EDITORS’ PICK
Association Between Pulmonary Embolism and COVID-19 in Emergency Department Patients Undergoing Computed Tomography Pulmonary Angiogram: The PEPCOV International Retrospective Study
Yonathan Freund, Marie Drogrey, Óscar Miró et al. 811

ORIGINAL CONTRIBUTIONS
Diagnostic Performance of Emergency Physician Gestalt for Predicting Acute Appendicitis in Patients Age 5 to 20 Years
Laura E. Simon, Mamata V. Kene, E. Margaret Warton et al. 821

The Effect of Patient Observation on Cranial Computed Tomography Rates in Children With Minor Head Trauma
Sonia Singh, Stephen J. C. Hearps, Meredith L. Borland et al. 832

Centers for Disease Control and Prevention Recommendations for Hepatitis C Testing: The Need to Adopt Universal Screening in an Appalachian Emergency Department
Elena M. Wojcik, Melinda J. Sharon, Stephen M. Davis et al. 844

Shared Decision Making for Syncope in the Emergency Department: A Randomized Controlled Feasibility Trial
Marc A. Probst, Michelle P. Lin, Jeremy J. Sze et al. 853

The Utility of Focused Assessment With Sonography for Trauma Enhanced Physical Examination in Children With Blunt Torso Trauma
Aaron E. Kornblith, Jahanara Graf, Newton Addo et al. 866

Care Transition Decisions After a Fall-related Emergency Department Visit: A Qualitative Study of Patients’ and Caregivers’ Experiences
Cameron J. Gettel, Kelsey Hayes, Renee R. Shield et al. 876

SYSTEMATIC REVIEWS (WITH OR WITHOUT META-ANALYSES)
Can Emergency Physicians Accurately Rule Out a Central Cause of Vertigo Using the HINTS Examination? A Systematic Review and Meta-analysis
Robert Ohle, Renee-Anne Montpellier, Virginie Marchadier et al. 887

Contents continued inside.
RESEARCH METHODS AND STATISTICS

Confidence at 100%: Characteristics of Likelihood Ratio Confidence Intervals in the Emergency Medicine Diagnostics Literature
Ari B. Friedman, Aric W. Berning, Keith A. Marill et al.

RESEARCH LETTERS

Antacid Monotherapy Is More Effective in Relieving Epigastric Pain Than in Combination With Lidocaine: A Randomized Double-blind Clinical Trial
Jaimee Warren, Blake Cooper, Anton Jermakoff et al.

Emergency Department Visits for Serious Diagnoses During the COVID-19 Pandemic
Howard S. Kim, Daniel S. Cruz, Michael J. Conrardy et al.

Percentage of Studies Presented at a Conference Prior to Publication in Emergency Medicine Journals
Michael Gottlieb, Dayle Davenport, Jessen Schiebout et al.

HOT OF THE PRESS

Hot Off the Press: Troponin Testing and Coronary Syndrome in Geriatric Patients With Nonspecific Complaints: Are We Overtesting?
Justin Morgenstern, Corey Heitz, Chris Bond et al.

THE BRASS TACKS: CONCISE REVIEWS OF PUBLISHED EVIDENCE

Accuracy of Physical Examination and Imaging Findings for the Diagnosis of Elevated Intracranial Pressure
Brit Long, Alex Koyfman, Michael Gottlieb

Medical Expulsive Therapy (Alpha Blockers) for Urologic Stone Disease
John C. Conway, Benjamin W. Friedman

COMMENTARY – INVITED

Cultivating a Better Understanding of COVID-19 Amidst a Shifting Landscape
Christopher Kabrhel

It's Time to Normalize Observation for Pediatric Minor Head Trauma
Robert L. Cloutier

CORRESPONDENCE - UNSOLICITED LETTERS TO THE EDITOR

Letter to the Editor: D-dimer and Presumptive Diagnosis of COVID-19
Estabraq Al-Najar, Fahd Alrumaih, Danah Arab et al.

CORRESPONDENCE - RESPONSE TO LETTERS TO THE EDITOR

Response: D-dimer Triage for COVID-19
Chenghong Li, Zhenguo Zhai, Bruce L. Davidson

COMMENTARY – UNSOLICITED

The Antiracist, Propatient Pledge of Emergency Medicine
Ian B. K. Martin, Stephen Hargarten

REFLECTIONS

Battle Scars and Resilience at the Health Care Frontline
Bernard P. Chang, George A. Bonanno

The Son Also Rises
Jameson Mannix

Orange Alert
Abdul Rehman

ERRATUM

Erratum
<table>
<thead>
<tr>
<th>ASSOCIATE EDITORS</th>
<th>DEPARTMENT EDITORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elizabeth Alperrn, MD, MSCE</td>
<td>Academic Social Media Editor</td>
</tr>
<tr>
<td>Children’s Hospital of Philadelphia Philadelphia, Pennsylvania</td>
<td>Michael Gottlieb, MD, RDMS</td>
</tr>
<tr>
<td>Harrison J. Alter, MD, MS</td>
<td>Rush University Medical Center Chicago, Illinois</td>
</tr>
<tr>
<td>Highland Hospital - Alameda Health System Oakland, California</td>
<td>The Biros Section on Research Ethics</td>
</tr>
<tr>
<td>Brigitte M. Baumann, MD, MSCE</td>
<td>Shellee L. Asher, MD</td>
</tr>
<tr>
<td>Cooper University Hospital Camden, New Jersey</td>
<td>Albany Medical College Albany, New York</td>
</tr>
<tr>
<td>Steven B. Bird, MD</td>
<td>Director, Resident Member of the Editorial Board Program</td>
</tr>
<tr>
<td>University of Massachusetts Medical School Worcester, Massachusetts</td>
<td>Mark B. Mycyk, MD</td>
</tr>
<tr>
<td>Anna Marie Chang, MD</td>
<td>Cook County Hospital Chicago, Illinois</td>
</tr>
<tr>
<td>Thomas Jefferson University Hospitals Philadelphia, PA</td>
<td>CME Editor Corey R. Heitz, MD</td>
</tr>
<tr>
<td>Robert Cloutier, MD, MCR</td>
<td>Lewis Gale Medical Center Salem, Virginia</td>
</tr>
<tr>
<td>Oregon Health &amp; Science University Portland, Oregon</td>
<td>Editor of Infographics Kirsty Challen, BSc, MBChB, MRes, PhD</td>
</tr>
<tr>
<td>Wendy C. Coates, MD</td>
<td>Lancashire Teaching Hospitals United Kingdom</td>
</tr>
<tr>
<td>Harbor-UCLA Medical Center Torrance, California</td>
<td>Media Reviews Peter E. Sokolove, MD</td>
</tr>
<tr>
<td>D. Mark Courtney, MD, MSCI</td>
<td>Medicine University of California, San Francisco</td>
</tr>
<tr>
<td>Northwestern University Chicago, Illinois</td>
<td>Reflections and Resident Portfolios Brian Zink, MD</td>
</tr>
<tr>
<td>Deborah B. Diercks, MD, Msc</td>
<td>Brown University Providence, Rhode Island</td>
</tr>
<tr>
<td>The University of Texas Southwestern Medical Center Dallas, Texas</td>
<td></td>
</tr>
<tr>
<td>Richard T. Griffe, MD, MPH</td>
<td></td>
</tr>
<tr>
<td>Washington University St. Louis, Missouri</td>
<td></td>
</tr>
<tr>
<td>Kennon Heard, MD, PhD</td>
<td></td>
</tr>
<tr>
<td>University of Colorado School of Medicine Aurora, Colorado</td>
<td></td>
</tr>
</tbody>
</table>
EDITORIAL OFFICES

