.37)</td>
</tr>
<tr>
<td><strong>TIMI score ≥ 6</strong></td>
<td>3 (18,895)</td>
<td>2.8 (0.8 to 9.6)</td>
<td>99.6 (98.5 to 99.9)</td>
<td>6.69 (3.58 to 12.50)</td>
<td>6.53 (3.53 to 12.08)</td>
<td>0.98 (0.95 to 1.01)</td>
</tr>
<tr>
<td><strong>Death</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HEART score ≥ 4</strong></td>
<td>7 (9,338)</td>
<td>95.0 (87.2 to 98.2)</td>
<td>34.2 (28.7 to 40.2)</td>
<td>9.97 (3.64 to 27.33)</td>
<td>1.45 (1.32 to 1.58)</td>
<td>0.14 (0.06 to 0.38)</td>
</tr>
<tr>
<td><strong>HEART score ≥ 7</strong></td>
<td>5 (8,092)</td>
<td>48.4 (31.7 to 65.4)</td>
<td>91.9 (88.4 to 94.3)</td>
<td>10.56 (5.80 to 19.24)</td>
<td>5.94 (4.17 to 8.45)</td>
<td>0.56 (0.41 to 0.78)</td>
</tr>
<tr>
<td><strong>MI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HEART score ≥ 4</strong></td>
<td>9 (13,032)</td>
<td>97.5 (93.7 to 99.0)</td>
<td>40.5 (33.6 to 47.9)</td>
<td>26.34 (10.55 to 65.76)</td>
<td>1.64 (1.46 to 1.84)</td>
<td>0.06 (0.03 to 0.15)</td>
</tr>
<tr>
<td><strong>HEART score ≥ 7</strong></td>
<td>5 (9,407)</td>
<td>42.5 (28.9 to 57.3)</td>
<td>96.9 (94.5 to 98.3)</td>
<td>22.88 (18.93 to 27.66)</td>
<td>13.58 (10.33 to 17.85)</td>
<td>0.59 (0.47 to 0.75)</td>
</tr>
<tr>
<td><strong>Coronary revascularization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HEART score ≥ 4</strong></td>
<td>6 (8,391)</td>
<td>89.7 (87.2 to 91.8)</td>
<td>41.8 (39.4 to 44.2)</td>
<td>6.27 (4.83 to 8.14)</td>
<td>1.54 (1.47 to 1.62)</td>
<td>0.25 (0.20 to 0.31)</td>
</tr>
<tr>
<td><strong>HEART score ≥ 7</strong></td>
<td>5 (8,092)</td>
<td>30.0 (20.2 to 42.1)</td>
<td>94.5 (91.2 to 96.6)</td>
<td>7.33 (4.19 to 12.84)</td>
<td>5.43 (3.40 to 8.66)</td>
<td>0.74 (0.64 to 0.86)</td>
</tr>
</tbody>
</table>

MACE = major adverse cardiac events; MI = myocardial infarction; TIMI = Thrombolysis in Myocardial Infarction.
engagement, while simultaneously reducing resource utilization. Therefore, the utilization of shared clinical decision aids (particularly with a focus of conveying risk to patients) may increase the utilization of the HEART score by clinicians, appropriately reassure patients, and reduce unnecessary downstream testing in low-risk patients.

Finally, we performed several subgroup analyses to evaluate the accuracy of the HEART score in different clinical contexts. We found that there was no substantial difference in prediction of MACE if the ECG was interpreted by a cardiologist or by an ED physician. While the original studies on the accuracy of the HEART score had ECG interpretation performed by a cardiologist, the primary application of the HEART score is likely to occur in the ED. Therefore, the absence of any substantial impact on prognostic accuracy of the HEART score from ED physician ECG interpretation should reinforce its use among ED clinicians. Furthermore, while the large majority of included studies used conventional troponin T and I assays for computation of the HEART score, a few of the more recent studies included a high-sensitivity troponin assay, although exclusion of these recent studies had minimal impact upon prognostic accuracy.

LIMITATIONS
This review has several strengths. It included a comprehensive search of multiple databases, clear inclusion and exclusion criteria, evaluation of multiple thresholds for each risk score, and multiple subgroup and sensitivity analyses. However, there are important limitations. Some studies were deemed to have potential high risk of bias due to inappropriate exclusion of low-risk patients, although our sensitivity analysis excluding these studies demonstrated similar performance. Additionally, we were unable to meta-analyze AUROC values, as these were not uniformly reported in most studies. However, in the evaluation of a decision instrument such as the HEART score, it can be argued that the summary estimates of sensitivity and specificity are more representative than AUROC, which does not incorporate the relative clinical consequences of false-negative and false-positive diagnoses.66 While minimizing false positives is important in reducing downstream testing that may be of limited value in patients with chest pain, false negatives can result in death or disability. These scenarios should not be treated equally, which is an assumption of AUROC comparisons.66 Importantly, none of the included studies compared accuracy of the HEART score to clinician gestalt. ED clinical decision instruments are rarely compared to clinician gestalt and are often not superior.67 While the performance of the HEART score

Figure 3. Hierarchical summary receiver operating characteristic curves, the bivariate summary points of (specificity, sensitivity), and the 95% confidence regions (dotted lines) of the summary points for (A) HEART score above the low-risk threshold (≥4) and (B) TIMI score above the low-risk threshold (≥2). FN = false negative; FP = false positive; TIMI = Thrombolysis in Myocardial Infarction; TN = true negative; TP = true positive.
among low-risk patients is reassuring, it is unclear whether the score identifies less patients as “low risk” compared to clinician gestalt and may ultimately lead to further downstream testing. That said, clinical gestalt should be utilized to determine pretest probability, which can then be influenced by the HEART score. In this way, clinical gestalt and decision instruments should be viewed as complementary and not competing. Finally, there was minimal evidence in the included studies surrounding the impact of the HEART score on resource utilization. Deployment of the HEART score in a stepped-wedge randomized trial did not demonstrate such savings in resource utilization, although this study was limited by clinicians who were uncomfortable discharging patients with HEART scores below the low-risk threshold. Therefore, future work should focus on how use of the HEART score may impact resource utilization both in the ED and following admission or discharge.

**CONCLUSION**

Our systematic review and meta-analysis demonstrates that the HEART score has excellent sensitivity for identifying low-risk chest pain patients at risk of short-term major adverse cardiac events, robustly supported by findings in external validation studies across a variety of populations, settings, and study designs. A HEART score above the low-risk threshold (≥4) was associated with high sensitivity for short-term major adverse cardiac events and particularly short-term mortality and acute myocardial infarction. Our findings support the use of the HEART score among clinicians for risk stratification of patients presenting with chest pain.

**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.13649/full

Data Supplement S1. Supplemental material.
A Case–Control Study of Sonographic Maximum Ovarian Diameter as a Predictor of Ovarian Torsion in Emergency Department Females With Pelvic Pain

Gavin Budhram, MD, Tala Elia, MD, Jeff Dan, MD, Michele Schroeder, MD, Golien Safain, MD, Walter Schlech, MD, Jennifer Friderici, MS, Alex Knee, MS, Magalie Anthouard, MD, and Elizabeth Schoenfeld, MD, MS

ABSTRACT

Background: Color and power Doppler ultrasound are commonly used in the evaluation of ovarian torsion but are unreliable. Because normal-sized ovaries are unlikely to cause torsion, maximum ovarian diameter (MOD) could theoretically be used as a screening test in the ED. Identification of MOD values below which torsion is unlikely would be of benefit to providers interpreting radiology department or point-of-care pelvic ultrasound.

Objectives: The objective was to determine if sonographic MOD can be used as a screening tool to rule out torsion in selected patients.

Methods: Via a retrospective case–control study spanning a 14-year period, we examined the ultrasound characteristics of patients with torsion and age-matched controls, all presenting to the emergency department with lower abdominal pain and receiving a radiology department pelvic ultrasound for “rule-out torsion.” Standardized data collection forms were utilized. Distributions of MOD were compared and sensitivity, specificity, and likelihood ratios were calculated for multiple cutoffs.

Results: We identified 92 cases of surgically confirmed ovarian torsion and selected 92 age-matched controls. In postmenarchal patients the sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio of 3- and 5-cm MODs were 100% (96%–100%), 30% (20%–41%), 1.4 (1.3–1.7), and 0 and 91% (83%–97%), 92% (83%–97%), 11.2 (5.5–22.9), and 0.09 (0.04–0.19), respectively. The 5-cm MOD, however, excluded an additional 52 of 84 (62%) postmenarchal patients.

Conclusions: A threshold MOD of 5 cm on pelvic ultrasound may be useful to rule out ovarian torsion in postmenarchal females presenting with lower abdominal and pelvic pain.

Ovarian torsion is relatively uncommon, accounting for only 3% of gynecologic emergencies. However, the initial clinical presentation of sudden severe pain, vomiting, palpable adnexal mass,
leukocytosis, and fever is nonspecific and overlaps with many more common clinical entities such as appendicitis, diverticulitis, renal colic, ruptured ovarian cyst, tuboovarian abscess, and others. Atypical presentations are common, with many patients reporting bilateral pain, mild rather than severe pain, intermittent pain, or lack of tenderness on examination. Up to 50% of patients are initially misdiagnosed. Early diagnosis and treatment are necessary to relieve the torsion, restore blood flow, and salvage the ovary and fallopian tube, particularly in women desire to maintain fertility. Given this combination of frequent atypical presentations and significant consequences for missed diagnosis, radiology department ultrasound evaluation for ovarian torsion is common in women presenting to the emergency department (ED) with pelvic pain, despite the rarity of the disease. This evaluation is time-consuming, expensive, and low yield.

In the evaluation of ovarian torsion, overreliance on Doppler is a common pitfall. Studies have found normal Doppler findings in 45% to 61% of torsion cases. In patients with ovarian torsion the most common finding is asymmetric ovarian enlargement, usually to greater than 5 cm. In adults the incidence of ovarian torsion without an accompanying ovarian mass of greater than 3 cm is extremely rare. In contrast, torsion in premenarchal females has been frequently reported with normal-sized ovaries and is thought to be due to the especially mobile uterine adnexa in children. It is unknown whether ovarian torsion could be ruled out in postmenarchal females solely by evaluation of maximum ovarian diameter (MOD) and without Doppler evaluation.

Although the American College of Emergency Physicians recognizes the identification of intrauterine pregnancy as a core application of point-of-care pelvic ultrasound, the evaluation of adnexal pathology is categorized as an “adjunct or emerging application,” because Doppler evaluation requires more expertise. If a size cutoff with appropriate accuracy could be identified, clinicians able to obtain accurate ovarian diameter measurements at the bedside could theoretically forego radiology department evaluation with Doppler in select patients, potentially improving a number of patient-centered outcomes such as length of stay, discomfort, and cost. Nevertheless, identification of threshold diameter values below which torsion is extremely unlikely would be of benefit both to providers utilizing radiology department ultrasound and to those comfortable with point-of-care ultrasound (POCUS) evaluation of the adnexa. We sought to evaluate the test characteristics of MOD with the hypothesis that either 3 or 5 cm may have an appropriately discriminatory sensitivity to rule out ovarian torsion without an accompanying Doppler evaluation.

**METHODS**

**Study Design and Setting**
This was a retrospective case–control study over a 14-year period from January 1, 2000, to December 31, 2014, in a large tertiary care ED with a combined adult and pediatric volume of 100,000 to 120,000 patients/year. This study was approved by the hospital institutional review board.

**Selection of Participants**
All female patients age 2 to 100 presenting to the ED during the study period with a chief complaint of “pelvic” or “lower abdominal” pain and who received a radiology department transvaginal and/or transabdominal pelvic ultrasound with an indication of “rule-out torsion” were identified for inclusion in the study. These patients were identified by screening the ED database for female patients presenting with International Classification of Diseases (ICD-9) codes 625.9 (pelvic pain), 789.0 (abdominal pain), 789.03 (right lower abdominal pain), and 789.04 (left lower abdominal pain). The group was then further screened via the Current Procedural Terminology codes 76857 (transabdominal pelvic ultrasound, not pregnant), 76815 (“transabdominal pelvic ultrasound, pregnant”), 76830 (“transvaginal pelvic ultrasound, not pregnant”), and 76817 (“transvaginal pelvic ultrasound, pregnant”). The case and control groups were identified from this larger group. The torsion (case) group was defined as patients from this group who had a surgically confirmed diagnosis of ovarian torsion (ICD-9 code 620.5). The control group was matched 1:1 based on 5-year age strata (2–4, 5–9, 10–14, 15–19, 20–24, 25–29, etc.) and drawn from the same group as above, with any diagnosis other than ovarian torsion. Controls were included sequentially in each age strata until matching was achieved. Patients were excluded from either group if both ovaries were not completely visualized and measured or if no ultrasound report was available. Patients were also excluded from the torsion group if manual chart review did not confirm torsion via operative findings. Patients were
excluded from the control group if torsion was identified on subsequent visits within 3 months, and these patients were included in the torsion group.

**Study Protocol**

Selection of the study population and cohorts was performed by a departmental data analyst who was not an investigator in the study and who was blinded to the study purpose and hypothesis. Via manual chart review, data were abstracted from paper records for patients seen from 2000 to 2008 and from electronic records for patients seen from 2008 to 2014 (Cerner, Cerner Corp.). Standardized data abstraction forms were used and information entered into REDCap (Vanderbilt). All data abstractors underwent a 1-hour training session and reviewed five practice charts each under the supervision of the principal investigator. Abstractors were not blinded to the study group, nor to the study hypothesis. Ten percent of the charts were selected via random sampling to calculate inter-rater reliability with regard to MOD.

**Measures**

We recorded demographic information as well as relevant aspects of the patient’s presentation such as fever, location of pain, and palpable pelvic mass. From ultrasound results, MOD and arterial/venous Doppler waveform information was obtained bilaterally. Surgical diagnosis was obtained from operative notes. Menarchal status was documented as identified in the chart, but if menarchal status was not noted patients younger than 12 were considered premenarchal, as median age of menarche in the United States is 12.22

The primary outcome was the association between ovarian torsion and a MOD of >3 cm as assessed by ultrasound in postmenarchal patients. Secondarily, we also sought to examine this association in premenarchal patients.

**Data Analysis**

Our sample size was limited by the cohort of patients with ovarian torsion since 100% of eligible cases were included. However, we did wish to report test characteristics of the 3- and 5-cm threshold values; therefore, estimating the width of the 95% confidence interval (CI) around these estimates was of importance. Based on previously published data and our ED’s yearly volume, we anticipated approximately 100 eligible cases (and therefore 100 controls) to be included in the study. Given this, the width of the 95% CI around estimated test characteristics of 0.50 to 0.90 would range from 20 to 13 points, respectively. With a reduction in sample size, these CIs would be wider.

Using the torsion and control cohorts, sensitivity, specificity, and likelihood ratios for the 3- and 5-cm threshold values were derived. Data analysis was performed using Stata Data Analysis and Statistical Software v15 (StataCorp LLC). A priori, we decided to analyze premenarchal subjects separately from postmenarchal subjects, as ovarian torsion is thought to occur more frequently with normal-sized ovaries in children.

**RESULTS**

After exclusions (Figure 1), 92 cases were identified with both surgically confirmed ovarian torsion and an ultrasound report. Ninety-two age-matched control patients who received a pelvic ultrasound to rule out torsion, but in whom torsion was not diagnosed, were also selected. Two patients with torsion presented twice for pelvic pain and were found to have MOD > 5 cm and normal Doppler studies. After gynecology consultation these patients were discharged, but presented again and found to have torsion at surgery. These patients were included in the torsion cohort and their index visit was excluded.

The characteristics of the included patients are seen in Table 1. Cases and controls were matched for age (median age = 28 years). Cases were slightly more likely to be white versus other race. Torsion patients were more likely to present with nausea and vomiting (50.8% vs. 19.6%) and a palpable pelvic mass (19.6% vs. 0%). Table 2 shows the final diagnoses in the control cohort.

Maximum ovarian diameter and Doppler findings by case and control status, in both pre- and postmenarchal patients, is summarized in Table 3. Zero patients in the postmenarchal torsion group had a MOD < 3 cm, while seven of 81 (8.6%) had MOD ≤ 5 cm. Among patients with torsion, 30 of 77 (39%) had normal arterial and venous Doppler signal. Doppler information was reported in a majority but not all of the ultrasound reports.

Table 4 shows test characteristics of both threshold MOD values as well as abnormal Doppler signals. In postmenarchal patients, the sensitivity for a 3-cm threshold value was 100% (95% CI = 95.5–100%) but the specificity was 30% (95% CI = 20.8–41.1%). Sensitivity decreased to 91% (95%
CI = 83%–96.5%) using a 5-cm MOD and specificity increased to 92% (95% CI = 83.9%–96.7%). In the postmenarchal group, the negative likelihood ratios for the 3- and 5-cm threshold values were 0 and 0.09 (95% CI = 0.04–0.19), respectively. Abnormal Doppler signal was a very specific finding for torsion in our study group (98%, 95% CI = 92.1%–99.7%), but not sensitive (61%, 95% CI = 49.2%–72%).

Figure 2 illustrates the MOD between the torsion and control cohorts in the premenarchal and postmenarchal populations. Regarding the 3-cm MOD, no patients (either postmenarchal or premenarchal) were found to have torsion with largest ovarian diameter < 3 cm. In the premenarchal group, torsion tended to occur within a smaller range of ovarian diameter, between 4 and 7 cm. In the postmenarchal group, the range was much larger, between 3 and 16 cm.

Figure 3 shows the distribution of ovarian sizes in the postmenarchal population. Two distinctive but overlapping distributions of MOD are identified between the torsion and control cohorts, with the torsion cohort heavily skewed toward larger MOD. Although not one of our predefined outcome measures, an MOD value of <4 cm was only found in two of 80 postmenarchal patients with torsion. Of the 10% of charts selected to assess for inter-rater reliability with regard to MOD, there was complete agreement between data abstractors and the second reviewer resulting in $\kappa = 1.0$. 
DISCUSSION

Over a 14-year period, the ED treated 1,351,204 patients and ovarian torsion was only diagnosed in 103 patients, approximately 0.007% of all ED visits. As this hospital is a referral center, it is possible that the true population prevalence is even lower. Although torsion is historically a poorly recognized clinical entity with difficult diagnosis, our findings are consistent with the largest study of ovarian torsion done to date by Houry and Abbott, who found an overall prevalence of 0.006%.

The most common grayscale sonographic abnormality in torsion is asymmetric ovarian enlargement, usually to greater than 5 cm, frequently due to an underlying mass. It is extremely rare in adults, the incidence of ovarian torsion without an accompanying ovarian mass greater than 3 cm. 

Lee et al. found ovarian enlargement in 100% of 32 torsion cases, with single largest diameter ranging from 5 to 33 cm. Houry and Abbott described a mean ovarian size of 9.5 cm in a series of 87 torsion cases, with 89% measuring greater than 5 cm. In 2010, Huchon

Table 1
Baseline Patient Characteristics

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Control (n = 92)</th>
<th>Case (n = 92)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>28 (15–41)</td>
<td>28 (15–40)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>36 (39.1)</td>
<td>50 (54.4)</td>
</tr>
<tr>
<td>Black</td>
<td>13 (14.1)</td>
<td>8 (8.7)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>14 (15.2)</td>
<td>13 (14.1)</td>
</tr>
<tr>
<td>Other</td>
<td>8 (8.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Refused/unknown</td>
<td>21 (22.8)</td>
<td>21 (28.6)</td>
</tr>
<tr>
<td>Reproductive status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenarchal</td>
<td>8 (8.7)</td>
<td>12 (13.0)</td>
</tr>
<tr>
<td>Postmenarchal</td>
<td>84 (91.3)</td>
<td>80 (87.0)</td>
</tr>
<tr>
<td>Signs/symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fever</td>
<td>3 (3.3)</td>
<td>7 (7.6)</td>
</tr>
<tr>
<td>Nausea/vomiting</td>
<td>18 (19.6)</td>
<td>55 (59.8)</td>
</tr>
<tr>
<td>Location of pain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right lower quadrant</td>
<td>27 (29.4)</td>
<td>46 (50.0)</td>
</tr>
<tr>
<td>Left lower quadrant</td>
<td>16 (17.4)</td>
<td>22 (23.9)</td>
</tr>
<tr>
<td>Bilateral lower quadrants</td>
<td>32 (34.7)</td>
<td>21 (22.3)</td>
</tr>
<tr>
<td>Suprapubic</td>
<td>17 (18.4)</td>
<td>3 (3.3)</td>
</tr>
<tr>
<td>Palpable pelvic mass</td>
<td>0 (0.0)</td>
<td>18 (19.6)</td>
</tr>
</tbody>
</table>

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

Table 2
Clinical Diagnoses in the Control Cohort

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal/pelvic pain NOS</td>
<td>36 (74)</td>
</tr>
<tr>
<td>Ovarian cyst</td>
<td>19 (21)</td>
</tr>
<tr>
<td>Pelvic inflammatory disease</td>
<td>11 (12)</td>
</tr>
<tr>
<td>Hemorrhagic/ruptured ovarian cyst</td>
<td>8 (9)</td>
</tr>
<tr>
<td>Uterine fibroids</td>
<td>7 (8)</td>
</tr>
<tr>
<td>Colitis</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Constipation</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Gastroenteritis</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (5)</td>
</tr>
</tbody>
</table>

Data are reported as n (%). NOS = not otherwise specified.

Table 3
US Findings by Case Status

<table>
<thead>
<tr>
<th></th>
<th>Torsion</th>
<th>Control</th>
<th>&lt;3 cm premenarchal</th>
<th>&gt;3 cm</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤3 cm</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;3 cm</td>
<td>11</td>
<td>0</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>11</td>
<td>6</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3 cm postmenarchal</td>
<td>0</td>
<td>26</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;3 cm</td>
<td>81</td>
<td>60</td>
<td>141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>81</td>
<td>86</td>
<td>167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤3 cm total cohort</td>
<td>0</td>
<td>32</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;3 cm</td>
<td>92</td>
<td>60</td>
<td>152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>92</td>
<td>92</td>
<td>184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5 cm premenarchal</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;5 cm</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>11</td>
<td>6</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5 cm postmenarchal</td>
<td>7</td>
<td>79</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;5 cm</td>
<td>74</td>
<td>7</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>81</td>
<td>86</td>
<td>167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5 cm total cohort</td>
<td>11</td>
<td>85</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;5 cm</td>
<td>81</td>
<td>7</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>92</td>
<td>92</td>
<td>184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal arterial waveform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>45</td>
<td>1</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>34</td>
<td>88</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>79</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal venous waveform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>40</td>
<td>1</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>36</td>
<td>88</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>76</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal arterial or venous waveform</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>47</td>
<td>2</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>30</td>
<td>87</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>77</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

US = ultrasound.
et al. even used a logistic regression analysis on a cohort of 142 patients with acute pelvic pain to derive a scoring system for torsion based on five independent predictors for torsion: cyst > 5 cm on ultrasound, pain duration < 8 hours, vomiting, spontaneous unilateral abdominal or lumbar pain, and absence of leukorrhea and metrorrhagia. They then prospectively evaluated a group of 35 women with pelvic pain and assigned them to a high-risk versus low-risk categories using the scoring tool, showing 0% probability of torsion in the low-risk group and 75% probability in the high-risk group.

In our study, the 3-cm threshold value demonstrated excellent sensitivity (100%) for torsion in the postmenarchal and total cohorts but was nonspecific. The 5-cm threshold value was somewhat less sensitive.

Table 4
Test Characteristics

<table>
<thead>
<tr>
<th>Test Characteristics</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR (+)</th>
<th>LR (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 cm</td>
<td>100% (96.1%–100%)</td>
<td>35% (25.1%–45.4%)</td>
<td>1.5 (1.3–1.8)</td>
<td>0</td>
</tr>
<tr>
<td>5 cm</td>
<td>88% (79.6%–93.9%)</td>
<td>92% (84.9%–96.9%)</td>
<td>11.6 (5.7–23.7)</td>
<td>0.13 (0.07–0.23)</td>
</tr>
<tr>
<td>3 cm</td>
<td>100% (95.5%–100%)</td>
<td>30% (20.8%–41.1%)</td>
<td>1.4 (1.3–1.7)</td>
<td>0</td>
</tr>
<tr>
<td>5 cm</td>
<td>91% (83%–96.5%)</td>
<td>92% (83.9%–96.7%)</td>
<td>11.2 (5.5–22.9)</td>
<td>0.09 (0.04–0.19)</td>
</tr>
</tbody>
</table>

LR(+) = positive likelihood ratio; LR(-) = negative likelihood ratio.

Figure 2. Distribution of ovarian diameter (cm) in control (left) and torsion (right) patients. Box plots represent 25th (Q1)–75th (Q3) percentile, with 50th percentile noted as white midline. Whiskers represent the distribution’s lower inner fence (Q1 – (1.5 × interquartile range)) and upper inner fence (Q3 + (1.5 × interquartile range)). Dots represent values outside of the inner fences and are considered outliers.

Figure 3. Distribution of mean ovarian diameter in postmenarchal patients.
(88% in the total cohort, 91% in the postmenarchal cohort), but was much more specific (92% in both the total and the postmenarchal cohorts). Of note, in our postmenarchal control population a large proportion of patients (52/84, 62.0%) had MOD > 3 and < 5 cm. Using a 5-cm rather than a 3-cm MOD in these patients would have excluded torsion in these additional 52 patients. Given the previously reported extremely low prevalence of ovarian torsion in the general population, our findings suggest that 5 cm might be the more useful threshold value in patients presenting with pelvic pain.

Our study confirms the poor sensitivity of Doppler in ruling out ovarian torsion that has been described in previous studies. Assessment of arterial and venous blood flow in the ovaries can be technically difficult due to body size, interposed bowel gas, or other patient factors. Even if properly identified and measured Doppler findings can still be unreliable due to the dual blood supply from the uterine and ovarian arteries and also because torsion may be intermittent.

Our findings suggest that screening decisions for patients with lower abdominal pain should be based on ovarian size rather than Doppler findings. Clinicians who rely on a radiology department ultrasound should be reassured by normal-sized ovaries and cautious when ovaries are larger than 5 cm even if arterial and venous waveforms are normal. Physicians who are comfortable with POCUS evaluation of the adnexae for torsion may be justified in solely measuring ovarian diameter without Doppler.

Clinicians should be careful not apply these findings to premenarchal patients. Ovarian torsion in patients with normal-sized ovaries has been reported in the pediatric literature and is thought to occur in up to 25% of cases of adnexal or ovarian torsion. Our findings in premenarchal patients are consistent with previous studies. Hypotheses for the etiology of torsion of normal adnexa include excessively mobile mesovaria or fallopian tubes, congenitally long pelvic ligaments, tubal spasm, or abrupt changes in intraabdominal pressure. Among premenarchal girls with torsion in our study, eight of 12 had a MOD > 5 cm and 12 of 12 had a MOD > 3 cm. This suggests that although a cutoff of 3 cm may be valid in this population, further study is needed and the clinician should still maintain a high index of suspicion for torsion even with normal-sized ovaries.

LIMITATIONS

Due to the rare nature of ovarian torsion and our exclusion criteria, only 92 patients with torsion over the 14-year period were identified. Therefore, CIs around some of our findings are large.

Our study also has limitations inherent to a retrospective case–control study design. Selection bias is a possibility as charts were retrospectively pulled for analysis. We attempted to compensate for spectrum bias by including all patients who received a pelvic ultrasound to rule out torsion, so that our study population would reflect the actual spectrum of illness in clinical practice. In addition, we tried to make sure our control population was appropriately selected to include the same presenting complaint, symptoms, and receiving the same investigational studies as the torsion cohort.

Since only cases that received an ultrasound for rule-out torsion were included, the reported numbers for sensitivity and specificity may artificially be increased (incorporation bias). Torsion was diagnosed surgically in 18 of 121 and by computed tomography in four of 121 cases, all in patients who did not receive an ultrasound. Only six of 92 patients in the control cohort went on to receive surgery, subjecting our study to “double-criterion-standard” bias, which may artificially elevate the sensitivity and specificity.

As this study included only radiology department ultrasounds, it is difficult to draw conclusions regarding POCUS examinations. Given the large number of examinations needed for assessment of this relatively rare diagnosis, and the completeness and searchability of our POCUS database over the 14-year time period, we were unable to directly study POCUS for the detection of torsion. Large multicenter studies would be needed to prospectively evaluate the clinician’s ability to rule out torsion at the bedside. Also, while this study demonstrates a distinct and useful size difference between torsed and nontorsed ovaries, it does not address the clinician’s ability to reliably visualize ovaries and measure ovarian diameter. To our knowledge no studies to date have assessed this skill. Data abstractors in this study were not blinded to study group or study hypothesis; however, inter-rater reliability among a 10% sample of the study population showed 100% agreement with the reported MOD in both groups.

In addition to measurement of length/width/height, ovarian volume (defined as length × width × height × 0.5) is a commonly used sonographic means of describing ovarian size. The upper limit of normal for ovarian
volume is 20 cm.\textsuperscript{3,29} We chose to use MOD rather than ovarian volume because this variable was more commonly identified in the radiology report and because it is simpler for novice sonographers to measure a maximum diameter than to calculate a volume. It is possible that volume is a more useful measurement.

**CONCLUSIONS**

A threshold maximum ovarian diameter of 5 cm on pelvic ultrasound may be useful to rule out ovarian torsion in postmenarchal females presenting with lower abdominal and pelvic pain. Doppler ultrasound is not sensitive enough to be used as a rule-out test for ovarian torsion.

**References**

Adoption and Utilization of an Emergency Department Naloxone Distribution and Peer Recovery Coach Consultation Program

Elizabeth A. Samuels, MD, MPH, MHS, Janette Baird, PhD, Eunice S. Yang, and Michael J. Mello, MD, MPH

ABSTRACT

Objective: Rising rates of opioid overdose deaths require innovative programs to prevent and reduce opioid-related morbidity and mortality. This study evaluates adoption, utilization, and maintenance of an emergency department (ED) take-home naloxone and peer recovery coach consultation program for ED patients at risk of opioid overdose.

Methods: Using a Reach Effectiveness Adoption Implementation Maintenance (RE-AIM) framework, we conducted a retrospective provider survey and electronic medical record (EMR) review to evaluate implementation of a naloxone distribution and peer recovery coach consultation program at two EDs. Provider adoption was measured by self-report using a novel survey instrument. EMRs of discharged ED patients at risk for opioid overdose were reviewed in three time periods: preimplementation, postimplementation, and maintenance. Primary study outcomes were take-home naloxone provision and recovery coach consultation. Secondary study outcome was referral to treatment. Chi-square analysis was used for study period comparisons. Logistic regression was conducted to examine utilization moderators. Poisson regression modeled utilization changes over time.

Results: Most providers reported utilization (72.8%, 83/114): 95.2% (79/83) provided take-home naloxone and 85.5% (71/83) consulted a recovery coach. There were 555 unique patients treated and discharged during the study periods: 131 preimplementation, 376 postimplementation, and 48 maintenance. Postimplementation provision of take-home naloxone increased from none to more than one-third (35.4%, \( p < 0.001 \)), one-third received consultation with a recovery coach (33.1%, 45/136), and discharge with referral to treatment increased from 9.16% to 20.74% (\( p = 0.003 \)). Take-home naloxone provision and recovery coach consultation did not depreciate over time.

Conclusions: ED naloxone distribution and consultation of a community-based peer recovery coach are feasible and acceptable and can be maintained over time.

From the Department of Emergency Medicine, Alpert Medical School of Brown University (EAS, JB, MM), Providence, RI; and the Geisinger Commonwealth School of Medicine (ESY), Scranton, PA.

Received March 1, 2018; revision received June 29 and July 6, accepted July 14, 2018.

Presented at the American College of Emergency Physicians Scientific Assembly, Boston, MA, October 27, 2015; and the Society for Academic Emergency Medicine Annual New England Regional Meeting, Boston, MA, April 1, 2015.

This study was supported by a National Institute on Drug Abuse and Emergency Medicine Foundation Training Award in Drug Abuse and Addiction.

The authors have no potential conflicts to disclose.

Author Contributions: EAS, MUM, and JB conceived the study and designed and piloted the survey; EAS and MUM supervised the conduct of the EMR review and data collection; EAS undertook obtaining institutional review board approval and participant recruitment; EAS and ESY conducted EMR data collection; EAS and JB supervised and conducted data analysis; EAS drafted the manuscript; and all authors contributed substantially to its revision. EAS takes responsibility for the paper as a whole.

Supervising Editor: Ula Hwang, MD.

Address for correspondence and reprints: Elizabeth A. Samuels, MD, MPH, MHS; e-mail: elizabeth_samuels@brown.edu.


© 2018 by the Society for Academic Emergency Medicine
doi: 10.1111/acem.13545
Unintentional opioid overdose deaths have increased 200% in the past two decades. Although emergency departments (EDs) are caring for an increasing number of people with opioid use disorder (OUD) and opioid overdose, a minority are referred to an outpatient treatment program or admitted to inpatient detox. Following an overdose, individuals are at higher risk of death, but some studies have also shown increased enrollment in OUD treatment. Each overdose event and related ED visit, therefore, presents a critical opportunity to prevent not only future overdose death, but also engagement in treatment.

Community opioid overdose education and naloxone distribution (OEND) programs have shown that lay people, including intravenous drug users, can reliably administer naloxone for overdose rescue and a reduction in opioid overdose mortality. Researchers evaluating OEND programs have observed a decline, not an increase, in opioid use among those receiving take home naloxone, as well as cost-effectiveness in high-risk populations. In light of this evidence, the Centers for Disease Control and Prevention have recommended naloxone distribution and linkage to peer recovery coaches to provide addiction treatment navigation. In response to escalating opioid overdose deaths, in 2014, Rhode Island (RI) ED physicians collaborated with the RI Department of Health, the RI Department of Behavioral Healthcare, Developmental Disabilities, and Hospitals; and the Anchor Recovery Community Center, a community-based peer recovery organization, to implement these recommendations in an ED setting.

Previous investigations have demonstrated provider willingness to provide patients with take-home naloxone, patient factors influencing acceptance of take-home naloxone, and feasibility of ED OEND with ED-based health promotion advocates, but provider adoption of ED naloxone distribution into clinical practice is unknown. Similarly, peer recovery coaches have also been shown to be an effective component of outpatient addiction treatment services navigation and support, but evaluation of their utilization in the ED is limited. This study aims to evaluate the adoption, utilization, and maintenance of an ED OEND program that uses community-based peer recovery coaches for addiction treatment navigation for ED patients with OUD and those treated after opioid overdose.

**METHODS**

**Intervention**

From 2010 to 2014, the rate of opioid overdose death in RI increased from 10.5 to 19.8 deaths per 100,000 people. In 2014, the RI Department of Health released new regulations enabling direct provider-to-patient naloxone distribution. RI ED physicians pharmacists, public and behavioral health professionals, and members of Anchor Recovery Community Center (Anchor), a community-based peer recovery organization, subsequently collaborated to design and implement an ED OEND program, the Lifespan Opioid Overdose Prevention (LOOP) Program, in two RI EDs in September 2014. LOOP provides ED patients at risk of opioid overdose 1) take-home intranasal naloxone and patient education for overdose rescue and 2) recovery coach consultation for addiction treatment support and navigation after the ED visit. The two affiliated hospitals where LOOP was implemented were a Level I trauma center (Site A) with approximately 110,000 annual adult ED visits and a community hospital (Site B) with approximately 50,000 annual ED visits. At the time of program implementation, Site A cared for a median of 44 opioid overdoses a month and Site B cared for a median of eight opioid overdoses a month. Both had social work and psychiatry available for ED consultation, but neither hospital provided specialized ED, inpatient, or outpatient addiction treatment services.

Key hospital stakeholders from the departments of pharmacy, social work, emergency medicine, psychiatry, nursing, risk management, legal services, and hospital administration participated in program design and implementation at both hospitals. Hospital administrators provided financial support for this initiative, purchasing contents of the take-home naloxone rescue kits as a community service. Terms of recovery coach consultation were outlined in a mutually agreed upon memorandum of understanding between the hospitals and Anchor.

Providers could order a take-home naloxone rescue kit and recovery coach consultation through an EMR order set. Provision of take-home naloxone included patient education about overdose prevention, response, and naloxone administration for overdose reversal through an educational video, bilingual printed instructions, and when available, in-person counseling by a recovery coach. Take-home naloxone kits included two doses of 2 mg of intranasal...
naloxone, a mucosal atomizer device, and pictorial and verbal assembly and administration instructions in English and Spanish (see Data Supplement S1, Appendix A, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13545/full). Kit contents were purchased by the hospital administration, assembled by the inpatient hospital pharmacy, stored in ED medication dispensing machines, and then retrieved by the ED nurse and given to the patient prior to ED discharge.

Recovery coaches were individuals in addiction treatment for 2 years or longer, had completed a 36-hour peer recovery coach training, and were employed by Anchor. Coach hiring, training, and supervision were conducted by Anchor. All coaches also underwent additional HIPAA training. Due to initial funding limitations, during the study period coaches were available Friday 8 PM to Monday 8 AM, when it was assumed that the highest volume of patients with opioid-related ED visits would present to the ED. To consult a recovery coach, ED providers would place an EMR order for consultation and ED secretaries would page an on-call coach through an answering service. Coaches arrived in the ED within 30 minutes of consultation. Using motivational interviewing techniques and a stages of change behavioral framework; coaches assessed patients’ readiness to seek treatment; identified risk factors for recurrent overdose; and provided naloxone teaching, individualized support, and addiction treatment navigation at the time of and after their ED visit.

Prior to LOOP implementation, the study principal investigator (PI) educated all ED providers and staff about program services and protocols at residency didactic conferences, faculty meetings, nursing change of shift roll calls, e-mail announcements, and signs posted in each ED work area. Updates were sent to prescribers every 3 to 6 months about overall program utilization and treatment linkage among patients receiving a recovery coach.

**Study Design**

Using an adapted Reach Effectiveness Adoption Implementation Maintenance (RE-AIM) framework, we conducted a retrospective mixed-methods evaluation of program adoption, utilization, and maintenance. Provider adoption was assessed through a novel, retrospective provider survey administered 7 months after program implementation in March 2015. At the time of the study, both EDs were staffed by a total of 165 providers, including attending and resident physicians and advance practice providers (APPs). Most providers worked at both sites. Program utilization was assessed through a retrospective electronic medical record (EMR) review of ED patients who were treated and discharged after an opioid overdose or who the ED provider documented as having opioid misuse or OUD from January 2014 to August 2015. Patients’ first ED visits during the study period were reviewed; subsequent visits were excluded. Patients admitted, who expired, who left against medical advice, or who eloped were excluded from the analysis since LOOP was intended for ED patients being discharged and required provider evaluation for provision of take-home naloxone, recovery coach consultation, or referral to treatment (Figure 1).

Emergency department visits were evaluated in three distinct, a priori defined study time periods (Figure 1): preimplementation, January to February 2014; postimplementation, the 6 months following program implementation, September 2014 to February 2015; and maintenance, the 12th month after program implementation, August 2015. Time periods were selected to account for seasonal variation in ED visits for opioid overdose and to ensure an adequate count of records to include in the analysis. Patient data were not collected in between the time periods.

Based on the known frequency of ED visits for opioid overdose and OUD, we estimated that 600 patients would be eligible for take-home naloxone and/or recovery coach consultation postimplementation and assumed that during the preimplementation period, less than 10% of eligible patients would be given take-home naloxone or discharged with referral to treatment. We anticipated this would increase by at least 10% to 20%, a small to moderate effect size, postimplementation, with no more than a 10% decrease during the maintenance period. Using a binomial test of difference in proportions between the start and peak of implementation, we estimated a power of 0.80, \( \alpha = 0.05 \), to test for differences in implementation by including a minimum of 100 patients in each study period.

**Measurements**

Primary study outcomes were take-home naloxone distribution and recovery coach consultation. Provision of naloxone and recovery coach consultation were measured through review of documented EMR orders. Secondary study outcome was discharge with referral
Referral to treatment was defined as provider documented discussion with an outpatient treatment provider and/or a documented follow-up plan at a specific treatment program in the provider’s note and/or patient discharge instructions. Confirmation of outpatient treatment enrollment was outside the scope of this study and therefore not conducted.