Peer Review Coordinator
Taylor Bowen
tbowen@saem.org
Tel 434.218.3366

Journal Manager
Stacey Roseen
sroseen@saem.org
Tel 847.813.9823 x207

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

Copyright and Copying (in any format)
Copyright © 2020 Society for Academic Emergency Medicine. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA, www.copyright.com, provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising or promotional purposes, for republication, for creating new collective works or for resale. Permissions for such reuse can be obtained using the RightsLink “Request Permissions” link on Wiley Online Library. Special requests should be addressed to permissions@wiley.com

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

For submission instructions, subscription and all other information visit http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1553-2712
Association Between Pulmonary Embolism and COVID-19 in Emergency Department Patients Undergoing Computed Tomography Pulmonary Angiogram: The PEPCOV International Retrospective Study

Yonathan Freund, MD, PhD, Marie Drogrey, MD, Óscar Miró, MD, PhD, Alessio Marra, MD, Anne-Laure Féraud-PIerssens, MD, PhD, Andrea Penaloza, PD, PhD, Barbara A. Lara Hernandez, MD, Sebastien Beaune, MD, PhD, Judith Gorlicki, MD, Prabakar Vaittinada Ayar, MD, Jennifer Truchot, MD, PhD, Barbara Pena, MD, Alfons Aguirre, MD, PhD, Florent Fémy, MD, Nicolas Javaud, MD, PhD, Anthony Chauvin, Md, PhD, Tahar Chouihed, MD, PhD, Emmanuel Montassier, PD, PhD, Pierre-Géraud Claret, MD, PhD, Céline Occelli, MD, Mélanie Roussel, MD, Fabien Brigant, MD, Sami Ellouze, MD, PhD, Pierrick Le Borgne, MD, Said Laribi, MD, PhD, Tabassome Simon, MD, PhD, Olivier Lucidarme, MD, PhD, Marine Cachanado, MsC, and Ben Bloom, MD, PhD

the IMPROVING EMERGENCY CARE FHU Collaborators

A related article appears on page 925.

From the Sorbonne Université; the Emergency Department, Hôpital Pitié-Salpêtrière, Assistance Publique–Hôpitaux de Paris (AP-HP), APHP.SU, Paris, France; the Emergency Department, Hospital Clinic, Barcelona, Catalonia, Spain; the Emergency Department, Centro EAS–Emergenza Alta Specializzazione, ASST Papa Giovanni XXIII Hospital, Bergamo, Italy; the Charles Lemoyne–Saguenay Lac Saint-Jean Research Center on Health Innovations (CR CSIS), Sherbrooke University, Longueuil, Québec, Canada; the Emergency Department, European Georges Pompidou hospital, APHP, Paris, France; the Service des Urgences, Cliniques Universitaires Saint-Luc, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; the Emergency Medicine Section, Pontificia Universidad Católica de Chile, Santiago, Chile; the Emergency Department, Hôpital Ambroise-Paré, APHP, Boulogne, INSERM UMR 1144, Université Paris Centre, Paris; the Emergency Department, SAMU 93, Avicenne University Hospital, APHP, P.H.U.P.S.S.D; the Emergency Department, Hospital General Universitario de Alicante, Alicante, Spain; the Emergency Department, Hospital del Mar, Barcelona, Catalonia, Spain; the Emergency Department, Georges Pompidou European Hospital, APHP, Université de Paris, Paris.

Received May 22, 2020; revision received June 18, 2020; accepted June 18, 2020.

No funding for this research was received. The sponsor was Assistance Publique–Hôpitaux de Paris. The sponsor had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript or the decision to submit for publication.

The authors declare have no potential conflicts of interest to declare.

The authors had no role in the conduction and analysis of the study.

The authors declare they have no conflict of interest with this study. There were no funding for this study.

See appendix 1 for IMPROVING EMERGENCY CARE FHU Collaborators group

Supervising Editor: Jeffrey A. Kline, MD.

Address for correspondence and reprints: Yonathan Freund, MD, PhD; e-mail: yonathan.freund@aphp.fr.


doi: 10.1111/acem.14096

ISSN 1553-2712
ABSTRACT

Background: There have been reports of procoagulant activity in patients with COVID-19. Whether there is an association between pulmonary embolism (PE) and COVID-19 in the emergency department (ED) is unknown. The aim of this study was to assess whether COVID-19 is associated with PE in ED patients who underwent a computed tomographic pulmonary angiogram (CTPA).

Methods: A retrospective study in 26 EDs from six countries. ED patients in whom a CTPA was performed for suspected PE during a 2-month period covering the pandemic peak. The primary endpoint was the occurrence of a PE on CTPA. COVID-19 was diagnosed in the ED either on CT or reverse transcriptase–polymerase chain reaction. A multivariable binary logistic regression was built to adjust with other variables known to be associated with PE. A sensitivity analysis was performed in patients included during the pandemic period.