Programmatic reach, effectiveness, implementation, and maintenance were assessed through a retrospective EMR review conducted in accordance to the accepted...
standards and reported using Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. Records were initially selected for review through an EMR search by ED informatics specialists. EMR fields (chief complaint, history of present illness, home medications, orders, discharge diagnosis, discharge instructions, and discharge prescriptions) were searched for keywords related to opioid overdose and opioid use identified by the study team (see Data Supplement S1, Supplemental Table 1).

All EMRs at the two study EDs during the study period were screened for study inclusion. Patient search lists were merged to remove duplicates. Each record from the initial screen was reviewed by a study research assistant (RA) and selected for inclusion if the patient was treated after an accidental, nonfatal opioid overdose or who the emergency medicine provider identified in their documentation as having opioid misuse or OUD, including lapses in medication for OUD, polysubstance use including opioids, intravenous drug use, or recreational use of prescription or illicit opioids. Patients were not formally assessed for OUD at the time of their ED visit or during the retrospective EMR review. An opioid overdose was defined as opioid use resulting in decreased mental status or respiratory depression necessitating the use of naloxone prior to or during the ED visit. Record selection was limited to these groups because it represented the most obvious population to offer LOOP services. Visits were excluded for patients presenting with non-opioid-related intoxication or overdose; those stable on medication for OUD; patients treated for a problem not related to opioids, suicidality, or substance use; and patients who were pregnant, incarcerated, or less than 18 years of age.

A coding manual and standardized data collection instrument were developed by senior study staff for data extraction. Data extractors were RAs with prior experience reviewing and extracting EMR data. They were not involved in program development and were blinded to provider education, quality improvement initiatives, and study objectives. All data extractors were provided uniform training in data collection. Null values were assigned for data fields not recorded or documented. The primary data extractor reviewed all records identified by information services (IS) and selected ED visits for data extraction meeting inclusion criteria. Data extraction underwent regular validation checks by senior study personnel. The PI reviewed 20% of extracted records to monitor accuracy and consistency for record selection for study inclusion. Any discrepancies were resolved by the PI and used for retraining. Ten percent of all reviewed records were also reviewed by a second data extractor. The study PI trained, supervised, and reviewed both data extractors in the same manner.

All repeat visits were removed prior to analysis. To evaluate programmatic reach, we measured the proportion of the target population receiving take-home naloxone or recovery coach consultation at their first ED visit during the study period. Recovery coach utilization proportions were calculated for patients presenting to the ED during available hours, Friday 8 PM to Monday 8 AM. Comparisons between the pre- and postimplementation periods were made to determine program effectiveness on discharge with referral to treatment. Implementation moderators were identified by assessing the impact patient and ED factors on take-home naloxone distribution, recovery coach consultation, and discharge with referral to treatment. Programmatic maintenance was determined examining utilization trends over time through the postimplementation and maintenance periods.

Adoption was measured by self-report using a novel survey instrument. Since providers work in teams of attending physicians along with residents and/or APPs, we were unable to retrospectively observe individual provider clinical decision making. Survey questions were developed and reviewed by a panel of content experts and tested for understandability with EM clinicians excluded from survey participation. Questions covered provider reported LOOP utilization, knowledge of overdose risk factors, and program-specific utilization barriers. The questionnaire (see Data Supplement S1, Appendix B) was an anonymous, 15-minute, online survey administered 7 months after program implementation on a Qualtrics interface. All non-per-diem physicians and APPs at study EDs were eligible to participate. They were e-mailed a link to the online survey and received a $20 gift card upon completion. Study protocols were reviewed and approved by the Rhode Island Hospital Institutional Review Board.

Data Analysis
Survey results were analyzed as proportions of responses. All responses were included in the analysis regardless of survey completion. Extracted EMR data were reviewed, cleaned, and imported into STATA v14.2 (StataCorp). Inter-rater agreement was calculated for three data categories: patient demographics, visit
characteristics (i.e., reason for visit, visit disposition), and study outcome variables. To determine inter-rater reliability, Kappa scores were calculated and averaged across each extraction category.

In conducting descriptive statistical analyses, chi-square testing was used to compare patient demographics, visit characteristics, linkage to treatment, take-home naloxone distribution, and recovery coach consultation between study periods. Fischer exact testing was used where appropriate. Recovery coach consultation was only assessed during advertised available hours (Friday 8 PM-Monday 8 AM). The median length of stay (LOS) in the ED of discharged patients was compared by services utilized using a Kruskal-Wallis test. To adjust for illness severity, patients requiring repeat naloxone administration in the ED and/or who had a documented oxygen requirement were excluded from the LOS comparison.

Logistic regressions were conducted to identify moderators of take-home naloxone distribution, recovery coach consultation, and referral to treatment. A priori variables included in the regression model included all patient demographic variables and ED visit characteristics (see Data Supplement S1, Supplemental Table 3). Subgroup chi-square and logistic regression analyses were similarly performed for patients who presented after an opioid overdose.

To evaluate utilization changes over time, we conducted a Poisson regression using SAS/STAT v 9.3 PROC GENMOD to model the effect of time, site, and time by site interaction. Each outcome—take-home naloxone, recovery coach consultation, and referral to treatment—was modeled separately. The model was designed to compare utilization rates in the pre- and post-implementation periods and estimate growth and patterns of utilization over the first 6 months of implementation and utilization rate deterioration over time. Parsimonious models included time, clinical site, and additional patient level characteristics (sex, race, opioid overdose) to evaluate the effect of these covariates on services uptake. The Akaike Information Criterion and the Bayesian Information Criterion were used to assess the overall fit of the predictive models and the scaled deviance criteria were examined for values greater than 1, indicating possible overdispersion necessitating model adjustment.

To assess change in naloxone distribution, recovery coach consultation and referral to treatment across the three phases of the program adoption (preimplementation, post-implementation, and maintenance), we used the Cochrane Armitage trend test, to assess change in these program elements across the time periods.

RESULTS

Provider Adoption: Provider Survey Results
Among 165 providers, 114 (69.1%) participated (Data Supplement S1, Supplemental Table 2). Half were attending physicians, 35.8% resident physicians, and 11.9% APPs. The majority reported utilizing LOOP (72.8%, 83/114). Of these, nearly all (95.2%, 79/83) reported providing take-home naloxone and 85.5% (71/83) consulted a recovery coach. Most utilizing providers (83.1%, 69/83; on a scale of 1 to 5, mean = 4.3, 95% confidence interval [CI] = 4.1–4.5) reported offering LOOP services most of the time or always for patients who had an opioid overdose and over half (55.4%, 46/83; on a scale of 1 to 5, mean = 3.6, 95% CI = 3.4–3.9) reported using it most of the time or always for patients requesting addiction treatment. Few had difficulty providing take-home naloxone (3.6%, 3/83). Less than one-third had difficulty consulting a recovery coach (32.5%, 27/83). The most commonly cited barriers to recovery coach consultation were attempted contact outside of available hours (85.2%, 23/27) and the patient wanting to leave the ED prior to recovery coach arrival (51.9%, 14/27).

Reach and Effectiveness: EMR Review Results
A total of 5,630 records were reviewed for study inclusion. Figure 1 shows flow of study participants. Primary reasons for record exclusion was intoxication or overdose not related to opioid use. Reviewers had high inter-rater agreement in each category, demographic agreement (96.2%, κ = 0.93), visit characteristics agreement (92.0%, κ = 0.76), and outcome variable agreement (92.2%, κ = 0.60).

There were 555 unique individuals meeting eligibility criteria: 131 preimplementation, 376 postimplementation, and 48 maintenance (Table 1). Most were younger than 50 years of age (83.4%, 463/555), were male (63.6%, 353/555), and had Medicaid (38.0%, 322/555; Table 1). There was a higher proportion of uninsured individuals in the preimplementation period. Demographics did not otherwise differ significantly between study periods (Table 1) nor by services provided (Table 2). Visit numbers and services utilization did not differ by day of week nor time of day.
After LOOP implementation, naloxone distribution increased from none to more than one-third (35.4%, 133/376, \( p < 0.001 \)), more than one-third received consultation with a recovery coach when one was available (33.1%, 45/136), and discharge with referral to treatment increased from 9.16% (12/131) to 20.74% (78/376, \( p = 0.003 \); Table 3). Most patients receiving a recovery coach also received take-home naloxone (88.9%, 40/45; Table 2). When recovery coaches were available, very few people got take-home naloxone without a coach (4.8%, 2/42; Table 2). During the postimplementation and maintenance periods, there were 48 additional recovery coach consultations outside of available hours (Table 2). These were not included in the analysis.

Length of stay was not significantly different between study periods or with LOOP utilization.
postimplementation, median LOS for those receiving usual care was 340.3 (interquartile range [IQR] = 246.0–560.0) minutes, 275.5 (IQR = 207.2–388.3) minutes when patients were given naloxone alone, and 319.8 (IQR = 236.3–427.2) minutes when patients received recovery coach consultation with or without take-home naloxone.

Nearly half (44.9%, 249/555) of the study sample was treated and discharged after an opioid overdose (Table 1). Provision of take-home naloxone to opioid overdose patients increased from none to over half (56.5%, 91/161); nearly half received consultation with a recovery coach during available hours (49.1%, 28/57) and discharge with referral to treatment.

Table 2
Postimplementation Services by Patient Demographics

<table>
<thead>
<tr>
<th>All Patients (N = 376)</th>
<th>Total (n = 376)</th>
<th>No Services (n = 232)</th>
<th>Take-home Naloxone Alone (n = 41)</th>
<th>Recovery Coach Alone (n = 11)*</th>
<th>Recovery Coach and Naloxone (n = 92)*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–29</td>
<td>155 (41.2)</td>
<td>86 (37.1)</td>
<td>21 (51.2)</td>
<td>4 (36.4)</td>
<td>44 (47.8)</td>
<td></td>
</tr>
<tr>
<td>30–50</td>
<td>159 (42.3)</td>
<td>100 (43.1)</td>
<td>14 (34.2)</td>
<td>7 (63.6)</td>
<td>38 (41.3)</td>
<td></td>
</tr>
<tr>
<td>51+</td>
<td>62 (16.5)</td>
<td>46 (19.8)</td>
<td>6 (14.6)</td>
<td>0</td>
<td>10 (10.9)</td>
<td>0.13</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>241 (64.1)</td>
<td>140 (60.3)</td>
<td>25 (61.0)</td>
<td>8 (72.3)</td>
<td>68 (73.9)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>134 (35.6)</td>
<td>91 (39.2)</td>
<td>16 (39.0)</td>
<td>3 (27.3)</td>
<td>24 (26.1)</td>
<td></td>
</tr>
<tr>
<td>Not specified</td>
<td>1 (0.3)</td>
<td>1 (0.4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.24</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>302 (80.3)</td>
<td>189 (81.5)</td>
<td>32 (78.1)</td>
<td>10 (90.9)</td>
<td>71 (77.2)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>31 (8.2)</td>
<td>17 (7.3)</td>
<td>6 (14.6)</td>
<td>0</td>
<td>8 (8.7)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>4 (1.1)</td>
<td>2 (0.9)</td>
<td>0</td>
<td>0</td>
<td>2 (2.2)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>24 (10.3)</td>
<td>24 (10.3)</td>
<td>3 (7.3)</td>
<td>1 (9.1)</td>
<td>10 (10.9)</td>
<td></td>
</tr>
<tr>
<td>Not documented</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 (1.1)</td>
<td>0.69</td>
</tr>
<tr>
<td>Insurance status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninsured</td>
<td>48 (12.8)</td>
<td>2 (9.5)</td>
<td>8 (19.5)</td>
<td>1 (9.1)</td>
<td>17 (18.5)</td>
<td></td>
</tr>
<tr>
<td>Medicaid</td>
<td>229 (60.9)</td>
<td>146 (62.9)</td>
<td>21 (51.2)</td>
<td>5 (45.5)</td>
<td>57 (62.0)</td>
<td></td>
</tr>
<tr>
<td>Medicare</td>
<td>46 (12.2)</td>
<td>36 (15.5)</td>
<td>5 (12.2)</td>
<td>2 (18.2)</td>
<td>3 (3.3)</td>
<td></td>
</tr>
<tr>
<td>Worker’s comp</td>
<td>1 (0.3)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 (1.1)</td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>52 (13.8)</td>
<td>28 (12.1)</td>
<td>7 (17.1)</td>
<td>3 (27.3)</td>
<td>14 (15.2)</td>
<td>0.05</td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>295 (78.5)</td>
<td>172 (75.0)</td>
<td>32 (78.1)</td>
<td>9 (81.8)</td>
<td>80 (87.0)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>81 (21.5)</td>
<td>58 (25.0)</td>
<td>9 (22.0)</td>
<td>2 (18.2)</td>
<td>12 (13.0)</td>
<td>0.11</td>
</tr>
<tr>
<td>Day of week</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monday</td>
<td>57 (15.2)</td>
<td>36 (15.5)</td>
<td>8 (19.5)</td>
<td>3 (27.3)</td>
<td>10 (11.0)</td>
<td></td>
</tr>
<tr>
<td>Tuesday</td>
<td>61 (16.2)</td>
<td>38 (16.4)</td>
<td>6 (14.6)</td>
<td>2 (18.2)</td>
<td>15 (16.3)</td>
<td></td>
</tr>
<tr>
<td>Wednesday</td>
<td>55 (14.6)</td>
<td>32 (13.8)</td>
<td>11 (26.8)</td>
<td>0</td>
<td>12 (13.0)</td>
<td></td>
</tr>
<tr>
<td>Thursday</td>
<td>47 (12.5)</td>
<td>27 (11.6)</td>
<td>9 (22.0)</td>
<td>0</td>
<td>11 (12.0)</td>
<td></td>
</tr>
<tr>
<td>Friday</td>
<td>60 (16.0)</td>
<td>36 (15.5)</td>
<td>6 (14.6)</td>
<td>2 (18.2)</td>
<td>16 (17.4)</td>
<td></td>
</tr>
<tr>
<td>Saturday</td>
<td>55 (14.6)</td>
<td>40 (17.2)</td>
<td>1 (2.4)</td>
<td>2 (18.2)</td>
<td>12 (13.0)</td>
<td></td>
</tr>
<tr>
<td>Sunday</td>
<td>41 (10.9)</td>
<td>23 (9.9)</td>
<td>0</td>
<td>2 (18.2)</td>
<td>16 (17.4)</td>
<td>0.65</td>
</tr>
<tr>
<td>Time of day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 AM-3 PM</td>
<td>124 (33.0)</td>
<td>83 (35.8)</td>
<td>8 (19.5)</td>
<td>2 (18.2)</td>
<td>31 (33.7)</td>
<td></td>
</tr>
<tr>
<td>3 PM-11 PM</td>
<td>184 (48.9)</td>
<td>106 (45.7)</td>
<td>27 (65.9)</td>
<td>7 (63.6)</td>
<td>44 (47.8)</td>
<td></td>
</tr>
<tr>
<td>11 PM-7 AM</td>
<td>68 (18.1)</td>
<td>43 (18.5)</td>
<td>6 (14.6)</td>
<td>1 (18.2)</td>
<td>17 (18.5)</td>
<td>0.30</td>
</tr>
<tr>
<td>Overdose</td>
<td>249 (44.9)</td>
<td>79 (30.5)</td>
<td>36 (67.9)</td>
<td>8 (66.7)</td>
<td>73 (73.0)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Data are reported as n (%). *58 consultations were made outside of posted available hours during the postimplementation period.
increased from 1.9% (1/53) to 14.9% (24/161, \( p = 0.01 \); Table 3).

**Implementation Moderators**

Complete logistic regression results are detailed in Data Supplement S1, Supplemental Table 3. Overall, odds of receiving take-home naloxone were higher during an overnight shift (odds ratio [OR] = 4.04, 95% CI = 1.3, 12.53), if the patient received out of hospital naloxone (OR = 3.46, 95% CI = 1.13, 10.64), or if the patient received consultation with a recovery coach (OR = 107.98, 95% CI = 32.05, 363.83)). Odds of getting take-home naloxone decreased when patients had documented use of prescription opioids (OR = 0.16, 95% CI = 0.04, 0.68). Odds of receiving recovery coach consultation were significantly lower for patients already prescribed methadone (OR = 0.17, 95% CI = 0.03, 0.90)).

Overall, patients had increased likelihood of referral to treatment if they were treated at site A (OR = 15.8, 95% CI = 2.8–89.2), received a psychiatry consult (OR = 2.6, 95% CI = 1.3–5.4), or were on a prescription psychotropic (OR = 2.7, 95% CI = 1.2–5.7). Patients had decreased odds of referral to treatment if they were between 30 and 50 years of age (OR = 0.47, 95% CI = 0.24, 0.93).

For patients seen and treated after an opioid overdose, odds of take-home naloxone distribution were higher on the 3pm-11pm shift (OR 4.98 [95% CI 1.03, 24.15]) and when patients received recovery coach consultation (OR 104.50 [95% CI = 14.66, 745.13]). Overdose patients were less likely to get take-home naloxone if they were between 30–50 years of age (OR 0.15 [95% CI = 0.03, 0.83]) or used prescription opioids (OR 0.04 [95% CI = 0.00, 0.42]). Similar to the overall study sample, opioid overdose patients had decreased odds of recovery coach consultation if they were prescribed methadone (OR = 0.09, 95% CI = 0.01, 0.87).

Odds of referral to treatment for opioid overdose patients approached zero when patients received out-of-hospital naloxone, were on a psychotropic medication, used prescription opioids, or had concurrent alcohol use. Odds of referral to treatment were higher when patients were privately insured, treated during the 3 PM-11 PM shift, received psychiatry consultation, used benzodiazepines or heroin, or were prescribed a sedative hypnotic, but CIs were very wide.

**Maintenance**

In the maintenance period, there was no overall significant depreciation in discharge with take-home naloxone (56.5% vs. 51.4%, \( p = 0.583 \)), recovery coach consultation (49.12% vs. 41.7%, \( p = 0.638 \)), or referral to treatment (14.9% vs. 8.6%, \( p = 0.32 \); Table 3). Poisson regression models were conducted to evaluate

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Specialty Consultation, LOOP Utilization, and Referral to Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Discharged Patients (N = 555)</td>
<td>Patients Discharged After Opioid Overdose (n = 249)</td>
</tr>
<tr>
<td>Preimplementation (n = 131)</td>
<td>Postimplementation (n = 376)</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Psychiatry consultation</td>
<td>98 (17.7)</td>
</tr>
<tr>
<td>Social work consultation</td>
<td>35 (6.3)</td>
</tr>
<tr>
<td>Take-home naloxone</td>
<td>153 (27.6)</td>
</tr>
<tr>
<td>Recovery coach consultation**</td>
<td>51 (32.9)†</td>
</tr>
<tr>
<td>Discharge with referral to treatment</td>
<td>95 (17.1)</td>
</tr>
</tbody>
</table>

Data are reported as n (%).
LOOP = Lifespan Opioid Overdose Prevention
*p values reflect comparison of Preimplementation and postimplementation Periods
**During hours of availability, Friday 8pm to Monday 8am
†out of N = 155
‡out of N = 136
§out of N = 17
***out of N = 99
††out of N = 57
‡‡out of N = 12
changes in the number of naloxone recovery kits dispensed, recovery coach consultations, and patients discharged with referral to treatment over time. These models had appropriate fit indices, and the values of the scaled chi-square statistic for all models was <1, indicating that no adjustment for overdispersion was necessary. Table 4 shows the results of this regression. For ease of interpretation of these results, the exponents of the parameter estimates of the model were calculated to provide incidence rate ratios (IRRs) and we report on those that were significant in the regression model. The IRRs for the effect of time on dispensing of naloxone were 1.19 (95% CI = 1.11, 1.26) and 1.16 (95% CI = 1.07, 1.26) for recovery coach consultation. This indicates that there was a 19% increase in the rate of naloxone dispensing and a 16% increase in recovery coach consultation after LOOP implementation. There were no effects of LOOP implementation on discharge with referral to substance use treatment. Patients who presented with an overdose had a greater frequency of take-home naloxone provision and recovery coach consultation at baseline and through the maintenance period, but overall had lower counts of discharge with referral to treatment compared to patients not seen for an opioid overdose.

Figure 2 demonstrate rates of take-home naloxone provision (Figure 2A), recovery coach consultation (Figure 2B), and discharge with referral to treatment (Figure 2C) over the three study periods. These figures show the percentage of patients presenting to the ED with OUD and/or after an opioid overdose. Cochrane Armitage trend test was significant for take-home naloxone distribution (Z = 7.34, p < 0.001), recovery coach consultation (Z = 6.68, p < 0.001), and discharge with referral to treatment (Z = 3.22, p < 0.01). However, as can be seen in Figure 2A, discharge with referral to treatment showed a significant downward trend in the maintenance phase.

**DISCUSSION**

The ED is on the front lines of the opioid overdose epidemic and, as part of the medical safety net and key access point to the health care system, has an essential role in preventing opioid overdose death and facilitating referral to addiction treatment.8 We found our ED naloxone distribution and community recovery coach consultation program to be overall acceptable to ED providers, had adequate reach and adoption, and was utilized consistently over time.

**Figure 2.** Proportion patients receiving take home naloxone, recovery coach consultation, and referral to treatment after program implementation. (A) Change in naloxone rescue kit distribution; (B) change in recovery coach consultation; (C) change in discharge with linkage to treatment.

**Table 4**

Regression Model of the Effects of Time and Site on LOOP Utilization

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Predictor</th>
<th>Estimate</th>
<th>Wald 95% Confidence Interval</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take-Home Naloxone</td>
<td>Site</td>
<td>−0.18</td>
<td>−1.10, 0.75</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>0.17</td>
<td>0.10, 0.23</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Site * Time</td>
<td>−0.02</td>
<td>−0.21, 0.16</td>
<td>0.8</td>
</tr>
<tr>
<td>Recovery Coach Consultation</td>
<td>Site</td>
<td>−0.19</td>
<td>−1.27, 0.90</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>0.15</td>
<td>0.07, 0.23</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>Site * Time</td>
<td>−0.08</td>
<td>−0.31, 1.16</td>
<td>0.53</td>
</tr>
<tr>
<td>Discharge with Linkage to Treatment</td>
<td>Site</td>
<td>0.9</td>
<td>0.12, 1.67</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>0.05</td>
<td>−0.05, 0.14</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Site * Time</td>
<td>0.02</td>
<td>−0.15, 0.19</td>
<td>0.83</td>
</tr>
</tbody>
</table>
without overall significant depreciation 1 year from implementation. Our findings have several important implications for EDs implementing similar programs and highlights areas for future research.

Many hospitals and EDs currently lack the infrastructure to provide overdose prevention and addiction treatment navigation and/or treatment initiation. In our program, purchase and storage of take-home naloxone was the primary cost assumed by the hospital. Small community hospitals may face financial barriers not encountered at larger institutions. Furthermore, while naloxone is covered by insurance, many states do not allow for direct provider to patient distribution, which can create a barrier to take-home naloxone provision at the time of the ED visit.42

Partnering with a community-based organization reduced many programmatic and cost barriers to providing addiction treatment support, navigation, and after-ED follow up. Although the recovery coaches were not located on site, LOS for this patient population did not increase significantly with utilization. Initially, recovery coach availability was limited by cost constraints. After demonstrating similar demand throughout the week, intervention acceptability by providers and patients, and program feasibility, Anchor was able to secure additional funding to increase availability in October 2015 to 24 hours a day, 7 days a week. In 2017, offering consultation with a peer recovery coach became a regulatory requirement for all RI EDs.44

Despite successful uptake and expansion, utilizing an external agency for services navigation has some challenges. Not being hospital-based creates potential breakdown of communication that may result in care gaps. Issues may also arise if hospital and community agency policies and care standards are not well aligned. Services alignment and quality assurance requires regular communication, mutual feedback, and collaboration between each hospital site and the community agency. Other EDs have implemented similar programs, primarily in areas with high rates of opioid overdose and high population density.44 Use of on-call peer support may have limited effectiveness in rural areas; however, there may be a role for telehealth or on-site peer placement as a means to provide access to peer support notwithstanding large geographic distances.45,46

Providers reported increased utilization for patients presenting after opioid overdose, which is consistent with observed utilization patterns in the EMR review. Interestingly, we found that while recovery coach consultation increased overall, it peaked at month 4 after implementation and decreased during the maintenance period. Being a retrospective study, we were unable to assess factors contributing to this decline, such as provider failure to offer consultation, lack of patient interest, and prior establishment of linkage with a coach. Similarly, while referral to treatment increased overall, we observed a downward trend over time. This decline may be due to treatment availability; patient willingness to accept services; provider and health system interest; resident and staff turnover; need for more frequent provider education about available services; or coach-level factors such as staff turnover, consultation variability, and changing availability of community resources.

Further investigation is needed to better understand the factors resulting in a decline of recovery coach consultation and referral to treatment, including implementation moderators associated with decreased services uptake: age, out-of-hospital naloxone administration, prescription opioid use, and prescription of a psychotropic or methadone. These factors may reflect preexisting treatment engagement, lack of identification of overdose risk by patients and/or providers, or patient readiness to engage in treatment.

LIMITATIONS

This study is subject to several limitations. Although the survey response rate was sufficient for a sample of emergency medicine providers,47 responses may be subject to reporting, recall, and desirability bias, therefore overreporting LOOP adoption and minimizing utilization difficulties.

Initial EMR screening parameters were designed to emphasize sensitivity. We sought to minimize EMR selection and misclassification bias by developing and piloting our data collection instrument prior to data extraction and conducting uniform screening and extraction of records during the study period using preestablished search terms and criteria. Despite clear criteria, given the limitations of retrospective EMR review,48 we may not have been able to identify all appropriate records for patients who were “at risk” for overdose, and patient encounters in which opioid overdose was the primary issue were likely disproportionately overrepresented. This can be mitigated in future prospective studies by including systematic patient assessment for OUD.

Information analyzed was also limited to patient report and EMR documentation by clinicians and staff,
which is subject to recording errors or omissions. Given the retrospective study design, we were unable to fully determine implementation fidelity, specifically whether patients left with take-home naloxone, if they demonstrated understandability of how to use naloxone, whether there was a family member or friend present for teaching, subsequent use of naloxone for overdose reversal, or the result of the conversation with the recovery coach other than what was documented by the ED provider. Finally, as a study limited to two hospitals in the same city, study results may lack generalizability to hospitals of different size and different patient and provider composition or in regions not as severely impacted by the opioid overdose epidemic.

CONCLUSIONS

This study demonstrates feasibility, acceptability, and utilization maintenance of an ED overdose education and naloxone distribution program with consultation of a community-based peer recovery coach for treatment navigation. While community overdose education and naloxone distribution programs have observed a reduction in overdose mortality, the impact of ED naloxone distribution on mortality is unknown and difficult to measure. Future studies are needed to evaluate the effectiveness of ED naloxone distribution and recovery coach patient navigation on successful linkage to treatment, repeat overdose, and overdose death.

For their collaboration, participation, and assistance during intervention implementation and study execution, the authors acknowledge and thank the Anchor Recovery Community Center, Francesca L. Beaudoin, MD, PhD, Kari Beck, BSN, RN, Rebecca Boss, MA, Sarah Bowman, MPH, Jeffrey Bratberg, PharmD, John Brister, Gary Bubly, MD, FACEP, Kristen Bunnell, PharmD, BCPS, Holly Cekala, Marc Chatterton, Melissa Conroy, MSW, Suzanne DunBriggs, JD, RN, BSN, Linda Dykstra, RN, MSN, CEN, Natalija Farrell, PharmD, Dennis Ferrante, Jonathan Goyer, Traci Green, PhD, MSc, Jason Hack, MD, Glenn Hebel, MD, Nathan Hudepohl, MD, MPH, Ilee Jenoumi, MD, MBA, Jennifer Koziol, MPH, Kristi Mangioca, MA, Christopher Matkovic, MD, Michelle McKenzie, MPH, John B. Murphy, MD, George O’Toole, Christina Parnagian, David Portelli, MD, Laura Ruhland, MD, Latha Sivaprasad, MD, FACP, FHM, and Brian Zink, MD.

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.13545/full

Data Supplement S1. Supplemental material.
Impact of Medical Scribes on Provider Efficiency in the Pediatric Emergency Department

Luke C. Addesso, Mark Nimmer, Alexis Visotcky, MS, Raphael Fraser, PhD, and David C. Brousseau, MD, MS

ABSTRACT

Objectives: Today’s emergency department (ED) providers spend a significant amount of time on medical record documentation, decreasing clinical productivity. One proposed solution is to utilize medical scribes who assist with documentation. We hypothesized that scribes would increase provider productivity and increase provider satisfaction without affecting patient experience or nursing satisfaction.

Methods: We conducted an observational pre-post study comparing ED prescribe and postscribe clinical productivity metrics for 18 pediatric emergency medicine physicians, two general pediatricians, and two nurse practitioners working in the 12-bed nonurgent area of the pediatric ED. Productivity metrics included patients per hour (pts/hr), work relative value units per hour (wRVUs/hr), and visit duration measured for 1 year pre- and postscribe implementation. Cross-sectional satisfaction surveys were administered to patient families, providers, and nurses during the initial scribe rollout.

Results: Overall, 24,518 prescribe and 27,062 postscribe visits were analyzed. Following scribe implementation, overall provider efficiency increased by 0.24 pts/hr (11.98%, \( p < 0.001 \)) and 0.72 wRVUs/hr (20.14%, \( p < 0.001 \)). The largest efficiency increase (0.36 pts/hr, 0.96 wRVUs/hr) occurred in January–March, when ED census peaked. Patient visit duration was 53 minutes in both the prescribe and the postscribe periods. During initial scribe implementation, 80% of parents of patients without a scribe rated the visit as very good/great compared to 84% with a scribe (\( p = 0.218 \)). Of the 34 providers surveyed, 88% preferred working with a scribe. A majority of providers (82%) felt that their skills were used more effectively when working with a scribe, decreasing their likelihood of experiencing burnout. Of the 43 nurses surveyed, 51% preferred scribes and 47% were indifferent.

Conclusions: Medical scribes increased ED efficiency without decreasing patient satisfaction. Providers strongly favored the use of scribes, while nurses were indifferent. The next steps include a cost analysis of the scribe program.
to direct patient care. With increasing ED volumes, this results in prolonged wait times and decreased satisfaction. Adequate documentation is particularly cumbersome for ED physicians, given the frequent interruptions they experience throughout a shift. Thus, to improve physician productivity, it is important to seek out ways to reduce the responsibilities of ED physicians that are not direct patient care.

One proposed approach to increase physician productivity is the utilization of medical scribes. A medical scribe is a staff member whose primary role is to assist physicians with accurate documentation using the electronic medical record (EMR) system. Scribes accompany the physician during patient encounters to document patient history, physical examination findings, procedures performed, and plans for testing and treatment. Real-time scribe documentation also allows the physician to focus on the patient and establish rapport. Scribes offer further assistance by retrieving results of laboratory and imaging studies, as well as facilitating consults with other providers. Relieving the physician of these responsibilities has been shown to significantly increase physician job satisfaction.

A preliminary study has demonstrated the potential of scribes to significantly reduce physician time spent charting, allowing this time to be redirected toward direct patient care. This is an important shift, as direct patient contact is one of the principal predictors of positive patient outcomes and patient satisfaction. However, studies have shown conflicting results regarding a scribe program’s effect on efficiency metrics, such as throughput times and number of patients seen per day. It is important to distinguish scribes’ effect on throughput times in EDs, as reduced ED crowding is associated with decreased patient mortality. Furthermore, there is still very little literature describing the effect of scribes on patient satisfaction and even less describing their effect in a pediatric ED.

The goal of this study was to assess the impact of implementing a scribe program in a large, urban, academic, pediatric ED. We hypothesized that the utilization of medical scribes would increase the efficiency of ED providers (measured in patients seen per hour [pts/hr] and provider generated work relative value units per hour [wRVUs/hr]), decrease patient visit duration, and increase the proportion of patient charts completed by the provider within 24 hours. We also hypothesized that the scribe program would increase provider satisfaction, with no change in nurse satisfaction or patient experience.

METHODS

Study Setting and Population

We conducted an observational pre-post study comparing ED prescribe and postscribe clinical productivity metrics. This study was conducted in the pediatric ED at a large, urban, academic medical center. The study focused on the 12-bed nonurgent area of the ED; this location is responsible for the care of approximately 25,000 of the total ED annual volume of 68,000 visits. This area predominantly treats patients with an Emergency Severity Index (ESI) of 4 or 5, although ESI level 3 patients are cared for in this area occasionally if needed. The area is open from 10 am–12 am daily with minor seasonal variation. Providers working in the ED included board-certified pediatric emergency medicine (PEM) physicians, general pediatricians, and nurse practitioners (NPs). Nurse-to-provider ratios are targeted at one nurse per provider. Residents/trainees are rarely scheduled in this area of the ED, which is one of the main reasons scribes were selectively introduced to the nonurgent area. Providers in this area are primarily caring for these patients and writing the notes themselves and therefore are more likely to benefit from a scribe compared to providers in the urgent area, who have residents to write the notes for them. The EMR platform used in this ED was Epic systems software. This project was determined to meet the criteria for a quality improvement project and was thus deemed exempt by the institutional review board.

Provider Efficiency

The primary outcome was provider efficiency. Two main aspects of provider efficiency were measured: number of patients seen per hour (pts/hr) and wRVUs generated per hour (wRVUs/hr). Efficiency data were collected on a quarterly basis. Baseline pre-scribe efficiency measurements were recorded for patients seen in the nonurgent area of the ED between July 1, 2015, and June 31, 2016. This pre-scribe period was composed of provider shifts without scribes, except for scribe training shifts. As scribes were training, they were not in their operational role and were not fully functioning as scribes. Therefore, the effect of the scribe training shifts was considered to be negligible. Postscribe measurements included patients seen in the same nonurgent area between July 1, 2016, and June 31, 2017. During the postscribe period, the goal was to have every provider shift accompanied by a scribe, but this was not 100%
successful due to real-life events such as scribe illness, scribe turnover, or an immediate need to increase provider coverage due to high volumes. Only providers who worked a shift during both data collection periods for a given quarter were included in the study. Efficiency metrics in the urgent section of the ED were also tracked to act as a control. No scribes were used in the urgent area during the pre or post period. The number of patients seen was abstracted from the hospital billing data warehouse and the number of hours worked was extracted from an online scheduling service. We used the electronic schedule data to systematically identify which providers worked shifts within the date parameters of our pre- and postscribe periods.

Two other process measures were assessed. We measured the proportion of patient charts completed by the provider, measured as median and interquartile range, as well as percent completed within 8 and 24 hours. We also evaluated visit duration, defined as the length of time between when a provider was assigned to a patient and the patient’s departure from the ED.

**Satisfaction and Experience Surveys**

The secondary outcomes for the study were patient experience and both provider and nurse satisfaction with the scribes. To gauge the effect of scribes on patient experience, parents or guardians (hereafter termed parents) with children in the ED were surveyed cross-sectionally during their ED visit. Surveys were administered at the end of the patient ED visit while the patient was waiting for discharge paperwork. Surveys were filled out electronically on a handheld device provided by the researcher and took less than 1 minute to complete. The purpose of these surveys was to ensure that patient experience was not negatively impacted by the introduction of scribes before continuing with the study. These surveys were administered during the initial rollout of the scribe program (June 1, 2016–August 31, 2016), when scribe coverage was not yet universal. Surveys were administered during predetermined times to all current patients, whether seen by a provider with a scribe or without a scribe. The predetermined times for survey administration were chosen by identifying the shifts where there would be at least one provider with a scribe and one provider without a scribe. This way, both groups of patients surveyed were exposed to similar waiting times prior to enrollment. Due to real-world events such as illness and unavailability of staff, surveys were administered during most, but not all, time periods that met this criteria.

The first part of the survey, assessing overall patient experience, was given to all families (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.13544/full], Patient survey A). Parents with children seen by providers with a scribe received additional questions to understand their specific reaction to scribe usage in the ED (Data Supplement S1, Patient survey B). Survey questions used a five-point Likert scale, with the exception of a single question asking parents to rate their ED visit using: great, very good, good, could have been better, or very disappointing. Patient experience was considered positive if “great” or “very good” was selected in response to this question. Questions were adapted and modified from the Consumer Assessment of Healthcare Providers and Systems developed by the Agency for Healthcare Research and Quality. All surveys were completed by a parent over the age of 18. Parents requiring a translator during their ED visit were excluded from these surveys.

Providers and nurses who worked one or more shifts with a scribe during the initial rollout of the scribe program were asked to complete a provider or nurse survey (hereafter termed as the initial satisfaction survey) to quantify their earliest reaction to scribe usage. Surveys were distributed via e-mail within 3 days of the shift. Each participant meeting these criteria was eligible to complete a maximum of one initial satisfaction survey. The surveys asked providers and nurses to respond to statements using a five-point Likert scale (Data Supplement S1).

After 1 full year of scribe usage in the ED, all respondents to the initial satisfaction survey were asked to complete the same survey once more (hereafter termed the delayed satisfaction survey). The purpose of this survey was to determine if providers and nurses still felt the same way about scribes after a year of working together. For all surveys in this study, responses that left one or more questions unanswered were still included in the results.

The surveys used in this study were piloted with patient families, providers, and nurses to ensure the questions addressed the concerns of each group. The patient survey was informally piloted by asking approximately 30 patient families about their understanding of the survey and any concerns they wanted addressed that were not included in the survey. Patient families
requested an explanation of a scribe’s duties prior to taking the survey. The provider survey was piloted with the 15 providers present at a PEM section meeting. These providers were told that the surveys were meant to measure provider satisfaction with scribe usage. Providers were asked to read the survey, respond with questions they had, and ask for clarification if any part of the survey was not clear. Provider concerns focused on the accuracy of the note, the quality of care they were providing, and how effectively their skills as a provider were being utilized. In a similar way, the nursing survey was piloted with approximately 10 to 20 nurses present at a nursing leadership meeting. They were asked to read the survey, respond with any questions, and ask for clarification when needed. Concerns from nursing staff centered around the effect scribes would have on nursing workflow. All surveys were revised in accordance with feedback from each survey population.

**Scribes**

For this study, scribes were recruited, hired, trained, and managed by a professional medical scribe company. Scribes were paid by the scribe management company, who then sent a monthly invoice to the institution for their services. Training consisted of a minimum of 50 hours of training for each scribe, at least 40 of which took place in the pediatric ED. No scribe had any prior experience. Scribes were students interested in a health care profession who either were currently attending or had recently graduated from local 4-year universities. Responsibilities during a shift included composing a history of present illness, completing the review of systems and certain parts of the physical examination, recording imaging results, documenting times of updates and consults, and writing discharge instructions for each patient.

**Data Analyses**

Efficiency measures for both the nonurgent and urgent areas were modeled separately using repeated-measures analysis of variance, adjusting for quarter, to evaluate the effect of scribes over time. Study data were collected and managed using REDCap (Research Electronic Data Capture) tools. REDCap is a secure, Web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry, 2) audit trails for tracking data manipulation and export procedures, 3) automated export procedures for seamless data downloads to common statistical packages, and 4) procedures for importing data from external sources. Analysis was completed using SAS v 9.4 (SAS Institute).

A sample size calculation was performed for the patient experience survey using an alpha of 0.05, power of 0.80, and assuming a patient seen with scribe to patient seen without scribe ratio of 1.60. The calculation showed that 650 patients surveyed would detect a difference of 9% or greater between the two groups’ experience scores. Patient experience surveys were analyzed using nonparametric Wilcoxon tests. For providers and nurses who completed both the initial and the delayed survey, paired t-tests were used to detect any significant change in responses.

**RESULTS**

**ED Efficiency**

There were 22 providers included in the efficiency portion of this study: 18 PEM physicians, two general pediatricians, and two NPs. All 24,518 prescribe and 27,062 postscribe visits to the nonurgent area were analyzed. The total ED census during the postscribe year was 68,899, a 4.1% increase from the prescribe year. This difference was considered to be negligible.