Results: A total of 3,358 patients were included, of whom 105 were excluded because COVID-19 status was unknown, leaving 3,253 for analysis. Among them, 974 (30%) were diagnosed with COVID-19. Mean (±SD) age was 61 (±19) years and 52% were women. A PE was diagnosed on CTPA in 500 patients (15%). The risk of PE was similar between COVID-19 patients and others (15% in both groups). In the multivariable binary logistic regression model, COVID-19 was not associated with higher risk of PE (adjusted odds ratio = 0.98, 95% confidence interval = 0.76 to 1.26). There was no association when limited to patients in the pandemic period.

Conclusion: In ED patients who underwent CTPA for suspected PE, COVID-19 was not associated with an increased probability of PE diagnosis. These results were also valid when limited to the pandemic period. However, these results may not apply to patients with suspected COVID-19 in general.

COVID-19 is currently one of the greatest worldwide threats to public health and a challenge for researchers and physicians. The reported mortality ranges from 0.1% to 8% depending on the disease severity. COVID-19 viral pneumonia is associated with hypoxia, a hyperinflammatory state, and coagulopathy. High rates of elevated D-dimers have also been reported in case series, which may be associated with worse outcomes. Rapid identification of patients with COVID-19 who are at risk of pulmonary embolism (PE) may improve prognosis by early initiation of anticoagulant therapy.

In the emergency department (ED), the diagnostic strategy for PE is well established. Several clinical decision rules (CDRs) have been validated to safely limit the use of radiative imaging studies (especially computed tomography pulmonary angiogram [CTPA], considered as the criterion standard). These CDRs are based on a Bayesian approach that combines pretest probability (i.e., suspected PE prevalence in the studied population estimated by a score or physician gestalt) with the D-dimer result to stratify risk and guide indication for CTPA. Application of the Pulmonary Embolism Rule Out Criteria (PERC) may safely exclude PE in patients with low clinical probability. Other CDR such as the Wells or revised Geneva scores (RGS) are also recommended, and the recent YEARS protocol may allow the D-dimer threshold to be raised while still safely limiting the use of CTPA. All these rules were validated before the COVID-19 pandemic, and their safety is based on estimated PE prevalence in the studied population. Since COVID-19 is reportedly associated with an increased risk of thromboembolic events, and the validity of these CDR is unknown, it is possible that during this pandemic, conventional ED diagnostic strategies for PE are unsafe. Furthermore, since decision rules are derived in specified populations, a rule that applies to a stratified subpopulation of ED patients may also not extrapolate to the general population. Conversely, even if COVID-19 causes an increase in the risk of venous thromboembolism in the general population, this may not trans-
late into a higher risk among ED patients with suspicion of PE. The aim of our study was to assess if COVID-19 was associated with PE in ED patients who underwent a CTPA for suspected PE and to assess whether RGS-based diagnostic strategy was safe in this period.

METHODS

Design
This was a multicenter retrospective study in 26 centers from France, Spain, Belgium, Italy, Chile, and Canada. The study was approved by the steering committee of Assistance Publique–Hôpitaux de Paris. Local ethics committees in all participating countries approved the study. Due to its retrospective nature on deidentified data, informed consent was waived in all participating countries.

Patients and Data
All patients who underwent a CTPA for suspected PE during the study period were included. The COVID-19 peaks of ED visits ranged from late approximately March to mid-April 2020. Since we wanted to include patients from both prior to and during the COVID-19 pandemic, the overall study period was comprised between February 1 to April 10, 2020. The study period slightly differed in the different centers as the COVID-19 pandemic peak occurred at different periods in the participating centers. In each country, we considered that the COVID-19 epidemic started when more than 100 patients were diagnosed positive, respectively, on March 4, 6, 15, 15, 18, and 24 in Italy, Spain, France, Chile, Belgium, and Quebec.

All CTPA procedures performed for ED patients during this period were collected, and data from the ED visits were collected. Patients with no COVID-19 status or inconclusive CTPA for the diagnosis of PE or in whom CTPA was performed for a reason other than “suspicion of PE” were excluded.

Study data were obtained from the electronic health system of each center by local study investigators. Baseline characteristics, risk factors for PE and items from conventional CDR were collected.

Objectives and Endpoint
The primary objective of this study was to assess whether COVID-19 was independently associated with PE in ED patients that underwent a CTPA. Secondary objectives included the validation of conventional diagnostic strategies using a combination of the revised Geneva score (RGS) and D-dimer in this period (Table 1) and to assess the diagnostic performance of D-dimer among COVID-19 patients. The primary endpoint was the presence of a PE on CTPA. Each CTPA was analyzed and interpreted by senior radiologist.

In cases of inconclusive CTPAs, the patient was excluded from analysis. COVID-19 status was defined with the following rule:

- Negative during the prepandemic period (defined as first 100 diagnosed cases in the country).
- Positive if reverse transcriptase–polymerase chain reaction (RT-PCR) was positive.
- Positive if lung CT showed evidence of a COVID-19 lesion, i.e., ground-glass opacities or crazy paving.\(^{11}\)

In cases where RT-PCR was not performed and CT was indeterminate or with nonspecific abnormalities, the patient was excluded because his/her COVID-19 status could not be determined. In the absence of positive RT-PCR, a patient with CT interpreted as “unlikely COVID-19” was considered as non–COVID-19. In these patients, severity of CT lesions was not reported. Severity of COVID-19 lesions was graded as recommended by French Society of Radiology and other reports as moderate (extent < 25), extended (25%–<50%), severe (50%–75%), and critical (>75%).\(^{12,13}\)

A simplified RGS was calculated using data collected during the ED visit (Table 1). Due to the retrospective nature of collected data, we merged the two items “unilateral lower-limb pain” and “pain on lower-limb deep venous palpation and unilateral edema” as “clinical signs of deep venous thrombosis,” with a weighting of 2 points. D-dimer, C-reactive protein, and leukocytes were also collected. CT findings were classified according to their probability of COVID-19 (likely, compatible, unlikely) and their severity (mild, moderate, severe, critical).\(^{14}\)

Data Analysis
Baseline characteristics were expressed as number (%) for categorical variables and mean (standard deviation [SD]) or median (interquartile range [IQR]) for continuous variables, depending on their distribution. Separate bivariate analyses were performed to determine the unadjusted association between PE and the following known risk factors: age, sex, heart rate, previous thromboembolic event, hemoptysis, clinical signs of deep venous thrombosis, estrogen intake, and surgery/immobilization within 4 weeks. In addition, severity of
COVID-19 symptom, period (before/after pandemic onset) and country of admission were studied too. A direct multivariable binary logistic regression model was built taking into account all known risk factors: age, sex, heart rate, previous thromboembolic event, hemoptysis, clinical signs of deep venous thrombosis, estrogen intake, and surgery/immobilization within 4 weeks (p-value inferior or equal to 0.2 in univariate analysis or forced into the model) and in addition center of admission. Multicollinearity was investigated using in first correlation matrix and in second tolerance and variation inflation parameters. Due to violation of linearity in the logic for age as a continuous variable, it was categorized in the model using quartile values. Receiver operating characteristic curve and Youden index were used to calculate the optimal cutoff of D-dimers for PE diagnosis. Several goodness-of-fit tests were performed to determine model performance (Hosmer-Lemeshow test, standard Pearson test, Osius test, McCullagh test, informative matrix [IM] test, and unweighted sum-of-squares test).

To validate the safety of a RGS-based diagnostic strategy, we compared the rate of PE diagnosed in patients with low to intermediate RGS and D-dimer below the age-adjusted threshold (i.e., 500 µg/mL under 50 years and age × 10 over 50 years) in patients with and without COVID-19.15,16

Since physicians’ threshold for ordering a CTPA may have changed during the pandemic period, we ran a sensitivity analysis limited to patients in this period. p-values < 0.05 were considered significant. SAS V.9.4 software (SAS Institute Inc., Cary, NC) was used for statistical analyses.

**RESULTS**

During the study period, 3,358 patients were included in the 26 participating EDs, 52% of whom were included during the pandemic period. Among them, COVID-19 status could not be determined in 105 patients (Figure 1). A total of 3,253 were analyzed. The mean (±SD) age was 61 (±19) years and 1,695 (52%) were women. Baseline characteristics are reported in Table 2. There was no difference in patient characteristic between the two periods pre/post pandemic (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.14096/full). A total of 974 patients were diagnosed with COVID-19; 530 (54%) were diagnosed with both positive RT-PCR and CT, 370 (38%) on CT only, and 74 (8%) only with RT-PCR.

Pulmonary embolism was diagnosed in 500 patients (15%); 148 patients (15%) had a COVID-19 diagnosis and 352 (15%) did not (difference = 0.3%, 95% confidence interval [CI] –3% to 3%; unadjusted odds ratio [OR] = 0.98, 95% CI = 0.78 to 1.19) (Table 3). Among the 500 patients with PE, 59 (15%) were isolated subsegmental. A RGS of 5 or more (considered as high risk) was associated with a higher risk of PE (23% vs 14%, difference = 9%, 95% CI = 1.5% to 15.7%).

In the multivariable logistic regression analysis, there was no association between COVID-19 and PE (adjusted OR = 0.98, 95% CI = 0.761 to 1.26) (Table 4). Hosmer-Lemeshow p-value was 0.40. The other goodness-of-fit tests had a p-value > 0.05 except for IM test (p = 0.01) There was no period effect between pre- and postpandemic onset date (OR = 1.02, 95% CI = 0.83 to 1.24, p = 0.72). The sensitivity analysis limited to the pandemic period reported similar results: COVID-19 had an adjusted OR of 1.10 (95% CI = 0.79 to 1.52) for the risk of PE (Data Supplement S1, Table S2). Another sensitivity analysis after excluding the 59 isolated subsegmental PEs reported that COVID-19 had an adjusted OR of 0.97 (95% CI = 0.74 to 1.26) for the risk of PEs. The area under the receiving operator characteristics curve of D-dimer for the diagnosis of PE was 0.79 (95% CI = 0.76 to 0.81) for the general population and 0.81 (95% CI = 0.77 to 0.85) in COVID-19 patients (Figure 2).

A total of 207 patients had a nonhigh clinical probability and D-dimer below the age-adjusted threshold, among whom four had a PE: one in a COVID-19 patient and three in non–COVID-19 patients. The
percentages of false negatives for the RGS with age-adjusted threshold were, respectively, 1.4 and 2.2%.

**DISCUSSION**

In this multicenter retrospective study, we report that COVID-19 is not associated with increased risk of PE diagnosis among ED patients who underwent a CTPA. The risk of PE was of 15% in both groups, and the adjusted OR of COVID-19 for PE was 1.01 (95% CI = 0.81 to 1.27).

This result is in contrast to recent reports and case series that highlighted a higher risk of thromboembolism in COVID-19 patients. In our study, we focused on ED patients, who are, by definition, at the beginning of the part of their hospitalized stage of disease. Recent reports suggested an increased risk of thromboembolic events in admitted patients, both to wards and to the intensive care unit. These patients may be at higher risk because they were identified both at a later stage of the disease and also after a potential period of immobilization. Furthermore, these previous reports were not comparative, and therefore no definitive conclusion of increased risk could be made. It is likely that COVID-19 is associated with higher risk of PE in the general population, but our reports suggest that this is not the case among ED patients with suspicion of PE. This is in line with studies including pregnant women, who in the general population are at increased risk of thromboembolic events. However, in ED patients with suspected PE, pregnancy was not reported to be associated with higher risk of PE.

We included patients based on whether a CTPA was performed in the ED. This is because the diagnostic strategy to rule out PE in the ED is based on a Bayesian approach, where the workup (especially regarding the order of a CTPA) depends on the pretest probability, which is dependent on PE prevalence in the studied population. We conducted this study to assess if COVID-19 was associated with PE, because if confirmed, it could have led to a change in the diagnostic strategy. Our results suggest that the current strategy may be safe during COVID-19 pandemic because the pretest probability of PE does not seem to depend on the COVID-19 status. Furthermore, only one “low-risk” (nonhigh RGS and D-dimer below age-adjusted threshold) patient was diagnosed with a subsegmental PE among COVID-19 patients. However, since we only included patients that had a CTPA performed, it is possible some patients with a nonhigh RGS and low D-dimer had a PE missed in the ED because a CTPA was not ordered.