Overall, providers exhibited an increase in efficiency after the scribe program was introduced. Average pts/hr increased from 1.97 to 2.21, an increase of 0.24 pts/hr or 11.98% (p < 0.0001). Similar improvements were noted in provider generated wRVUs/hr. Average provider wRVUs/hr increased from 3.55 to 4.27, an increase of 0.72 wRVUs/hr or 20.14% (p < 0.0001; Table 1).

The largest quarterly increase in efficiency (+0.36 pts/hr [18.23%], +0.96 wRVUs/hr [26.43%]) occurred in Quarter 1 (January 1–March 31), which also was the quarter with the highest total ED census (19,757). The quarter that showed the least improvement was Quarter 4 (October 1–December 31), with increases of 0.17 pts/hr (7.70%) and 0.55 wRVUs/hr (14.48%). After the addition of scribes, the proportion of the total ED census being sent to nonurgent area increased from 36.85% to 38.95%.

Evaluation of chart completion data revealed no changes with the addition of the scribe program. The prescribe median time to chart completion was 0 hours (interquartile range [IQR] = 0–2 hours), compared to a median of 1 hour (IQR = 0–4 hours) in the postscribe period. The proportion of charts
completed within 8 hours of discharge was 86% in the prescribe period and 85% in the postscribe period, with 24-hour chart completions of 90% and 89%, respectively (all not significant). Patient visit duration was unchanged, measuring 53 minutes in both the prescribe and postscribe years.

Efficiency metrics were also tracked in the urgent section of the ED as a control. In this area, there was no change in the pts/hr (−0.02 [−0.68%], p = 0.83) or wRVUs/hr (+0.07 [+1.69%], p = 0.29). Similar to the nonurgent area, chart completion was unchanged.

**Patient Experience**

During the scribe rollout (June 1, 2016–August 31, 2016), 651 patient experience surveys were administered. The patient population included 398 patients who were seen by a provider with a scribe and 253 patients who were seen by a provider without one. Of the 665 parents who were asked to take the survey, eight chose not to participate (98% response rate) and six required a translator, excluding them from the study. Patient experience was positive for 83.76% of patients who had a scribe in the room and for 80.08% of patients who did not have a scribe. There was no statistically significant difference between the two groups (Table 2). When patients who had a scribe were asked if they would rather be seen by a provider with a scribe or without one, 42.96% chose the provider with a scribe, 47.74% were neutral, and 9.30% chose a provider without a scribe (Table 3).

**Provider Satisfaction**

During the scribe rollout period, 36 providers worked at least one shift with a scribe, deeming them eligible to complete the initial provider satisfaction survey. Responses were recorded from 34 of these providers (94% response rate), reflecting the views of 17 PEM physicians, two general pediatricians, six advanced practice providers (NPs and physician assistants), and nine PEM fellows.

Overall, the initial provider satisfaction survey showed that 78.13% of providers felt a scribe improved the quality of patient care. Scribes also

| Table 1 | Efficiency in the Nonurgent Area |
|-----------------|-----------------|-----------------|
| Quarter         | pts/hr | wRVUs/hr |
| Year            |         |         |
| Prescribe       | 1.97    | 3.55    |
| Postscribe      | 2.21    | 4.27    |
| Change          | +0.24   | +0.72   |
| p-value*        | 0.0001  | 0.0001  |
| Q1              |         |         |
| Prescribe       | 1.99    | 3.64    |
| Postscribe      | 2.36    | 4.60    |
| Change          | +0.36   | +0.96   |
| Q2              |         |         |
| Prescribe       | 2.07    | 3.72    |
| Postscribe      | 2.25    | 4.53    |
| Change          | +0.18   | +0.80   |
| Q3              |         |         |
| Prescribe       | 1.76    | 3.10    |
| Postscribe      | 2.00    | 3.63    |
| Change          | +0.24   | +0.53   |
| Q4              |         |         |
| Prescribe       | 2.08    | 3.78    |
| Postscribe      | 2.25    | 4.33    |
| Change          | +0.17   | +0.55   |

pts/hr = patients per hour; wRVUs/hr = work relative value units per hour.

| Table 2 | Patient Experience Comparison: Surveys Were Conducted During Same Shifts for Patients Seen With and Without Scribes |
|-----------------|-----------------|-----------------|
| Question        | Scribe Status   | n*   | Disagree/Strongly Disagree | Neutral | Agree/Strongly Agree | p-value† |
| Q1. How was your emergency department visit today?‡ | No scribe | 246  | 3 (1.22%) | 46 (18.70%) | 197 (80.08%) | 0.218  |
|                | Scribe         | 394  | 1 (0.25%) | 63 (15.99%) | 330 (83.76%) |
| Q2. I felt the provider knew the important information to care for my child | No scribe | 253  | 3 (1.19%) | 7 (2.77%) | 243 (96.05%) | 0.793  |
|                | Scribe         | 395  | 5 (1.27%) | 9 (2.28%) | 381 (96.46%) |
| Q3. I felt the provider listened carefully and showed respect for what I had to say | No scribe | 253  | 2 (0.79%) | 6 (2.37%) | 245 (96.84%) | 0.261  |
|                | Scribe         | 395  | 5 (1.27%) | 2 (0.51%) | 388 (98.23%) |
| Q4. I felt the provider explained things to me in a way that was easy to understand | No scribe | 253  | 1 (0.40%) | 7 (2.77%) | 245 (96.84%) | 0.512  |
|                | Scribe         | 394  | 5 (1.27%) | 4 (1.02%) | 385 (97.72%) |

*A total of 398 “scribe” and 253 “no scribe” surveys were completed. The variation in the n-values can be attributed to surveys that were submitted with some questions left unanswered.

†Wilcoxon rank-sum tests were used to detect significant differences between responses in the “scribe” and “no scribe” groups.
‡Q1 used a modified Likert scale (Very disappointing/could have been better, good, very good/great).
raised job satisfaction among providers, as 81.82% of survey respondents reported that their skills were being more effectively utilized, and 88.24% expressed their preference of working a shift with a scribe rather than without one. Over half (58.82%) of providers thought that scribes increased the accuracy and completeness of charting, while 35.29% did not perceive any improvement in this category (Table 4).

The delayed provider satisfaction survey, administered 1 year after scribe implementation, was completed by 27 of the 34 eligible providers (79% response rate). There was no significant change in provider responses compared to the initial survey.

### Nursing Satisfaction

The response rate for the initial nurse satisfaction survey was 70%, reflecting the views of 43 of the 61 nurses who were sent a survey. Most nurses did not feel that scribes made a significant impact on their job, with “neutral” being the most common response for three of the five questions. While most nurses did not appreciate any clear benefit from the scribes, 76.64% of respondents indicated that scribes did not negatively impact nursing flow in any way. Overall 51.16% of respondents said they would rather work a shift with a scribe than without one and 46.51% were indifferent (Table 5). The delayed nurse satisfaction

---

**Table 3**

Patient Reaction to Scribes

<table>
<thead>
<tr>
<th>Question</th>
<th>n*</th>
<th>Disagree/Strongly Disagree</th>
<th>Neutral</th>
<th>Agree/Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. The provider seemed focused on me and spent the majority of the time talking to me or my child</td>
<td>397</td>
<td>9 (2.27%)</td>
<td>5 (1.26%)</td>
<td>383 (96.47%)</td>
</tr>
<tr>
<td>Q2. Having a scribe in the room made me feel uncomfortable</td>
<td>397</td>
<td>348 (87.66%)</td>
<td>28 (7.05%)</td>
<td>21 (5.29%)</td>
</tr>
<tr>
<td>Q3. It is important to me that information about my history and physical exam is documented during the encounter</td>
<td>396</td>
<td>7 (1.77%)</td>
<td>41 (10.35%)</td>
<td>348 (87.88%)</td>
</tr>
<tr>
<td>Q4. I would rather be seen by a provider with a scribe than a provider without one</td>
<td>398</td>
<td>37 (9.30%)</td>
<td>190 (47.74%)</td>
<td>171 (42.96%)</td>
</tr>
</tbody>
</table>

*A total of 398 surveys were completed. The variation in the n-values can be attributed to surveys that were submitted with some questions left unanswered.*

**Table 4**

Provider Reaction to Scribes

<table>
<thead>
<tr>
<th>Question</th>
<th>Time</th>
<th>n†</th>
<th>Disagree/Strongly Disagree</th>
<th>Neutral</th>
<th>Agree/Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Having a scribe has decreased the amount of time I spend on indirect patient care activities (charting, telephone calls, retrieving test results)</td>
<td>Initial</td>
<td>33</td>
<td>1 (3.30%)</td>
<td>1 (3.03%)</td>
<td>31 (93.94%)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>27</td>
<td>1 (3.70%)</td>
<td>0 (0.00%)</td>
<td>26 (96.30%)</td>
</tr>
<tr>
<td>Q2. Having a scribe has increased the accuracy and completeness of my charting</td>
<td>Initial</td>
<td>34</td>
<td>2 (6.88%)</td>
<td>12 (35.29%)</td>
<td>20 (58.82%)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>27</td>
<td>1 (3.70%)</td>
<td>3 (11.11%)</td>
<td>23 (85.19%)</td>
</tr>
<tr>
<td>Q3. I feel my skills as a provider are more effectively utilized when I work with a scribe</td>
<td>Initial</td>
<td>33</td>
<td>0 (0.00%)</td>
<td>6 (18.18%)</td>
<td>27 (81.82%)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>27</td>
<td>0 (0.00%)</td>
<td>2 (7.41%)</td>
<td>25 (92.59%)</td>
</tr>
<tr>
<td>Q4. I would rather work a shift with a scribe than without one</td>
<td>Initial</td>
<td>34</td>
<td>1 (2.94%)</td>
<td>3 (8.82%)</td>
<td>30 (88.24%)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>27</td>
<td>0 (0.00%)</td>
<td>2 (7.41%)</td>
<td>25 (92.59%)</td>
</tr>
<tr>
<td>Q5. I think hiring a team of scribes for the emergency department was a good investment</td>
<td>Initial</td>
<td>34</td>
<td>0 (0.00%)</td>
<td>3 (8.82%)</td>
<td>31 (91.18%)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>27</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
<td>27 (100.0%)</td>
</tr>
<tr>
<td>Q6. Scribes improve my ability to deliver high quality patient care</td>
<td>Initial</td>
<td>32</td>
<td>2 (6.25%)</td>
<td>5 (15.63%)</td>
<td>25 (78.13%)</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>26</td>
<td>0 (0.00%)</td>
<td>2 (7.69%)</td>
<td>24 (92.31%)</td>
</tr>
</tbody>
</table>

*For providers who completed both an initial and delayed survey, paired t-tests detected no significant change between initial and delayed survey responses. All p-values were > 0.05.

†A total of 34 initial and 27 delayed surveys were completed. The variation in the n-values can be attributed to surveys that were submitted with some questions left unanswered.*
DISCUSSION

The results of this study support the hypothesis that the utilization of medical scribes increases ED provider efficiency and satisfaction while maintaining a high level of patient and nurse satisfaction. This was one of the first studies to examine the impact of scribes in a pediatric ED. As such, it provides valuable insight into the impact of implementing a scribe program in this unique setting and using scribes with different types of providers.

ED Efficiency

Our study showed increased efficiency with the addition of scribes to the nonurgent area of the ED. EDs are one of the most sensible places to target for efficiency improvements. The unpredictability of hour-to-hour ED census makes the efficiency status of an ED readily impressionable and responsive to changes such as the addition of a scribe program. Previous literature regarding the effect of scribe use in the ED has had conflicting results.\(^5\)\(^-\)\(^12\) Outside of the ED setting, scribes have been shown to have a similar impact. In an article from 2015, Bank and Gage\(^15\) reported that a scribe program in an outpatient cardiology clinic permitted providers to schedule additional patient visits each day, increasing efficiency by 9.6% pts/hr.

We focused our study on the nonurgent area of the ED. The unchanged efficiency in the urgent area of the ED, which never used scribes, suggested that the scribes were likely the key factor that drove up efficiency in the nonurgent area. The obvious benefit to increased ED efficiency is the ability to increase patient volume seen in the ED or reduce the number of providers staffing the ED at any given time, allowing for cost savings. Seasonality seemed to influence the degree to which scribes augmented productivity. For example, quarter 1 (January 1–March 31) was the period that saw the largest increase in efficiency. Quarter 1 also had the largest total ED census, suggesting that the scribes provide a more substantial benefit during times of increased ED volume, such as cold and flu season, when a higher number of patients are seeking care. The impact of scribes in this study was not delayed in onset, indicating that providers needed minimal time to adjust to the new workflow.

Notably, after the addition of scribes, there was an absolute 2.10% increase in the proportion of total ED patients sent to the nonurgent area. Over the course of a year, this equates to approximately 1400 patients being seen in the nonurgent area that would have normally been sent to the urgent area. Increased throughput in the nonurgent wing is beneficial, as it alleviates some of the burden on the rest of the ED. Approximately 60% of all patients presenting to this ED are classified as ESI 4 or 5. Allocating a greater fraction of these patients to the nonurgent wing ensures a higher likelihood of a room being available immediately on arrival for patients with higher-acuity chief complaints.

No significant changes were noted in patient visit duration, as measured by time from the patient being assigned to a provider until discharge. This implies that the observed productivity surge came as a result of scribes allowing providers to manage more patients simultaneously or spend less time charting between patients, which allowed them to sign up for the next

survey, administered 1 year later, showed no significant change in nurse opinion.

Table 5

<table>
<thead>
<tr>
<th>Question</th>
<th>Time</th>
<th>n</th>
<th>Disagree/Strongly Disagree</th>
<th>Neutral</th>
<th>Agree/Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Scribes enable better communication between nurses and providers</td>
<td>Initial 43</td>
<td>4</td>
<td>9.30%</td>
<td>28 (65.12%)</td>
<td>11 (25.58%)</td>
</tr>
<tr>
<td></td>
<td>Delayed 24</td>
<td>4</td>
<td>16.67%</td>
<td>13 (54.17%)</td>
<td>7 (29.17%)</td>
</tr>
<tr>
<td>Q2. Scribes improve my ability to deliver high quality patient care</td>
<td>Initial 43</td>
<td>3</td>
<td>6.98%</td>
<td>25 (58.14%)</td>
<td>15 (34.88%)</td>
</tr>
<tr>
<td></td>
<td>Delayed 24</td>
<td>5</td>
<td>20.83%</td>
<td>10 (41.67%)</td>
<td>9 (37.50%)</td>
</tr>
<tr>
<td>Q3. I would rather work a shift with a scribe than without one</td>
<td>Initial 43</td>
<td>1</td>
<td>2.33%</td>
<td>20 (46.51%)</td>
<td>22 (51.16%)</td>
</tr>
<tr>
<td></td>
<td>Delayed 24</td>
<td>0</td>
<td>0.00%</td>
<td>9 (37.50%)</td>
<td>15 (62.50%)</td>
</tr>
<tr>
<td>Q4. Scribes facilitate communication between providers and families</td>
<td>Initial 43</td>
<td>2</td>
<td>4.65%</td>
<td>28 (65.12%)</td>
<td>13 (20.23%)</td>
</tr>
<tr>
<td></td>
<td>Delayed 24</td>
<td>5</td>
<td>20.83%</td>
<td>10 (41.67%)</td>
<td>9 (37.50%)</td>
</tr>
<tr>
<td>Q5. Scribes negatively impact nursing workflow</td>
<td>Initial 43</td>
<td>3</td>
<td>76.64%</td>
<td>8 (18.60%)</td>
<td>2 (4.65%)</td>
</tr>
<tr>
<td></td>
<td>Delayed 24</td>
<td>2</td>
<td>91.67%</td>
<td>1 (4.17%)</td>
<td>1 (4.17%)</td>
</tr>
</tbody>
</table>

*For nurses who completed both an initial and delayed survey, paired t-tests detected no significant change between initial and delayed survey responses. All p-values were > 0.05.
patient faster. We did not measure wait times before being seen by the provider. The prescribe visit duration in this study, defined as provider to discharge, was much lower (53 minutes) than was reported in previous studies (173 and 231 minutes), likely due to the nonurgent nature of patient complaints.\(^\text{16,17}\) The low baseline may have limited the potential to shorten patient visit duration like other studies have reported.

Chart completion times were unchanged after the addition of scribes. We had hypothesized that charts would be completed faster, as the increased efficiency would theoretically afford providers extra time to complete the chart during their shift. However, it seems that the completion of patient charts was still being deferred until after the shift. Instead of using the extra time to complete notes, providers were opting to sign up for new patients right away.

**Patient Experience**
To the best of our knowledge, this study obtained the largest collection of patient experience data pertaining to a medical scribe program.\(^\text{18,19}\) Similar to previous studies, this study revealed that patient experience was largely the same between patients who encountered a scribe and patients who did not.\(^\text{20,21}\) This is important because it demonstrates that the productivity gain that scribes provide is not negated by a concurrent downturn in patient experience. While the positive patient experience proportion was shown to be sustained at approximately 80%, there were some novel patient concerns that arose over the course of the study. Of the patients who reported feeling uncomfortable with a scribe in the room, several cited their uncertainty of who the scribe was as their main concern. To remedy this going forward, it would be useful to educate health care providers on how to effectively introduce the scribe to patients.

**Provider Satisfaction**
Scribes were well received by providers. Providers showed a considerable preference for working with scribes rather than without one. Increasing physician satisfaction is particularly important in the ED, as PEM physicians have one of the highest burnout rates among medical specialties.\(^\text{22}\) One of the defining symptoms of burnout is a feeling of ineffectiveness.\(^\text{23}\) Providers in this study reported feelings of increased effectiveness when working with a scribe and are thus at a lower risk of burnout. One previous study that directly examined the effect of scribes on physician burnout agreed, concluding that physicians who used scribes felt more authentic or true to self, a known defense against burnout.\(^\text{24}\)

A majority of providers in this study felt that having a scribe resulted in more accurate documentation. This result is in accordance with previous studies, in which notes written by scribes were determined to be more detailed than notes written by providers.\(^\text{25,26}\)

**Nurse Satisfaction**
Nurses were predominantly indifferent to the scribe program. There was an initial concern at our site that the increased patient volumes in the nonurgent area would adversely affect nurse satisfaction; this was not found to be true. Despite an increase in pts/hr, nurse workload did not change significantly. To the best of our knowledge, this is the first study to record nurses’ impressions of scribes.

**LIMITATIONS**
This study has several limitations. First, it was limited to the nonurgent area of a pediatric ED that only rarely had trainees present, thus limiting the generalizability of these results to that setting. Different settings, such as urgent areas or areas with lots of trainees may not have the same benefit. We were also limited by the lack of a validated tool to assess patient, provider, and nurse reaction to scribes. While not valid, the surveys were piloted with all three survey populations to ensure that the questions correctly assessed their concerns. Finally, a true cost analysis was not performed. The result of a cost analysis would vary by institution, given that it would depend on variables such as the cost of the scribe program and resultant staffing changes.

**CONCLUSIONS**
Scribes were found to increase efficiency (number of patients per hour, work relative value units per hour) in the pediatric ED. Providers experienced increased satisfaction and expressed their preference for working with scribes, while patient experience and nurse satisfaction was preserved at high levels.

**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.13544/full Data Supplement S1. Appendix.
CME Information: Low-dose Magnesium Sulfate Versus High Dose in the Early Management of Rapid Atrial Fibrillation: Randomized Controlled Double-blind Study (LOMAGHI Study)

CME Editor: Corey Heitz, MD

Authors: Wahid Bouida, MD, Kaouthar Beltaief, MD, Mohamed Amine Msolli, MD, Noussaiba Azaiez, MD, Houda Ben Soltane, MD, Adel Sekma, MD, Imen Trabelsi, MSc, Hamdi Boubaker, MD, Mohamed Habib Grissa, MD, Mehdi Methemem, MD, Riadh Boukef, MD, Zohra Dridi, MD, Asma Belguith, MD, and Semir Nouira, MD

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

Educational Objectives
After reading the article, participants should be able to discuss the utility of different doses of magnesium giving during rate control of atrial fibrillation in the emergency department.

Activity Disclosures
No commercial support has been accepted related to the development or publication of this activity.

No conflicts of interest or financial relationships relevant to this article were reported.

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.
Low-dose Magnesium Sulfate Versus High Dose in the Early Management of Rapid Atrial Fibrillation: Randomized Controlled Double-blind Study (LOMAGHI Study)

Wahid Bouida, MD, Kaouthar Beltaief, MD, Mohamed Amine Msolli, MD, Noussaiba Azaiez, MD, Houda Ben Soltane, MD, Adel Sekma, MD, Imen Trabelsi, MSc, Hamdi Boubaker, MD, Mohamed Habib Grissa, MD, Mehdi Methemem, MD, Riadh Boukef, MD, Zohra Dridi, MD, Asma Belguith, MD, and Semir Nouira, MD

ABSTRACT

Objectives: We aim to determine the benefit of two different doses magnesium sulfate (MgSO₄) compared to placebo in rate control of rapid atrial fibrillation (AF) managed in the emergency department (ED).

Methods: We undertook a randomized, controlled, double-blind clinical trial in three university hospital EDs between August 2009 and December 2014. Patients > 18 years with rapid AF (>120 beats/min) were enrolled and randomized to 9 g of intravenous MgSO₄ (high-dose group, n = 153), 4.5 g of intravenous MgSO₄ (low-dose group, n = 148), or serum saline infusion (placebo group, n = 149), given in addition to atrioventricular (AV) nodal blocking agents. The primary outcome was the reduction of baseline ventricular rate (VR) to 90 beats/min or less or reduction of VR by 20% or greater from baseline (therapeutic response). Secondary outcome included resolution time (defined as the elapsed time from start of treatment to therapeutic response), sinus rhythm conversion rate, and adverse events within the first 24 hours.

Results: At 4 hours, therapeutic response rate was higher in low- and high-MgSO₄ groups compared to placebo group; the absolute differences were, respectively, 20.5% (risk ratio [RR] = 2.31, 95% confidence interval [CI] = 1.45–3.69) and +15.8% (RR = 1.89, 95% CI = 1.20–2.99). At 24 hours, compared to placebo group, therapeutic response difference was +14.1% (RR = 9.74, 95% CI = 2.87–17.05) with low-dose MgSO₄ and +10.3% (RR = 3.22, 95% CI = 1.45–7.17) with high-dose MgSO₄. The lowest resolution time was observed in the low-dose MgSO₄ group (5.2 ± 2 hours) compared to 6.1 ± 1.9 hours in the high-dose MgSO₄ group and 8.4 ± 2.5 hours in the placebo group. Rhythm control rate at 24 hours was significantly higher in the low-dose MgSO₄ group (22.9%) compared to the high-dose MgSO₄ group (13.0%, p = 0.03) and the placebo group (10.7%). Adverse effects were minor and significantly more frequent with high-dose MgSO₄.

Conclusions: Intravenous MgSO₄ appears to have a synergistic effect when combined with other AV nodal blockers resulting in improved rate control. Similar efficacy was observed with 4.5 and 9 g of MgSO₄ but a dose of 9 g was associated with more side effects.
Atrial fibrillation (AF) is the most frequent cardiac arrhythmia and its incidence increases with age. For the management of AF in emergency department (ED), the physician must decrease the ventricular rate (VR) with or without restoration of sinus rhythm. Several drugs are recommended such as calcium channel blockers, beta-blockers, and digoxin, but the ultimate one is unknown. The use of magnesium as an alternative drug or in addition to usual care has been previously investigated. The rationale for its use was based on its physiologic and pharmacologic properties to decrease the frequency of sinus node depolarization, to prolong the refractory period of the atrioventricular (AV) node. It acts as a calcium antagonist by inhibiting L-type calcium current in heart cells. If not treated promptly, rapid AF can be associated with significant complications including congestive heart failure, hypotension, and cardiac ischemia. Intravenous magnesium is safe and cheap and may have a synergistic effect with usual antiarrhythmic drugs. A previously published meta-analysis conducted by Onalan et al. suggested that intravenous magnesium compared to placebo or standard rate control agents is an effective and safe strategy for the acute management of rapid AF. However, most of the included trials had small sample size or were performed in post-cardiac surgery patients. In addition, the dose of magnesium used in previous studies varied widely, which could influence its efficacy as AF rate control may be dose dependent. We are unaware of any study comparing two different doses of intravenous magnesium on the outcome of VR of AF. Accordingly, a definitive conclusion regarding the benefit of MgSO₄ in the rapid management of rapid AF is still uncertain and available results could not be extrapolated to patients treated in the ED.

We hypothesize that intravenous magnesium may have synergistic action with currently used rate-control agents in the treatment of patients with rapid AF. In addition, this action would depend on the dose of magnesium. The aim of this study was to investigate the efficacy and tolerance of magnesium sulfate (MgSO₄), administered at two different doses, to reduce VR in patients admitted to ED with rapid AF.

**METHODS**

**Study Design and Setting**

This is a prospective randomized, controlled, double-blind study carried out in three EDs of tertiary referral Tunisian hospitals with annual census of 90,000 to 110,000 adult patients. There are a total of 14 senior doctors and 36 residents working in the three participating EDs. Patients were enrolled between August 2009 and December 2014. The trial was registered on ClinicalTrials.gov registry (NCT00965874) and approved by the human research ethics committees of the participating centers. The study was not supported by any funding organization.

**Selection of Participants**

Consecutive patients over 18 years old admitted to the ED for rapid AF (≥120 beats/min) were eligible for enrollment. Patients were ineligible in presence of arterial hypotension (systolic arterial pressure < 90 mm Hg) if they had impaired consciousness, renal failure (serum creatinine > 180 μmol/L), wide-complex ventricular response, or contraindication to MgSO₄. We also excluded patients with acute myocardial infarction, acute congestive heart failure (New York Heart Association functional class 3 or 4), sick sinus syndrome, or rhythm other than AF. Informed consent was obtained from the patients or their relatives. All the treating physicians working in the three participating EDs have the prerogative not to enroll a patient if they deemed the individual too unstable for the trial.

**Methods and Measurements**

On arrival in the emergency department, all patients were administered oxygen as needed and an intravenous line inserted. A detailed history was taken including associated illness details and clinical examination was performed. Any medications received by the patient within 24 hours of their visit were recorded. After initial assessment, standard laboratory tests were performed and baseline serum magnesium was measured. Patients were then randomized to receive one of three treatments: 4.5 g intravenous MgSO₄ in 100 mL of normal saline (low-dose group), 9 g intravenous MgSO₄ in 100 mL of normal saline (high-dose group), or 100 mL of intravenous normal saline (placebo group). Protocol treatments were administered within 30 minutes. Study packs were prepared by the pharmacy department of Fattouma Bourguiba University Hospital. Each contained vials of experimental or placebo treatment and patient identification code. Randomization using random-number tables was achieved by blocks of three packs (one for each arm) by a pharmacist not involved with patient enrolment, data collection, or data analysis. The
MgSO₄ and placebo solutions were identical in appearance. Physicians and patients were both blinded to the randomization, which was done by random number. Physicians did not wait for the serum magnesium results before they started the protocol. All patients were monitored with continuous electrocardiographic monitoring. Blood pressure, respiratory rate, and pulse arterial oxygen saturation were recorded every hour. The same methods were used to record the VR data. Additional AV nodal blocking agents given at the same time as MgSO₄ were left at the discretion of the treating physicians and not mandated by the study protocol. In the three participant EDs, usual-care antiarrhythmics were not given, nor was electrical cardioversion done unless AF onset was diagnosed with certainty as a recent event (<48 hours). Any adverse effects noted by the patient or physicians were recorded on case report forms. Common adverse effects including flushing, nausea, vomiting, headache, dizziness, and hypotension were specifically sought and recorded. Patients were managed in the ED and data collected until 24 hours after randomization. If the patient was discharged or admitted prior to the 24 hours, data were no longer collected and the patient was excluded from the study. At this point, if not already undertaken, a final decision regarding hospital admission or home discharge was made. The decision to discharge the patient was taken by the attending emergency physician.

**Outcome Measures**

Primary endpoints of the study were VR control within the first 4 hours defined as reduction of baseline VR to 90 beats/min or less or reduction of VR by at least 20% from baseline (therapeutic response). Only patients who maintained these changes until the end of the protocol were considered to have achieved therapeutic response. Secondary endpoints included elapsed time from start of treatment to therapeutic response (resolution time), sinus rhythm conversion rate, and adverse events defined as major if they required treatment discontinuation or caused death.

**Data Analysis**

Analysis was undertaken on an intention-to-treat basis. Patients were removed from analysis after randomization only if recruitment was an unequivocal protocol violation (i.e., no consent had been recorded or if they had previously been recruited) or if the patient withdrew from the trial prior to any treatments having been administered. In all other cases, participants were analyzed in accordance with the groups they were allocated to regardless of whether or not they actually completed their allocated treatment. The study was designed to test the superiority of adjunctive low-dose MgSO₄ over placebo group. We estimated that a sample size on the basis on the following assumptions: with 145 patients on control treatment and 145 patients on MgSO₄, there will be a 80% chance of detecting a significant difference at a one-sided 0.05 significance level. This assumes that the response rate of control treatment is 0.5 and the response rate of MgSO₄ treatment is 0.65. The sample size was inflated by 3% to account for missing data, attrition, and protocol violations. Patient characteristics and outcome measures were reported as means with standard deviations (SDs) or medians and 95% confidence intervals (CIs), as appropriate. Descriptive and inferential statistical analyses (Kruskal-Wallis, Mann-Whitney rank sum, or Friedman tests for continuous variables; Fisher’s exact or chi-square tests for categorical data) were performed as appropriate. Pairwise comparisons were used in our analysis with Bonferroni adjustment. Nonparametric statistical techniques were used for the continuous data, as these data were not normally distributed. The risk ratio (RR) and 95% CI were calculated. Data obtained in this study have been recorded and analyzed with the SPSS computer software (Version 17). A p-value of <0.05 level was used to determine significant differences.

**RESULTS**

**Characteristics of Study Subjects**

A study enrollment flow diagram is displayed in Figure 1. A total of 469 patients underwent randomization; of these, 19 were withdrawn from the study prior to receiving the study medications. Of the 19 withdrawals, 11 patients withdrew consent before treatment, seven patients did not receive study medication because it was not available, and one patient left the ED for a procedure in the cardiology department. A total of 450 patients ultimately received the study medications, 149 in the placebo group, 148 in the low-dose MgSO₄ group, and 153 in the high-dose MgSO₄ group. Summary demographic and clinical characteristics for patients in the three study groups are presented in Table 1. There were no
significant differences among the three treatment groups with respect to baseline demographic or clinical characteristics. Utilization of rate-control medications in the three groups was similar in the three groups. Digoxin was the most used rate-control agent as usual care (47.5%).

Main Results
All groups showed reductions in VR relative to baseline as shown in Figure 2. At each time point there is a similarity in VR decrease from baseline across both MgSO₄ study groups, which reached statistical significance at time 4 hours (Figure 2). The superiority of MgSO₄ treatment groups compared to placebo group in decreasing HR was significant at 4 hours and persisted during all the protocol period. Therapeutic response rates at 4 and 24 hours are summarized in Table 2. The absolute difference was significant between the low-MgSO₄ group and the placebo group (absolute difference = 20.5%, RR = 2.31, 95% CI = 1.45–3.69) and between the high-MgSO₄ group and the placebo group (absolute difference = 15.8%, RR = 1.89, 95% CI = 1.20–2.99; Figure 3). The difference was not significant between both MgSO₄ groups (absolute difference = 4.7%, RR = 0.81, 95% CI = 0.51–1.30). At 24 hours, the therapeutic response rate was significantly higher in the low-MgSO₄ group (absolute difference = 10.3%, RR = 3.22, 95% CI = 1.45–7.17; Figure 3). Mean resolution time was 8.4 ± 2.5 hours in the placebo group, 6.1 ± 1.9 hours for the low-dose group, and 5.2 ± 2.0 hours for the high-dose group;

<table>
<thead>
<tr>
<th>Group</th>
<th>Randomized</th>
<th>Enrolled</th>
<th>Excluded</th>
<th>Withdraw Consent</th>
<th>Study Medication Not Available</th>
<th>Left ED for Cath Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnesium High dose group</td>
<td>153</td>
<td>153</td>
<td>19</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Magnesium Low dose group</td>
<td>148</td>
<td>148</td>
<td>19</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Placebo group</td>
<td>149</td>
<td>149</td>
<td>19</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 1. Patients’ flow chart.
the difference was statistically significant only between the placebo and the MgSO₄ groups. Conversion to sinus rhythm at 4 hours was achieved, respectively, in 10 patients from the placebo group (6.7%), in 18 patients from the low-dose group (12.1%), and in 12 patients from the high-dose group (7.8%); the difference was not statistically significant between the three groups. At 24 hours, rhythm control was achieved, respectively, in 16 patients from the placebo group (10.7%), in 34 patients from the low-dose group (22.9%), and in 20 patients from the high-dose group (13.0%). The difference was statistically significant between the low-dose group and the placebo group (p = 0.02). The most frequent adverse effect was transient flushing reported in 25 patients. The other adverse effects are transient hypotension observed in four patients (two in the high-dose group, one in the low-dose group, and one in the placebo group) and bradycardia observed in three patients, one in each group (Table 3). There was no death reported during the study, and in no patient was the protocol treatment stopped because of an adverse effect.

DISCUSSION

In this study, intravenous MgSO₄ appears to have a synergistic effect when combined with other AV nodal blockers resulting in improved rate control. Similar efficacy was observed with the 4.5 and 9 g of MgSO₄ but a dose of 9 g was associated with more side effects. Based on our findings, it seems that the logical approach is to combine MgSO₄ with usual rate-control agents to obtain efficient and more rapid action.

In the management of AF in ED, the objective is to rapidly decrease VR with or without restoration of sinus rhythm. Several drugs such as calcium channel blockers, beta-blockers, and digoxin are now the standard of care of rapid AF. However, current evidence
regarding the optimal VR control agents is limited. The benefits of magnesium have been suggested for both rate and rhythm control acting synergistically with antiarrhythmic drugs. However, prior research in this issue has focused predominantly on patients whose arrhythmia followed cardiac or thoracic procedures. It is likely that the etiology and pathology of postoperative AF in such patients differs from that of ED population. Additionally, small population sizes of earlier investigations may have resulted in type 2 statistical errors. So far, there are only four randomized controlled double-blind trials assessing MgSO₄ for rate control of rapid AF in non-cardiac surgery patients.¹⁰,¹⁶ Only two trials were performed in ED setting. These trials evaluated different alternative drugs and different protocols with regard the dose (4 to 6 g) and the duration (2 to 6 hours). In the largest study including 199 ED patients, Davey and Teubner⁶ found that MgSO₄ added to standard treatment was more likely than placebo to achieve a pulse rate of less than 100 beats/min (65% vs. 34%) and more likely to convert to sinus rhythm (27% vs. 12%) within the 150 minutes of study protocol. Their results were close to those observed in our study within the 4-hour period with regard to rate control. Importantly, our study showed that superiority of MgSO₄ regarding rate and rhythm control continued until 24 hours with a faster onset. Additionally, we demonstrated that using MgSO₄ at a dose of 9 g was not associated with greater efficacy on rate control compared to 4.5 g. Perhaps the more limited response

![Figure 2](image-url)  
**Figure 2.** Mean heart rate in relation to time in patients treated with low-dose MgSO₄, high-dose MgSO₄, and placebo. Repeated heart rate monitoring showed a significant and greater reduction of heart rate in both magnesium groups compared to placebo. *p < 0.05 versus baseline; †p < 0.05 versus placebo. GP = group; MgSO₄ = magnesium sulfate.

![Figure 3](image-url)  
**Figure 3.** Mean (and standard error) absolute difference of therapeutic response (as reduction of baseline VR to 90 beats/min or less or reduction of VR by at least 20% from baseline) between MgSO₄ groups and placebo group 4 and 24 hours after the start of the study protocol. *p < 0.05 compared to placebo. MgSO₄ = magnesium sulfate; VR = ventricular rate.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Rate Response from Baseline</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placebo</td>
<td>Low-dose Magnesium</td>
</tr>
<tr>
<td>4 hours</td>
<td>43.6% (35.7%–51.6%)</td>
<td>64.2% (56.5%–71.9%)*</td>
</tr>
<tr>
<td>24 hours</td>
<td>83.3% (77.2%–89.2%)</td>
<td>97.9% (95.7%–100%)</td>
</tr>
</tbody>
</table>

*p ≤ 0.05 versus placebo.
to a higher dose of magnesium was due to the lower baseline serum magnesium levels but this was not the case in this study. Our findings could also suggest that electrophysiologic effects of magnesium are probably dose related. In fact, Christiansen et al.\textsuperscript{17} demonstrated in a dose–response study that 5 mmol intravenous magnesium induced prolongation of AV node conduction but no further prolongation was observed with higher doses. Accordingly, we think that there is no need to use high dose of MgSO\textsubscript{4} in rapid AF and that using a dose of 4.5 g as in the present study would be effective. Whether using lower doses would be as effective, the question should be specifically investigated. Although MgSO\textsubscript{4} has a relatively wide toxic therapeutic window, the risk of adverse effects is possible\textsuperscript{6,18} and potentially more frequent with high MgSO\textsubscript{4} dose as shown in our study. Another important question should be discussed. Can magnesium be used as a single first-line agent or as an adjunctive treatment in rapid AF? In a double-blind, placebo-controlled clinical trial, Chu et al.\textsuperscript{19} demonstrated that 10 mmol intravenous MgSO\textsubscript{4} was not different from placebo for reducing VR or conversion to sinus rhythm at 2 hours posttreatment in ED patients with AF of less than 48 hours’ duration. The improved efficacy of MgSO\textsubscript{4} when added to standard treatment might also reduce the need to use higher doses of these agents, which is worthwhile regarding their potential adverse effects.

**LIMITATIONS**

First, we did not define a priori standard treatment that was left to the discretion of the ED physician. Of note, evidence-based treatment of rapid AF is still not well defined and current guidelines are mainly based on results of small studies or expert opinions.\textsuperscript{20} It should also be highlighted that most of our patients received one or more of the recommended rate-control agents in this setting with similar repartition in the three protocol groups. Nonetheless, we acknowledge that digoxin is no longer a commonly used acute rate-control agent and was the most commonly used agent in this study, which may impact the generalizability of our results. Second, we excluded patients with hemodynamic instability, those with severe left ventricular dysfunction, and patients with acute AF associated with and/or other cardiovascular comorbidities such as myocardial infarction; such exclusion will limit the generalization of our findings to these patients. Third, we did not conduct data collection in the patients included once they left the ED after the protocol. As such, no information regarding longer-term variables or complications was readily available. Fourth, there is a lack of consensus regarding optimal rate control in acute AF. The 2014 AHA/ACC/HRS guidelines described heart rate control as a resting heart rate of less than 80 bpm for symptomatic management of rapid AF. An outcome goal less than 100 beats/min was used by other studies. In this study, a VR control of 90 beats/min was between these values. Fifth, in this study we did not try to correlate serum magnesium to clinical response. An adequately powered study is needed to establish this correlation and its clinical relevance. Finally, we acknowledge that the choice of MgSO\textsubscript{4} doses in this study may lack some objectivity. However, available data indicated that MgSO\textsubscript{4} posology differs according to indications and several dosage recommendations have been proposed. In the rate control of AF, MgSO\textsubscript{4} was used at a dose ranging from 1.2 to 10 g. The usual dose seems between 4 and 5 g. Based on these findings, we chose 4.5 g as the reference dose and 9 g as the high dose.

**CONCLUSIONS**

Intravenous magnesium sulfate appears to have a synergistic effect when combined with other atrioventricular nodal blockers resulting in improved rate control. Similar efficacy was observed with the 4.5 and 9 g of magnesium sulfate but a dose of 9 g was associated with more side effects.

The authors acknowledge all of our colleagues at the three emergency departments who contributed greatly to this study.