**LIMITATIONS**

This study has several limitations. First, patients were included only if a CTPA was performed in the ED. This means that all patients that had a suspicion of PE and a negative D-dimer and a nonhigh clinical
probability of PE were excluded. This inclusion bias limits our ability to conclude whether or not these results can apply to the whole ED population with suspicion of PE and moreover to the general population. Among included patients, it is possible that some underwent a CTPA for an alternate diagnosis such as aortic dissection. After screening all CTPA performed during the study period, the local investigator sought in the patient’s file whether the CTPA may have been performed outside a suspicion of PE and subsequently excluded him/her. However, we may have missed some CTPAs with no clear listed indication and have a subsequent inclusion bias. Furthermore, it is possible that during the COVID-19 pandemic, emergency physicians may have had a lower threshold for ordering a CTPA especially because COVID-19 has been reported to be associated to higher risk of PE and also because a lung CT was often performed to diagnose COVID-19. However, the patient’s baseline characteristic was similar between the two periods (Data Supplement S1, Table S1), and we found no period effect in the analysis (Data Supplement S1, Table S2). Furthermore, the sensitivity analysis restricted to patients included in the pandemic period reported similar results, with no association between COVID-19 and risk of PE (Data Supplement S1, Table S3). However, a bias may still exist since the potential risk of COVID-19–induced coagulopathy was not described at the beginning of the pandemic period but after a few weeks. This bias is limited because patients were included until April 10, before

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>PE (n = 500)</th>
<th>No PE (n = 2,753)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3,253</td>
<td>234 (45)</td>
<td>1,471 (53)</td>
</tr>
<tr>
<td>ED in France</td>
<td>3,253</td>
<td>358 (72)</td>
<td>2,276 (83)</td>
</tr>
<tr>
<td>Postpandemic period</td>
<td>3,253</td>
<td>252 (50)</td>
<td>1,642 (51)</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic respiratory insufficiency</td>
<td>3,248</td>
<td>42 (8)</td>
<td>276 (10)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>3,247</td>
<td>207 (42)</td>
<td>1,087 (40)</td>
</tr>
<tr>
<td>Chronic heart failure</td>
<td>3,248</td>
<td>61 (12)</td>
<td>476 (17)</td>
</tr>
<tr>
<td>Chronic kidney failure</td>
<td>3,245</td>
<td>26 (5)</td>
<td>133 (5)</td>
</tr>
<tr>
<td>ED presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chest pain</td>
<td>3,245</td>
<td>179 (36)</td>
<td>1,075 (39)</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>3,249</td>
<td>349 (70)</td>
<td>1,731 (63)</td>
</tr>
<tr>
<td>Syncope</td>
<td>3,245</td>
<td>60 (12)</td>
<td>384 (14)</td>
</tr>
<tr>
<td>Delay from onset to ED visit</td>
<td>2,883</td>
<td>3 (1–7)</td>
<td>3 (1–7)</td>
</tr>
<tr>
<td>Heart rate (beats/min)</td>
<td>2,893</td>
<td>97 (±20)</td>
<td>93 (±20)</td>
</tr>
<tr>
<td>Respiratory rate (breaths/min)</td>
<td>2,297</td>
<td>23 (±7)</td>
<td>23 (±7)</td>
</tr>
<tr>
<td>SpO₂ (%)</td>
<td>3,104</td>
<td>96 (93–98)</td>
<td>97 (94–99)</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>3,191</td>
<td>134 (±24)</td>
<td>137 (±24)</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>3,123</td>
<td>36.8 (±0.8)</td>
<td>37 (±0.9)</td>
</tr>
<tr>
<td>Risk factors for PE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estrogen use</td>
<td>3,236</td>
<td>12 (2)</td>
<td>78 (3)</td>
</tr>
<tr>
<td>Clinical signs of DVT</td>
<td>3,247</td>
<td>101 (11)</td>
<td>248 (9)</td>
</tr>
<tr>
<td>Surgery or trauma requiring immobilization within 1 month</td>
<td>3,246</td>
<td>53 (11)</td>
<td>168 (6)</td>
</tr>
<tr>
<td>Past history of PE or DVT</td>
<td>3,245</td>
<td>106 (21)</td>
<td>279 (10)</td>
</tr>
<tr>
<td>Hemoptysis</td>
<td>3,245</td>
<td>15 (3)</td>
<td>104 (4)</td>
</tr>
<tr>
<td>Active malignancy</td>
<td>3,246</td>
<td>68 (14)</td>
<td>374 (14)</td>
</tr>
<tr>
<td>Laboratory results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-dimer (ng/mL)</td>
<td>2,495</td>
<td>4,270 (1,730–10,000)</td>
<td>1,181 (762–2,105)</td>
</tr>
<tr>
<td>Leukocytes</td>
<td>3,106</td>
<td>10.6 (±4.6)</td>
<td>9.1 (±5.1)</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>2,758</td>
<td>23 (7–83)</td>
<td>14 (4–65)</td>
</tr>
</tbody>
</table>

Data are reported as mean (±SD), n (%), or median (IQR). DVT = deep venous thrombosis; IQR = Interquartile range; PE = pulmonary embolism.
physicians were aware of suspected COVID-19–induced coagulopathy. A sensitivity analysis with time forced in the model as a categorical variables (weeks of inclusion) reports similar results with no effect of time (Data Supplement S1, Table S4).

Second, defining the presence of COVID-19 in the ED may be difficult. We excluded 105 patients in whom COVID-19 status could not have been determined, but it is possible that other patients were wrongly classified. The reported sensitivity for the diagnosis of COVID-19 of RT-PCR ranges from 71 to 98% and 93 to 97% for lung CT.\(^2\),\(^2\) To mitigate this, in this study, patients were considered to have COVID-19 if one or the other of the tests was positive, which limited the risk of false negatives. In 38% of cases, the diagnosis of COVID-19 was only adjudicated on CT. This is in part caused by the limited availability of RT-PCR testing in France, and the longer turnaround time for RT-PCR results compared to CT. In these patients, the suboptimal specificity of CT could have led to false positives, and radiologists exhibited moderate performances in differentiating COVID-19 pneumonia from other viral pneumonia on lung CT.\(^2\),\(^2\)\(^3\) This limit is inherent to our design and represents a classification bias. However, this can be seen as a challenge faced in the day-to-day clinical practice of emergency medicine and patients with a suspected COVID-19. In our study, sensitivity of RT-PCR was 84% and false-negative rate was 23%, which is consistent with what has been reported in the literature.\(^2\)\(^0\) However, we cannot exclude that some COVID-19 patients may have had both false-negative PCR and false-negative CT.