**References**

1. American College of Cardiology/American Heart Association Task Force on Practice Guidelines/European Society of Cardiology Committee for Practice Guidelines and


Injury-based Geographic Access to Trauma Centers

Ran Wei, PhD, MS, N. Clay Mann, PhD, MS, Mengtao Dai, MS, and Renee Y. Hsia, MD, MSc

ABSTRACT

Objective: Previous studies examining access to trauma care use patient residence as a proxy for location and need for services, which could result in a flawed understanding of access to trauma centers. The objective of this study was to examine the geographic access of the U.S. population to trauma centers based on trauma incident locations.

Methods: We conducted a cross-sectional study using 9-1-1 emergency medical services activations associated with traumatic injury from the 2014 National Emergency Medical Services Information System and trauma centers participating in the 2014 American Hospital Association Annual Survey. The measures included the percentage of trauma incidents that could reach a trauma center within 60 minutes by ground ambulance, capacity-to-demand ratio for each trauma center, and overall trauma care accessibility ratio for each U.S. zip code.

Results: A total of 92.9% of all trauma incidents could be transported to an existing trauma center within 60 minutes by ground ambulance, and 85.3% could be transported to a Level I or II trauma center within this time frame in the 32 study states. While 94.7% of trauma incidents in the Northeast area could be transported to a Level I or II trauma center within a 60-minute driving time, the capacity-to-demand ratios of trauma centers in this region were low, indicating high utilization of those trauma center resources. By using the accessibility measure, we found that some Midwestern and Southern states had higher amounts of accessible trauma center resources relative to the number of injuries than Northeastern states.

Conclusions: These findings suggest that greater access to trauma care and significant variations can be observed throughout the 32 study states when using trauma incident location rather than patient residence to calculate access to trauma care. The proposed capacity-to-demand ratio and accessibility ratio can be applied to many other needs assessments in health care.

Trauma is one of the leading causes of death in the United States. Current research indicates that designated trauma centers significantly lower the risk of mortality and morbidity from trauma compared to nontrauma centers—with a landmark study showing a 25% reduction in 1-year mortality. A significant amount of research has been dedicated to evaluating geographic access to trauma centers. This notion is underscored by recent efforts by the American College of Surgeons Committee on Trauma to develop a needs-based assessment of trauma centers.

A common assumption for all of these studies is that the location of residence serves as an adequate proxy for the location of traumatic events. Based
on this assumption, the access of a traumatic injury to trauma centers can therefore be approximated by the proximity of a patient residence to trauma centers. However, a recent study has shown that there is, in fact, substantial geographic discordance between residential locations and trauma incident locations, with 63.5% of motor vehicle crashes, 59.3% of industrial accidents, 50.65% of mass casualty incidents, and 42.4% of other traumatic injuries occurring outside of a patient’s residence zip code.9 This study indicates that, for traumatic injury, the median network distance between residence and injury zip codes is as few 11 miles for patients older the 80 years and more than 15 miles for patients under 64 years of age. As a result, evaluating access to trauma center care using residence location could result in a flawed understanding of the distribution of trauma centers, providing imprecise information for trauma system planning.

To provide a more accurate assessment of trauma center resource allocation for health systems, it is therefore essential to employ trauma incident locations, rather than residence locations, to calculate demand for trauma centers. The goal of this article is to explore the U.S. population’s actual geographic access to trauma centers based on trauma incidents using the only national source of prehospital responder data available, the National Emergency Medical Services Information System (NEMSIS).

**METHODS**

**Study Design**

We incorporate three innovative perspectives in our study of geographic access to trauma incidents to trauma centers. First, we summarize the percentage of trauma incidents that could reach a trauma center within 60 minutes. Sixty minutes is chosen here not only due to the so-called notion of the “golden hour,”10 but also because it has been widely used in previous research to evaluate access to trauma centers from a home residence.1,5–7 While the actual binary categorization of whether or not a patient receives definitive care within the 60-minute window has been a subject of controversy,11,12 without question, earlier provision of health care is clearly linked to salvaging potentially preventable morbidity or mortality for patients with traumatic injuries.13–15 Second, we calculate a capacity-to-demand ratio for each trauma center, where the capacity of a trauma center is defined as the number of medical and surgical beds in the trauma center, and demand is estimated as the number of incidents within the 60-minute service zone of the trauma center. In addition to the geographic proximity, the capacity utilization of trauma centers can affect how accessible trauma resources are. Shortage of hospital beds is considered a major contributor for ambulance diversion,16–18 which is associated with impaired access, such as increased transport time and inability to receive treatment at preferred locations,19–21 and adverse outcomes, such as increased mortality.22,23 This capacity-to-demand ratio can approximately reflect the average degree of utilization and availability of trauma center resources when trauma patients arrive. Finally, we derive a trauma care accessibility measure for each incident zip code to evaluate the potential access of the zip code to trauma centers by summing the capacity-to-demand ratio of all trauma centers an injured patient in a given zip code can reach within 60 minutes. While the transport time from the zip code to the nearest trauma center seems to be a more intuitive accessibility measure, it does not capture the total number of accessible trauma center resources and the utilization capacity of accessible trauma centers. For example, the transport time to the nearest trauma center cannot distinguish zip codes that can access multiple trauma centers within 60 minutes from those that only have access to a single trauma center. This proposed accessibility measure takes into account those factors and reflects the overall accessibility to trauma center resources for each incident zip code.

**Data Sources**

NEMSIS is a unique national data set that archives the majority of 9-1-1–initiated emergency medical services (EMS) activations for emergency care in the United States. In 2014, the national EMS data set included 25,835,729 EMS activations from 45 states, representing approximately 72% of all EMS activations occurring in the United States.24 This data set is composed of patient care reports that are submitted by participating EMS agencies to state repositories; a subset of patient care data is then forwarded to the national EMS database. This subset of data contains 78 data elements, including clinical data as well as location and times associated with the EMS response and resulting care. A detailed description of the NEMSIS data can be found on the NEMSIS website (www.nemsis.org).

For this study, we chose to include only 9-1-1–initiated EMS activations that resulted in a patient
being treated and transported to a hospital by EMS. We included only activations dispatched for a possible injury or activations for which the provider’s primary or secondary impression of the patient’s condition was traumatic injury. We excluded medical transports to focus on first-responder calls rather than calls for an interhospital transfer and excluded all activations with a missing incident zip code. We included only states reporting that greater than 87% of all EMS activations are submitted to the NEMSIS National EMS registry, resulting in 32 study states. Eighty-seven percent was selected as the cutoff based on the observable distribution of record submissions by states, attempting to include a robust collection of states with near complete data capture. We included only activations where the transport mode from the scene was “Code 3” (i.e., “lights and sirens”) since we did not want to include minor injuries or nonurgent requests for assistance and our primary goal was to isolate true traumatic injuries that would require the expedient provision of definitive care. Given data availability and historical use of resident zip codes as a proxy for injury location in assessing access to medical services,7,25 we used the zip code where the trauma incident occurred as the trauma location to be studied. A detailed inclusion/exclusion graph is provided in Figure 1.

To include trauma center facility-level information in our study, we obtained the 2014 American Hospital Association Annual Survey, which sample includes general, acute, and short-stay hospitals and indicates whether a facility is designated as a trauma center. If a designated trauma center is identified, the survey also indicates the designated level of service (Level I through Level IV). The capacity of the trauma center, which is approximated by the number of general medical and surgical beds in the trauma center, may also be calculated from the survey. Finally, we captured the longitude and latitude coordinates of the identified trauma centers from the survey to calculate the transport time between the trauma center and an incident zip codes’ centroid.

### Service Zone Calculations and Validations

To evaluate whether a trauma incident could reach a trauma center within 60 minutes, we generated a

---

**Figure 1.** Constitution of the study population. NEMSIS = National Emergency Medical Services Information System.
60-minute service zone for each trauma center that encompasses all areas that could reach the center by driving within 60 minutes. Here we assume trauma centers can serve incidents occurring at its own states and neighboring states as long as the transportation time is less than 60 minutes. In other words, the service zone is not limited by political boundaries and could include areas from multiple states. Previous research in trauma center access also takes into account empirical time spent on the scene in addition to the transportation time.\(^1\)\(^5\)\(^6\)\(^7\) However, NEMSIS data suggest that there are significant variations in on-scene time across trauma incidents and the on-scene time and trauma center service zone might not be correlated. As a result, we validated the 60-minute service zones by using EMS activations that include both trauma incident zip code and destination hospital zip code from NEMSIS, not including on-the-scene time. Specifically, we derived the 60-minute service zone of the destination hospital and test if the incident is within the service zone. The total percentage of incident zip codes that are within the 60-minute service zones of the destination hospital zip codes is used to assess the validity of the 60-minute service zone.

We performed the service zone analysis with the ESRI ArcGIS Network Analyst Toolbox, where driving time is estimated using actual road network and speed limits. Once the 60-minute service zones of trauma centers were generated, the centroids of trauma incident zip codes were overlaid with service zones to determine if trauma incidents could be served by existing trauma centers within 60-minute ground transport time. We then used this to calculate the percentage of trauma incidents that were within 60-minute service zones of trauma centers.

### Capacity-to-demand Ratio Calculations

In addition to examining the percentage of trauma incidents that were within 60-minute service zones of trauma centers, we also calculated a capacity-to-demand ratio for each trauma center to approximate the utilization of trauma center resources. The capacity-to-demand ratio, denoted as \(R_i\), for each trauma center \(i\), was defined as

\[
R_i = \frac{B_i}{N_i},
\]

where \(B_i\) is the number of medical and surgical beds in the trauma center, and \(N_i\) represents the number of incidents within the 60-minute service zone of trauma center \(i\). The \(B_i\) was used as a proxy for capacity of trauma center as done in a previous study.\(^26\)\(^–\)\(^28\) The number of trauma incidents that can be potentially served by the trauma center was used to estimate the demand. A large capacity and small demand result in large capacity-to-demand ratio, indicating a potential underutilization of trauma center capacity, while small capacity but large demand led to small capacity-to-demand ratio, suggesting a potential shortage of trauma center capacity.

### Zip Code Accessibility Calculations

Once the capacity-to-demand ratios of trauma centers were determined, we summed up the capacity-to-demand ratios of all trauma centers that could be reached within 60 minutes from each zip code \(j\) and used the sum to represent the spatial accessibility of zip code \(j\), denoted as \(A_j\):

\[
A_j = \sum_{i \in \Omega_j} R_i,
\]

where \(\Omega_j\) represents the set of trauma centers that can be reached within 60 minutes from zip code \(j\). Such accessibility measures have been utilized in previous studies to evaluate the potential access to various medical services.\(^29\)\(^30\) The summation of capacity-to-demand ratios of all accessible trauma centers takes into account both the number of accessible trauma centers and the capacity utilization condition of accessible trauma centers, reflecting the overall accessibility to trauma center resources for each incident zip code.

### Sensitivity Analysis

While the service zone of a trauma center is not limited by political boundaries and could include areas from multiple states, we only take into account trauma centers that are in the 32 study states due to incomplete NEMSIS data in other states. Thus, non-participating states are not included in the calculation of 60-minute access percentage and accessibility score calculations. For example, when evaluating whether a zip code in Rhode Island is within the 60-minute service zone of a trauma center, we accounted for trauma centers in Rhode Island and Connecticut but those in Massachusetts are not included. We performed a sensitivity analysis to judge the extent to which this could alter the 60-minute access percentage and the zip code accessibility. Specifically, the 60-minute service zone is generated for all trauma...
centers in the continental United States; the number and percentage of trauma incidents that could reach a trauma center in the continental United States within 60 minutes are then calculated and compared with those where only trauma centers in the 32 study states are considered. To assess its impact in the zip code accessibility calculations, we exclude all zip codes that are within 60-minute service zones of trauma centers in the excluded states and then for each state test the null hypothesis as to whether the mean accessibility score in this scenario is equal to that where all zip codes are included.

RESULTS

A total of 424,670 EMS activations met study inclusion criterion (Figure 1). When comparing study-appropriate EMS activations from states included (and excluded) from the analysis, study states are relatively similar to all states submitting to NEMSIS, expecting in racial and geographic distributions (Table 1). The study sample is more likely white and underrepresents states in the Midwest and Western regions of the United States. The number of study EMS activations missing a transport mode from the scene was relatively few (3.2%).

Table 1
Descriptive Characteristics of Study Sample From Included and Excluded States

<table>
<thead>
<tr>
<th></th>
<th>Included Sample</th>
<th>Excluded Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-17</td>
<td>42,809 (10.1)</td>
<td>14,948 (9.8)</td>
</tr>
<tr>
<td>18-64</td>
<td>252,470 (59.5)</td>
<td>92,804 (60.8)</td>
</tr>
<tr>
<td>65-79</td>
<td>60,628 (14.7)</td>
<td>21,271 (13.9)</td>
</tr>
<tr>
<td>80+</td>
<td>65,131 (15.3)</td>
<td>22,837 (14.8)</td>
</tr>
<tr>
<td>(Missing)</td>
<td>3,632 (0.9)</td>
<td>1,079 (0.7)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>228,403 (53.8)</td>
<td>81,262 (53.2)</td>
</tr>
<tr>
<td>Female</td>
<td>195,189 (46.0)</td>
<td>71,109 (46.6)</td>
</tr>
<tr>
<td>(Missing)</td>
<td>1,078 (0.2)</td>
<td>368 (0.2)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>231,299 (54.5)</td>
<td>62,454 (40.9)</td>
</tr>
<tr>
<td>Black</td>
<td>75,341 (17.7)</td>
<td>30,898 (20.2)</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>2,014 (0.5)</td>
<td>1,353 (0.9)</td>
</tr>
<tr>
<td>Asian</td>
<td>3,325 (0.8)</td>
<td>2,806 (1.8)</td>
</tr>
<tr>
<td>Other</td>
<td>11,222 (2.6)</td>
<td>16,991 (11.1)</td>
</tr>
<tr>
<td>(Missing)</td>
<td>101,569 (23.9)</td>
<td>38,237 (25.0)</td>
</tr>
<tr>
<td><strong>U.S. Census Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>167,815 (39.5)</td>
<td>0</td>
</tr>
<tr>
<td>Midwest</td>
<td>38,122 (9.0)</td>
<td>111,802 (73.2)</td>
</tr>
<tr>
<td>South</td>
<td>215,342 (50.7)</td>
<td>1,336 (0.9)</td>
</tr>
<tr>
<td>West</td>
<td>11,472 (2.7)</td>
<td>39,601 (25.9)</td>
</tr>
<tr>
<td>(Missing)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Cause of Injury</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor vehicle traffic crash</td>
<td>70,920 (16.7)</td>
<td>22,926 (15.0)</td>
</tr>
<tr>
<td>Motor vehicle nontraffic crash</td>
<td>9,518 (2.2)</td>
<td>3,062 (2.0)</td>
</tr>
<tr>
<td>Stabbing/cutting assault</td>
<td>6,232 (1.5)</td>
<td>1,066 (0.7)</td>
</tr>
<tr>
<td>Firearm assault</td>
<td>5,792 (1.4)</td>
<td>2,439 (1.6)</td>
</tr>
<tr>
<td>Other</td>
<td>167,306 (39.4)</td>
<td>63,643 (41.7)</td>
</tr>
<tr>
<td>(Missing)</td>
<td>164,902 (38.8)</td>
<td>59,603 (39.0)</td>
</tr>
<tr>
<td><strong>Transport mode from scene</strong></td>
<td>T = 2,378,128</td>
<td>T = 504,075</td>
</tr>
<tr>
<td>Lights and sirens</td>
<td>424,670 (17.6)</td>
<td>152,739 (30.3)</td>
</tr>
<tr>
<td>(Missing)</td>
<td>76,733 (3.2)</td>
<td>11,615 (2.3)</td>
</tr>
</tbody>
</table>

Data are reported as n (%).
The 2014 American Hospital Association Annual Survey identified 1,709 designated trauma centers in the United States, with 890 general trauma centers and 17 pediatric trauma centers in the 32 study states. Among those 890 general trauma centers, 144 were designated Level I, 241 were Level II, and the remaining 505 were Level III or IV as shown in Figure 2.

Among 424,670 EMS activations meeting inclusion criteria (and had an incident zip code recorded), 92.9% patients \( (n = 394,605) \) could be transported to an existing trauma center (any level) within a 60-minute driving time by a ground ambulance (Figure 3A), and 85.3% \( (n = 362,227) \) could be transported to a Level I or II trauma center within a 60-minute driving time (Figure 3B). Due to the relative completeness of NEMSIS data in northeastern states, the access of trauma centers in the Northeast is highlighted to make comparison with previous studies. There are 165,497 EMS activations in the Northeast region and 94.7% \( (n = 156,659) \) could be transported to a Level I or II trauma center within a 60-minute driving time.

Among EMS activations treating injuries meeting inclusion criteria, 243,010 activations recorded both an incident zip code and destination hospital zip code, which were used to validate the 60-minute service zone. A total of 99.3% of these activations contained incident zip codes that were within a 60-minute service zone of the destination hospital, indicating that the 60-minute driving time is a reasonable estimate for delineating service zones for trauma centers.

Among the 890 trauma centers, four of them (one in Indiana, one in Kentucky, one in Oklahoma, and one in Colorado) do not provide the number of general medical and surgical beds information, and 11 of them (three in Kentucky, five in Louisiana, one in Oklahoma, one in Idaho, and one in Utah) have zero traumatic incidents identified within the 60-minute service zone. These 15 trauma centers are therefore excluded from the capacity-to-demand ratio calculation to ensure that the ratio is valid and divisible. The capacity-to-demand ratios of trauma centers range from 0.00 to 23.33 and are classified into five subsets with each containing 175 centers. Values nearer to 0.00
Figure 3. Service areas of trauma centers: (A) all trauma centers and (B) Level I or II only.
indicate a potential shortage in trauma center resources (small capacity compared to demand) and values nearer 23.33 indicate a potential underutilization of trauma center resources due to a large capacity versus demand ratio. The distributions of Level I, Level II, and Level III or IV trauma centers in each subset are shown in Figure 4. Thirty-two percent of Level I trauma centers have capacity-to-demand ratios falling between 0.00 and 0.02, whereas 22% of Level II and 16% of Level III or IV trauma centers have ratios in this category. This is because Level I trauma centers are serving the largest number of trauma incidents in comparison with Level II and Level III or IV. It is also interesting to note that 26% of Level II trauma centers have capacity-to-demand ratios in the highest category, 0.36 to 23.33, in comparison to 12% of Level I and 20% of Level III or IV. The distribution of trauma centers in the Northeast region (n = 150) in each subset shows that 61.3% of trauma centers in the Northeast region have capacity-to-demand ratios in the lowest category, 0.00–0.02, and 4% of trauma centers in the Northeast region have capacity-to-demand ratios in the highest category, 0.36 to 23.33. While many trauma centers are clustered in the Northeast area (Figure 2), the capacity-to-demand ratios of those centers are relatively low. This is because of the more dense number of trauma incidents occurring in those areas.

The accessibility score for each zip code is displayed in Figure 5, where the scores vary from 0.00 to 32.08 and are classified into five subsets of equal size. A large accessibility score indicates a higher amount of trauma center resources (capacity) which are accessible within 60 minutes relative to number of injuries (demand), while a small score suggests a lower amount of accessible trauma center resources relative to number of injuries. The zip code accessibility score varies significantly across the 32 states, where low to medium scores are observed in most Northeastern states, and high scores are observed in some Midwestern states like Minnesota and Kentucky and Southern states like Mississippi and Louisiana. Significance testing also show that the accessibility scores in Midwestern and Southern region are significantly larger than those in Northeastern region (p < 0.001). Most zip codes in the five studied western states have zero accessibility scores. It is important to note that the accessibility score is calculated based on a binary service coverage of trauma centers. That is, a zip code is either considered to be completely covered by a trauma center if the centroid of the zip code is within the 60-minute service zone of the trauma center; otherwise, any portion of the zip code is considered to be not served by the trauma center. This could cause potential bias and sharp accessibility score transitions in Figure 5 when the area of zip code is large as partial coverage might not be considered. Table 2 summarizes the average accessibility score and average expected EMS transport time to the nearest trauma center from zip code centroids by state. When only Level I and II trauma centers are taken into account, New Mexico has the lowest average accessibility score whereas Louisiana...
has the highest average accessibility score. Yet, both New Mexico and Louisiana have long average transport times to the nearest Level I or II trauma center. This is because while 383 of 473 zip codes in Louisiana have an accessibility score of zero, the rest of the zip codes within the state, especially in the New Orleans metropolitan area, have very high accessibility scores due to a concentration of several Level I or II trauma centers with high capacity but low number of traumatic incidents in their 60-minute service zones. When all trauma centers are taken into account, the lowest accessibility score is observed in Rhode Island and highest score in Mississippi. However, it is interesting to note that both Rhode Island and Mississippi have short average transport times to the nearest trauma center in this scenario. This corresponds to the high and low number of traumatic incidents in the 60-minute service zone of trauma centers in Rhode Island and Mississippi, respectively.

The sensitivity analysis shows that the inclusion of all trauma centers in the continental United States increased the 60-minute access percentage by less than 1% (n = 994) for the 424,670 trauma incidents in the 32 study states. Among the 18,145 zip codes in the 32 study states, 9.9% (n = 1,799) zip codes are also within 60-minute service zones of trauma centers in excluded states. A sensitivity analysis of the impact in zip code accessibility, after excluding these zip codes, shows that the mean accessibility scores of Pennsylvania and New Jersey become significantly different from those where all zip codes are included (p < 0.05), whereas the other 30 states do not demonstrate significant differences.

**COMMENT**

While many previous studies have examined geographic access to trauma centers, they use residence locations as a proxy for trauma incident locations. To our knowledge, this study is the first to explicitly examine geographic access to trauma centers using trauma incident locations, and provides several important new findings.

First, 92.9% of trauma incidents in the 32 studied states can be transported to an existing trauma center.
within an hour by ground ambulance, and 85.3% of them can be transported to a Level I or II trauma center within this time frame. In addition, 94.7% of trauma incidents in the Northeastern states can be transported to a Level I or II trauma center within an hour by ground ambulance. A previous national study in 2005 based on residence locations showed that 56.4% of population in the United States and 73.2% of population in the Northeast have access to Level I or II trauma centers within an hour by ground ambulance. The discrepancies of 28.6% in the United States and 21.5% in the Northeast is likely not only attributable to the different temporal period, but also due to the use of trauma incident locations rather than residence locations. As most trauma incidents occur in urban areas that have more trauma center resources, trauma centers will likely appear to provide better service coverage if considering trauma incident locations when compared with home residences.

Second, the capacity-to-demand ratios of trauma centers suggest significant variations in the utilization of trauma center resources throughout different levels of trauma centers and different states in the United States. While Level I trauma centers usually have the

<table>
<thead>
<tr>
<th>State</th>
<th>No. of Zip Codes</th>
<th>Level I and II Only</th>
<th>All Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accessibility Score</td>
<td>Transport Time to Nearest TC (min)</td>
<td>Accessibility Score</td>
</tr>
<tr>
<td>Alabama</td>
<td>599</td>
<td>0.21 (0.18–0.24)</td>
<td>56 (54–59)</td>
</tr>
<tr>
<td>Arkansas</td>
<td>572</td>
<td>0.32 (0.27–0.37)</td>
<td>82 (78–85)</td>
</tr>
<tr>
<td>Colorado</td>
<td>460</td>
<td>0.48 (0.41–0.55)</td>
<td>83 (76–90)</td>
</tr>
<tr>
<td>Connecticut</td>
<td>269</td>
<td>0.18 (0.17–0.19)</td>
<td>22 (20–24)</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>25</td>
<td>0.16 (0.16–0.16)</td>
<td>4 (4–5)</td>
</tr>
<tr>
<td>Florida</td>
<td>885</td>
<td>0.2 (0.19–0.21)</td>
<td>39 (37–41)</td>
</tr>
<tr>
<td>Georgia</td>
<td>691</td>
<td>0.16 (0.14–0.17)</td>
<td>57 (54–60)</td>
</tr>
<tr>
<td>Idaho</td>
<td>252</td>
<td>0.1 (0.07–0.13)</td>
<td>199 (178–220)</td>
</tr>
<tr>
<td>Indiana</td>
<td>687</td>
<td>0.72 (0.67–0.77)</td>
<td>52 (50–54)</td>
</tr>
<tr>
<td>Kentucky</td>
<td>713</td>
<td>0.76 (0.68–0.84)</td>
<td>69 (66–72)</td>
</tr>
<tr>
<td>Louisiana</td>
<td>473</td>
<td>1.98 (1.58–2.38)</td>
<td>132 (125–139)</td>
</tr>
<tr>
<td>Maine</td>
<td>383</td>
<td>0.23 (0.2–0.26)</td>
<td>48 (44–52)</td>
</tr>
<tr>
<td>Maryland</td>
<td>409</td>
<td>0.09 (0.09–0.1)</td>
<td>34 (32–37)</td>
</tr>
<tr>
<td>Michigan</td>
<td>898</td>
<td>0.59 (0.55–0.63)</td>
<td>50 (47–53)</td>
</tr>
<tr>
<td>Minnesota</td>
<td>860</td>
<td>0.73 (0.6–0.85)</td>
<td>87 (83–92)</td>
</tr>
<tr>
<td>Mississippi</td>
<td>384</td>
<td>1.33 (1.17–1.48)</td>
<td>41 (39–44)</td>
</tr>
<tr>
<td>Missouri</td>
<td>985</td>
<td>0.32 (0.29–0.35)</td>
<td>73 (70–76)</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>245</td>
<td>0.08 (0.07–0.09)</td>
<td>37 (34–40)</td>
</tr>
<tr>
<td>New Jersey</td>
<td>533</td>
<td>0.17 (0.17–0.18)</td>
<td>19 (17–20)</td>
</tr>
<tr>
<td>New Mexico</td>
<td>284</td>
<td>0.02 (0.01–0.03)</td>
<td>180 (169–192)</td>
</tr>
<tr>
<td>New York</td>
<td>1561</td>
<td>0.19 (0.18–0.19)</td>
<td>39 (38–41)</td>
</tr>
<tr>
<td>North Carolina</td>
<td>721</td>
<td>0.24 (0.22–0.26)</td>
<td>53 (51–56)</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>591</td>
<td>0.15 (0.12–0.17)</td>
<td>89 (84–94)</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>1478</td>
<td>0.19 (0.18–0.2)</td>
<td>43 (41–44)</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>70</td>
<td>0.06 (0.06–0.06)</td>
<td>18 (15–20)</td>
</tr>
<tr>
<td>South Carolina</td>
<td>380</td>
<td>0.14 (0.13–0.16)</td>
<td>48 (43–53)</td>
</tr>
<tr>
<td>Utah</td>
<td>226</td>
<td>0.49 (0.4–0.58)</td>
<td>76 (66–86)</td>
</tr>
<tr>
<td>Vermont</td>
<td>246</td>
<td>0.2 (0.17–0.23)</td>
<td>53 (50–56)</td>
</tr>
<tr>
<td>Virginia</td>
<td>825</td>
<td>0.18 (0.16–0.19)</td>
<td>49 (47–51)</td>
</tr>
<tr>
<td>West Virginia</td>
<td>581</td>
<td>0.09 (0.08–0.11)</td>
<td>68 (65–70)</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>718</td>
<td>0.39 (0.31–0.47)</td>
<td>51 (49–54)</td>
</tr>
<tr>
<td>Wyoming</td>
<td>141</td>
<td>0.4 (0.22–0.57)</td>
<td>83 (76–91)</td>
</tr>
</tbody>
</table>

Data are reported as mean (95% CI). TC = trauma center.
greatest capacity, there is a much larger number of incidents occurring within their 60-minute service zone compared with Level II and Level III or IV. This results in a higher percentage of Level I trauma centers having low capacity-to-demand ratios and potential shortages in resources, compared with Level II and Level III or IV centers. It is well known that there are an abundance of trauma centers in the Northeast area, and many studies show that this region has the best geographic access to trauma center resources. However, the occurrence (i.e., density) of trauma incidents in this region is also very high. The low capacity-to-demand ratios of trauma centers in this region indicate high utilization of those trauma center resources. It is important to take into account the demand and existing utilization of trauma center resources when evaluating the allocation of trauma center resources.

Finally, the accessibility analysis provides a novel measure to compare areas with large accessibility scores versus those with low accessibility scores. While at first glance, it might be assumed that large accessibility scores equate to “better” coverage, another interpretation could be that large accessibility scores could also point to potential underutilization (or overcapacity) of trauma resources in certain areas. Further research in this area could answer questions related to whether reallocation of trauma center resources might be necessary to improve population access to trauma centers and, ultimately, population health.

These findings are particularly relevant given that tools to determine the adequacy of the provision of trauma services are under development. In 2015, the American College of Surgeons Committee on Trauma convened a consensus conference to discuss the development of a “Needs-Based Assessment of Trauma Systems” tool to determine the number of trauma centers that each region would require based on a number of factors including population size, transport times, and number of severely injured patients hospitalized at trauma and nontrauma centers. These inputs are based on where patients already receive care and could be considered endogenous, given that patients only receive care where facilities exist. The factors in our model (i.e., incident location and transport time) are independent of the current location of trauma centers and therefore allow for a more de novo determination of need, which can then be matched with the current provision of services to identify where gaps may exist. As more and more states submit EMS data to NEMSIS, a complete assessment of trauma center access using the traumatic incident locations for the entire United States will be possible.

On a larger level, the accessibility measures we propose in this article can be more broadly extracted to other assessments of need in health care. The broader discussion of appropriate provision and regionalization of emergency care, as well as other types of time-sensitive care (e.g., cardiac emergencies, burns, or pediatric critical care), have all previously been based on residence location and may also benefit from a perspective that includes these calculations that are based in operational engineering and extracted to health care settings. Because such tools do not yet exist, our findings provide a test case for determining adequacy of service provision by both accounting for demand as well as capacity.

There are several limitations associated with our study. First, 32 of the available 46 states submitting to NEMSIS are included in the study to ensure near complete capture of all related EMS activations. This criterion selectively limited the sample from some geographic regions and underrepresented some racial groups. While the inclusion of other states could lead to an evaluation of access to trauma centers for the entire United States, a near complete set of EMS activations are not currently available for all states. As a result, only 32 study states were included in the calculation of 60-minute access percentage and accessibility score calculations. However, the sensitivity analysis shows that the inclusion of all trauma centers only increased the 60-minute access percentage by less than 1% and the mean accessibility scores of 30 states do not have significant difference. Second, the criteria we employ to identify incidents that may need care at trauma centers may not reflect the actual injury severity. We do include primary and secondary impression from the provider as well as call information to identify traumatic events and limit our study to those requiring “lights and sirens” transport from the scene, but certainly there may be traumatic injuries where 9-1-1 is not called or where the provider impression is incorrect. There may also be cultural (regional, racial/ethnic, age, sex) differences in who initiates a 9-1-1 call, even though we tried to limit our study to include calls for more severe types of injury where there would be less discretion about whether 9-1-1 should be called. Our findings could therefore be conservative if the incidents are underestimates in certain regions. The third limitation is associated with the use of zip codes as incident locations. This is due to NEMSIS
data limitations and we attempt to mitigate potential biases by relying on travel time from centroids to represent the natural distribution of all trauma incidents within each zip code area. Another limitation in service zone delineation is that driving time is estimated using real road network and speed limits but the actual speeds of ground ambulance vehicles could vary given the actual traffic conditions encountered. Finally, we include only trauma incidents that use ground ambulance for transport. Taking aeromedical transport into account would likely enhance our estimation of trauma care access. However, given the paucity of detailed data for trauma requiring aeromedical transport, our findings can be considered conservative in this regard.

CONCLUSION

This study using trauma incident locations finds that 92.9% of trauma incidents can be transported to an existing trauma center within an hour by ground ambulance, and 85.3% can be transported to a Level I or II trauma center in the 32 study states. Our calculations of the capacity-to-demand ratio reveal significant variations in the utilization of trauma center resources throughout the 32 states. We derive an accessibility measure that integrates both the number of accessible trauma centers and the utilization status of accessible trauma centers to help identify areas that have high amounts of accessible trauma center resources relative to the number of injuries and areas that have low amounts of accessible trauma center resources relative to the number of injuries.

The authors thank Sarah Sabbagh, MPH, for her editorial assistance.

References

26. Rivara FP, Nathens AB, Jurkovich GJ, Maier RV. Do trauma centers have the capacity to respond to disasters? J Trauma Acute Care Surg 2006;61:949–53.
Referring Hospital Characteristics Associated With Potentially Avoidable Emergency Department Transfers

Kristin N. Ray, MD, MS, Jennifer R. Marin, MD, MSc, Joyce Li, MD, MPH, Billie S. Davis, PhD, and Jeremy M. Kahn, MD, MS

ABSTRACT

Objective: Many emergency department (ED) transfers of children may be avoidable. Identifying hospital-level variables associated with avoidable transfers may guide system-level interventions to improve pediatric emergency care. We sought to examine hospital characteristics associated with ED transfers deemed “probably avoidable” in a large state Medicaid program.

Methods: We performed a retrospective cohort study using 2009 to 2013 claims data for Pennsylvania Medicaid beneficiaries. We categorized all ED transfers of children < 17 years old as “probably avoidable,” “possibly avoidable”, or “unavoidable” based on ultimate disposition and procedures (including subspecialty consultations) at the receiving ED. Using descriptive statistics and multivariable regression, we examined hospital characteristics associated with probably avoidable transfers.

Results: Among 2,839,379 pediatric visits to EDs across 158 Pennsylvania hospitals, 20,304 resulted in transfer. Among these, 3,764 (18.5%) were categorized as probably avoidable and 6,091 (30.0%) as possibly avoidable transfers. In adjusted analysis, compared to hospitals with no pediatric-specific capabilities, probably avoidable transfers were less likely from referring hospitals with pediatric-specific EDs and no other pediatric-specific capabilities (adjusted odds ratio [aOR] = 0.38, 95% confidence interval [CI] = 0.21–0.71) and from referring hospitals with pediatric-specific EDs and inpatient capabilities (aOR = 0.36, 95% CI = 0.20–0.64). Probably avoidable transfers were more likely from referring hospitals in large metropolitan areas (aOR = 2.64, 95% CI = 1.46–4.80) compared to those in rural areas.

Conclusions: Among pediatric emergency transfers in a large state Medicaid program with a nearly 20% probably avoidable transfer rate, there was significant hospital-level variation in the proportion of probably avoidable transfers. Transfers from hospitals in large metropolitan areas and transfers from hospitals without pediatric-specific capabilities had increased odds of being probably avoidable transfers, such that these hospitals may represent targets for interventions to reduce these transfers.
Children account for over 23 million emergency department (ED) visits per year. Over 70% of ED visits by children occur at nonpediatric hospitals. Some of these hospitals may be underprepared to care for children with pediatric emergencies, lacking pediatric-specific infrastructure, equipment, and personnel. In these cases, a cornerstone in emergency care of children is interfacility transfer to an ED or hospital capable of providing a higher level of pediatric care when needed. Despite the potential benefits of such transfers, there are also potential drawbacks, including associated travel, time, and monetary costs for both payers and families. If transfers are not associated with improved outcomes, they may ultimately represent low-value care.

In many cases, transfer to a pediatric facility is necessary for specialized care. However, as many as 55% of transferred patients are discharged home from the receiving ED after little or no additional interventions, suggesting that some transfers may be avoidable. Attention to these transfers has been part of efforts to improve pediatric emergency care. While prior studies have described patient and clinical characteristics of transfers that may be avoidable, hospital characteristics and hospital-level variation have not been examined to the same degree. Understanding hospital-level characteristics and variation may help inform strategies to reduce unnecessary transfers.

To better understand this issue, we sought to determine the prevalence of pediatric emergency transfers that may have been avoidable, to assess hospital-level variation in potentially avoidable transfer rates, and to examine hospital characteristics associated with these transfers. We focused on a large state Medicaid program since Medicaid is a major payer for pediatric health care, and nearly a quarter of pediatric Medicaid beneficiaries visit the ED annually. We hypothesized that identifiable hospital characteristics would be associated with “probably avoidable” transfers and that variation in the proportion of transfers that were probably avoidable would exist across hospitals even after accounting for patient characteristics.

**METHODS**

**Study Design**

We performed a retrospective cohort study using 2009 to 2013 administrative claims data for Pennsylvania Medicaid beneficiaries. Pennsylvania is the sixth largest U.S. state, with Pennsylvania Medicaid covering 38% of Pennsylvania children. These data, which are available through an ongoing partnership between the Pennsylvania Department of Human Services and our institution, include final action claims for all fee-for-service and managed care Medicaid beneficiaries. These data were augmented with data on hospital characteristics from the American Hospital Association and the Centers for Medicare & Medicaid Services’ Healthcare Cost Report Information System. This study used de-identified data and was judged to be exempt by our institution’s institutional review board.

**Study Population**

We included all Pennsylvania Medicaid beneficiaries aged 2 days through 16 years, inclusive, with an ED encounter identified using previously published methods. We included ED visits initiated at Pennsylvania general and pediatric short-stay hospitals.

**Identifying ED Transfers**

We first classified all ED visits based on the patients’ disposition from the presenting hospital: discharge to home, admission to the presenting hospital, and transfer to another hospital. Transfers to another hospital were identified based on an initial ED visit followed by an ED or inpatient claim at a different hospital on the same day or one day apart with an intervening ambulance claim. We excluded transfers to psychiatric hospitals. Because claims are date-stamped but not time-stamped, we ranked EDs based on annual pediatric visit volume and assigned transfer direction as moving from lower pediatric volume to higher pediatric volume EDs.

**Defining “Probably Avoidable” and “Possibly Avoidable” Transfers**

We then categorized transfers into three mutually exclusive groups (“probably avoidable,” “possibly avoidable,” or “unavoidable”) based on 1) patient disposition and 2) procedures/consultations at the receiving ED.

1) Patient disposition after transfer was categorized as discharged from the receiving ED, admitted to the receiving hospital for < 2 days, or either admitted for ≥ 2 days or died. Prior studies of potentially avoidable transfers have similarly distinguished brief admissions from more prolonged hospitalizations.

2) Procedures after transfer were identified using International Classification of Diseases, Version 9.0, Clinical Modification (ICD-9-CM) procedure codes. We focused on procedures (i.e., resource utilization)
because classification of ED visits by resource utilization provided higher concordance and agreement with full expert chart review than other methods (i.e., diagnoses, triage score) in prior work. While some prior studies have examined whether any procedures occurred after transfer, we sought to refine this approach by differentiating between procedures that require transfer and procedures with the potential to be performed without transfer. We first identified all procedure codes (n = 1,286) at receiving hospitals. Two board-certified pediatric emergency medicine physicians with experience in several pediatric and general EDs independently rated the likelihood that each procedure would be routinely available at any ED. We used three categories: “always receivable” at any ED (e.g., x-ray wrist, bilirubin level); “potentially receivable” at any ED (e.g., treat fracture radius and ulna, computed tomography [CT] scan head/brain); and “never receivable” at nonspecialized ED (e.g., continuous mechanical ventilation 96+ hours, remove ventricle shunt). To categorize procedures, we considered not only overall availability but also whether they can be performed in a manner that is effective, safe, timely, and pediatric-specific. For example, receipt of an effective, safe, timely, and pediatric-specific head CT scan requires having a CT scanner, having pediatric-specific CT protocols, having age-appropriate immobilization/sedation options, and having timely pediatric radiologic interpretation. Given the need for all of these components for effective, safe, timely, and pediatric-specific head CT, this procedure was categorized as potentially receivable at any ED but not always receivable. We chose this conservative definition in recognition of the numerous factors that may limit the ability to provide pediatric-specific care at referring EDs to avoid overestimation of the number of probably avoidable transfers. Concordance between the two reviewers after initial review was adequate (weighted kappa = 0.66), with discordancess resolved through iterative discussion.

Additionally, because consultations are another intervention that can prompt transfer, we identified pediatric subspecialty consults using professional claims. We identified Current Procedural Terminology codes used for subspecialty consultations (99201–99215, 99241–99245, 99281–99285, 99221–99223, 90792) occurring on the day of transfer or first day after transfer. We then used the physician’s National Provider Identifier code and associated taxonomies to further identify instances where the performing provider was a pediatric subspecialist (i.e., pediatric cardiologist, pediatric orthopedist), excluding pediatric emergency medicine physicians and diagnostic specialties (i.e., pediatric radiology, pediatric pathology). Because pediatric subspecialist consultation can be provided in some hospitals via telemedicine, pediatric subspecialty consults were categorized as potentially receivable but not always receivable.

After determining disposition and procedures/consultations, transfers were determined to be probably avoidable if the child 1) was discharged from the ED and 2) received either no procedure or only procedures always receivable at general EDs (Figure 1). These probably avoidable transfers are the focus of our main analysis.

For the purpose of sensitivity analysis examining the impact of a more inclusive definition of “avoidable,” we further classified the remaining transfers as either unavoidable or possibly avoidable. The transfer was unavoidable if the child received at least one procedure classified as never receivable at general EDs, was admitted for ≥2 days posttransfer, or died (Figure 1). The remaining transfers were possibly avoidable, with these transfers including either receipt of a potentially receivable procedure or brief admission (<2 days).

Patient, Visit, and Hospital Characteristics

Patient sex, age, race, and visit year were determined from Medicaid enrollment files. Using primary diagnosis code at the presenting hospital, a published, publicly available pediatric diagnosis grouping system was applied that assigned both diagnosis group and illness severity.

Hospital characteristics were determined using 2010 to 2011 American Hospital Association data and 2005 to 2011 Centers for Medicare & Medicaid Services’ Healthcare Cost Report Information System data. We identified five levels of hospital pediatric-specific capabilities: 1) pediatric ED, pediatric inpatient care, and pediatric intensive care unit (PICU) (termed “highly pediatric-resourced hospitals,” including children’s hospitals); 2) pediatric ED and pediatric inpatient care; 3) pediatric inpatient care only; 4) a pediatric ED only; or 5) no pediatric ED, inpatient, or PICU (“no pediatric-specific capability”). We also identified hospital size, annual ED volume, metropolitan statistical area (MSA) size, and straight-line distance from each hospital to the closest highly pediatric-resourced hospital.
Data Analysis
We used descriptive statistics to summarize patient and hospital characteristics of ED visits stratified by disposition (i.e., discharged, admitted, transferred) and of ED transfers stratified by transfer type (i.e., probably avoidable, possibly avoidable, and unavoidable). We used chi-square tests to determine the significance of differences between these groups.

We then performed two multivariable analyses. First, to determine patient and hospital characteristics independently associated with any transfer, we used multivariable logistic regression with clustering of standard errors by hospital. The dependent variable was transfer (vs. nontransfer, including discharge from or admission to the presenting hospital). Second, to determine patient and hospital characteristics associated with probably avoidable transfer after adjusting for other characteristics, we used multivariable logistic regression with clustering of standard errors by hospital, with probably avoidable transfer (vs. possibly avoidable and unavoidable transfers) as the dependent variable. As a sensitivity analysis for this second model, we also fit a regression model using probably avoidable or possibly avoidable transfer (vs. unavoidable transfer) as the dependent variable. Independent variables in each model, which were selected a priori based on our conceptual model of variables thought to impact ED visit disposition, included hospital pediatric-specific capabilities, ED volume, MSA size, distance to closest highly pediatric-resourced hospital, child age, sex, race/ethnicity, Medicaid eligibility category, illness severity code, diagnosis group, and year. All hypothesized variables were retained in the final models, and the area under the receiver operating curves for our main models indicated fair to good discrimination (>0.70). We excluded visits and transfers initially presenting to highly pediatric-resourced hospitals due to infrequent transfers from these hospitals. Finally, we used indirect standardization to estimate individual hospital-specific risk-adjusted transfer rates and hospital-specific risk-adjusted probably avoidable transfer rates from separate models using only patient/visit characteristics (and not including hospital characteristics). Use of indirect standardization, based on the ratio between observed and expected transfer rates at each hospital, allowed comparison of hospital transfer rates after adjustment for the characteristics of each hospital’s patient population. Analyses were conducted in Stata 14 (StataCorp) using an alpha level of 0.05 to test for statistical significance.

RESULTS
Characteristics of ED Visits and Transfers
Among 2,839,379 ED visits by children < 17 years to 158 hospitals in Pennsylvania, 20,304 visits resulted in transfer to another ED (0.7%; 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.13519/full). Unadjusted, the proportion of ED visits resulting in transfer was higher among children presenting to hospitals with a pediatric ED only compared to those presenting to highly pediatric-resource hospitals or hospitals with pediatric inpatient unit but no pediatric ED (p < 0.001). The proportion of ED visits resulting in transfer was also higher for children presenting to hospitals within 30 miles of the closest highly pediatric-resource hospital compared to hospitals > 60 miles from the closest highly pediatric resource hospital (p < 0.001).

### Probably Avoidable and Possibly Avoidable Transfers

After disposition and procedures (including consultations) at the receiving hospital for the 20,304 transfers were considered, 3,764 (18.5%) were categorized as probably avoidable, based on discharge from the receiving ED and claims indicating either no procedures (n = 1,344) or only procedures determined to be always receivable at general EDs (n = 2,420; Figure 1, white bars). An additional 6,091 (30%) were categorized as possibly avoidable (Figure 1, light gray bars) and 10,449 (51.5%) were categorized as unavoidable (Figure 1, dark gray bars) after consideration of disposition and procedures after transfer.

Patient and hospital characteristics by transfer type are shown in Table 1. Among transfers from hospitals with no pediatric-specific capabilities, 23.4% were probably avoidable, compared to only 8.3% of transfers from hospitals with a pediatric ED but no pediatric inpatient care or PICU (p < 0.001). Hospitals with lower ED volume had more probably avoidable transfers compared to hospitals with higher ED volume (p < 0.001). Hospitals in large MSAs also had increased probably avoidable transfers compared to hospitals in small and medium MSAs.

### Factors Associated With Overall Odds of Transfer

Factors associated with overall likelihood of transfer (vs. discharge or admission to the presenting hospital) adjusting for patient and hospital characteristics are shown in Table 2. Children presenting to hospitals with a pediatric ED but no pediatric inpatient unit or PICU had higher odds of transfer (adjusted odds ratio aOR = 1.68, 95% confidence interval [CI] = 1.03–2.72) compared to those presenting to hospitals with no pediatric-specific capabilities. Additionally, children presenting to hospitals in large metropolitan areas had increased odds of transfer (aOR = 3.49, 95% CI = 2.04–5.96) compared to children presenting to hospitals in small metropolitan areas. Children presenting to hospitals > 60 miles from a highly pediatric-resource hospital had lower odds of transfer (aOR = 0.68, 95% CI = 0.47–0.97) compared to ED visits to hospitals within 30 miles of a highly pediatric-resource hospital.

### Factors Associated With Odds of Probably Avoidable Transfer

Factors associated with probably avoidable transfer (vs. possibly avoidable or unavoidable transfer), adjusting for patient and hospital characteristics, are shown in Table 2. Compared to transfers from hospitals with no pediatric-specific capabilities, there was decreased odds of probably avoidable transfer from hospitals with a pediatric ED but no other pediatric-specific capabilities (aOR = 0.38, 95% CI = 0.21–0.71) or a pediatric ED and inpatient care but no PICU (aOR = 0.36, 95% CI = 0.20–0.64). Transfers from hospitals in large metropolitan areas had increased odds of probably avoidable transfer (aOR = 2.64, 95% CI = 1.46–4.80).

Sensitivity analysis including possibly avoidable transfers with probably avoidable transfers as the dependent variable generated similar results (Data Supplement S2, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13519/full), suggesting that our findings are robust to a range of definitions for transfers that may be avoidable.

Using indirect standardization to estimate individual hospital-specific risk-adjusted transfer percentages and hospital-specific risk-adjusted probably avoidable transfer percentages, adjusting for patient and visit characteristics (but not hospital characteristics), we found significant variation in both risk-adjusted transfer rates and risk-adjusted probably avoidable transfer rates across hospitals (Figure 2).

### DISCUSSION

In a large statewide sample with nearly 20% probably avoidable transfers, we found odds of probably avoidable transfers were significantly increased for transfers
<table>
<thead>
<tr>
<th>Table 1</th>
<th>Hospital and Patient Characteristics of Transfers by Transfer Type</th>
<th>Probably Avoidable ( (n = 3,764 \text{ [18.5%]} ) )</th>
<th>Possibly Avoidable ( (n = 6,091 \text{ [30.0%]} ) )</th>
<th>Unavoidable ( (n = 10,449 \text{ [51.5%]} ) )</th>
<th>p-value</th>
</tr>
</thead>
</table>

**Presenting Hospital Characteristics**

Pediatric-specific capability

- No pediatric ED, inpatient, or ICU: 2,218 (23.4), 2,861 (30.2), 4,383 (46.3) <0.001
- Pediatric ED only: 304 (6.3), 1,759 (34.1), 2,268 (51.7)
- Pediatric inpatient only: 826 (18.9), 908 (30.7), 2,644 (60.4)
- Pediatric inpatient + ED: 123 (9.2), 419 (31.2), 802 (59.7)
- Pediatric inpatient + ED + ICU (“highly pediatric resourced”): 233 (20.0), 144 (30.7), 352 (48.3)

ED volume

- Small: 1,539 (22.7), 1,945 (28.7), 3,290 (48.6) <0.001
- Medium: 1,601 (22.0), 1,871 (25.7), 3,807 (52.3)
- Large: 580 (9.4), 2,253 (36.6), 3,329 (54.0)
- Missing: 44 (49.4), 22 (24.7), 23 (25.8)

Size (number of beds)

- >250: 1,601 (15.3), 3,506 (33.5), 5,374 (51.3) <0.001
- 100-250: 1,827 (22.4), 2,132 (26.2), 4,187 (51.4)
- <100: 336 (20.0), 453 (27.0), 888 (53.0)

MSA size (population)

- Small (<100,000): 196 (15.9), 270 (21.9), 765 (62.1) <0.001
- Medium (100,000-999,999): 367 (10.1), 733 (20.1), 2,550 (69.9)
- Large (≥1,000,000): 3,201 (20.8), 5,088 (33.0), 7,134 (46.3)

Critical access hospital (%)

- Not critical access hospital: 3,747 (18.7), 6,028 (30.1), 10,285 (51.3) <0.001
- Critical access hospital: 17 (7.0), 63 (25.8), 164 (67.2)

Distance to closest highly pediatric-resourced hospital (miles)

- 0 to 30: 2,222 (17.9), 4,356 (35.0), 5,861 (47.1) <0.001
- >30-60: 852 (18.2), 1,102 (23.6), 2,726 (58.3)
- >60: 457 (18.6), 489 (19.9), 1,510 (61.5)
- N/A (hospital is highly pediatric-resourced): 233 (32.0), 144 (19.8), 352 (48.3)

**Patient/Visit Characteristics**

Age (years)

- <1: 818 (20.1), 1,120 (27.6), 2,125 (52.3) <0.001
- 1 to <2: 644 (24.9), 825 (32.0), 1,113 (43.1)
- 2 to <5: 914 (22.7), 1,428 (35.4), 1,837 (41.9)
- 5 to <11: 812 (18.7), 1,485 (34.1), 2,055 (47.2)
- 11 to <17: 576 (10.9), 1,233 (23.4), 3,469 (65.7)

Sex

- Male: 2,130 (18.6), 3,607 (31.5), 5,712 (49.9) <0.001
- Female: 1,634 (18.5), 2,484 (28.1), 4,737 (53.5)

Race

- White: 2,130 (21.7), 2,574 (26.2), 5,107 (52.1) <0.001
- Black: 1,250 (16.5), 2,607 (34.3), 3,734 (49.2)
- Other: 384 (13.2), 910 (31.4), 1,608 (55.4)

Medicaid eligibility category

- Medical eligibility: 1,704 (16.9), 2,870 (28.4), 5,532 (54.7) <0.001
- Income eligibility or missing: 2,060 (20.2), 3,221 (31.6), 4,917 (48.2)

(Continued)
from hospitals lacking pediatric-specific EDs and from hospitals in large metropolitan areas. Further, we found wide variation in risk-adjusted rates of probably avoidable transfers across EDs. This variation suggests that these transfers are related in part to hospital-level factors such as available resources and proximity of alternatives. Understanding such factors is particularly relevant given recent findings that hospitalized care for children is becoming increasingly concentrated.20

Specifically, our findings can help guide improvements in organization of the pediatric emergency care system, including efforts to develop regional systems of care designed to optimize care for children by strengthening pediatric-specific capabilities and matching patient needs to hospital resources.10

We found that hospitals with pediatric EDs but no other pediatric-specific capabilities (i.e., pediatric inpatient beds or PICUs) were more likely to transfer pediatric patients. However, transfers from these hospitals were less likely to be probably avoidable compared to facilities with no pediatric-specific capabilities. This finding aligns with previous work that potentially avoidable transfers occur less commonly when patients are transferred from EDs staffed by physicians with pediatric training.21,22 Together, this body of work suggests that expanding local pediatric expertise and resources may reduce both overall transfers and probably avoidable transfers. This is consistent with recognition of the difficulty and uncertainty that emergency physicians phase when risk stratifying children in the ED.23 Strategies to address this problem may include pediatric physician and nursing champions in nonpediatric EDs and active educational outreach by pediatric ED specialists to community providers23 to facilitate ongoing pediatric-specific education and knowledge translation. Telemedicine offers another potential means of extending pediatric expertise to additional EDs. Prior studies report that telemedicine is of interest to parents experiencing transfers,24 may improve evaluation quality,25 and may reduce illness severity among children requiring transfer,26 although numerous barriers still face the uptake and sustainability of pediatric emergency telemedicine programs.27,28

We also found that presenting to an ED in large metropolitan areas was associated with both increased odds of transfers and increased odds of probably avoidable transfer. Proximity to a highly pediatric-resource hospital was also associated with increased

Table 1 (continued)

<table>
<thead>
<tr>
<th>Illness severity score</th>
<th>Probably Avoidable (n = 3,764 [18.5%])</th>
<th>Possibly Avoidable (n = 6,091 [30.0%])</th>
<th>Unavoidable (n = 10,449 [51.5%])</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (least severe)</td>
<td>42 (38.5)</td>
<td>31 (28.4)</td>
<td>36 (33.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>759 (30.5)</td>
<td>769 (30.9)</td>
<td>960 (38.6)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2,307 (19.1)</td>
<td>3,765 (31.2)</td>
<td>5,999 (49.7)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>524 (12.6)</td>
<td>1,219 (29.3)</td>
<td>2,413 (58.1)</td>
<td></td>
</tr>
<tr>
<td>5 (most severe)</td>
<td>48 (6.5)</td>
<td>144 (19.6)</td>
<td>544 (73.9)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>84 (11.3)</td>
<td>163 (21.9)</td>
<td>497 (66.8)</td>
<td></td>
</tr>
</tbody>
</table>

Primary diagnosis category*<0.001

| Respiratory diseases | 624 (14.2) | 1,465 (33.2) | 2,319 (54.6) |         |
| Trauma               | 874 (28.7) | 1,218 (40.0) | 951 (31.3)   |         |
| Gastrointestinal diseases | 576 (21.7) | 925 (34.8) | 1,156 (43.5) |         |
| Neurologic diseases  | 263 (17.3) | 490 (32.3) | 766 (50.4)   |         |
| Psychiatric and behavioral diseases and substance abuse | 27 (1.9) | 35 (2.5) | 1,364 (95.7) |         |
| Other, including missing | 1,400 (19.3) | 1,958 (27.0) | 3,893 (53.7) |         |

Year 0.003

<table>
<thead>
<tr>
<th>Year</th>
<th>n (row %)</th>
<th>n (row %)</th>
<th>n (row %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>791 (19.5)</td>
<td>1,137 (28.1)</td>
<td>2,121 (52.4)</td>
</tr>
<tr>
<td>2010</td>
<td>633 (19.0)</td>
<td>976 (29.2)</td>
<td>1,729 (51.8)</td>
</tr>
<tr>
<td>2011</td>
<td>630 (17.3)</td>
<td>1,177 (31.8)</td>
<td>1,881 (50.9)</td>
</tr>
<tr>
<td>2012</td>
<td>747 (17.7)</td>
<td>1,331 (31.6)</td>
<td>2,135 (50.7)</td>
</tr>
<tr>
<td>2013</td>
<td>954 (19.1)</td>
<td>1,470 (29.4)</td>
<td>2,583 (51.6)</td>
</tr>
</tbody>
</table>

All results are reported as n (row %). Row percentages provided to better allow comparison of transfers across hospital and patient characteristics. Percentages may sum to >100% due to rounding.

ICU = intensive care unit; MSA = metropolitan statistical area.

*Due to space constraints, only the five most common diagnosis categories are specifically listed, with all others grouped as “other.”
Table 2
Adjusted Hospital and Patient Characteristics Associated With ED Transfers and With Probably Avoidable Transfers

<table>
<thead>
<tr>
<th>Presenting Hospital Characteristics</th>
<th>Odds of Any Interfacility Transfer ((n = 2,130,478))</th>
<th>Odds of Probably Avoidable Interfacility Transfer ((n = 19,575))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pediatric-specific capability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pediatric ED, inpatient or ICU</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Pediatric ED only</td>
<td>1.68(^1) (1.03-2.72)</td>
<td>0.38(^1) (0.21-0.71)</td>
</tr>
<tr>
<td>Pediatric inpatient only</td>
<td>0.54(^2) (0.40-0.73)</td>
<td>1.11 (0.77-1.60)</td>
</tr>
<tr>
<td>Pediatric inpatient + ED</td>
<td>0.55 (0.29-1.02)</td>
<td>0.36(^1) (0.20-0.64)</td>
</tr>
<tr>
<td><strong>ED volume</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Medium</td>
<td>1.14 (0.84-1.54)</td>
<td>0.94 (0.64-1.38)</td>
</tr>
<tr>
<td>Large</td>
<td>0.98 (0.60-1.58)</td>
<td>0.57 (0.29-1.12)</td>
</tr>
<tr>
<td>Missing</td>
<td>1.32 (0.74-2.35)</td>
<td>2.44(^2) (1.66-3.60)</td>
</tr>
<tr>
<td><strong>MSA size (population)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small ((&lt;100,000))</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Medium (100,000–999,999)</td>
<td>1.15 (0.69-1.92)</td>
<td>0.84 (0.47-1.48)</td>
</tr>
<tr>
<td>Large ((\geq1,000,000))</td>
<td>3.49(^2) (2.04-5.96)</td>
<td>2.64(^2) (1.46-4.80)</td>
</tr>
<tr>
<td><strong>Distance to closest hospital with pediatric ED, ward, and ICU (miles)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-30</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>&gt;30-60</td>
<td>0.9 (0.69-1.18)</td>
<td>0.86 (0.61-1.21)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>0.68(^1) (0.47-0.97)</td>
<td>1.39 (0.96-2.01)</td>
</tr>
<tr>
<td><strong>Patient/Visit Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>1 to &lt;2</td>
<td>0.75(^1) (0.70-0.81)</td>
<td>1.26(^1) (1.10-1.44)</td>
</tr>
<tr>
<td>2 to 5</td>
<td>0.64(^1) (0.59-0.69)</td>
<td>1.1 (0.97-1.24)</td>
</tr>
<tr>
<td>5 to &lt;11</td>
<td>0.54(^1) (0.49-0.60)</td>
<td>0.87 (0.75-1.02)</td>
</tr>
<tr>
<td>11 to &lt;17</td>
<td>0.47(^1) (0.41-0.54)</td>
<td>0.55(^1) (0.43-0.69)</td>
</tr>
<tr>
<td>Female</td>
<td>0.86(^1) (0.83-0.88)</td>
<td>1.02 (0.93-1.12)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Black</td>
<td>1.02 (0.90-1.17)</td>
<td>0.73(^1) (0.59-0.91)</td>
</tr>
<tr>
<td>Other</td>
<td>0.92 (0.81-1.04)</td>
<td>0.66(^1) (0.54-0.79)</td>
</tr>
<tr>
<td><strong>Medicaid eligibility category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical eligibility</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Income eligibility</td>
<td>0.99 (0.94-1.04)</td>
<td>1.17(^1) (1.06-1.30)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.53 (0.23-1.25)</td>
<td>1.62 (0.36-7.30)</td>
</tr>
<tr>
<td><strong>Illness severity score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (least severe)</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>2</td>
<td>3.43(^2) (2.57-4.58)</td>
<td>0.61 (0.33-1.12)</td>
</tr>
<tr>
<td>3</td>
<td>15.15(^1) (11.73-19.56)</td>
<td>0.37(^1) (0.21-0.65)</td>
</tr>
<tr>
<td>4</td>
<td>57.29(^1) (43.04-76.24)</td>
<td>0.22(^1) (0.13-0.38)</td>
</tr>
<tr>
<td>5 (most severe)</td>
<td>192.15(^1) (140.9-262.1)</td>
<td>0.12(^1) (0.06-0.22)</td>
</tr>
<tr>
<td>Missing</td>
<td>28.89(^2) (21.61-38.62)</td>
<td>0.17(^2) (0.09-0.31)</td>
</tr>
<tr>
<td><strong>Primary diagnosis category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Allergic, immunologic, and rheumatologic diseases</td>
<td>0.61(^1) (0.45-0.82)</td>
<td>1.39 (0.70-2.77)</td>
</tr>
<tr>
<td>Child abuse</td>
<td>0.64(^1) (0.47-0.87)</td>
<td>3.88(^1) (2.07-7.26)</td>
</tr>
<tr>
<td>Circulatory and cardiovascular diseases</td>
<td>1.48(^1) (1.10-2.01)</td>
<td>0.38(^2) (0.19-0.77)</td>
</tr>
</tbody>
</table>

(Continued)
odds of transfer, but had no significant association with odds of probably avoidable transfer. Because larger MSAs have a higher concentration of EDs and hospitals, density of EDs rather than distance specifically to a highly pediatric-resourced hospital may contribute to this finding. Physicians or parents may be more interested in transfer if alternative hospitals are closer and thus more convenient. Alternatively, in a prior survey, physicians practicing with local access to pediatric specialists were less comfortable managing some pediatric conditions compared to those practicing without local access to pediatric specialists, which could also contribute to increased transfers for hospitals in large metropolitan areas and close to highly pediatric-resourced hospitals. Relatedly, some hospitals in large metropolitan areas may have chosen intentionally to reduce their own pediatric facilities, resources, and staffing in light of the presence of alternative EDs for patients, which may also contribute to this finding. Given the increase in odds of both transfers and probably avoidable transfers, our results suggest that strategies targeting EDs in large MSAs may have the greatest impact on reducing probably avoidable transfers.

In response to this finding, one potential strategy is to increase pediatric-specific expertise in these EDs through methods noted above, such as local champions, educational outreach, and telemedicine. An alternative approach, particularly for large metropolitan areas with multiple EDs, is to guide children to hospitals best prepared to receive and triage them. We found that hospitals vary significantly in the proportion of probably avoidable transfers coming from them, even after excluding the most highly pediatric-resourced hospitals. This variability across EDs highlights the potential value of efforts to better identify

### Table 2 (continued)

<table>
<thead>
<tr>
<th>Diseases of the eye</th>
<th>Odds of Any Interfacility Transfer (n = 2,130,478)</th>
<th>Odds of Probably Avoidable Interfacility Transfer (n = 19,575)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENT, dental, and mouth diseases</td>
<td>0.63 (0.50–0.79)</td>
<td>0.86 (0.50–1.48)</td>
</tr>
<tr>
<td>Endocrine, metabolic, and nutritional diseases</td>
<td>6.10 (4.77–7.81)</td>
<td>2.51 (0.15–0.44)</td>
</tr>
<tr>
<td>Fluid and electrolyte disorders</td>
<td>3.17 (2.40–4.19)</td>
<td>0.74 (0.44–1.14)</td>
</tr>
<tr>
<td>Gastrointestinal diseases</td>
<td>1.91 (1.71–2.14)</td>
<td>1.84 (0.94–3.61)</td>
</tr>
<tr>
<td>Genital and reproductive diseases</td>
<td>0.68 (0.50–0.91)</td>
<td>0.81 (0.49–1.35)</td>
</tr>
<tr>
<td>Hematologic diseases</td>
<td>6.71 (4.79–9.40)</td>
<td>0.18 (0.08–0.42)</td>
</tr>
<tr>
<td>Musculoskeletal and connective tissue diseases</td>
<td>1.68 (1.34–2.10)</td>
<td>0.81 (0.49–1.35)</td>
</tr>
<tr>
<td>Neurologic diseases</td>
<td>2.58 (2.25–2.97)</td>
<td>0.68 (0.50–0.92)</td>
</tr>
<tr>
<td>Neoplastic diseases (cancer, not benign neoplasms)</td>
<td>4.49 (2.02–10.0)</td>
<td>0.25 (0.03–1.93)</td>
</tr>
<tr>
<td>Psychiatric and behavioral diseases and substance abuse</td>
<td>2.49 (1.39–4.47)</td>
<td>0.06 (0.03–0.10)</td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td>2.41 (1.95–2.97)</td>
<td>0.45 (0.29–0.70)</td>
</tr>
<tr>
<td>Skin, dermatologic, and soft tissue diseases</td>
<td>1.62 (1.20–2.18)</td>
<td>0.34 (0.24–0.48)</td>
</tr>
<tr>
<td>Systemic states</td>
<td>1.04 (0.92–1.18)</td>
<td>0.69 (0.48–0.98)</td>
</tr>
<tr>
<td>Toxicologic emergencies (including environment)</td>
<td>1.38 (1.17–1.62)</td>
<td>0.47 (0.31–0.70)</td>
</tr>
<tr>
<td>Urinary tract diseases</td>
<td>0.66 (0.56–0.78)</td>
<td>0.66 (0.36–1.21)</td>
</tr>
<tr>
<td>Other, not categorized, missing</td>
<td>1.17 (0.99–1.38)</td>
<td>0.79 (0.55–1.13)</td>
</tr>
</tbody>
</table>

#### Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Odds of Any Interfacility Transfer</th>
<th>Odds of Probably Avoidable Interfacility Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>2010</td>
<td>0.89 (0.84–0.95)</td>
<td>1.08 (0.92–1.28)</td>
</tr>
<tr>
<td>2011</td>
<td>0.97 (0.91–1.03)</td>
<td>0.92 (0.78–1.09)</td>
</tr>
<tr>
<td>2012</td>
<td>1.15 (1.06–1.25)</td>
<td>0.99 (0.83–1.19)</td>
</tr>
<tr>
<td>2013</td>
<td>1.33 (1.12–1.58)</td>
<td>1.08 (0.88–1.32)</td>
</tr>
</tbody>
</table>

Data are reported as adjusted OR (95% CI). Odds of transfer (vs. discharge or admission) and odds of probably avoidable transfer (vs. possibly avoidable or unavoidable transfer), adjusting for listed variables.

ENT = ear-nose-throat; ICU = intensive care unit; MSA = metropolitan statistical area.

*Visits and transfers initially presenting to highly pediatric-resourced hospitals (i.e., hospitals with pediatric inpatient, ED, and ICU) were excluded due to infrequent transfers from these hospitals.

†Significance at the 0.05 threshold.

‡Significance at the 0.01 threshold.
nonpediatric hospitals that have the necessary resources, personnel, and protocols to manage and triage pediatric patients\(^4\) and to incorporate these hospitals into regionalization efforts. For example, some states have formal efforts to categorize ED based on pediatric-specific criteria,\(^{30-32}\) such as “emergency department approved for pediatrics” status. In participating regions of California, emergency medical services (EMS) agencies preferentially transport critically ill children to the closest ED that has emergency department approved for pediatrics status.\(^30\) In New Jersey, emergency department approved for pediatric status is mandated for all EDs.\(^32\) The impact of such pediatric-specific designation strategies on avoidable transfers is not yet known. Ultimately, reducing the proportion of potentially avoidable transfers will require both regional and hospital-level approaches, as potential interventions (i.e., pediatric verification, EMS triage protocols, hospital transfer agreements, patient and provider outreach, telemedicine) require substantial collaborations.

**LIMITATIONS**

First, as a claims analysis, we did not have access to clinical data, so there is the potential for unmeasured confounders that our analysis was not able to identify. For example, we could not identify patients with family request or primary care physician request factoring into transfer decisions. However, in prior work, such requests only accounted for a small number of transfers.\(^11\) Further, use of claims allowed us to assess patterns of care for over a million children across multiple years, and to capture transfers between all hospitals, rather than only those to a specific referral center. Second, our categorization of procedures was based on expert consensus rather than assessment of resources available at specific referring hospitals. Despite this limitation, by more fully recognizing potential resources at referring hospitals, this categorization improves on prior approaches that focused simply on whether any procedures occurred after transfer and is in keeping with prior work demonstrating the value of focusing on ED resource utilization in visit categorization.\(^{15,33,34}\) Our sensitivity analysis indicates that our results regarding the association between hospital characteristics and transfers that may be avoidable were robust to both conservative and liberal algorithms to define these transfers, including consideration only of always receivable procedures versus always receivable and potentially receivable procedures. Third, identification of pediatric subspecialty ED consultations via claims only captured subspecialty consultations billed by attending subspecialists. As a result, we likely underestimated subspecialty consultations among ED transfers, thereby potentially overestimating probably avoidable transfers. We note that only a minority of visits statewide resulted in transfer, but such transfers warrant ongoing attention as they consume an increased proportion of time, resources,
and expense for families, EDs, and insurers. Finally, our analysis exclusively evaluated the Medicaid population. However, Medicaid is the largest single insurer of children in Pennsylvania, and the desire for cost containment in Medicaid should make this analysis particularly relevant to state governments.

CONCLUSION

In conclusion, among pediatric patients with Medicaid in Pennsylvania with a nearly 20% probably avoidable transfer rate, there was significant hospital-level variation in the proportion of probably avoidable transfers. Specifically, larger metropolitan statistical area size increased the odds of a probably avoidable transfer, while hospitals with the presence of a pediatric-specific ED experienced lower odds of probably avoidable transfers. These results indicate that interfacility transfers are driven in part by referring hospital characteristics and support the need for targeted initiatives and interventions to support optimal care and triage of pediatric patients.

References

medicine among emergency physicians. Acad Emerg Med 2017;24:769–75.

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.13519/full

Data Supplement S1. Hospital and patient characteristics of ED visits by disposition.

Data Supplement S2. Sensitivity analysis: adjusted hospital and patient characteristics associated with probably/possibly avoidable transfers.
ABSTRACT

Objective: The objective was to determine if ascending aorta (AscAo) diameters measured by noncontrast computed tomography (CT) allow for meaningful discrimination between patients with and without type A aortic dissection (TAAD), ideally with 100% sensitivity.

Methods: This study was a retrospective analysis of cases of TAAD, as well as controls, undergoing evaluation for TAAD with CT aortography, presenting to 21 emergency departments within an integrated health system between 2007 and 2015. AscAo diameters were determined using axial noncontrast CT images at the level of the right main pulmonary artery by two readers. AscAo diameters were additionally normalized for age, sex, and body surface area (assessed by a Z-score, which is the number of standard deviations between the observed and expected AscAo diameters). Overall model discrimination was assessed using the area under the receiver operating characteristic curve (AUC). Comparative discrimination was assessed using both the change in AUC (ΔAUC) and the continuous net reclassification index (NRI).

Results: A total of 230 cases of TAAD and 325 controls were included in the study. The median ages for cases and controls were 65 and 62 years, and the median AscAo diameters were 50 and 35 mm, respectively. The raw and normalized AscAo diameters demonstrated similarly excellent discrimination (AUCs of 0.96 vs. 0.97, respectively; ΔAUC = 0.01, p = 0.09) and an NRI of 0.30 (95% confidence interval [CI] = 0.13–0.47), both indicating small incremental improvements in classification with the use of the normalized AscAo measures. A raw AscAo diameter of 34 mm and a normalized Z-score of 1.84 both yielded 100% sensitivity for TAAD, with respective specificities of 35% (95% CI = 29.6%–40.2%) and 67% (95% CI = 61.7%–72.2%).

Conclusions: Nearly all patients with TAAD appear to have enlarged AscAo diameters as measured by noncontrast CT, whereas most patients with suspected but absent TAAD have relatively normal AscAo diameters. Both raw and normalized AscAo measures provided relatively comparable discriminatory value. If validated, these data may be useful in adjudicating risk among patients with suspected TAAD in whom a criterion standard test is unavailable, nondiagnostic, or contraindicated.

Proximal thoracic aortic dissections (Stanford classification type A) occur with an incidence of four per 100,000 person-years and an observed average mortality of 25%. While patients with specific structural and genetic diseases are at higher risk for type A aortic dissection (TAAD), prospective registries have
found that the majority of cases occur in patients without these known predispositions, but with similar resultant mortality rates. Unfortunately patients with acute TAAD have highly variable clinical presentations, leaving clinicians to rely on gestalt and the liberal use of advanced imaging modalities, in particular high-dose computed tomography (CT) angiography, magnetic resonance imaging (MRI), and transesophageal echocardiography (TEE), all of which have reported sensitivities for TAAD around 98%. Given the highly morbid consequences of misdiagnosis, very low (i.e., less than 2%) testing thresholds are commonly promoted with correspondingly low diagnostic yields. However, despite liberal screening recommendations, rates of initial misdiagnosis range from 14% to 31%. As such, TAADs represent a high-risk diagnostic clinical challenge.

Fortunately, there are anatomic factors that may reliably identify at-risk patients and direct more appropriate use of advanced imaging. Analysis of one single-center (177 patients) and one multicenter registry (591 patients) of patients with acute TAAD revealed that only 5% to 10% of patients in these studies had maximal ascending aorta (AscAo) diameters less than 4.0 cm. Conversely, nearly all healthy men and women have maximal AscAo diameters less than 4.0 cm (upper 95th percentiles between 3.8–4.2 and 3.4–3.9 cm, respectively). Along these lines, predictive models using age, gender and body surface area (BSA) have been developed to further refine population norms for AscAo diameters. Using normalized measures based on these models might significantly improve the discriminative ability of an AscAo diameter to detect TAAD, as opposed to using a single raw threshold diameter.

We hypothesized that, among emergency department (ED) patients undergoing CT angiography to rule out TAAD, the use of AscAo measurements (as measured on noncontrast CT images) would allow for excellent discrimination between patients with and without TAAD, allowing for determination of a threshold with 100% sensitivity for TAAD as well as clinically useful specificity. We additionally hypothesized that normalization of these AscAo measurements for patient age, sex, and BSA would further improve predictive performance. Identification of such thresholds might allow for accurate TAAD risk stratification using AscAo measurements obtained by transthoracic echocardiography (TTE) or noncontrast CT, the latter being relevant for patients with contraindications to intravenous contrast or nondiagnostic CT angiography.

METHODS

Study Design, Setting, and Population
We undertook a retrospective chart review of ED patients treated within Kaiser Permanente Northern California (KPNC) between 2007 and 2015 to establish a case-control cohort. KPNC is an integrated health care delivery system that provided comprehensive medical care for over 3.9 million health plan members during the study period, with over 1 million ED visits annually among 21 EDs. The study was approved by the KPNC Institutional Review Board with a waiver of the requirement for informed consent and was funded by a Kaiser Permanente Community Benefits Program grant.

Study Protocol
Cases of TAAD were identified by electronically screening patient health records for ED or linked inpatient KPNC encounters with an associated International Statistical Classification of Diseases and Related Health Problems, ninth edition (ICD-9) or 10th edition (ICD-10), diagnostic code for thoracic aortic dissection (441.01 or I71.01), thoracoabdominal dissection (441.03 or I71.03), dissection of aorta, unspecified site (441.00 or I71.00), or dissection of other artery (443.29 or I77.79). Charts were then manually reviewed for case status confirmation both based on CT evidence of thoracic aortic dissection using the radiologist’s finalized interpretation in addition to documented clinical concern for acute thoracic aortic dissection by a consulting cardiothoracic or vascular surgeon, or in cases of death prior to consultation, based on ED physician documentation. Designation of Type A Stanford classification of the aortic dissection was based on radiologist’s finalized interpretation, or if not specified therein, based on the consulting cardiothoracic or vascular surgeon’s documentation.

Control patients were electronically selected from ED patients who underwent CT angiography of the chest during an index ED visit with a reason for referral of “aneurysm/dissection” in a planned 2:1 ratio of controls to cases. Random selection of charts was performed using the RAND function in Microsoft Excel 2016 to generate a pseudo-random value for each patient, which was then used to assign a sorting order for review. Manual review of control charts and images was performed to confirm a documented complaint of chest pain and that a CT protocol for aortic dissection was performed (based on the written radiology report); patients without complaints of chest pain or with alternatively (or inadequately) protocolled CT were excluded from the study. Only the first ED
encounter was included for analysis if multiple qualifying encounters were present for a given patient. In addition, controls were required to have continuously active KPNC health plan membership for the 30 days following the index ED visit, and were excluded from the control cohort if they had a subsequent diagnosis of acute aortic dissection or death in that 30-day period. Additional exclusion criteria for both cases and controls were a previous diagnosis of aortic dissection, pregnancy at the time of the index visit, presence of thoracic trauma, active incarceration, or unknown sex.

**Measurements**

The following variables were electronically abstracted from structured databases drawn from the electronic health record: age, sex, ethnicity, height, weight, past medical history (hypertension, diabetes, aortic stenosis), predisposing syndromes associated with TAAD (Marfan’s, Turner’s, fibromuscular dysplasia) and family history of aortic dissection or aneurysm. CT images were obtained using either 16- or 64-slice technology on a variety of platforms, most being performed without electrocardiogram synchronized image capture (cardiac-gating), as is commonplace under emergent diagnostic circumstances. Digitized CT images (iSite PACS, Koninklijke Philips NV) from the initial diagnostic study were analyzed by two emergency physicians (DGM and JAD) to obtain manual measurements of the AscAo diameter using the precontrast series. An axial slice at the level of the mid-right pulmonary artery was selected to obtain two separate measurements of the AscAo (anterior-posterior and left-anterior oblique) using outside-to-outside wall measurements. The maximum diameter of these two measures was used for analysis. Measurements were rounded to the nearest millimeter. One author (JAD) analyzed all cases and a second (DGM) analyzed all controls. A subset of 70 randomly selected case CTs were reviewed independently by both authors to assess inter-rater reliability, and an additional subset of 20 randomly selected case and control CTs was analyzed twice by each respective reader in a blinded fashion to assess intra-rater reliability. Blinding to case versus control status during imaging review was not feasible, however, given that direct or indirect evidence of TAAD was frequently present on the precontrast series.

**Data Analysis**

Normalized values for AscAo diameter were calculated using a formula that accounts for individual patient age, sex, and BSA. In brief, following logarithmic transformation of age and AscAo diameters (for homogeneity of variance), the difference between expected and observed AscAo diameters was standardized to obtain a Z-score (i.e., the number of standard deviations between the observed and expected AscAo diameters). Both the raw and the normalized AscAo diameter values (the latter in terms of the Z-score) were then modeled independently against the outcome (case status) to determine relative discrimination as well as the sensitivity and specificity of various cut points.

To assess the comparative discrimination of raw versus normalized AscAo diameters, we used receiver operating characteristic curves, reporting the Youden index, area under the receiver operating characteristic curve (AUC), and change in AUC (ΔAUC), including subgroup analyses among older (age > 65 years) and larger (BSA of 2.0 m² or greater) patients. We also calculated the continuous net reclassification index (NRI) as an additional measure of incremental discrimination at the individual patient level. Marker calibration was assessed graphically using locally weighted scatterplot smoothed calibration curves as well as with the Hosmer-Lemeshow test. Overall marker performance was assessed using the scaled Brier score, calculated as [1 – observed Brier score/maximal possible Brier score], and expressed as a percentage which is comparable in scale and interpretation to Pearson’s R² statistic.