In addition, we found a center effect and investigated this. It transpired that French EDs was a protective factor for PE (adjusted OR = 0.61, 95% CI = 0.48 to 0.78), which suggests different practice patterns across countries. This may reflect the fact that heterogenous data sources were combined, especially in light of the fact that French EDs dominated the sample size. The multivariable model adjusted the results for this association, but whether this could affect the external validity of our results is unknown.

Last, as a retrospective study, although the case record form was standardized, there was no monitoring of data collection methods in the six countries and 26 sites. This was mitigated by making the data required as pragmatic and minimal as possible to satisfy the primary objective.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>COVID-19 Status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE (n = 500)</td>
</tr>
<tr>
<td>Signs of COVID-19</td>
<td></td>
</tr>
<tr>
<td>Very likely</td>
<td>82 (16)</td>
</tr>
<tr>
<td>Compatible</td>
<td>54 (11)</td>
</tr>
<tr>
<td>Unlikely</td>
<td>364 (73)</td>
</tr>
<tr>
<td>Extent of lesions n</td>
<td>127</td>
</tr>
<tr>
<td>Moderate</td>
<td>35 (28)</td>
</tr>
<tr>
<td>Extended</td>
<td>47 (37)</td>
</tr>
<tr>
<td>Severe</td>
<td>40 (31)</td>
</tr>
<tr>
<td>Critical</td>
<td>5 (4)</td>
</tr>
<tr>
<td>RT-PCR COVID-19</td>
<td></td>
</tr>
<tr>
<td>Performed</td>
<td>202 (40)</td>
</tr>
<tr>
<td>Positive</td>
<td>80 (16)</td>
</tr>
<tr>
<td>Confirmed COVID-19</td>
<td>148 (30)</td>
</tr>
</tbody>
</table>

Data are reported as n (%).

Extent of lesions: moderate < 25%, extended 25% to 50%, severe 50% to 75%, and critical > 75%.

PE = pulmonary embolism; RT-PCR = reverse transcriptase-polymerase chain reaction.

Figure 2. Receiving operator characteristic curves of D-dimer for diagnosis of PE in the emergency department. (A) Whole population: area under the curve = 0.79, 95% CI = 0.76 to 0.81. (B) COVID-19 patients: area under the curve = 0.81, 95% CI = 0.77 to 0.85.
CONCLUSION

In ED patients that had computed tomography pulmonary angiogram performed for suspected pulmonary embolism, COVID-19 was not associated with a higher risk of pulmonary embolism. These results suggest that conventional diagnostic strategies for pulmonary embolism in ED patients with suspected COVID-19 are safe.

References

by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. Radiol Cardiothorac Imag 2020;2:e200152.


Appendix 1.

IMPROVING EMERGENCY CARE FHU Collaborators group: Hélène Goulet, MD, Hôpital Tenon, APHP, Paris, France; Benoît Ghaye, MD, PhD, Service de Radiologie, Cliniques Universitaires Saint-Luc, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; Leslie Sgorlon, MD, Department of Anaesthesia Resuscitation Pain Emergency Medicine, Nîmes University Hospital, Nîmes, France; Julien Frandon, MD, Department of Radiology, Medical Imaging Group Nîmes, Nîmes University Hospital, Nîmes, France; Adrien Bassand, MD, Emergency Department, University Hospital of Nancy, Nancy, France; Romain Gillet, MD, Guilloz Imaging Department, Central Hospital, University Hospital of Nancy, Nancy, France; Pablo Aguilera, MD, Emergency Section Chair, Red de Salud UC Christus Health, Pontificia Universidad Católica de Chile, Santiago, Chile; Pedro Antonio Gondim Teixeira, MD, PhD, Guilloz Imaging Department, Central Hospital, University Hospital of Nancy, Nancy, France; Anne Laure Gautier, MD, Departments of Radiology, Georges Pompidou European Hospital, Assistance Publique–Hôpitaux de Paris; Jean-Marc Chauny, MD, Emergency Department, Research Centre, Hôpital du Sacré-Cœur de Montréal (CIUSS du Nord de l’île-de-Montréal), Montréal, Québec, Canada; Pascal Bilbault, MD, PhD, Emergency Department, Hôpitaux Universitaires de Strasbourg, and INSERM UMR 1260, Regenerative NanoMedicine (RNM), Fédération de Médecine Translationnelle (FMTS), University of Strasbourg, Strasbourg, France; Carlos Cardozo, MD, Emergency Department, Hospital Clínico, Barcelona, Catalonia, Spain; Pere Llorens, MD, PhD, Emergency Department, Hospital General Universitario de Alicante, Alicante, Spain; Lorenzo Della Bella, MD, Federico Zanardi, MD, Carlo Preti, MD, Cosentini Roberto, MD, Andrea Della Bella, MD, Carlo Pretti, MD, Emergency Department, Centro EAS–Emergenza Alte Specializzazioni, ASST Papa Giovanni XXIII Hospital, Bergamo, Italy; Samia Boussoar, MD, APHP–Sorbonne Universités, Pitié Salpêtrière Hospital, Radiology Department, Paris, France; Victoria Doncicu, MD, APHP–Sorbonne Universités, Pitié Salpêtrière Hospital, Radiology Department, Paris, France; Alban Redheuil, MD, PhD, APHP–Sorbonne Universités, Pitié Salpêtrière Hospital, Radiology Department and UPMC Université Paris 06, CNRS, INSERM, Laboratoire d’Imagerie Biomédicale, Paris, France; Léa Legay, MD, Emergency Department, Hôpital Saint-Louis, Assistance Publique–Hôpitaux de Paris, APHP, Paris, France; Olivier Peyrony, MD, PhD, Hôpital Saint-Louis, Assistance Publique–Hôpitaux de Paris, APHP, Paris, France; Ernesto Escobedo Carvajal, Emergency Medicine Section. Pontificia Universidad Católica de Chile, Santiago, Chile; Arthur David, Central Department of Radiology and Medical Imaging, CHU Nantes, Nantes, France; Bousson Valérie Service d’Imagerie Ostéo-Articulaire, Viscérale et Vasculaire, Hôpital Lariboisière, APHP.Nord, Paris, and Laboratoire de Biologie, Bioingénierie et Biomimétique Ostéo-Articulaires (B3OA), UMR CNRS 7052, INSERM U1271,
Université de Paris, Paris, France; Laurent Brunereau, MD, PHD, Service Imagerie, CHU de Tours; Sophie Dabin-Pouchard, IRC, Département de médecine d’Urgence, CHU de Tours, Tours, France; and Isabel Cirera, MD, PhD, Emergency Department, Hospital del Mar, Barcelona, Spain.

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.14096/full

Data Supplement S1. Supplemental material.
ABSTRACT

Objectives: Pediatric appendicitis remains a challenging diagnosis in the emergency department (ED). Available risk prediction algorithms may contribute to excessive ED imaging studies. Incorporation of physician gestalt assessment could help refine predictive tools and improve diagnostic imaging decisions.

Methods: This study was a subanalysis of a parent study that prospectively enrolled patients ages 5 to 20.9 years with a chief complaint of abdominal pain presenting to 11 community EDs within an integrated delivery system between October 1, 2016, and September 30, 2018. Prior to diagnostic imaging, attending emergency physicians enrolled patients with ≤5 days of right-sided or diffuse abdominal pain using a Web-based application embedded in the electronic health record. Predicted risk (gestalt) of acute appendicitis was prospectively entered using a sliding scale from 1% to 100%. As a planned secondary analysis, we assessed the performance of gestalt via c-statistics of receiver operating characteristic (ROC) curves; tested associations between gestalt performance and patient, physician, and facility characteristics; and examined clinical characteristics affecting gestalt estimates.

Results: Of 3,426 patients, 334 (9.8%) had confirmed appendicitis. Physician gestalt had excellent ROC curve characteristics (c-statistic = 0.83, 95% confidence interval = 0.81 to 0.85), performing particularly well in the low-risk strata (appendicitis rate = 1.1% in gestalt 1%–10% range, negative predictive value of 98.9% for appendicitis...
Physicians with ≥5 years since medical school graduation demonstrated improved gestalt performance over those with less experience (p = 0.007). All clinical characteristics tested, except pain <24 hours, were significantly associated with physician gestalt value (p < 0.05).

Conclusion: Physician gestalt for acute appendicitis diagnosis performed well, especially in low-risk patients and when employed by experienced physicians.

Pediatric abdominal pain with concern for acute appendicitis is a common clinical scenario in the emergency department (ED). Acute appendicitis symptoms overlap with other conditions, making the assessment challenging. Clinical prediction risk scores, such as the Pediatric Appendicitis Score (PAS), can aid in diagnosis. However, some scores assign a large proportion of patients to intermediate-risk categories, leading to the potential overutilization of computed tomography (CT) and ultrasound (US) imaging.

Physician gestalt can be defined as a physician’s implicit probability estimation based on a synthesis of provider experience and clinical perception in the absence of definitive diagnostic testing. Assessments of physician gestalt across various medical conditions, such as pulmonary embolism and acute coronary syndrome, demonstrate variable accuracy; for some conditions, such as pulmonary embolism, studies suggest that gestalt can perform similarly to clinical prediction rules. However, physician gestalt of diagnostic probability is rarely incorporated into risk-stratification tools. Additionally, although it has been shown to perform well in many scenarios, physicians do not always behave consistently with their reported gestalt. This may be due to concern for adverse consequences of a missed diagnosis and the limited number of validation assessments of physician gestalt performance.

To our knowledge, only a handful of studies have described the diagnostic performance of physician gestalt for acute appendicitis, and only one has assessed emergency physician gestalt exclusively in a pediatric population. This four-center Australian study in academic EDs (two tertiary pediatric centers and two mixed) reported reasonable diagnostic accuracy for emergency physicians (70%–82%) that did not vary with experience. Our investigation had a similar objective but in a U.S. community ED population with a secondary goal of providing data that could inform clinicians when gestalt is a reliable diagnostic tool and when to utilize other clinical decision support (CDS) tools, imaging, or consultants.

In this secondary analysis to a larger prospective cohort study, we sought to 1) characterize the diagnostic performance of general emergency physician gestalt for acute appendicitis in patients age 5 to 20.9 years presenting to a community ED with acute abdominal pain; 2) characterize the association between patient-, physician-, and facility-level characteristics and the receiver operating characteristic (ROC) curve characteristics of physician gestalt; and 3) examine clinical characteristics associated with gestalt assessments. We hypothesized that emergency physician gestalt would have a good c-statistic for predicting acute appendicitis and that more experienced physicians would show superior ROC curve characteristics.

METHODS

Study Design and Setting
Kaiser Permanente Northern California (KPNC) is a large, integrated health care delivery system that provides care to approximately four million members across 21 medical facilities with multiple clinics and ancillary services. KPNC members represent approximately 33% of the insured population in areas served and are comparable to the surrounding and statewide population with respect to age, sex, and race/ethnicity. KPNC utilizes a comprehensive integrated electronic health record (EHR; Epic, Verona, WI), fully implemented in 2009.

This study was conducted as a secondary analysis of a larger prospective study evaluating a CDS system for pediatric abdominal pain evaluation in 11 KPNC EDs (NCT02633735). This larger investigation consisted of a pre–post cluster-randomized trial of providing CDS with the pediatric Appendicitis Risk Calculator (pARC) score to providers. Detailed implementation methods of the larger study are reported elsewhere.

At study EDs, care was provided by board-certified or board-eligible emergency physicians. Table S1 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/ace.m.13931/full) shows facility-specific characteristics. All facilities had access to CT and US during regular business hours; however, after-hours US availability varied across facilities. Four of the study facilities had pediatric inpatient units.
Participant Selection
Treating emergency physicians enrolled eligible patients through a Web-based application embedded in the EHR. Patients were eligible if they were 5 to 20 years old with ≤5 days of right-sided or diffuse abdominal pain. These inclusion criteria were based on the original derivation/validation cohorts of the pARC. The age range, with an upper limit of 20 years, was chosen based on the inclusion criteria of the parent study. Exclusion criteria included abdominal trauma, known appendicitis or history of appendectomy, current pregnancy, or other uncommon chronic or confounding conditions described previously. To ensure that gestalt assessment was not influenced by imaging results, patients were excluded if enrollment occurred after ordering advanced abdominal imaging (US or CT). Only the first patient encounter between October 1, 2016, and September 30, 2018, was included in this analysis and enrollments made by providers listed as residents, students, or physician assistants were removed from the cohort post hoc.