To identify clinically useful cut points, we assessed for the maximal AscAo diameter with 100% sensitivity, as well as the cut point corresponding to the optimal overall AUC for each model (i.e., greatest summation of sensitivity and specificity, referable to the Youden index). We determined a priori that, using an optimal sensitivity point estimate of 100%, we would need 188 cases of TAAD to achieve a lower 95% confidence interval (CI) limit of 98% or greater. Inter- and intra-rater agreement were assessed using Bland Altman plots, reported as absolute and proportional (to the mean) limits of agreement, as well as using intraclass correlation coefficients (two-way mixed effects, consistency, and multiple raters model). CIs were calculated using Clopper-Pearson exact intervals, with the exception of those for both the AUC and the NRI, which were bootstrapped from 1000 random replications maintaining the original proportionality between cases and controls. P-values were calculated using chi-square and Wilcoxon signed-rank tests as indicated. All analyses were performed using Stata v.13 (StataCorp LLC, College Station, TX, USA).
RESULTS

We identified 230 cases of TAAD of 595 manually reviewed charts. An additional 325 controls were randomly selected out of a total of 14,306 potentially eligible patients, truncated at less than the planned 1:2 ratio of cases to controls due to time and budget limitations (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.13547/full], Figures e1 and e2). The median age for cases and controls was 65 and 62 years, respectively, with fewer patients carrying a diagnosis of diabetes mellitus in the case group (5% vs. 26%). The median raw AscAo diameters were 50 and 35 mm, respectively. Demographic and summary findings for the case and control cohorts are presented in Table 1.

Table 2 outlines the respective performance measures between the raw and normalized AscAo values. Receiver operating curve analysis (Figure 1) revealed similar AUCs for the raw and normalized measures (0.96 vs. 0.97, ΔAUC = 0.01, p = 0.09), and results were unchanged when restricting to patients aged 65 years and older (n = 265, 0.96 vs. 0.97, ΔAUC = 0.01, p = 0.052) or with a BSA of 2.0 m² or greater (n = 263, 0.97 vs. 0.97, ΔAUC = 0, p = 0.44). Likewise, the Youden indices were comparable at 0.83 and 0.82. In terms of discrimination at the individual patient level, the continuous NRI was 0.30 (95% CI = 0.03–0.47) when comparing the normalized to the raw AscAo diameters. Taken together with the ΔAUC results, this indicates a small but relatively insignificant improvement in risk classification at the individual patient level with the use of the normalized measures. Calibration for both the raw and the normalized AscAo measures was acceptable as assessed by calibration plots (Data Supplement S1, Figures e3 and e4) and the Hosmer-Lemeshow test. The scaled Brier scores were similar at 70 and 72%, indicating strong overall performance.

Table 3 summarizes the sensitivity, specificity, and likelihood ratios for cut points of interest. The maximal cut points with 100% sensitivity were a raw AscAo diameter of 34 mm, and a Z-score of 1.84, with respective specificities of 35% (95% CI = 29.6%–40.2%) and 67% (95% CI = 61.7%–72.2%). The optimal cut points (maximal sum of sensitivity and specificity) were a raw diameter of 42 mm and a Z-score of 2.95, with respective sensitivities of 93.5 and 91.0% and specificities of 93 and 91.0%.

Bland-Altman analysis indicated strong intra-rater reliability for cases and controls, with a bias of –0.3 and 0.1 mm, absolute limits of agreement of ±2.2 and ±1.7 mm, and relative limits of agreement of ±8.7 and ±4.8%, respectively (Data Supplement S1, Figures e5 and e6). Inter-rater reliability as assessed by Bland-Altman analysis was slightly less robust with a bias of –0.6 mm, absolute limits of agreement of ±4.3 mm, and relative limits of agreement of ±8.7%, but revealed a strong intra-class correlation coefficient of 0.95 (95% CI = 0.93–0.97; Data Supplement S1, Figure e7). These findings are similar to inter-observer variations reported between cardiothoracic radiologists measuring AscAo diameters using cardiac-gated CT images.21

Table 1
Case and Control Cohort Demographics and Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Case (n = 230)</th>
<th>Control (n = 325)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>65 (53–77)</td>
<td>62 (51–72)</td>
<td>0.05</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td>48</td>
<td>0.004</td>
</tr>
<tr>
<td>Diabetes</td>
<td>5</td>
<td>26</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hypertension</td>
<td>73</td>
<td>64</td>
<td>0.02</td>
</tr>
<tr>
<td>Aortic stenosis</td>
<td>3</td>
<td>3</td>
<td>0.85</td>
</tr>
<tr>
<td>Prior aortic valve surgery</td>
<td>2</td>
<td>2</td>
<td>0.85</td>
</tr>
<tr>
<td>Predisposing syndrome</td>
<td>3</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Family history of aortic pathology</td>
<td>6</td>
<td>1</td>
<td>0.002</td>
</tr>
<tr>
<td>Body mass index</td>
<td>27.1 (24–32)</td>
<td>29.3 (25.7–32.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AscAo diameter (mm)</td>
<td>50 (45–54)</td>
<td>35 (32–39)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Normalized AscAo diameter, Z-score</td>
<td>4.57 (3.61–5.41)</td>
<td>1.32 (0.66–2.12)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Data are reported as median (IQR) or %.
AscAo = ascending aorta; IQR = interquartile range
*Values calculated using chi-square and Wilcoxon signed-rank tests.
DISCUSSION

The objective of this study was to examine AscAo diameter thresholds, below which the probability of acute TAAD is negligible and/or exceedingly small. From an absolute diameter perspective, 100 and 93.5% of patients in this series had AscAo diameters of at least 34 and 42 mm, respectively. This is notable in comparison to a study of 4,039 asymptomatic adult patients undergoing noncontrast cardiac CT, in which the mean AscAo diameter was 33 mm and the upper limit of normal was 41 mm. We found that normalizing these raw AscAo measurements for sex, age, and BSA appeared to offer improved specificity since all cases registered a Z-score > 1.84, meaning that 96.7% of a disease-free reference population would be expected to have smaller normalized AscAo measurements. However, systems for generating normalized AscAo values (ideally via calculators integrated into the electronic health record) would be necessary to make their use clinical practical. It is also notable that most cases of TAAD in our study had AscAo diameters below the traditional cutoff of 55 mm for preventative aortic surgery (the median AscAo diameter among cases of TAAD was 50 mm), a finding consistent with prior registry reports.

How might these findings be practically applied for further study and clinical use? Foremost, it should be stated that identification of moderate levels of aortic dilation (40 to 55 mm), specifically among asymptomatic patients, does not indicate an imminent risk of TAAD. While it appears from these and other observational data that some dilation of the aorta is a nearly universal necessary risk factor for TAAD, the absolute incidence of TAAD is quite low at AscAo diameters less than 55 mm, with 5-year risk estimates of 0.1 and 0.4% for patients with diameters of 40 and 45 mm, respectively, underscoring the relative rarity of TAAD.

Figure 1. Receiver operating characteristic curves. AscAo = ascending aorta; AUROC = area under receiver operating characteristic.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Sensitivity, Specificity, and Likelihood Ratios for Cut Points of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw AscAo Diameter (mm)</td>
<td>Normalized AscAo Diameter (Z-score)</td>
</tr>
<tr>
<td>≥34</td>
<td>≥42</td>
</tr>
<tr>
<td>Sensitivity (%), 95% CI</td>
<td>100 (98.4–100)</td>
</tr>
<tr>
<td>Specificity (%), 95% CI</td>
<td>34.8 (29.6–40.2)</td>
</tr>
<tr>
<td>LR+ (95% CI)</td>
<td>1.5 (1.4–1.7)</td>
</tr>
<tr>
<td>LR− (95% CI)</td>
<td>0 (NR)</td>
</tr>
</tbody>
</table>

AscAo = ascending aorta; LR+ = positive likelihood ratio; LR− = negative likelihood ratio.

Table 2 | Performance Characteristic Comparison Between Raw and Normalized AscAo Measurements |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw AscAo Diameter</td>
<td>Normalized AscAo Diameter</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.96 (0.95–0.98)</td>
</tr>
<tr>
<td>Youden index</td>
<td>0.83</td>
</tr>
<tr>
<td>Brier score (max 0.242)</td>
<td>0.072</td>
</tr>
<tr>
<td>Scaled Brier score</td>
<td>70%</td>
</tr>
<tr>
<td>Hosmer-Lemeshow test</td>
<td>p = 0.28</td>
</tr>
<tr>
<td>Continuous event NRI</td>
<td>0.07</td>
</tr>
<tr>
<td>Continuous nonevent NRI</td>
<td>0.23</td>
</tr>
<tr>
<td>Continuous NRI (95% CI*)</td>
<td>0.30 (0.13–0.47)</td>
</tr>
</tbody>
</table>

AscAo = ascending aorta; AUC = area under receiver operating characteristic curve; IQR = interquartile range; NRI = net reclassification index.

*Bootstrapped estimates using 1,000 replications drawn separately from cases and controls based on an estimated disease prevalence of 0.4 per 10,000.
†p = 0.09.
Given very low estimated testing thresholds for TAAD, combined with the inherent imprecision of point estimates derived herein, this information should also not be used indiscriminately to determine which patients should undergo criterion standard testing (i.e., CT aortography, MRI, or TEE) for TAAD. However, these data could be applied (with further validation) to help modify TAAD risk estimates when those criterion standard imaging tests are either unavailable, nondiagnostic (e.g., due to inadequate contrast opacification), or contraindicated (renal failure, esophageal strictures, anaphylactic allergy, etc.). In these circumstances, two readily available alternatives (noncontrast CT and TTE) can provide accurate measures of AscAo diameter. While noncontrast CT may be the preferred option given ease of access, quick speed and minimal technical barriers, both TTE and TEE provide similarly accurate measurements of AscAo diameters compared to CT. TTE also has better inter-rater reliability compared to CT, even among centers of aorta excellence. Additionally, TTE-derived measures of AscAo diameters have been investigated as adjuncts to the aortic dissection detection risk score. However, since the current study findings are solely derived from CT measurements, they should not be directly extrapolated to TTE-derived measurements without further study.

From a feasibility standpoint, it is important to note that studies of directed transthoracic cardiac ultrasound examining AscAo diameters at the point of care have demonstrated notable differences in performance between echocardiographers and clinician sonographers. Two studies examining the ability of emergency physicians to detect AscAo dilation yielded sensitivities of 77 and 79% compared to CT, whereas two studies of formal echocardiography reported sensitivities of 87 and 92%, one of which found specific signs of TAAD in up to 97% of cases. As point-of-care ultrasound technology improves with time, it is likely that imaging quality, clinician experience, and resultant test sensitivity will improve concordantly, raising the likelihood that the findings from this study could be clinically validated in the context of TTE-directed diagnostic protocols. In the meanwhile, it is important to remember that the practical use of point-of-care TTE by clinician sonographers to detect AscAo dilatation is an endeavor that has fallen short of being clinically reliable to date.

LIMITATIONS

Given the retrospective nature of the study, cases were screened for inclusion using ICD codes. While we did additionally confirm the presence of acute TAAD both with confirmatory CT imaging and with a documented clinical concern for an acute process, we cannot exclude the possibility that some cases of acute TAAD were not included in the cohort. However, it seems unlikely that any missed cases due to incomplete diagnostic coding would be biased in terms of AscAo diameter. Also, while we did not reach our intended case-to-control ratio goal of 1:2 (0.5), in retrospect this was an excessive goal given our intention to maximize cut-point sensitivity, as well as the high AUC of the AscAo diameters, circumstances under which the optimal case to control ratio (for a given sample size) is actually above 1. A key limitation of our data is the variability in AscAo size that occurs throughout the cardiac cycle, given that cardiac-gated CT image capture was rarely employed in this study, as is typical under emergent diagnostic circumstances. Studies have reported average changes in AscAo diameters of 2 mm or 3% between end-diastole and end-systole, and guidelines thus recommend the use of CT detector-arrays with > 64 rows in conjunction with cardiac gating for imaging capture. Using echocardiography or CT, the minimal AscAo diameter is typically obtained at end-diastole and the maximal AscAo diameter is typically observed during systole, although professional guidelines recommend measurement at end-diastole, largely by precedent. Since a lack of cardiac gating would be expected to overestimate the size of the aorta compared to a measurements obtained at end-diastole, the application of the cutoffs suggested in this study should arguably be limited to measurements obtained at end-diastole, thereby reducing the likelihood of false positives in the diagnosis of TAAD. This nuance also highlights the importance of developing standard interchangeable methodologies for obtaining AscAo diameter measurements across imaging modalities, which are as of yet unestablished.

Ideally, the measurements of AscAo diameters in this study would have been determined by experts in...
cardiothoracic radiology. However, given that our measures of intra- and inter-rater reliability are similar to those seen among expert raters in terms of percent variation, the added utility of such a design is debatable. More impactful perhaps was the general unavailability of double oblique reconstructions, which can provide true anatomic perpendicular measures of the AscAo and are recommended for use in obtaining aortic measures, though not consistently used in clinical practice. For this reason, our measures were aortic measures, though not consistently used in clinical practice. For this reason, our measures were taken at the level of the right main pulmonary artery, a level at which measurements obtained by standard axial cuts do not vary in diameter by more than 1 to 2 mm or 2% on average between anatomically aligned double oblique reconstructions, and thus are also within established intra- and inter-rater ranges of variation.

Finally, while the use of the normalized measurements appeared to offer superior specificity at the 100% sensitivity cut point (67% for the normalized vs. 35% for the raw measure), this finding may be spurious, given the relative equality in AUCs between the two measures, as well as similar observed sensitivities and specificities at the optimal cut points. Also, we used a specific formula derived from echocardiographic measures in a reference population, and thus the Z-score data in this study (at least in terms of absolute values) is not generalizable to other published formulas predicting AscAo size. Finally, the BSA values used to calculate the normalized measurements were derived from the electronic health record and may have been inaccurate. Accordingly, the raw cutoffs presented in this study are arguably best suited for further investigation and validation, although further research on normalized values should ideally be conducted in tandem.

CONCLUSION

Patients with type A aortic dissection appear to have uniformly enlarged ascending aorta diameters as measured by noncontrast computed tomography, whereas most patients with suspected but absent type A aortic dissection have relatively normal ascending aorta diameters. These data may be useful in adjudicating risk among patients with suspected type A aortic dissection for whom a criterion standard test is unavailable, non-diagnostic, or contraindicated. Further validation is needed prior to clinical adoption, including studies using data from alternative imaging modalities (e.g., transthoracic echocardiography).

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.13547/full

Data Supplement S1. Supplemental material.
ABSTRACT

Background: Dementia is underrecognized in older adult emergency department (ED) patients, which threatens operational efficiency, diagnostic accuracy, and patient satisfaction. The Society for Academic Emergency Medicine geriatric ED guidelines advocate dementia screening using validated instruments.

Objectives: The objective was to perform a systematic review and meta-analysis of the diagnostic accuracy of sufficiently brief screening instruments for dementia in geriatric ED patients. A secondary objective was to define an evidence-based pretest probability of dementia based on published research and then estimate disease thresholds at which dementia screening is most appropriate. This systematic review was registered with PROSPERO (CRD42017074855).

Methods: PubMed, EMBASE, CINAHL, CENTRAL, DARE, and SCOPUS were searched. Studies in which ED patients ages 65 years or older for dementia were included if sufficient details to reconstruct 2 x 2 tables were reported. QUADAS-2 was used to assess study quality with meta-analysis reported if more than one study evaluated the same instrument against the same reference standard. Outcomes were sensitivity, specificity, and positive and negative likelihood ratios (LR+ and LR–). To identify test and treatment thresholds, we employed the Pauker-Kassirer method.

Results: A total of 1,616 publications were identified, of which 16 underwent full text-review; nine studies were included with a weighted average dementia prevalence of 31% (range, 12%–43%). Eight studies used the Mini Mental Status Examination (MMSE) as the reference standard and the other study used the MMSE in conjunction with a geriatrician’s neurocognitive evaluation. Blinding to the index test and/or reference standard was inadequate in four studies. Eight instruments were evaluated in 2,423 patients across four countries in Europe and North America. The Abbreviated Mental Test (AMT-4) most accurately ruled in dementia (LR+ = 7.69 [95% confidence interval {CI} = 3.45–17.10]) while the Brief Alzheimer’s Screen most accurately ruled out dementia (LR– = 0.10 [95% CI = 0.02–0.28]). Using estimates of diagnostic accuracy for AMT-4 from this meta-analysis as one trigger for more comprehensive geriatric vulnerability assessments, ED dementia screening benefits patients when the prescreening probability of dementia is between 14 and 36%.
Cognitive dysfunction in older adults includes mild cognitive impairment, dementia, and delirium in addition to traumatic brain injury, intoxication, and central nervous system infections also encountered in younger populations. Delirium is an acute and reversible disturbance in attention with multiple potential precipitants, while mild cognitive impairment is an early form of Alzheimer’s disease with memory and language problems only manifest with formal testing.3 On the other hand, dementia represents a chronic neurodegenerative disease that impairs executive functioning, memory, orientation, and judgment.1 Total United States expenditures for dementia care in 2010 were estimated at $157 to $215 billion,3 while globally costs were estimated at $604 billion in 2010 and projected to increase to $1 trillion by 2018.4 Dementia has the potential to negatively influence effective emergency care. For example, unrecognized dementia is associated with diagnostic inaccuracy as clinicians evaluate a patient’s acute complaints and the cause of their symptoms.5 Dementia is also associated with increased use of the ED,6 prolonged ED length of stay, increased admission rates,7 prolonged hospitalizations, incident delirium,8 fall risk,9 and higher mortality,7 as well as subsequent ED returns10 and hospital readmissions.11,12 Therefore, older adults with dementia represent a vulnerable ED population that will present with increasing frequency over coming decades. In response, nurse and physician leaders worldwide are increasingly advocating a more dementia friendly ED, which includes multiple strategies such as limiting psychotropic or anxiolytic medications that can worsen confusion or agitation, adapting pain assessment measures to accommodate dementia patients, and ensuring follow-up mechanisms after ED discharge.13 Whether these strategies or other approaches improve dementia patient outcomes is largely unknown and quite unexplorable until dementia can be accurately identified in ED settings.

Primary care providers often fail to diagnose dementia, so reliance on the past medical history alone during an episode of emergency care can miss over 90% of cases.14–16 While the traditional emergency medicine process appropriately prioritizes the identification of acutely life-threatening illness or injury, dementia is a chronic neurodegenerative process without a cure, and its key symptom (cognitive impairment) is often underrecognized in the ED.15 Most ED staff worldwide do not screen for cognitive dysfunction, even in the geriatric-specific ED.17–19 Nonetheless, ascertaining older adults’ baseline cognitive status is a core competency for emergency medicine residency graduates,20 and screening for dementia and delirium are prominent recommendations in the American College of Emergency Physicians-Society for Academic Emergency Medicine geriatric emergency department guidelines.21 Efficient dementia screening in ED settings relies on ultrabrief instruments that are simultaneously reliable and accurate, psychometrically valid, acceptable to ailing patients, acknowledged as value-added by health care providers, and available during episodes of care without requiring extra resources or equipment to administer. Although several dementia screening instruments from myriad health settings have been described, no prior systematic review has quantitatively evaluated the diagnostic accuracy of these instruments in the ED.22 The primary objective of this meta-analysis was to identify and summarize the pooled diagnostic test characteristics (sensitivity, specificity, and likelihood ratios [LRs]) for dementia screening instruments in the ED. A secondary objective was to assess the pretest probability of dementia along with test and treatment thresholds using the Pauker-Kassirer method, based on estimates of sensitivity, specificity, diagnostic risks, and treatment benefits derived from this meta-analysis.23

**METHODS**

**Study Design**

We conducted a systematic review and meta-analysis of original research studies that reported data on the diagnostic accuracy of dementia in older ED patients. The design and reporting of this systematic review conform to the Preferred Reporting Items for Systematic
Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies. Studies were included if they described adults aged 65 years or older, evaluated in the ED setting with an index test for dementia and compared with an acceptable reference standard for dementia. A priori determinants of acceptable reference standards included the Mini Mental Status Examination (MMSE) or more formal neuropsychological evaluation by qualified individuals (psychiatrist, neurologist, geriatrician) using Diagnostic Statistical Manual of Mental Disorders (DSM-V) criteria. For inclusion, studies had to provide sufficient detail on the dementia screening test and reference standard to construct two-by-two tables. We contacted the authors of potentially appropriate studies if they did not report sufficient detail to reconstruct two-by-two tables. When multiple studies reported diagnostic accuracy on the same or overlapping patient populations (same site, same time period), the publication with the largest sample size was selected for inclusion. We elected to define “disease positive” as an abnormal reference standard using the threshold defined in the original studies, whereas “disease-negative” patients were those with normal results on the reference standard. This systematic review was registered with PROSPERO (CRD42017074855).

Search Strategy
The published literature was searched using strategies created by a medical librarian (SF) for the concepts of emergency department, people 60 and older, screening, dementia and diagnosis. These strategies were established using a combination of standardized terms and key words and were implemented in PubMed, Embase.com, EBSCO Cumulative Index for Nursing and Allied Health (CINAHL), Wiley Cochrane Central Register of Controlled Trials (CENTRAL), Wiley Database of Abstracts of Reviews of Effects (DARE), Wiley Cochrane Database of Systematic Reviews, and clinicaltrials.gov. All searches were completed in March 2014 and were limited to English using database supplied limits. The search was updated in June 2018. Due to a change in database access, Scopus was used in place of Embase. All previous databases were searched again. All results were exported to EndNote. We used the automatic duplicate finder in EndNote and duplicates were assumed to be accurately identified and removed. Full search strategies are provided in Data Supplement S1 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13573/full). Two authors (DK, LS) also conducted bibliographic searches of research abstracts presented at scientific meetings published in Academic Emergency Medicine, Annals of Emergency Medicine, Canadian Journal of Emergency Medicine, and Journal of the American Geriatrics Society. Data from each study were abstracted by one author (CRC).

Individual Evidence Quality Appraisal
Two authors (JB, MAL) independently used the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) for systematic reviews to evaluate the risk of bias for the identified studies. Several a priori conditions were used to evaluate individual study’s risk of bias and degree of applicability.

- If the study enrollment occurred anywhere other than an ED (for example, if enrollment included patients in a dementia clinic or on a hospital ward in addition to the ED) then the results were assessed as “low applicability.”
- If the same assessor collected elements of both the index test and the reference standard, blinding was considered inadequate.
- Establishing the presence or absence of dementia required the evaluation of an expert in cognitive assessment (generally a geriatrician, neuropsychologist, or neurologist) using DSM-V criteria (or the equivalent of DSM-V for earlier studies). If the MMSE was the sole reference standard, the study was deemed high risk of bias.

QUADAS-2 inter-rater agreement was quantified using a kappa analysis with qualitative descriptors previously described by Byrt. Discrepancies were resolved after review by a third author (CRC).

Data Analysis
One author (CRC) computed meta-analysis estimates when one or more studies evaluated the same
dementia screening test against the same reference standard. No consensus exists about whether a fixed-effects or random-effects model is more appropriate for diagnostic meta-analyses, although some evidence indicates significant between-study heterogeneity in diagnostic studies which implies that a random-effects model is more appropriate. Therefore, we generated combined estimates for diagnostic accuracy using a random-effects model (MetaDiSc Version 1.4, Hospital Universitario Ramón y Cajal). The DerSimonian-Laird random effects model was used to quantify statistical interstudy heterogeneity via the index of inconsistency ($I^2$), Cochrane’s Q, and tau-square. Tau represents the estimated standard deviation of underlying effects across studies, while $I^2$ estimates the proportion of total variability in point estimates attributable to heterogeneity. We also report pooled estimates of dichotomous positive LRs (LR+) and negative LRs (LR−) from the random-effects model. Because of the small number of studies and uncertain interpretation for diagnostic meta-analyses, publication bias was not evaluated.

**Test–Treatment Threshold**

Universal ED dementia screening is likely not feasible, nor does the United States Preventive Services Task Force support such widespread screening. However, a subset of ED patients with previously unrecognized dementia might benefit from screening using sufficiently accurate instruments. Identifying which patients might benefit and what interventions might benefit them requires methods distinct from diagnostic meta-analysis. We used the Pauker-Kassirer method to estimate thresholds for further dementia testing or referral for dementia treatment. This technique is based on seven variables: false-negative and false-positive rates of the diagnostic test assessed as well as sensitivity and specificity, risk of the diagnostic test, potential risk of treatment in false-positive patients, and benefits of treatment in true-positive cases. Few of these risks or benefits have been formally evaluated in ED settings for dementia screening, so our estimates are arbitrary. Accordingly, we provide an interactive calculator to empower readers to recompose thresholds with different estimates of screening test accuracy or anticipated risks and benefits that may be more acceptable to individual clinician’s patient populations and clinical setting (Data Supplement S2, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13573/full).

**RESULTS**

A total of 1,616 unique citations were identified through a search of PubMed, EMBASE, CINAHL, CENTRAL, and DARE and 1,588 were excluded after review of their title and abstract revealed inapplicability to our inclusion criteria. A total of 16 studies underwent full-text review and nine were included in the current analysis (Figure 1). The seven excluded studies enrolled duplicate patients, provided insufficient details to reconstruct two-by-two tables or did not assess a brief screening instrument. Included studies are detailed in Table 1 and the eight dementia screening instruments are summarized in Data Supplement S3 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13573/full). Studies were conducted between 2003 and 2016, enrolling a total of 2,423 patients. Four studies occurred in the United States, two in Canada, two in Ireland, and one in Scotland. Enrollment criteria ages ranged from over age 65 to over 75 and all excluded critically ill patients. Non–English-speaking patients were excluded with the exception of Wilding et al., which also included those speaking French. One study was a randomized trial comparing two different screening instruments, while the other eight studies were prospective cross-sectional, convenience sampling investigations.

QUADAS-2 assessment for risk of bias and applicability demonstrated poor inter-rater agreement for the assessment of predefined thresholds for the index test, real-world administration of the index test, acceptability of the reference standard, sufficient interval between the index test and the reference standard, and uniformity of reference standard testing (Table 2). However, the poor inter-rater reliability observed for the index test applicability, reference standard acceptability, between-test interval, and reference test uniformity reflect the paradox of high agreement and low kappa. Each of those QUADAS-2 questions had agreement in eight of nine studies (agreement = 89%) between reviewers but had two zero cells in the kappa two-by-two table. This paradox is a recognized limitation of kappa as a quantitative measure of inter-rater agreement. This paradox does not explain the low kappa observed for predefined thresholds for the index test since the agreement was only 66% for this assessment, so the low kappa likely reflects our failure to define acceptable index test thresholds before the QUADAS-2 reviews ensued.
Blinding to the reference standard for the individual interpreting the index test was uncertain or inadequate in four studies, while blinding to the index test by the reference standard evaluator was inadequate in six studies. The reference standard was the MMSE for all but two studies, but the threshold to differentiate “dementia” from “no dementia” was ≤ 23 in five studies\textsuperscript{16,34--37} and ≤ 24 in two studies.\textsuperscript{39,41} The other two studies used MMSE in combination with other instruments and assessments as the reference standard for dementia as detailed in Table 1.\textsuperscript{38,40} O’Sullivan et al.\textsuperscript{40} was the only study to use DSM-V criteria and a formal neurocognitive assessment by a geriatrician. The weighted mean prevalence of dementia was 30.7% and ranged from 12% in Barbic et al.\textsuperscript{41} to 43% in Schofield et al.\textsuperscript{36} Wilding et al.\textsuperscript{39} used trained geriatric nurses to administer the screening test, while all other studies employed research assistants for that task. Only four studies reported adherence to any iteration of the Standard for Reporting Diagnostic Accuracy Studies (STARD) reporting criteria.\textsuperscript{50,51}

**Screening Instruments**

Five instruments were assessed in multiple studies permitting meta-analysis of diagnostic accuracy: Abbreviated Mental Test-4 (AMT-4), caregiver Alzheimer’s Disease-8 (cAD8), Ottawa 3DY (O3DY), Short Blessed Test (SBT), and the Six Item Screener (Figure 2). Significant statistical heterogeneity was noted for each measure of diagnostic accuracy for every instrument with the exception of the Six Item Screener LR– and cAD8 LR+ and LR−. The AMT-4 demonstrated the highest pooled LR+ (7.69 [95% confidence interval {CI} = 3.47–17.10]), while the O3DY pooled LR− 0.17 (95% CI 0.05–0.66) and the SBT pooled LR− 0.18 (95% CI = 0.09–0.39) most accurately reduce the probability of dementia. Three instruments were only evaluated in single studies: Animal Fluency,\textsuperscript{39} Brief Alzheimer’s Screen,\textsuperscript{37} and the Mini Cog\textsuperscript{34} (Table 3). The Brief Alzheimer’s Screen was more accurate than the O3DY and Short Blessed Test to reduce the probability of dementia but requires more time to administer and complex algebraic computations to interpret.
<table>
<thead>
<tr>
<th>Study</th>
<th>Location</th>
<th>No. Patients</th>
<th>Mean or Median Age (Years)</th>
<th>Exclusion Criteria</th>
<th>Study Design</th>
<th>Dementia Screeners Assessed</th>
<th>Criterion Standard</th>
<th>Prevalence of Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbic 2018</td>
<td>St. Paul’s Hospital, Vancouver British Columbia, over 5-month period in 2016</td>
<td>117</td>
<td>82</td>
<td>Age &lt; 75, Canadian Triage Score Level 1, sensory deficits prohibiting communication, acute confusion or hallucinations, non-English speaking, previous diagnosis of dementia, nursing home resident, inability to provide or lack of consent</td>
<td>Prospective, cross-sectional, convenience sampling</td>
<td>SBT, O3DY</td>
<td>MMSE ≤ 24</td>
<td>MMSE ≤ 24 in 12% O3DY agreed with MMSE in 58%, while SBT did so in 61.5% of cases Kappa for O3DY 0.64 and for SBT 0.63</td>
</tr>
<tr>
<td>Carpenter 2011a</td>
<td>Barnes-Jewish Hospital, St. Louis, MO, Jun 2009-Mar 2010</td>
<td>163</td>
<td>78</td>
<td>Age &lt; 65; receipt of antiemetic, benzodiazepine, or narcotic prior to cognitive assessment; non-English speaking; critical illness as judged by emergency physician; inability to provide consent in absence of caregiver to consent</td>
<td>Prospective, cross-sectional, convenience sampling</td>
<td>BAS, SBT, O3DY, cAD8</td>
<td>MMSE ≤ 23</td>
<td>MMSE ≤ 23 in 37% Abnormal cognitive screening noted by O3DY in 66%, BAS in 65%, cAD8 in 55%, and SBT in 43%</td>
</tr>
<tr>
<td>Carpenter 2011b</td>
<td>Barnes-Jewish Hospital Emergency Department, St. Louis, MO, Jul 2008-Apr 2009</td>
<td>319</td>
<td>76</td>
<td>Age &lt; 65; receipt of antiemetic, benzodiazepine, or narcotic prior to cognitive assessment; non-English speaking; critical illness as judged by emergency physician; inability to provide consent in absence of caregiver to consent</td>
<td>Prospective, cross-sectional, convenience sampling</td>
<td>SIS, cAD8</td>
<td>MMSE ≤ 23</td>
<td>MMSE ≤ 23 in 35% Review of past medical history noted “dementia” in only 6% of patients</td>
</tr>
<tr>
<td>Dyer 2016</td>
<td>Tallaght Hospital Emergency Department, Dublin, Ireland, Jun-Aug 2014</td>
<td>196</td>
<td>78</td>
<td>Age &lt; 65, too unwell, unable or unwilling to consent</td>
<td>Prospective naturalistic study, convenience sampling</td>
<td>AMT-4</td>
<td>Either a positive delirium screen using CAM-ICU or cAD8 ≥ 2 or MMSE ≤ 26 with negative AD8</td>
<td>50% had abnormal result on one or more of the criterion standards (13% CAM-ICU positive delirium, 23.5% abnormal cAD8, 14% MMSE ≤ 26) 74% with abnormal results had no prior formal diagnosis of dementia</td>
</tr>
<tr>
<td>O’Sullivan 2017</td>
<td>Mercy University Hospital Emergency Department, Cork, Ireland, Jun-Nov 2015</td>
<td>368</td>
<td>77</td>
<td>Age &lt; 70, intoxicated, poor English skills, medical instability, severe intellectual disability, refusal or inability to consent without family to assent</td>
<td>Prospective, nonconsecutive sampling</td>
<td>SBT</td>
<td>Geriatrician assessment using DSM-V criteria including researcher collected MMSE, DRS-R98, and IQCODE</td>
<td>19.6% of the 368 with SBT screening categorized as dementia by geriatrician assessment 21.7% screened positive for delirium using 4-AT</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Location</th>
<th>No. Patients</th>
<th>Mean or Median Age (Years)</th>
<th>Exclusion Criteria</th>
<th>Study Design</th>
<th>Dementia Screeners Assessed</th>
<th>Criterion Standard</th>
<th>Prevalence of Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schofield 2010</td>
<td>Glasgow Royal Infirmary Accident and Emergency Department, Feb–Aug 2007</td>
<td>520</td>
<td>77</td>
<td>Age &lt; 65, nonverbal, learning disability, non-English speaking without interpreter</td>
<td>Prospective, nonconsecutive sampling</td>
<td>AMT4</td>
<td>MMSE ≤ 23</td>
<td>MMSE ≤ 23 in 43% Due to pain or eyesight problems 13.5% of enrolled patients unable to complete the MMSE Mean completion times for AMT4 and MMSE were 4.7 and 11.6 minutes, respectively</td>
</tr>
<tr>
<td>Wilber 2005</td>
<td>Akron City Hospital Emergency Department, fall 2003</td>
<td>75 for SIS, 75 for Mini-Cog</td>
<td>75</td>
<td>Age &lt; 65, nonverbal, learning disability, non-English speaking, medically unstable, prescreening medications “that could affect their mental status”</td>
<td>Prospective, randomized cross-sectional convenience sampling</td>
<td>SIS</td>
<td>Mini-Cog</td>
<td>MMSE ≤ 23 in 24% of SIS group and 21% of Mini-Cog cohort SIS required &lt; 1 minute to complete Mini-Cog clock drawing test alone required a median of 1.5 minutes to complete</td>
</tr>
<tr>
<td>Wilber 2008</td>
<td>EDs of Akron City Hospital, Barnes-Jewish Hospital, and Cleveland Clinic, Jan 2006–Jan 2007</td>
<td>352</td>
<td>77</td>
<td>Age &lt; 65, non-English speaking; receipt of opioids, antiremetics, or benzodiazepines prior to cognitive assessment; medically unstable</td>
<td>Prospective, nonconsecutive sampling</td>
<td>SIS</td>
<td>MMSE ≤ 23</td>
<td>MMSE ≤ 23 in 32% Statistically nonsignificant differences in sensitivity and specificity were observed across sites</td>
</tr>
<tr>
<td>Wilding 2016</td>
<td>The Ottawa Hospital, Jan–Aug 2010</td>
<td>238</td>
<td>82</td>
<td>Age &lt; 75, medically unstable, preexisting dementia diagnosis, overt cognitive impairment, residence outside Ottawa, non-English or non-French language, hearing or visual impairment</td>
<td>Prospective, nonconsecutive sampling</td>
<td>O3DY Animal Fluency Test</td>
<td>MMSE ≤ 24</td>
<td>MMSE ≤ 24 in 13% MMSE and O3DY agreed in 75.6% MMSE and Animal Fluency Test agreed in 46% using cutoff &lt; 15 and 76% using cutoff &lt; 10</td>
</tr>
</tbody>
</table>

AMT-4 = Abbreviated Mental Test 4; 4-AT = delirium screen; BAS = Brief Alzheimer’s Screen; cAD8 = Caregiver Alzheimer’s Dementia 8; CAM-ICU = Confusion Assessment Method for the Intensive Care Unit; DRS-R98 = Delirium Rating Scale Revised 98; IQCODE = Informant Questionnaire on Cognitive Decline in the Elderly; MMSE = Mini Mental State Examination; O3DY = Ottawa3DY; SBT = Short Blessed Test; SIS = Six Item Screener; DSM-V = Diagnostic & Statistical Manual of Mental Disorders.
Barbic et al.\textsuperscript{41} reported moderate inter-rater reliability for O3DY ($\kappa = 0.64$) and for the SBT ($\kappa = 0.63$). Carpenter et al.\textsuperscript{37} evaluated multiple instruments head to head and observed that the O3DY and Brief Alzheimer’s Score categorized more patients with dementia (66 and 65\%, respectively) than did the cAD8 (55\%) or SBT (43\%). In terms of feasibility, Schofield et al.\textsuperscript{36} noted that the AMT-4 required a mean of 4.7 minutes to complete compared with 11.6 minutes for the MMSE. Wilber et al.\textsuperscript{34} reported that the Six Item Screener is usually completed in less than 1 minute compared with the Mini-Cog requiring a mean of 1.5 minutes to complete. Furthermore, many patients could not complete the Mini-Cog because of pain or intravenous line in their dominant arm or because they did not have their corrective lens in the ED.

The MMSE is inaccurate for identifying mild cognitive impairment and some favor the Montreal Cognitive Assessment (MoCA) instead.\textsuperscript{42,52,53} When compared against the MoCA, a “normal” Brief Alzheimer’s Screen, SBT, or cAD8 do not accurately reduce the probability of mild cognitive impairment. An abnormal cAD8 and SBT significantly increase the probability of mild cognitive impairment in African Americans in one unpublished urban study in the United States.\textsuperscript{42} However, the MoCA categorized 93\% of African American patients in this study as mild cognitive impairment compared with 63\% of Caucasians, which are both substantially higher than population norms. Health literacy also impacts the diagnostic accuracy of ED dementia screening for some patients. One urban United States study reported the cAD8 is significantly better to rule out dementia for patients with health literacy levels beyond 12th grade than are the Brief Alzheimer’s Screen or SBT.\textsuperscript{43}

### Test–Treatment Threshold

In developing the test–treatment threshold for the older adult with possible newly diagnosed dementia in the ED setting, we were most interested in exploring the scenario of expedited outpatient referral for definitive diagnostic testing that typically requires lengthy neurologic testing followed by advanced neuroimaging and cerebrospinal fluid analysis.\textsuperscript{54} Since Alzheimer’s disease is the most common dementia subtype and currently has no cure, the hypothetical benefits of diagnosing this disorder include potential disease modifying medications to slow the rate of cognitive decline\textsuperscript{55,56} as well as the opportunity for patients to voice end-of-life and other medical care preferences.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbic \textsuperscript{2018}\textsuperscript{41}</td>
<td>Nonconsecutive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Carpenter \textsuperscript{2011a}\textsuperscript{37}</td>
<td>Nonconsecutive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Dyer \textsuperscript{2016}\textsuperscript{47}</td>
<td>Nonconsecutive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>O’Sullivan \textsuperscript{2017}\textsuperscript{40}</td>
<td>Nonconsecutive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Schofield \textsuperscript{2010}\textsuperscript{36}</td>
<td>Uncertain</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Wilber \textsuperscript{2005}\textsuperscript{34}</td>
<td>Consecutive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Wilber \textsuperscript{2008}\textsuperscript{35}</td>
<td>Nonconsecutive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Wilding \textsuperscript{2008}\textsuperscript{39}</td>
<td>Nonconsecutive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

\textsuperscript{41} Barbic et al.\textsuperscript{41} reported moderate inter-rater reliability for O3DY ($\kappa = 0.64$) and for the SBT ($\kappa = 0.63$). Carpenter et al.\textsuperscript{37} evaluated multiple instruments head to head and observed that the O3DY and Brief Alzheimer’s Score categorized more patients with dementia (66 and 65\%, respectively) than did the cAD8 (55\%) or SBT (43\%). In terms of feasibility, Schofield et al.\textsuperscript{36} noted that the AMT-4 required a mean of 4.7 minutes to complete compared with 11.6 minutes for the MMSE. Wilber et al.\textsuperscript{34} reported that the Six Item Screener is usually completed in less than 1 minute compared with the Mini-Cog requiring a mean of 1.5 minutes to complete. Furthermore, many patients could not complete the Mini-Cog because of pain or intravenous line in their dominant arm or because they did not have their corrective lens in the ED.

The MMSE is inaccurate for identifying mild cognitive impairment and some favor the Montreal Cognitive Assessment (MoCA) instead.\textsuperscript{42,52,53} When compared against the MoCA, a “normal” Brief Alzheimer’s Screen, SBT, or cAD8 do not accurately reduce the probability of mild cognitive impairment. An abnormal cAD8 and SBT significantly increase the probability of mild cognitive impairment in African Americans in one unpublished urban study in the United States.\textsuperscript{42} However, the MoCA categorized 93\% of African American patients in this study as mild cognitive impairment compared with 63\% of Caucasians, which are both substantially higher than population norms. Health literacy also impacts the diagnostic accuracy of ED dementia screening for some patients. One urban United States study reported the cAD8 is significantly better to rule out dementia for patients with health literacy levels beyond 12th grade than are the Brief Alzheimer’s Screen or SBT.\textsuperscript{43}

### Test–Treatment Threshold

In developing the test–treatment threshold for the older adult with possible newly diagnosed dementia in the ED setting, we were most interested in exploring the scenario of expedited outpatient referral for definitive diagnostic testing that typically requires lengthy neurologic testing followed by advanced neuroimaging and cerebrospinal fluid analysis.\textsuperscript{54} Since Alzheimer’s disease is the most common dementia subtype and currently has no cure, the hypothetical benefits of diagnosing this disorder include potential disease modifying medications to slow the rate of cognitive decline\textsuperscript{55,56} as well as the opportunity for patients to voice end-of-life and other medical care preferences.
while still able. We were unable to find any research evidence with which to estimate the benefits or harms of ED dementia screening to guide downstream prescribing of disease modifying medications or goals of care. Another potential benefit is to use cognitive vulnerability identified by abnormal dementia screening as a decision point to initiate comprehensive geriatric assessment (CGA) in the ED or following the ED visit.57,58

CGA includes a structured diagnostic and therapeutic approach to identify a frail older person’s medical, functional, cognitive, and social capabilities and limitations with the ultimate objective to identify and manage geriatric vulnerabilities within the scope of the patient’s goals of care. A multidisciplinary approach to CGA is required and most of this approach will occur outside of the ED via a “frailty unit”, specialist ward, or mobile assessment team.59,60 While a recent Cochrane review exploring CGA for hospitalized patients demonstrates high-quality evidence demonstrating increased likelihood of remaining in home in the next year (16 trials, 6,799 patients) without improving dependence, mortality, or hospital readmissions, these CGA interventions did not occur in the ED.57 A before/after evaluation exploring the benefits of CGA in the ED demonstrated a significant reduction in hospital admissions for patients over age 65 from 59.5% to 53.0%.59 For the purposes of our test–treatment analysis, we used this 6.5% absolute risk reduction (59.5–53.0) as our proposed benefit for dementia screening in the ED (BRx).

The risks of dementia screening include unnecessary angst and additional testing with concomitant...
costs for false-negative results. Another risk is delay to potentially lifesaving interventions for acute illness or injury while dementia screening is performed. No ED studies have quantified this risk, so we estimate a 0.5% cumulative risk of dementia screening in this setting (Rt). The risks of treatment with CGA in patients without dementia (Rrx) have not been quantified either in the ED setting or in Cochrane reviews of hospitalized or surgical patients, so we estimate a 2% risk for the test–treatment equation.

The AMT-4 had the highest LR+ and when used in the test–treatment formula with pooled sensitivity 74% (Ppos/d), probability of a negative result in patients with dementia (Pneg/d - 1 - 0.74 = 0.26), pooled specificity 88% (Ppos/nd), and probability of a positive result in patients without dementia (Pneg/nd - 1 - 0.88 = 0.12), the test threshold is 14.7% and the treatment threshold is 36.5% (Figure 3). Since the weighted mean prevalence of abnormal dementia screening results in ED patients with predominantly previously unrecognized dementia in this meta-analysis is 30.6%, the test–treatment threshold based on these assumptions indicates that the potential harms and benefits of screening geriatric patients for dementia in the ED favor dementia screening. Since the O3DY is a simpler instrument that has been studied in more ED settings and more accurately reduces the probability of dementia than does the AMT-4 (pooled LR = 0.17 vs. 0.31), using the pooled sensitivity and specificity of the O3DY in the test–treatment equation yields a test threshold of 18% and a treatment threshold of 43%, which are not significantly different than the AMT-4 estimate. Test and treatment threshold estimates provide a quantitative context on which to
evaluate the value of process changes that incorporate dementia screening into ED patient care. An Excel file is provided as an online supplement to this manuscript for readers to recalculate these thresholds using alternative estimates of risk, benefit, and diagnostic accuracy as new evidence emerges (Data Supplement S2).

DISCUSSION

This meta-analysis provides ED clinicians with comparative diagnostic accuracy results on which to base emergency medicine dementia screening protocols, while also highlighting research priorities for future investigators. A 2016 Cochrane review quantified the accuracy of the MMSE as a dementia screening instrument for inpatient and primary care settings, noting a sensitivity of 85% and specificity of 90% at thresholds of not more than 24. However, the MMSE is copyright protected and is a time-consuming and cumbersome screening instrument for the fast-paced ED environment. The MMSE may yield false positives in lower socioeconomic and limited health literacy populations, while exhibiting false negatives in highly educated groups. The issue of wrongly labeling sociodemographic populations such as economically challenged or lower literate or non-English

Figure 2. Continued

![Graphs showing sensitivity, specificity, positive LR, and negative LR for dementia screening protocols across different studies.](image-url)
groups as higher risk for dementia based on the MMSE is particularly problematic since these groups may be disproportionately represented in some EDs. Profiling ethnic and socioeconomic disparities in future ED dementia screening studies is one approach to mitigate this bias, as highlighted at the 2003 Academic Emergency Medicine consensus conference. Our meta-analysis adds to the Cochrane review by summarizing the quantity and quality of emergency medicine research to guide clinicians, educators, and guideline developers creating evidence-based protocols to improve health care outcomes for ED patients with dementia and to provide guidance and support to their relatives.

Overcoming skepticism regarding the value of screening in the ED, weighed against the dangers of continual mission creep, is essential as the role of emergency medicine expands. Why should emergency providers screen for dementia if this chronic neurodegenerative process is incurable? Potential motivators include structural and process quality indicators (QIs) for ED dementia patients to reduce practice variability. Structural QIs include local ED policies for management of older persons with recognized dementia that include family/care partners in medical decision making during an episode of care, as well as adaptation of pain assessment approaches in dementia patients. Process QIs include tracking the proportion
of older ED patients with documented dementia assessment with Health Insurance Portability and Accountability Act–compliant notification of family to solicit collateral history when dementia is suspected. In addition, multiorganizational guidelines in the United States and Canada advise emergency providers in any adult ED to incorporate and document baseline cognitive function in the initial assessment of all aging adults. Routinely screening older adults for dementia and documenting this assessment using the same validated and psychometrically

### Table 3
Diagnostic Accuracy of Dementia Instruments from Single Studies

<table>
<thead>
<tr>
<th>Screening Instrument</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>Positive LR (95% CI)</th>
<th>Negative LR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal fluency &lt; 10</td>
<td>63 (45–77)</td>
<td>78 (76–81)</td>
<td>2.9 (1.8–4.0)</td>
<td>0.48 (0.28–0.73)</td>
</tr>
<tr>
<td>Animal fluency &lt; 15</td>
<td>91 (75–98)</td>
<td>39 (37–40)</td>
<td>1.5 (1.2–1.6)</td>
<td>0.24 (0.06–0.68)</td>
</tr>
<tr>
<td>Brief Alzheimer Screen</td>
<td>95 (87–99)</td>
<td>52 (48–55)</td>
<td>2.0 (1.6–2.2)</td>
<td>0.10 (0.02–0.28)</td>
</tr>
<tr>
<td>Mini-Cog</td>
<td>75 (51–91)</td>
<td>85 (78–89)</td>
<td>4.9 (2.4–8.3)</td>
<td>0.30 (0.10–0.62)</td>
</tr>
</tbody>
</table>

LR = likelihood ratio.
A robust instrument would allow downstream providers to rapidly differentiate acute from chronic cognitive dysfunction and contemplate delirium earlier in the ED episode of care. Identifying dementia should not delay access to care. While obtaining consent for time-dependent emergencies is often unnecessary, most ED decisions are not time dependent, and recognizing the presence or absence of dementia is essential for establishing capacity while engaging in shared decision making regarding tests, interventions, and disposition decisions.\textsuperscript{71} In addition, dementia severity varies between patients and none of the ED screening instruments evaluates dementia severity or decisional capacity.\textsuperscript{72}

Understanding the diagnostic accuracy of ultrabrief dementia screening instruments is also relevant for medical educators. Approximately one-third of senior emergency medicine residents lack confidence in the recognition and management of cognitive disorders, despite the emphasis on assessing for dementia and delirium as a core competency by the American Board of Emergency Medicine in 2010.\textsuperscript{20,73} Evidence-based educators should emphasize that not all dementia screening instruments evaluate dementia severity or decisional capacity.\textsuperscript{72}

Implications for Future Research
This review highlights multiple lessons that future investigators can employ to improve the overall quality and reproducibility of ED dementia research. Current labels for dementia screening instruments elicit confusion in communicating across fields and countries. The SBT was originally described in 1983 as the Orientation-Memory-Concentration Test\textsuperscript{74} and in the emergency medicine literature has also been called the Quick Confusion Scale\textsuperscript{45,75} and the 6-Item Cognitive Impairment Test.\textsuperscript{40} Establishing a uniform

\textbf{Box 1}
\textbf{Applying These Results in the Clinical Setting}

A 75-year old community-dwelling female presenting to the ED would have a pretest probability for dementia of 30% based on this systematic review. An abnormal AMT4 (pooled LR+ = 7.69) would increase the probability that this patient has dementia to 77\% (95\% CI = 60\%–88\%). The Six Item Screener (pooled LR+ = 3.53) would increase the posttest probability of dementia for this patient to 60\% (95\% CI = 50\%–69\%). The Brief Alzheimer’s Screen demonstrated the lowest LR− to most accurately reduce the probability of dementia (LR− = 0.10), decreasing the probability in this patient to 4\% (95\% CI = 0.8\%–11\%). Clinicians should explain the purpose of ED dementia screening and emphasize the necessity of definitive testing for abnormal results since these are not designed as stand-alone diagnostic tests.
nomenclature or index of synonymous instruments would permit investigators and journal editors to more clearly compare future studies with prior research. In addition, reporting guidelines for dementia diagnostic accuracy studies exist, but none of the studies used them.50 Only four studies used any version of the general diagnostic STARD reporting guidelines.51 Adherence to Enhancing the Quality and Transparency of Health Research (EQUATOR) Network reporting guidelines reduces intra- and interspecialty variability in communicating scientific methods, results, and clinical implications within the context of previous knowledge, yet guidelines like STARD are too often neglected in emergency medicine.76,77

Other methodologic issues require clarification and likely standardization across dementia studies. Some medications like benzodiazepines, antiemetics, and opioids often cloud individual’s sensorium, but only a few dementia studies specifically excluded patients receiving those medications prior to cognitive assessment with the index test or criterion standard. The illness or injury leading to the ED visit may also skew cognitive screening toward “abnormal.” In other words, the test–retest cognitive test reproducibility may be suboptimal when comparing cognitive testing in the ED while unwell with the same testing (same instrument) weeks later after recovery.78 The order of administration of the index test and the reference standard may also skew observed accuracy via recall bias.79,80 For example, a three-item recall or stating months backward may be performed more easily by patients the second time they are tested. If an index test and reference standard share a similar component such as stating the months of the year backward, the index test could be interpreted as a false negative or false positive depending on the order of administration as was observed in Wilber et al.35 Additionally, few studies used objective assessments to differentiate delirium from dementia. Although the overall recognition of cognitive dysfunction, whether acute or chronic, is the most important assessment of decisional capacity in aging adults, differentiating dementia and delirium is an important epidemiologic issue and essential for interventions striving to improve outcomes. A variety of instruments exist to identify delirium.21,81 Few studies explore the ethical concerns surrounding research of potentially cognitively impaired individuals in emergency settings. Although all studies appropriately excluded critically ill or injured patients, ethical guidelines for emergency research recommending subject assent and care partner consent exist and merit elaboration in future ED dementia screening research.82

Longterm, multidisciplinary acceptance of ED dementia screening instruments will ultimately require comparison against reference standards respected by a vast array of stakeholders including neurologists, psychologists, and geriatricians. Unfortunately, the chaotic ED is not conducive to time-consuming testing that adheres to DSM-V criteria, nor do most ED research laboratories have the expert personnel required to provide such neuropsychological testing around the clock. Every study in this meta-analysis used the MMSE, which is not an acceptable reference standard. Feasible alternatives to the MMSE include the MoCA52 and the St. Louis University Mental State examination,83 although preliminary unpublished ED evaluation in urban U.S. settings indicate over 90% of patients are stratified as abnormal using the MoCA implying a need to adapt population norms for these tests.42

As demonstrated in our test–treatment assumptions, quantifiable benefits for ED dementia screening are lacking, although not unimaginable. We estimated benefits at the level of the patient, but caregiver and societal benefits are also conceivable. For example, identification of possible dementia could be linked with assessment of caregiver strain, which when unaddressed has been linked with dissatisfaction on medical care patient surveys.69,84 Identifying dementia risk could trigger subsequent assessment of caregiver strain via linkage to telephone follow-up,85 community care coordination programs like “Partners in Dementia Care,”86 or other initiatives that could simultaneously improve the process and outcomes of care for patients with previously unrecognized dementia. Societal benefits for ED dementia screening include assessment of driving safety for older adults, whereby early recognition might prevent future accidents.87

In diagnostics a hierarchy of evidence exists. Quantifying accuracy alone is a lower tier of value than research demonstrating improved outcomes of importance to patients and families.88 Demonstrating benefits for ED dementia screening may require randomized trials. Such trials should evaluate both beneficial patient-centric outcomes (which are not necessarily traditional mortality or ED returns89) and potential adverse consequences of angst, additional medical expenses, and inconvenience for false-positive screening results. In addition, dementia is not a monomorphic disease, but rather exists in spectrums of severity with multiple subtypes. Rarely, dementia is a symptom of a reversible
disease like hypothyroidism or depression.92 Future ED diagnostic and interventional researchers should begin to stratify dementia by severity. The Clinical Dementia Rating scale exists in the Alzheimer’s research community but has never been assessed in ED settings.91 Instead, ED investigators stratify dementia severity by nonvalidated gradients of worsening MMSE scoring, which is likely skewed by language barriers, literacy levels, and sensory impairments. Finally, implementation researchers should also evaluate the feasibility and accuracy of dementia screening using technology like smart phones and iPads while patients and caregivers await care in waiting rooms or other times of ED delays.92 The majority of older patients are comfortable with this technology and more automated dementia screening processes not reliant on the nurse or physician personnel would promote knowledge translation.93

LIMITATIONS

The ability of this meta-analysis to accurately and reliably delineate the diagnostic role of appropriate ED screening instruments for dementia is limited by the small number of studies available. In addition, these studies used a variety of criteria to establish the diagnosis of dementia. Many of the studies used the same nonclinical personnel to collect the variables for both the screening test being assessed and the reference standard, which increases the possibility of incorporation bias that can skew observed estimates of both sensitivity and specificity upward.94 Only one study used a reference standard that incorporated DSM criteria and evaluation by an expert in cognitive evaluation. Although clinicians are unlikely to apply DSM criteria to rule in or rule out dementia, diagnostic researchers in ED settings can and should rely on DSM criteria. Unfortunately, no studies evaluated or sufficiently hypothesized about the potential value of ED dementia screening for patients, caregivers, society, or ED operational flow. Therefore, our test–treatment results and discussion derive from arbitrary estimates for risks of dementia screening, as well as non–evidence-based assumptions of intervention potential benefits and harms once dementia is assumed based on ED screening. Acknowledging these arbitrary assumptions, this article includes an interactive Excel file to empower readers to adjust our estimates when higher-quality evidence emerges. Finally, a minority of included studies adhered to the STARD reporting criteria, which is likely associated with the significant heterogeneity observed between the individual studies.96

CONCLUSIONS

Despite the frequency and anticipated societal challenges associated with aging population’s dementia-related cognitive dysfunction, little diagnostic research exists to guide geriatric ED dementia screening protocols. Existing research is limited by inadequate reference standards. Acknowledging these limitations, the AMT-4 most accurately rules in dementia, while the BAS most accurately rules out dementia. Future ED dementia screening accuracy research should use DSM criteria, standard names for the same instruments, and the same thresholds to enhance the quality of subsequent studies. Based on largely arbitrary assumptions of risk and benefit, our test–treatment threshold calculations indicate that ED dementia screening as a marker of vulnerability to guide initiation of CGA would be beneficial to a subset of geriatric patients.

The authors acknowledge responses from J. Stephen Huff, MD, seeking additional study details.

References


41. Barbic D, Kim B, Salehmohamed Q, Kemplin K, Carpenter CR, Barbic SP. Diagnostic accuracy of the Ottawa 3DY and Short Blessed Test to detect cognitive dysfunction in geriatric patients presenting to the emergency department. BMJ Open 2018;8:e019652.


87. Carr DB, Ott BR. The older adult driver with cognitive impairment: “It’s a very frustrating life”. JAMA 2010;303:1632–41.

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.13573/full
- Data Supplement S1. Supplemental material.
- Data Supplement S2. Interactive calculator.
- Data Supplement S3. Appendix of ED dementia screening instruments.
Hot Off the Press: Validation of the Pediatric NEXUS II Head Computed Tomography Decision Instrument for Selective Imaging of Pediatric Patients with Blunt Head Trauma

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

ABSTRACT
This is a validation of a preplanned secondary analysis of the NEXUS II Head computed tomography (CT) decision instrument, focusing on the pediatric population. A total of 1,018 patients less than 18 years old, who underwent head CT imaging for blunt head trauma, were included. The decision instrument correctly identified all patient (27/27) who required neurosurgical intervention and 48 of 49 patients who had clinically significant head injury on CT imaging.

BACKGROUND
Blunt head trauma is a very common presenting complaint to U.S. emergency departments (EDs), with over 2 million pediatric visits annually.1 While only 10% of patients are ultimately diagnosed with traumatic brain injury (TBI) and 1% with clinically important injury, CT use is still on the rise.2,3 A number of clinical decision instruments (CDIs) have been developed to help with decision making on these patients, with PECARN being recommended by the American Academy of Pediatrics.4 The authors of this study used a subgroup of the NEXUS Head CT DI to evaluate its utility in pediatric patients.

ARTICLE SUMMARY
This is a planned secondary analysis of the NEXUS Head CT DI derivation and validation studies. Patients less than 18 years who underwent head CT imaging at four hospital EDs were included. The NEXUS head CT DI is considered negative if all the following are true: 1) no evidence of skull fracture, 2) no scalp hematoma, 3) no neurologic deficits, 4) normal level of alertness, 5) normal behavior, 6) no persistent vomiting, and 7) no coagulopathy. The primary outcome was need for neurosurgical intervention (death due to head injury, craniotomy, elevation of skull fracture, intubation related to head injury, or...
intracranial pressure monitoring within 7 days). The secondary outcome was clinically significant head injury on CT imaging.

QUALITY ASSESSMENT

This was a well-done, preplanned validation study. The patients were enrolled from multiple ED settings. Although the primary outcome was appropriate, it was a composite outcome, and not all the outcomes are equal. The rule was not compared to any other CDIs, nor to clinical gestalt. Furthermore, this is a validation study, but not an implementation study, so we do not know whether this rule will actually decrease imaging when applied. The final total number of patients with the primary outcome was small, so confidence intervals (CIs) were relatively wide but similar to those for the PECARN rule, but the total number of patients included was large. The criteria included in the DI have been previously found to be reliably assessed by clinicians. Secondary outcomes were appropriate. In addition, only patients who received a CT were included; those who did not were excluded. Nobody in a 368-patient verification bias cohort, including 118 children, were ultimately diagnosed with an intracranial injury.

KEY RESULTS

There were 1,018 patients less than 18 years old who received head CT scans. This cohort included 27 patients (2.7%) who required neurologic intervention, and 49 patients (4.8%) had significant intracranial injuries.

- Primary outcome: Need for neurosurgical intervention.
  - Sensitivity: 100% (95% CI = 87.2%–100%)—27 of 27.
  - Specificity: 33.3% (95% CI = 30.3%–36.3%)—330 patients of 991 who did not require intervention were classified as low-risk status by the Pediatric NEXUS II Head CT DI.
  - NPV: 100% (95% CI = 99.6%–100%)—None of the 991 low-risk patients required neurosurgical intervention.
- Secondary outcome: Clinically significant head injury.
  - Sensitivity: 98.0% (95% CI = 89.1%–99.9%)—48 of 49 patients with significant injury were identified by the Pediatric NEXUS II Head CT DI.
  - Specificity: 34.0% (95% CI = 31.0%–37.0%)—329 of 969 patients who did not have significant injury were classified as low risk by the Pediatric NEXUS Head CT DI.
  - NPV: 99.7% (95% CI = 98.3%–100%)—329 of 330 low-risk patients were absent of a clinically significant head injury.

AUTHORS’ COMMENTS

A properly powered study with tighter CIs around the point estimate for the patient-oriented outcome of need for neurosurgical intervention and an implementation study to examine the rule’s true impact on imaging are desirable before the Pediatric NEXUS II Head CT DI is widely adopted.

TOP SOCIAL MEDIA COMMENTARY

Ian Stiell (@EMO_daddy)
We have a wealth of decision instruments for peds head injuries. Don’t forget the Canadian CATCH rule as well, validation just published in CMAJ.

Chris Bond (@socmobem)
Have CATCH and NEXUS II had any effect on your practice?

Shawn Dowling (@shawnkdowling)
They haven’t changed my practice. I would CT way more children if I followed CATCH CATCH2 and I’m pretty sure I haven’t missed a significant TBI in peds. I use this paper to justify using clinical gestalt—https://goo.gl/mG9Mwc.

Justin Morgenstern (@First10EM)
Nice thing about high sensitivity, low specificity tools is that you can use 2 on a patient, and if they pass one, they are almost certainly low risk. Although for this topic I still think gestalt combined with a preference for watchful waiting over CT is the way to go.
Twitter Poll

What decision instrument do you use to decide on head CT for kids with blunt head trauma? #SGEMHOP #FOAMed

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

Pediatric NEXUS II in blunt head trauma
4-site prospective observational cohort: preplanned secondary analysis
Patients <18y with blunt head trauma undergoing CT head
Excl: penetrating trauma, injury >24h previous, non-trauma indication for CT

1018 patients
688 high risk
330 low risk

27 neurosurgical intervention
0 NS intervention
Sens 100% (87-100%) spec 33% (30-36%) NPV 100% (99.6-100%)

48 significant injury
1 significant injury
Sens 98% (89-99%) spec 34% (31-37%) NPV 99.7% (98-100%)

Gupta doi 10.1111/acem.13431

Pediatric NEXUS decision instrument: any = high risk

Evidence of basilar or depressed skull #
Scalp hematoma
Neurologic deficit
Abnormal alertness
Abnormal behavior
Persistent (>1) vomiting
Cocanutpathy

SGEMHOP #225
TAKE-TO-WORK POINTS

Traumatic brain injury will remain a common complaint in pediatric patients presenting to the ED. Multiple clinical decision instruments exist to help with clinician decision making. NEXUS II Head CT DI shows promise in helping reduce unnecessary head CTs in patients who may have received them, but more work is needed before the instrument is ready for prime time.

References


ABSTRACT
Physician-assisted death (PAD) has long been a strongly debated moral and public policy issue in the United States, and an increasing number of jurisdictions have legalized this practice under certain circumstances. In light of changing terminology, laws, public and professional attitudes, and the availability of published data about the practice, we review key concepts and terms in the ongoing PAD debate, moral arguments for and against PAD, the current legal status of PAD in the United States and in other nations, and data on the reported experience with PAD in those U.S. jurisdictions where it is a legal practice. We then identify situations in which emergency physicians (EPs) may encounter patients who request PAD or have attempted to end their lives with physician assistance and consider EP responses in those situations. Based on our analysis, we offer recommendations for emergency medical practice and professional association policy.

S
ince the announcement by Dr. Jack Kevorkian in 1990 that he had used his “suicide machine” to assist the death of a patient with Alzheimer disease, physician-assisted death (PAD) has been a hotly debated moral and public policy issue in the United States (US), and an increasing number of jurisdictions have legalized the practice under certain circumstances. In light of changing terminology, laws, public and professional attitudes, and more published data about the practice, ethicists and professional medical organizations have re-examined physician obligations, though not specifically from the perspective of emergency physicians (EPs). This article will provide a brief review of:

1. Key concepts and terms in the ongoing PAD debate;
2. Moral arguments for and against PAD;
3. The current legal status of PAD in the United States and in several other nations; and
4. Data on the reported experience with PAD in those U.S. jurisdictions where it is a legal practice.

The article will then identify situations in which EPs may encounter patients who request PAD or have attempted to end their lives with physician assistance and will consider EP responses in those situations.

KEY CONCEPTS AND TERMS
As is the case with other morally controversial issues such as abortion and the rationing of health care, the choice of terms to refer to the practices in question in PAD is itself a matter of considerable debate. Despite disagreement about what terms should be used, there is general agreement that physician assistance in dying...
can occur in two major ways. First, physicians can provide patients with a prescription for lethal medication and instruct them on how to take that medication. Patients can fill the prescription and then ingest the lethal dose of drug(s) to end their lives at a time of their own choosing. Second, physicians themselves can perform an action that directly causes a patient’s death, most commonly injection of a lethal drug dose. Several nations have legalized both of these practices for certain patients; in the United States, however, the jurisdictions that have legalized PAD to date permit only the first of the above practices.

Over roughly the first 15 years of the current US debate over PAD, the two terms most commonly used in the medical, bioethics, and legal literature were “physician-assisted suicide” for the former practice and “euthanasia” for the latter practice. Within the past decade, U.S. proponents of the practice called physician-assisted suicide have argued that this term should be abandoned, to avoid inappropriate associations with negative psychological and moral attitudes toward suicide. Multiple terms with more neutral or positive connotations have been proposed to replace it, including “physician-assisted death,” “physician aid in dying,” “hastening death,” and “medical assistance in dying.” Opponents of PAD, in contrast, continue to use the term physician-assisted suicide, presumably because they are convinced that it more clearly indicates the nature of this action and that the negative moral connotations of this term are appropriate. The remainder of this article will use the terms “physician-assisted death” (PAD) and “physician-assisted suicide” interchangeably to refer to the first of the two practices described above, that is, physician provision of a prescription for a lethal medication with instructions on how to take that medication.

**MORAL ARGUMENTS FOR AND AGAINST PAD**

Over the past quarter-century, multiple commentators have articulated moral arguments both for and against PAD. Defenders of PAD typically conclude that this practice should be a lawful option for willing patients and their physicians, and critics maintain that PAD should be prohibited by law. The basic arguments are well known and have not changed significantly over the course of the debate.

Major arguments offered by proponents of PAD include the following:

- **Respect for autonomy:** As autonomous moral agents, people approaching the end of life should be free to choose to die in an effective, painless, and nondisfiguring way, with the aid of a willing physician.
- **Consistent treatment:** Patients whose survival is dependent on life-prolonging medical treatments may choose to hasten their deaths by discontinuing those treatments. Therefore, patients approaching the end of life who are not dependent on life-prolonging medical treatments should also be able to hasten their deaths through ingestion of a lethal medication prescribed by a willing physician.
- **Relief of suffering:** Patients should be able to receive physician aid in dying to prevent foreseen and undesired significant suffering inflicted by a terminal illness or lethal injury.

Major arguments offered by opponents of PAD include the following:

- **Sanctity of life:** Because all human life is sacred, human beings should not choose to hasten their deaths. Therefore, they should not take their own lives and should not seek or receive the assistance of physicians in committing suicide.
- **Abuse:** Physician-assisted suicide should be prohibited because this practice is susceptible to serious abuse, most notably manipulating or coercing vulnerable patients to take their own lives or eroding the safeguards to encompass patients who lack decisional capacity.
- **Physician integrity:** Physician-assisted suicide should be prohibited because this practice undermines the integrity of medicine as a healing profession that is dedicated to caring for patients with advanced illness and is opposed to killing. The legal practice of physician-assisted suicide will erode patients’ trust that their physicians are committed to protecting and extending life.

In the now protracted debate over PAD, scholars have analyzed these and other arguments and offered multiple responses to them. Proponents of PAD, for example, argue that the statutory conditions and procedures for access to PAD are sufficient to prevent abuses of this practice. Opponents of PAD argue that palliative and supportive care practices are able to relieve the physical and emotional suffering of patients approaching the end of life and so PAD is not needed to prevent that suffering.
THE LEGAL STATUS OF PAD

In a pair of eagerly awaited 1997 decisions, the U.S. Supreme Court rejected lower-court assertions of a constitutional right to physician-assisted suicide and authorized states to enact legislation either permitting or prohibiting this practice.\(^1\) By late 2018, eight U.S. jurisdictions (seven states and the District of Columbia) had legalized the practice of PAD. Oregon was the first state to do so, in 1997, followed by Washington and Montana in 2009, Vermont in 2013, California in 2015, Colorado and the District of Columbia in 2016, and Hawaii in 2018.\(^1\) With the exception of Montana, where the right to PAD was established by a state supreme court decision, statutes in each of these jurisdictions impose a similar set of conditions on access to PAD. To be eligible for PAD, persons must:

1. Be residents of the jurisdiction in which their request for PAD is made;
2. Be at least 18 years old;
3. Be able to make and communicate health care decisions;
4. Have a terminal illness, that is, an illness that physicians judge will lead to death within 6 months; and
5. Be able to self-administer and ingest the prescribed medication.\(^1\)

Two physicians must certify that patients meet all of the above eligibility criteria. Patients must make two oral requests for PAD to their physicians; these requests must be at least a set number of days apart. They must also make a written request for PAD. Additionally, Hawaii requires a mental health screening. If all of these conditions are met, physicians may provide a prescription for a lethal medication; they may also refuse to honor requests for PAD.\(^1\)

In addition to these U.S. jurisdictions, six other nations (Belgium, Canada, Colombia, Luxembourg, Switzerland, and the Netherlands) have legalized PAD.\(^1\) All of these nations except Switzerland have also legalized the practice of physician-administered euthanasia. On November 29, 2017, Victoria became the first Australian state to enact legislation legalizing PAD.\(^1\)

U.S. EXPERIENCE WITH PAD

Public health departments in three states (Oregon, Washington, and California) have published annual reports summarizing patient data reported by physicians who have provided prescriptions for lethal medications under their PAD statutes.\(^1\) Oregon’s most recent report provides the most extensive data on PAD; it includes information about the 1,967 people who received prescriptions for lethal medications and the 1,275 who died from ingesting lethal medications in Oregon from 1998 through 2017.\(^1\) Of the patients who died from ingesting lethal medications during this period:

- 96.3% were white;
- 98.7% had private or governmental health insurance;
- 90.2% were enrolled in hospice;
- 72.5% had at least some college education; and
- 77.9% had cancer as their underlying illness.

These patients’ most common reasons for requesting PAD, as assessed by their prescribing physicians, were loss of autonomy (90.9%), loss of ability to engage in activities that make life enjoyable (89.5%), and loss of dignity (75.7%). Inadequate pain control was identified as a concern in 25.8% of these patients. The rate of PADs in Oregon has increased over the years since its legalization, but in 2017 it was still less than 0.4% of total deaths in that state. Data on PAD from Oregon and other states may inform evaluation of the potential for abuse of this practice.

WHAT ROLE FOR EMERGENCY PHYSICIANS?

Given the ongoing controversy over the practice of PAD, morally responsible EPs may interpret their duties in different ways. In light of this, and of the expanding legalization of the practice, EP responses in the following situations deserve consideration:

Patients Who Request PAD

As described above, dying with physician assistance in those U.S. jurisdictions where this is a lawful option is a multistep process, including three separate requests for PAD and satisfaction of multiple eligibility criteria. The large majority of patients seeking PAD, therefore, are likely to approach physicians with whom they have an ongoing relationship, such as their primary care physicians or oncologists. Patients’ personal physicians are also likely to be in the best position to determine whether their patients meet all of the eligibility criteria for PAD. Because they do not typically provide continuing medical care, EPs are unlikely to receive requests for PAD. It is nevertheless possible
that a person may present to a hospital emergency department (ED) and request assistance in dying from an EP on duty in that setting. How should the EP respond?

To honor such a request in accordance with statutory requirements, the EP would need to determine that the patient satisfies all of the eligibility criteria for PAD, obtain confirmation from a second physician that the patient meets those eligibility criteria, arrange for an additional oral and a written request from the patient after the prescribed waiting periods, and be personally willing to honor the patient’s request. Because the mission and setting of the ED are not well suited for this process, we believe that the overwhelming majority of EPs would choose not to honor requests for PAD. If, however, an EP is willing and can competently undertake all of the required steps in this process, we know of no in-principle reason why he or she should be prohibited from doing so, in jurisdictions where PAD is permitted. Nevertheless, in keeping with standard practice for patients with progressed suicidal ideation, EPs may initiate psychiatric evaluation and seek palliative care consultation whenever indicated.

Patients Who Have Ingested Lethal Medication in Accord with the PAD Law of That State or Jurisdiction

The Oregon Public Health Division’s most recent 2017 summary of PAD outcomes reports 34 “complications” among the 1,275 deaths from ingesting lethal medications in Oregon since 1998 and an additional seven patients who regained consciousness after ingesting lethal medications. If these data are representative, complications after ingesting lethal medications are very infrequent. If a complication does occur, however, patients may present to the ED for assistance. How should EPs respond to this situation?

Emergency physicians’ initial involvement may occur if families or patients call emergency medical services (EMS) first responders when the patient does not die after ingesting medication prescribed to cause death. If patients have portable orders for end-of-life care, such as physician orders for life-sustaining treatment (POLST) or portable do not attempt resuscitation (DNAR) orders, EPs and EMS directors should advise EMS responders to respect these portable orders regarding interventions ranging from invasive life-prolonging measures to noninvasive respiratory support to palliative interventions only. If called to the scene, EMS should generally transport patients in this circumstance to facilities to optimize palliative interventions, unless arrangements for available and sufficient hospice care are already in place. Orders for limited interventions and palliative interventions only remain subject to rescission by patients with decision-making capacity or their legally authorized surrogates; in that case, aggressive EMS interventions to preserve life are appropriate.

In a 2001 analysis of the moral problem posed by patients arriving in the ED after ingesting a prescribed lethal medication, Moskop and Iserson identified three options: 1) aggressive intervention to preserve life, 2) palliative care only, and 3) assistance in completing the suicide. These authors cautioned EPs against active assistance in completing the suicide by providing additional lethal medication in this circumstance. EPs have only limited information about these patients, and they are unable to carry out the required statutory steps for providing these drugs. Moskop and Iserson advised that EPs should provide only comfort care if they have reliable evidence that the patient’s intention is to complete the dying process but should take active measures to preserve life if they have reliable evidence that patients have changed their minds and no longer desire to end their lives. If evidence regarding the patient’s intention is lacking or ambiguous, EPs should treat the patient promptly with the degree of aggressiveness that they judge to be medically indicated.

Patients Who Have Ingested Lethal Medication in Accord with PAD Law of That Jurisdiction but Who Present in a Neighboring Jurisdiction

These circumstances will occur if residents of a jurisdiction that has legalized PAD cross borders to a nearby hospital located in a jurisdiction that has not approved PAD, POLST, or other portable medical orders designed to guide interventions. Although most states do not provide reciprocity for other states’ portable orders (and so don’t provide immunity from liability for honoring those orders), we maintain that these orders provide significant evidence of the patient’s wishes and so can be used as a reliable guide to treatment.

Patients Without Decision-making Capacity or Surrogates

For patients appearing without decision-making capacity or accompanying surrogates, EPs should intervene
as they would for other patients by providing beneficial treatment until they can clarify goals of treatment. Because they lack information about the patient’s wishes, EPs should pursue appropriate life-sustaining treatments that may be discontinued when more information is available.

**RECOMMENDATIONS**

**Advocacy**

Multiple U.S. medical professional societies condemn the practice of PAD and oppose its legalization. Others have adopted a neutral stance on this issue, and none has formally endorsed PAD. There are areas where EP professional organizations should be at the forefront of advocating for change and for patients’ rights. Since PAD is not a primary EM issue, however, we propose that EP organizations maintain a position of neutrality, following many other professional societies and state medical boards.

**EP Participation**

The ED is not a practice setting conducive to complying with the requirements of current PAD laws. For example, Oregon specifies that the prescribing physician should be a “physician who has primary responsibility for the care of the patient and treatment of the patient’s terminal disease.” Rarely does the EP function as the primary physician for a patient. Most patients who choose PAD are dying of cancer, and generally EPs are not experts in estimating life expectancy in these disease processes. Moreover, the single encounter model of the ED would not enable the required review process and waiting periods in each state’s law. For all of these reasons, we recommend that EP professional organizations not support EP participation except in special circumstances. Individual EPs should be free to express and defend their views on PAD.

**EP Treatment after Ingestion**

Although reports from states with a legal practice of PAD provide evidence that significant complications of PAD are rare, EPs should still consider the scenario of a patient who presents to the ED after an incomplete or failed PAD attempt. In this situation, the EP is not the legally authorized prescriber of the lethal medication, and administration of IV medications with the intent to cause death would be euthanasia, not PAD, and would violate state homicide laws. For these reasons, we conclude that the EP should not administer additional lethal medication in the ED. The EP should, however, respect patient wishes as indicated by advance directives and portable medical orders, such as portable DNAR orders and POLST-paradigm orders. EPs should respond to a “failed” attempt at PAD with appropriate comfort care measures and support for family members, even though comfort care measures might hasten death as an unintended consequence of the effort to provide comfort. In the event that the patient “changes his or her mind,” supportive and invasive care would be appropriate. We recommend caution in circumstances where the patient is unable to speak for him or herself and a family member insists that “everything must be done.” Ethically appropriate surrogate decisions should reflect patient choices even though the surrogate may have statutory authority to rescind portable medical orders.

**SUMMARY**

As additional states choose to legalize physician-assisted death, emergency physicians are well advised to consider their professional role in the care of patients who, with support of their primary physicians, choose this course of action. Because the ED is not well suited to support the practice of physician-assisted death under current guidelines, we recommend that emergency physicians not provide this service, absent special circumstances. We argue that emergency physicians should not provide lethal medication to complete physician-assisted death in the ED, but that they should honor patients’ treatment wishes, including wishes expressed in advance directives and portable physician orders. Finally, because most emergency physicians will not provide physician-assisted death, we recommend that emergency physician professional organizations not endorse a position for or against the morality or legalization of physician-assisted death.

**References**


Epinephrine for Out-of-hospital Cardiac Arrest

Kyle Kelson, MD, and Ian S. deSouza, MD

Summary
Epinephrine increases the probability of Return of Spontaneous Circulation (ROSC) without improving survival or favorable long-term neurological outcome.

Source

Benefits in NNT
One in seven adult patients treated with epinephrine for OHCA achieved ROSC prior to hospital arrival. There was no benefit for survival to hospital discharge or survival at 1 month.

Benefits in percentage
Fourteen percent of adult patients treated with epinephrine for OHCA achieved ROSC prior to hospital arrival. There was no benefit for survival to hospital discharge or survival at 1 month.

Harms in NNT
One in 83 adult patients treated with epinephrine for OHCA had a worse long-term neurologic outcome.

Harms in percentage
A total of 1.2% of adult patients treated with epinephrine for OHCA had a worse long-term neurologic outcome.

Color recommendation
Red (no benefits)

Efficacy Endpoint(s)
Prehospital ROSC, survival to hospital discharge, survival at 1 month.

Harm Endpoint(s)
Long-term neurologic outcome, defined as cerebral performance category (CPC) score of 1 or 2 (corresponding to independence in Activities of Daily Living).

WHO WAS IN THE STUDIES
A total of 655,853 patients from 13 observational studies and one randomized, controlled trial (RCT) involving patients who experienced out-of-hospital cardiac arrest (OHCA).

Narrative
The original data supporting the use of epinephrine for cardiac arrest is rooted in a poorly controlled canine study1 from the 1960s. The Advanced Cardiac Life Support (ACLS) recommendation that it “may be reasonable” to use epinephrine in cardiac arrest continues in contemporary practice. Yet, when the ACLS guidelines2 are read carefully, they state that “for both survival to discharge and survival to discharge with good neurological outcome, there was no benefit” to receiving epinephrine.

Physiologically, epinephrine is theorized to promote peripheral vasoconstriction, thereby increasing diastolic pressure and coronary perfusion. Epinephrine also increases myocardial work and metabolic demand and may worsen tachydysrhythmia. Despite the purported physiologic benefits, it appears that epinephrine use for OHCA increases the rate of ROSC but does not increase the chance of survival (Table 1). Epinephrine may even worsen the neurologic outcome in patients who do survive. Epinephrine use for OHCA may therefore extend suffering and increase end-of-life health care costs due to intensive care and prolonged hospitalization without clear patient-centered, long-term benefits. Patients who survive to hospital discharge...
may be more likely to be dependent on others for care. This presents an obvious ethical quandary, as patient and family preferences may differ greatly with regard to life-prolonging therapies.

**CAVEATS**

The source meta-analysis³ incorporated 13 observational studies and only one RCT.⁴ Jacobs et al.⁴ studied 601 patients and showed that epinephrine for OHCA was effective at increasing rate of ROSC (odds ratio [OR] = 3.4, 95% confidence interval [CI] = 2.0 to 5.6) and survival to hospital admission (OR = 2.3, 95% CI = 1.4 to 3.6) but lacked a statistically significant effect on survival to hospital discharge (OR = 2.2, 95% CI = 0.7 to 6.3). The observational trials included in the meta-analysis³ attempted to control for potential confounders (i.e., time to cardiopulmonary resuscitation [CPR]) by propensity-matching individual study subjects, although bias can never be fully accounted for in any observational study. Similarly, the use of random-effects methods cannot fully control for heterogeneity when reporting pooled effects in meta-analyses, and three-fourths of reported outcomes had exceedingly high heterogeneity (I² = 96%). The source of heterogeneity may be patient characteristics, cointerventions, or trial-level and random-effects methods only attempt to adjust for between-trial variability, which can unintentionally inflate the effect of small studies on the pooled results.⁵ Propensity-matched outcomes in individual studies were generally in agreement with the pooled outcomes in this meta-analysis³ with the exception of survival at 1 month. The pooled results for 1-month survival may be confounded by the negative effect of smaller studies that were in favor of withholding epinephrine.

Specific effect modifiers in the included studies such as the timing and dosing of epinephrine administration may have influenced treatment effects on resuscitation outcomes. Earlier epinephrine administration has been associated with more favorable outcomes.⁶⁻⁸ The “standard” 1 mg of epinephrine given in cardiac arrest is not weight-based and therefore can have a differential physiologic effect depending on the individual patient.⁹ The meta-analysis³ included some trials that used “high dose” (0.1 to 0.2 mg/kg) epinephrine, which may have contributed negatively to pooled outcome results. Epinephrine administered at higher doses may be harmful¹⁰ and is not recommended by current guidelines.²

Loomba et al.³ defines positive neurologic outcome as a CPC of 1 or 2, which represents mild or moderate cerebral dysfunction but ability to independently perform activities of daily living. This is a validated method of measuring neurologic outcome; however, trials included in the systematic review³ measured the CPC at different time points (hospital discharge, 1 month, etc.), resulting in significant heterogeneity in the data. The “recovery time” prior to CPC assessment is an important confounder when determining neurologic outcome as are post-ROSC interventions such as targeted temperature management and urgent cardiac catheterization, none of which were adjusted for in the individual trials of this meta-analysis.³ Therefore, the results of the meta-analysis pertaining to this outcome should be interpreted with caution. The data discussed in Loomba et al.³ are primarily applicable to cardiac arrests that occur out of hospital, and conclusions may not apply to patients that experience cardiac arrest in the setting of an emergency department, hospital floor, intensive care unit, or operating room. In these more controlled hospital settings, epinephrine is more likely to be given earlier along with prompt defibrillation and high-quality CPR, and therefore its use for in-hospital cardiac arrests may result in different outcomes. Finally, this systematic review³ analyzed only adult patients, and its conclusions should not be applied to the pediatric population.

**Table 1**

<table>
<thead>
<tr>
<th>Benefits and Harms Associated With the Use of Epinephrine in Prehospital Cardiac Arrest Between Epinephrine and No Epinephrine Groups*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Epinephrine</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Prehospital ROSC</td>
</tr>
<tr>
<td>Survival at 1 month</td>
</tr>
<tr>
<td>Survival at hospital discharge</td>
</tr>
<tr>
<td>Good neurologic outcome (CPC score 1 or 2)</td>
</tr>
</tbody>
</table>

*Source: Loomba, et al.¹
CPC score = cerebral performance category score; NNH = number needed to harm; NNT = number needed to treat.

Loomba et al.³ defines positive neurologic outcome as a CPC of 1 or 2, which represents mild or moderate cerebral dysfunction but ability to independently perform activities of daily living. This is a validated method of measuring neurologic outcome; however, trials included in the systematic review³ measured the CPC at different time points (hospital discharge, 1 month, etc.), resulting in significant heterogeneity in the data. The “recovery time” prior to CPC assessment is an important confounder when determining neurologic outcome as are post-ROSC interventions such as targeted temperature management and urgent cardiac catheterization, none of which were adjusted for in the individual trials of this meta-analysis.³ Therefore, the results of the meta-analysis pertaining to this outcome should be interpreted with caution. The data discussed in Loomba et al.³ are primarily applicable to cardiac arrests that occur out of hospital, and conclusions may not apply to patients that experience cardiac arrest in the setting of an emergency department, hospital floor, intensive care unit, or operating room. In these more controlled hospital settings, epinephrine is more likely to be given earlier along with prompt defibrillation and high-quality CPR, and therefore its use for in-hospital cardiac arrests may result in different outcomes. Finally, this systematic review³ analyzed only adult patients, and its conclusions should not be applied to the pediatric population.
Although in the past, there may have been barriers to performing true RCTs that must withhold potentially lifesaving, guideline-recommended, “standard of care” from a control group, a double-blind, RCT (PARAMEDIC-2)\textsuperscript{11} was recently published and largely agrees with the meta-analysis.\textsuperscript{3} After 8,014 patients were enrolled and followed in pragmatic fashion, PARAMEDIC-2 demonstrated that epinephrine increases survival to hospital admission and 30 days, with an NNT of 112 to prevent one death at 30 days. However, the additional proportion that received epinephrine and survived did so with severe neurologic disability (defined as 4 or 5 on the modified Rankin scale). There was no evidence of benefit in the proportion of patients who survived to hospital discharge with a favorable neurologic outcome (OR = 1.18, 95% CI = 0.86–1.61), and the lack of effect persisted at 3 months. This confirms that the use of epinephrine for OHCA is not patient-centered (as defined by patients and the public in preparation for the PARAMEDIC-2 trial) and will cause prolonged suffering in a severely disabled state.

In summary, we chose a color recommendation of “red” for epinephrine administration in OHCA. There is no patient-centered benefit and probable harm due to increased survival with worse long-term neurologic function.

The Brass Tacks series are the results of collaboration between Academic Emergency Medicine journal and the evidence-based medicine website www.TheNNT.com.

References

Video Laryngoscopy vs. Direct Laryngoscopy

Abdullah Bakhsh, MBBS, FAAEM, and Michael Ritchie, MD

**Review color recommendation**  
Green (benefit > harm)

**Summary heading**  
Compared with direct laryngoscopy, video laryngoscopy had Fewer failed intubations and less likelihood of airway trauma and postintubation hoarseness

**Benefits in NNT**  
NNT of 14 for preventing failed intubation  
NNT of 11 for preventing postintubation hoarseness  
NNT of 37 for preventing airway trauma

**Benefits in percentages**  
7% lower rate of failed intubations  
9% lower risk of postintubation hoarseness  
3% lower risk of airway trauma

**Harms in NNT (NNH)**  
No identifiable harms (no difference in the rate of hypoxia, sore throat, or mortality between the two groups)

**Harms in percentages**  
No identifiable harm

**Efficacy endpoints**  
Failed intubations, first successful attempts, airway trauma, postintubation hoarseness, and improved visualization

**Harm endpoints**  
Hypoxia, sore throat, and mortality

**Who was in the studies?**  
7,044 adult patients (16 years and older) who required tracheal intubation across 64 trials

**NARRATIVE**

Tracheal intubation is a critical step in securing the patient’s airway in a variety of emergent and nonemergent settings. Difficulties and complications may arise with this procedure, and alternative laryngoscopes that use video technology have been designed to improve visibility when airway difficulty is predicted or encountered. These devices may be flexible or rigid in design for the purpose of assisting in intubations, especially expected difficult intubations. Video laryngoscopes have been advertised as being able to reduce difficulty, failure, trauma, and other complications compared with direct laryngoscopy. Rigid video laryngoscopy uses a blade to retract the soft tissues and transmits a video image to a screen attached to the end of the handle or to a monitor. This design enables a lighted view of the larynx without direct “line of sight” and is also referred to as indirect laryngoscopy. In the Cochrane review discussed here, video laryngoscopy is compared to direct laryngoscopy in the tracheal intubation of adult patients.

The Cochrane review included randomized control trials of both parallel and crossover design. No simulation or manikin studies were included. Participants were aged 16 years and older who required tracheal intubation electively for scheduled surgery, as well as participants requiring emergent intubation in the emergency department (ED) or the intensive care unit (ICU). The included trials compared the use of a video laryngoscope (VLS) of any model versus direct laryngoscopy with a Macintosh blade.

Nine types of VLS designs were used in the 64 included studies: GlideScope, Pentax AWS, C-MAC (to include DCI laryngoscope), McGrath, X-lite,
C-MAC D-blade, Airtraq, Truview EVO2, and CEL-100. Most studies compared the use of GlideScope, Pentax AWS, C-MAC, and McGrath. Some designs of Airtraq and Truview EVO2 can be used with and without a camera attachment, so only those studies in which used with a camera were included.

The meta-analysis showed statistically significant decrease in number of failed intubations when VLS was used (odds ratio [OR] = 0.35, 95 confidence intervals [CI] = 0.19–0.65, absolute risk difference [ARD] = 7%, number needed to treat [NNT] =14). However, the rate of successful first attempt intubation between the groups was not statistically significant (OR = 1.27, 95% CI = 0.77–2.09).

Subgroup analyses carried out by type of scope revealed no significant difference in the number of failed intubations when the GlideScope, Pentax, or McGrath were compared with the Macintosh blade. The result for failed intubation remained statistically significant in favor of the C-MAC device in this analysis. Another subgroup analysis was performed based on predicted or known difficulty airways. This subgroup analysis revealed that fewer failed intubations occurred when a VLS was used in predicted or known difficult airways (OR = 0.28, 95% CI = 0.15–0.55, ARD = 7%, NNT = 14, n = 830 patients).

The systematic review also demonstrated statistically significant reduction in likelihood of laryngeal/airway traumas (22 trials, OR = 0.68, 95% CI = 0.48–0.96, ARD = 3%, NNT = 37) and fewer incidences of postoperative hoarseness (six trials, OR = 0.57, 95% CI = 0.36–0.88, ARD = 9%, NNT 11) when a VLS device was used.

Additionally, the Cochrane review analyzed intubation difficulty scale (IDS) and airway visualization. IDS scores were recorded in seven studies with 0 representing no difficulty. VLS increased the likelihood of a reported intubation difficulty scores of 0. Airway visualization was evaluated using Cormack-Lehane views. Achieving a Cormack-Lehane grade 1 views was also more likely with VLS.

**CAVEATS**

The use of video laryngoscopy was not adequately explored in the emergency setting. Of 64 studies included in the meta-analysis, only three studies included participants requiring emergency intubation (one in the ICU, one in the ED, and one in an out-of-hospital setting). Therefore, the findings of this systematic review might not be generalizable to emergency settings.

All studies were subject to a high level of bias due to the inability to blind personnel to the type of laryngoscope used with each participant. As a result, the Cochrane authors downgraded the evidence for each outcome by one level for study limitations. Failed intubation, proportion of successful first attempts, and sore throat outcomes were assessed to be moderate-quality evidence, whereas outcomes of hypoxia, serious respiratory complications, and mortality were downgraded to very-low-quality evidence for imprecision. Additionally, a large number of studies with substantial heterogeneity that reported time to tracheal intubation were downgraded to very-low-quality evidence.

Most studies used an experienced anesthetist to perform laryngoscopies. However, it was not always clear from the papers whether anesthetists had equivalent experience with both devices. In light of improved patient centered outcome and relative safety, we assign a color recommendation of green (benefits > harms) to this intervention.

**Reference**

The HEART Score is yet another clinical prediction tool to identify emergency department (ED) chest pain patients who are at low risk for a short-term major cardiac event (MACE). The acronym is clever—HEART: history (H), electrocardiogram (ECG, E), age (A), risk factors (R), and troponin (T). Each of the five variables is assigned 2 points for convincing evidence, 1 point for partial evidence, and 0 points for no convincing evidence. Another older clinical prediction tool, the TIMI (Thrombolysis in Myocardial Infarction) risk score, also includes history, ECG, age, risk factors, and cardiac biomarker findings, but the scoring system is less finely divided; each variable has two values, not three. For example, HEART assigns 2 points for troponin more than three times normal limit, 1 point for troponin one to three times the normal limit, and 0 points for troponin of not more than the normal limit. In contrast, TIMI assigns 1 point for “positive cardiac biomarker” and 0 points otherwise.

In this issue, Fernando et al. present a systematic review of test accuracy studies that evaluated the HEART score as a predictor of MACE. They identified 29 studies (28 distinct cohorts) assessing the sensitivity and specificity of a HEART Score > 3 for predicting MACE in ED patients who present with chest pain. In seven of the cohorts, the authors were able to compare the performance of HEART > 3 to TIMI > 1. Overall, the studies comprised more than 44,000 ED visits. In the studies evaluating both the HEART and the TIMI scores, the HEART score performed better. The authors conclude that “the HEART score has excellent performance for prediction of MACE in ED chest pain patients and should be the primary clinical decision instrument used for the risk stratification of this patient population.” This conclusion may be premature. The HEART Score does appear to discriminate better than the TIMI Score, but the authors present no evidence that it discriminates better than other scores or an experienced clinician.

In the 1980s, we saw ED chest pain decision rules from Tierney et al., Goldman et al., and Pozen et al. The Tierney rule did not even perform better than clinicians in its own derivation set. Sensitivity and specificity were 81 and 86% for the rule and 87 and 78% for clinicians. A comparison of the Goldman and Pozen rules to the admitting decisions of family medicine and primary care internal medicine residents found that neither rule “could reduce unnecessary admissions without seriously increasing the rate of inappropriate discharges.” In 1998, the Pozen rule (also called ACI-TIPI—acute cardiac ischemia time-insensitive predictive instrument) was evaluated in a clinical trial. During the intervention periods, the rule-produced probability of ACI was printed across the top of ECGs for ED patients with chest pain. This did not affect sensitivity for ACI, nor did it increase the specificity of attending physicians or supervised residents, only of unsupervised residents. Now, in addition to HEART, we have the Emergency Department Assessment of Chest Pain Score (EDACS) in original and simplified versions, the Vancouver Chest Pain Rule, and others. Which rule is best, and is it better than the judgment of an experienced clinician?

In 1954, Meehl, famously showed that simple prediction rules tend to perform better than trained professionals making subjective judgements, and we are aware of the report by Backus et al. that reweighting the elements of the HEART score resulted in only moderate improvement. However, we find it difficult
to believe that a score which gives the same weight to age $\geq 65$ years as it does to troponin of more than three times normal discriminates better than a clinician who weights the abnormal troponin more highly. All other factors being the same, a 69-year-old woman with a normal troponin has the same HEART score as that of a 43-year-old man with a troponin of more than three times the normal limit. According to a strict interpretation of the HEART score, if they both have a family history of coronary artery disease and an ECG showing nonspecific repolarization abnormality, they both have a HEART score of 4, which is “positive” using the cutoff of 3. If the HEART score is the “primary clinical decision instrument used for risk stratification,” then both patients should be treated the same way, whether that means sending them both to urgent cardiac catheterization or admitting them both for monitoring and serial troponins.

Despite Meehl and his successors, is it possible that the experienced clinician may be able to preserve the HEART Score’s high sensitivity for MACE while admitting and catheterizing fewer patients? One possible reason is that HEART itself requires clinician judgment. The patient history is rated as “slightly suspicious” (0 points), “moderately suspicious” (1 point), or “highly suspicious” (2 points), where “suspicious elements include: retrosternal pain, pressure, radiation to jaw/left shoulder/arms, duration of 5–15 minutes, exertional, diaphoresis, nausea/vomiting, relief of symptoms with sublingual nitrates.” One can see how two different interviewers might disagree on how many points to assign. They might also disagree on how many points to assign for risk factors, as no explicit definition of each is provided.

Before we accept HEART as the winner of the chest pain decision rule derby, we would like to see it compared to other rules and most especially to the judgment of experienced clinicians who can weight variables differently and consider additional elements in a more nuanced way.

Andrew Worster MD, MSc
Emergency Medicine, McMaster University, Hamilton, Ontario, Canada

Supervising Editor: John H. Burton, MD

References
Bedside End-tidal Carbon Dioxide in Evaluation for Pulmonary Embolism

BACKGROUND

Pulmonary embolism (PE) is associated with approximately 100,000 deaths per year in the United States and the incidence of deep vein thrombosis/pulmonary embolism in the United States is estimated at more than 350,000 cases annually. The diagnosis of pulmonary embolism poses a diagnostic challenge in the emergency department (ED), despite validated decision rules, laboratory tests, and radiographic imaging. While a simple laboratory test would be ideal for diagnosis, the well-known D-dimer test is, at its best, only about 54% specific. The criterion standard pulmonary artery computed tomography angiography (CTA) has numerous downsides including a requirement of clinical stability for transport to radiology, administration of potentially nephrotoxic contrast agents, and radiation exposure to patients, some of whom may be pregnant. Hemodynamically significant PE increases pulmonary dead space and therefore increases the alveolar dead space fraction; however, calculating this involves invasive testing with ABG and a slightly cumbersome calculation. Studies suggest that it is possible to use ETCO₂ alone to screen for PE with the resultant increase in dead space causing the amount of exhaled ETCO₂ to be lower in patients with clinically significant PE as opposed to invasive ABG testing. However, these studies have typically included patients admitted to the hospital, which represent a fraction of those seen in the ED, and likely have a higher prevalence of PE. No study has prospectively evaluated real-time ETCO₂ in ED patients suspected of having PE. We sought to determine if ETCO₂ can rule out hemodynamically significant PE, hypothesizing that no patients with hemodynamically significant PE would have an ETCO₂ greater than 35 mm Hg. Our secondary hypothesis was that the mean ETCO₂ would be significantly lower in patients with PE versus those without PE. Full institutional review board (IRB) approval was obtained through MedStar Washington Hospital Center’s IRB prior to initiation of this study.

METHODS

Prospective observational study of a convenience sample of adults presenting to a tertiary care, academic, urban ED, which receives 87,000 visits annually, between June 6, 2017, and July 28, 2017, Monday through Friday from 8 AM until 10 PM. Study subjects were adults 18 or older in whom the clinical team suspected PE; had D-dimer, pulmonary CTA, or ventilation/perfusion nuclear (V/Q) scan ordered; and were deemed stable to consent. Patients who were pregnant, incarcerated, under arrest, unable to consent for themselves, or not fluent in English were excluded. Potential subjects were approached by one of two trained medical student research assistants who consented and enrolled patients prior to definitive testing.
Patients were placed on a standard model bedside nasal cannula (nasal CO₂ cannula by Spacelabs Medical Inc., 704-0014-00) using General Electric Dash 5000 monitors and their ETCO₂ was measured over three breaths by waveform capnography. The mean of the three measurements was calculated. All measurements, demographics, vital signs, and testing results were recorded using a standardized data collection sheet. After definitive testing, a physician chart reviewer extracted which patients had a PE and, of those, if they were clinically significant. A true positive was defined as CTA or V/Q consistent with PE. A true negative was defined as a negative D-dimer, negative CTPE or very low or low probability VQ scan that the clinical team did not treat. A PE was considered hemodynamically significant if it met a priori criteria based on previous studies. This included troponin elevated from baseline, newly elevated BNP, echocardiogram with evidence of right ventricular dysfunction, tachycardia (heart rate > 100 beats/min), hypoxia (oxygen saturation < 93% on room air), intermediate care or intensive care unit level of admission, or hypotension systolic < 100 mm Hg, diastolic < 60 mm Hg).

Sample size calculations were performed for a beta of 0.20 and alpha error ratio of 0.05, based on our previously calculated PE incident rate of 6% and the previously described standard deviation among ETCO₂ measurement in ED patients of 2.8 mm Hg. Using this method, the ideal sample size was calculated to be 68 patients. Demographic differences between the groups were tested using two-sample t-tests and nonparametric Wilcoxon rank-sum test for continuous measurements. Categorical variables were tested using Fisher’s exact test. Comparison between ETCO₂ in the two groups was made using robust regression models with iteratively reweighted least squares (IRLS) method, adjusting for COPD status, age, and sex. Presence of COPD was considered a possible confounder; a preplanned secondary analysis was performed excluding patients with COPD given the assumption that patients with COPD would have increased physiologic dead space and therefore deranged ETCO₂. Therefore, we used robust regression models to estimate the mean ETCO₂ difference between PE and non-PE groups adjusting for COPD status, age, and sex of the patients. IRLS method was used to estimate the difference between ETCO₂ means. It has a tendency to underestimate the standard errors therefore bootstrapped quantile regression was recommended with data resampling and was used to compare medians. Statistical analyses were conducted in Stata 14 (www.stata.com, StataCorp LP).

### RESULTS

A total of 69 patients with suspected PE were enrolled in the study. The median age of study subjects was 55 and 68% were female. Of the 69 subjects, nine (13%) had PE identified on CT or V/Q scanning; seven (10.1%) were hemodynamically significant. Patient characteristics did not vary significantly between patients with hemodynamically PE and those without with regards to sex, COPD, pulse oximetry, heart rate, or blood pressure. The primary hypothesis was rejected as two of the patients with significant PE had a mean ETCO₂ greater than 35 mm Hg (see Table 1). IRLS method showed that the mean (±SD) ETCO₂ level is significantly lower by 6.6 mm Hg for the PE group, 35.7 (±5.6) versus 29.1 (±7.7; Table 1).
p < 0.001). Bootstrapped quantile regression with 300 repetitions show that the median ETCO₂ was significantly lower by 7.0 units (35.7 mm Hg vs. 29.7 mm Hg, p = 0.04).

**DISCUSSION**

Emergency physicians are on the front line of diagnosis and treatment of PE and many considerations must be made in the workup of patients in whom this diagnosis is suspected. Previous studies have described significantly lower ETCO₂ in patients with massive PE and some have even generated receiver operator curves with specific cutoffs for ETCO₂ in patients with hemodynamically significant PE.¹⁰ Our study found that two patients with significant PE had an ETCO₂ above these previously estimated cutoffs, refuting our primary hypothesis. In this way, we continue to call into question the utility of ETCO₂ alone as a screening modality for PE. However, our secondary hypothesis, despite our small sample size, proved correct in that there appears to be an overall meaningful difference in ETCO₂ in patients with hemodynamically significant PE versus those who do not. Previous studies have indicated a high degree of sensitivity when dead space is combined with D-dimer to diagnose PE; while further studies are warranted, our study suggests that ETCO₂ can aid in the diagnosis of PE. This will be most helpful in patients in whom bedside testing is more appropriate, for example, those with hemodynamic instability making CT scan unobtainable.

**LIMITATIONS**

This is a single-center study performed at an urban tertiary academic center using convenience sampling and accordingly has resultant limitations regarding its external validity. Enrollment hours and the notification process were designed to be as reflective as possible of the ED population given these restraints. The relative incidence of hemodynamically significant PE in our population was 10.1%. While this is consistent with previous studies of PE, the number of patients in this group was much lower than the number of patients without PE who were enrolled, necessitating use of nonparametric statistical analysis. Additionally, we required subjects to consent for themselves, potentially limiting the acuity of our cohort and possibly resulting in the exclusion of the more critically ill subjects likely to have increased dead space and therefore even lower ETCO₂.

Marianne C. Wallis, MD
(Mariannewallis@gmail.com)

Matthew D. Wilson, MD
Emergency Department,
Medstar Washington Hospital Center,
Georgetown University School of Medicine,
Washington, DC

Mihriye Mete, PhD
Epidemiology and Biostatistics,
Medstar Research Institute, Hyattsville, MD

Lydia Koroshetz
Rui Soares
Georgetown University School of Medicine,
Washington, DC

Munish Goyal, MD
Emergency Department,
Medstar Washington Hospital Center,
Georgetown University School of Medicine,
Washington, DC

Supervising Editor: John H. Burton, MD

**References**


Dyspnea is the second leading cause for visit among the 20 million annual U.S. emergency department (ED) visits in patients aged ≥65 years old and is most commonly caused by pneumonia, acute exacerbations of chronic obstructive pulmonary disease (COPD), and heart failure (HF) in older adults. Despite its frequency and substantial associated morbidity and mortality, diagnosing the cause of dyspnea in an older adult poses substantial challenges to the emergency physician. Approximately 20% of older adults with dyspnea have a diagnosis missed in the ED and 31% of adult ED patients with dyspnea are treated for more than one of the common causes of dyspnea (pneumonia, COPD, and HF). ED physician diagnostic uncertainty is common, manifesting either as treatment for multiple conditions (cotreatment) or as documentation of multiple primary diagnoses (codiagnosis). ED diagnostic uncertainty is associated with increased admission rates, longer hospital lengths of stay, and increases in death or rehospitalization within 1 year.

We sought to characterize the population of patients presenting to U.S. EDs with dyspnea. We describe the rates of cotreatment, codiagnosis, and diagnostic uncertainty, stratified by age and focusing on those with pneumonia, COPD, and HF.

The National Hospital Ambulatory Medical Care Survey (NHAMCS) is conducted annually using previously described methods to describe ambulatory emergency care at U.S. hospitals. Data from calendar years 2010 to 2014, the most recent available, were included. This study was exempt from institutional review board review.

Visits by individuals aged ≥18 years old with a complaint of dyspnea were identified using NHAMCS reason for visit codes (1415.0, 1420.0, 1425.0, 1430.0, 1430.1, and 1403.2). To allow comparison with previous literature and exclude patients who had clear etiologies that can also cause dyspnea (e.g., atrial fibrillation with rapid ventricular response), analyses of cotreatment, codiagnosis, and diagnostic uncertainty were limited to the subset of patients with ED diagnoses of pneumonia, COPD, or HF.

Treatments for pneumonia, COPD, and HF were determined based on ED medications administered that were distinct for one of these conditions using the drug categories in NHAMCS and following prior work. Pneumonia treatment included penicillin, cephalosporin, fluoroquinolone, macrolide, vancomycin, tetracycline, aminoglycoside, or carbapenem antibiotics. Treatment for COPD included glucocorticoids, and for HF, loop diuretics, vasodilators, or positive inotropes. ED and hospital diagnoses of pneumonia, COPD, HF, and dyspnea not otherwise specified (NOS) were defined by ICD-9-CM codes.

The primary outcomes were the proportion of ED visits with codiagnosis, cotreatment (treatment or diagnosis for one or more of pneumonia, COPD, and HF) and diagnostic uncertainty (codiagnosis, cotreatment, or a lone diagnosis of dyspnea NOS).
included treatment as well as diagnosis, as ED documentation of diagnoses can be incomplete and may not represent whether the treating physician felt a condition was present. In admitted patients, we describe the proportion of agreement between ED and hospital diagnoses.

Descriptive statistics were calculated, stratified by age ≥ 65 years. Confidence intervals and p-values are not reported as statistical significance would not correlate with clinical significance given large weighted sample sizes. Standard NHAMCS weighting procedures were utilized to obtain nationally representative estimates.

Data management was conducted using SAS 9.4 (SAS Institute, Inc.) and STATA 15 (StataCorp).

From 2010 to 2014, the NHAMCS contained 34,832,195 weighted visits by adults with dyspnea characteristics of the study population are shown in Table 1. Older adults had more comorbidities (including HF) and higher admission rates. Among groups of ICD-9 diagnoses, older adults had greater proportions of heart disease (26% vs. 8%) and pneumonia (15% vs 8%) with less acute respiratory infection (<5% vs. 8%). There were similar proportions of COPD and allied conditions (27 and 31%), but there was a shift from an asthma predominance in younger to COPD in older adults.

Using specific ICD-9 codes, of all ED visits by patients with dyspnea, 10% were diagnosed with pneumonia, 13% with COPD, and 10% with HF with all diagnoses more common in older adults. Across all age groups, dyspnea NOS was a common diagnosis (35%) and was the only diagnosis in 28% of visits. Codiagnosis occurred in 3.5% of the entire study population and cotreatment in 10%, but both were more common in older than younger adults (Table 1).

Almost half (n = 6,021,403, 47%) of older adults were diagnosed with pneumonia, COPD, and/or HF versus 20% (n = 4,324,921) of younger adults. For patients with one of these three diagnoses, 68% of all adult visits diagnosed with pneumonia were treated, 54% for COPD and 64% for HF. Younger adults were less likely to be treated for pneumonia (61% vs. 74%) but more likely for COPD (56% vs. 53%) and HF (67% vs. 44%). In this subset of patients, cotreatment rates (21% across the ages) were similar between older and young adults as was sole diagnosis of dyspnea NOS (24%) but codiagnosis was higher (13% vs. 7.9%). In this subset, diagnostic uncertainty was present in almost half (45%) with similar rates between younger and older adults.

Among hospitalized patients with one of these three diagnoses, hospital diagnosis agreement with ED diagnosis of pneumonia, COPD, and HF was low. Overall, 47% of ED pneumonia diagnoses had a hospital discharge diagnosis of pneumonia. Rates were 56% for COPD and 54% for HF. Diagnosis agreement was similar in younger and older adults.

Diagnostic uncertainty in older ED patients with dyspnea is associated with increased admission rates, longer hospital lengths of stay, and increased death or rehospitalization within 1 year. In addition, ED treatment directed toward an incorrect diagnosis of dyspnea increased mortality from 11% to 25% in older adults with respiratory failure with similar results in the inpatient setting. Overall, older adults with dyspnea (all diagnoses) had greater rates of codiagnosis (6% vs. 2%) and cotreatment (15% vs. 6%) than younger adults.

Older adults are particularly susceptible to adverse effects of these issues due to their much higher rates of diagnosis with pneumonia, COPD, and/or HF than younger adults, particularly given their proven susceptibility and poor outcomes in the setting of uncertainty and inappropriate treatment. Although older and younger adults with these three diagnoses had similar rates of codiagnosis, cotreatment, and diagnostic uncertainty, the much larger proportion of older adults with one of the three diagnoses makes the problem particularly important in this population. Therefore, improving diagnostic accuracy and, as a result, appropriate treatment could have a large impact on morbidity and mortality among the half of older ED patients with dyspnea diagnosed with pneumonia, COPD, and HF.

The high rate of ED codiagnosis for these three conditions indicates a need to improve diagnostic accuracy. This is further supported by the poor agreement between ED and inpatient diagnoses although the reason for the discrepancy is unclear. With more information or further differentiation of the patient’s presentation, the inpatient physicians could be narrowing the diagnosis. Alternatively, the ED diagnosis fields could be inaccurate or incomplete leading to underestimate of codiagnosis.

Emergency department cotreatment in the three conditions was common in patients of all ages. This could demonstrate that when diagnostic uncertainty exists, physicians treat for several etiologies in the ED. Our estimates of cotreatment in older adults (pneumonia 29%, COPD 29%, HF 12%) are almost identical.
for pneumonia, higher for COPD, and much lower for HF when compared with the study by Dharmarajan et al.\textsuperscript{4} study of older inpatients (pneumonia 32%, COPD 19%, HF 38%). This may reflect differences between an ED and inpatient population where some ED patients are discharged, ED stays are short, or fewer treatments are provided in the ED. Patients could also develop additional diagnoses in the inpatient setting. Data are not available for outcome comparison to previous work, but it is known that inappropriate treatment for the etiology of dyspnea increases mortality in both the ED\textsuperscript{2} and inpatient settings.\textsuperscript{4} A prospective study would better study these outcomes in dyspneic ED patients.

Emergency department diagnostic uncertainty in dyspnea is likely multifactorial with patient, physician, and diagnostic factors. In older adults, diagnosis is complicated by atypical presentations,\textsuperscript{10–14} decreased sensation of dyspnea,\textsuperscript{15,16} multiple comorbidities,\textsuperscript{17} failure of clinical prediction rules,\textsuperscript{18–20} and decreased biomarker accuracy.\textsuperscript{21} As a result, diagnosing the etiology of dyspnea is challenging for emergency physicians.\textsuperscript{2,6–8,22} These contributing factors could not be examined. Future studies should use a prospective cohort where data can be collected directly from physicians. To improve diagnostic uncertainty and unnecessary cotreatment of ED patients with dyspnea, we need better rapid tests for diagnoses contributing to dyspnea such as biomarkers and algorithms specific to the older adult population.

Our approach has limitations. First, the NHAMCS data set has recognized limitations.\textsuperscript{23–25} Rather than

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Weighted Demographic Characteristics and Comorbidities of Visits by Adult Patients With Dyspnea by Age in Calendar Years 2010 to 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All ((n = 34,832,195))</td>
</tr>
<tr>
<td>Female</td>
<td>20,035,084 ((58))</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>22,582,963 ((65))</td>
</tr>
<tr>
<td>Black</td>
<td>6,434,534 ((18))</td>
</tr>
<tr>
<td>Other</td>
<td>1,035,532 ((3))</td>
</tr>
<tr>
<td>Missing</td>
<td>4,779,166 ((14))</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
</tr>
<tr>
<td>HF</td>
<td>5,792,147 ((17))</td>
</tr>
<tr>
<td>Diabetes</td>
<td>6,560,732 ((19))</td>
</tr>
<tr>
<td>Renal disease*</td>
<td>1,398,167 ((4))</td>
</tr>
<tr>
<td>Residence</td>
<td></td>
</tr>
<tr>
<td>Private residence</td>
<td>31,315,147 ((90))</td>
</tr>
<tr>
<td>Nursing home</td>
<td>1,624,362 ((5))</td>
</tr>
<tr>
<td>Homeless</td>
<td>91,286 ((0))</td>
</tr>
<tr>
<td>Other</td>
<td>487,906 ((1))</td>
</tr>
<tr>
<td>Missing/unknown</td>
<td>1,313,494 ((4))</td>
</tr>
<tr>
<td>Arrived by ambulance</td>
<td>10,080,009 ((29))</td>
</tr>
<tr>
<td>Admitted to hospital</td>
<td>11,887,096 ((34))</td>
</tr>
<tr>
<td>Diagnosed with</td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>3,429,809 ((10))</td>
</tr>
<tr>
<td>COPD</td>
<td>4,687,940 ((13))</td>
</tr>
<tr>
<td>HF</td>
<td>3,426,534 ((10))</td>
</tr>
<tr>
<td>≥1 of pneumonia, COPD, HF</td>
<td>10,346,324 ((30))</td>
</tr>
<tr>
<td>Dyspnea NOS</td>
<td>12,313,914 ((35))</td>
</tr>
<tr>
<td>Only dyspnea NOS</td>
<td>9,825,894 ((28))</td>
</tr>
<tr>
<td>Codiagnosis</td>
<td>1,111,439 ((3))</td>
</tr>
<tr>
<td>Cotreatment</td>
<td>3,327,737 ((10))</td>
</tr>
</tbody>
</table>

Data presented as \(n\) (%).

COPD = chronic obstructive pulmonary disease; HF = heart failure; NOS = not otherwise specified.

*Variable “EDDIAL” for calendar years 2010 to 2013; “CKD” and “ESRD” for 2014.
admitting hospital diagnosis, NHAMCS only includes hospital discharge diagnoses. Therefore, we may have overestimated the ED and hospital diagnosis disagreement; however, given our high disagreement rate and previous literature, disagreement is likely to exist.  

Older adults frequently present to the ED with dyspnea and suffer from increased rates of codiagnosis and cotreatment than younger adults overall. Half of older adults with dyspnea are diagnosed with pneumonia, chronic obstructive pulmonary disease, or heart failure (vs. 20% of younger adults) with high rates of codiagnosis, cotreatment, and diagnostic uncertainty. As ED care is known to profoundly impact subsequent care, improved ED diagnostic accuracy is necessary to improve morbidity and mortality.

Katherine M. Hunold, MD  
(Katherine.Buck@osumc.edu)  
Jeffrey M. Caterino, MD, MPH  
Department of Emergency Medicine, The Ohio State University, Columbus, OH

Supervising Editor: John H. Burton, MD

References


As a fourth-year emergency medicine resident, I have noticed that we don’t “save lives” as often as I expected. My nonmedical friends tell me, “I can’t imagine going to work and having someone’s life in my hands.” They are talking about what they see on TV, where a patient comes in to the emergency department (ED) coding and walks out skipping, a sore misunderstanding of our success rate with cardiac arrest. In that sense, we don’t save lives all that often at all.

I have grown numb to codes. A code is muscle memory at this point. Compressions, shocks, and medications. No matter what we do, or how fast, most codes end the same. And when the patient survives, it usually feels cruel to have brought someone back who has no hope of a meaningful life. After the code is called, the beeping stops, and we return to the normal buzz of the ED.

Two weeks ago, I walked into my shift and found out that it was the kind of morning where every person from every part of the ED was needed in the resuscitation bay because multiple critical patients had arrived at the same time. I led a code of an 86-year-old man who had been down for over an hour by the time he came in. It was hopeless from the beginning. We tried everything we could for a brief time, but there was no indication that we could save him. I called the code.

I stepped over to the next bay where Mrs. D was groaning and gasping, complaining she could not breathe, when she suddenly collapsed and lost pulses. Despite the fact that the entire team had been coding another patient in the next bay less than a minute before, everyone jumped into action without hesitation or fatigue. We quickly intubated, gave medications, and placed lines. She achieved return of spontaneous circulation several times but continued to lose pulses despite our efforts. On bedside ultrasound, we found a massively dilated right ventricle and discussed as a team the risks and benefits of giving thrombolytics for pulmonary embolism. We pushed the viscous fluid into an IV in her neck. She had return of spontaneous circulation, and this time she stabilized until she was whisked out of the ED by the critical care team.

Sometimes the story stops there. Or, I follow up by reading the notes from the ICU team in the medical record. Usually they write about a guarded prognosis, a discussion about tracheostomy, and if the patient does particularly well, a plan for placement in rehabilitation. But not for Mrs. D. Three days later, I went to her room. She sat in her chair, extubated, with her husband at her side and pictures from her grandchildren on her wall. Her husband thanked me for visiting and said that they would spend the afternoon watching “Crime Watch Daily” together, just like they do at home. Today, two weeks later, we gathered the entire team of doctors, nurses, and technicians who helped to save Mrs. D and went to her room to say goodbye as she packed up to go home. She will get to continue to spend the afternoons with her husband watching “Crime Watch Daily” before picking up her grandson from school. She hugged each of us and with tears in her eyes said, “thank you.”

So when others say they can’t imagine doing what we do, I just smile and shake my head, because they can’t imagine how it feels the times that we do save a life.

Jillian Nickerson, MD, MS
(jillian.nickerson@gmail.com)
Icahn School of Medicine at Mount Sinai,
New York, NY

Supervising Editor: Brian Zink, MD
ERRATUM

Re: SAEM Annual Meeting Abstracts

Abstract 45 (“Multicenter Analysis of Nonsupine Positioning During Endotracheal Intubation”) in the SAEM Annual Meeting Abstracts for 2018 (https://doi.org/10.1111/acem.13424) is missing two coauthors. Amy Kaji, MD, PhD of UCLA Harbor is the fifth author, and Andrea Fantegrossi, MPH of Brigham and Women’s Hospital is the sixth author. Also, Calvin Brown III, MD had his affiliation left off. He is at Brigham and Women’s Hospital.