To facilitate enrollment, promotional posters were placed in EDs, emergency physicians were sent automated text-message alerts when assigned a potentially eligible patient, and physicians received a small incentive ($5 gift card) for each completed enrollment. For the last 15 months of the study period, six of the 11 facilities also received CDS based on the pARC with care pathway recommendations (following gestalt entry) as part of the larger cluster-randomized trial. Other risk-stratification tools such as the Alvarado and PAS were not provided to, or routinely used by, our clinicians.

This study was approved by the KPNC Institutional Review Board with a waiver of informed consent. Patient safety was monitored by an independent data safety monitoring board.

Data Collection
Clinical variables of interest were identified based on previously reported associations with appendicitis and incorporation in validated risk scores. Data were collected from the EHR using automated data collection techniques and from physician-entered enrollment responses. Clinical characteristics entered by the emergency physician at the time of ED visit were based on predetermined definitions adapted from Kharbanda et al. (Table S2) and required for the pARC. Physicians prospectively entered gestalt on a continuous sliding scale of 1% to 100% after reporting the variables for the pARC but prior to ordering abdominal imaging (Figure S1). Gestalt could be entered before or after a white blood cell (WBC) count was determined. Gestalt estimates were not permitted post hoc.

Laboratory and abdominal imaging results were extracted from EHR data. Emergency physician data included age, sex, years since medical school graduation, and years as a KPNC physician. Facility characteristics included the presence of a pediatric inpatient unit and teaching hospital designation.

Outcomes
Our primary outcome was physician gestalt performance for the diagnosis of acute appendicitis. Patients were considered to have acute appendicitis if the diagnosis was made at the index ED visit or within 7 days. Appendicitis verification was performed via manual EHR review of operative and pathology reports with outcome definitions based on prior work by the study team. If the patient had a diagnosis of appendicitis in the EHR but no operative or pathology reports were available, the patient record was manually reviewed by a trained study abstractor. Patients transferred out of the KPNC system with an ED diagnosis of appendicitis (n = 6) were assumed to have appendicitis based on review of their encounter notes. As a subset of appendicitis cases, missed appendicitis was determined as a safety outcome and defined as appendicitis within 7 days after the initial ED enrollment and not part of the initial encounter or immediate transfer. All outcomes were reviewed by two trained study investigators with adjudication by a third investigator as needed. All cases of missed appendicitis were reviewed by four study investigators.

Secondary outcomes were analyzed to further assess the safety of physician gestalt assessment and included the rate of negative appendectomy and perforation. Negative appendectomy was defined as an appendectomy without a confirmed diagnosis of appendicitis based on operative or pathology notes. Perforation was defined as perforated appendicitis confirmed by operative and pathology notes.

Patients Not Enrolled
We assessed for potentially missed eligible patients via EHR database query and calculated the estimated appendicitis rate using principal diagnosis and appendectomy procedural codes in the missed eligible and excluded patient populations. Additionally, an audit was conducted at the start of the study to assess the characteristics of missed eligible patients.
Data Analysis
We generated initial predicted probabilities of appendicitis for each patient with a logistic model regressed on provider gestalt. We then ran logistic regression models of the outcome on the predicted probabilities to generate area under the curve (AUC) estimates and standard errors for each comparison group separately and compared the difference in AUC estimates using a chi-square distribution. A calibration plot was graphed and a Hosmer-Lemeshow test was used to determine goodness of fit. We compared differences in the c-statistics for physician gestalt by facility characteristics and by physician experience measures including age ($\leq 40$ vs. $>40$), years since medical school graduation ($<5$ vs. $\geq 5$), and years with the medical group ($<5$ vs. $\geq 5$). Age of 40 years was chosen based on median emergency physician age and experience cutoffs were based on a prior study in the same care setting.\(^26\) To analyze differences in these independent groups within the cohort, we compared differences in area under the ROC curves using chi-square tests with gestalt treated as a continuous variable.\(^27\) In addition to comparisons by facility and physician characteristics, we compared distributions of clinical characteristics across physician gestalt categories with chi-square tests for categorical variables and ANOVA for continuous variables. Gestalt categories of 1% to 10%, 11% to 49%, 50% to 89%, and 90% to 100% were chosen for descriptives purposes a priori because of their potential for clinical relevance. Test characteristics were calculated for the gestalt 1% to 10% category as a diagnostic predictor of appendicitis. A power analysis was conducted based on preliminary data and demonstrated that differences in c-statistics of 0.06 could be detected with 93% power with a sample size of 2,250 patients.\(^26\) All analyses were conducted in SAS version 9.4.

Sensitivity Analysis
As a planned sensitivity analysis, we assessed the c-statistic for gestalt after excluding cases where the WBC count was resulted prior to gestalt entry, determined using time stamps in the EHR.

RESULTS
We enrolled 3,426 patients (Figure 1) over the 24-month period; 436 physicians (mean age of 40.6 years, 60.6% male) completed enrollments. Physician gestalt estimates ranged from 1% to 97% (median = 18%, interquartile range = 5% to 43%). Of the eligible patients, 1,493 (43.6%) were in the physician gestalt category of 1% to 10%, 1,121 (32.7%) were 11% to 49%, 744 (21.7%) were 50% to 89%, and 68 (2.0%) were 90% to 100%. A total of 1,938 (56.6%) patients had a WBC count determined in the ED, 385 (11.2%) determined before gestalt entry, and 1,774 (51.8%) patients received US and/or CT imaging. Of those with low gestalt (1%–10%), 341 (22.8%) had imaging done in the ED (CT 1.7%, US 20.2%, both 0.9%). Sixty-six percent of patients in the 11% to 49% gestalt category received imaging (CT 7.3%, US 52.3%, both 6.3%).

Among eligible patients, 334 (9.8%) had confirmed acute appendicitis. Gestalt was found to be an excellent predictor of acute pediatric appendicitis with a c-statistic of 0.83 (95% CI = 0.81 to 0.85). Physician gestalt categorized 43.6% of patients in the low-gestalt category of 1% to 10% with an appendicitis rate of 1.1% and perforation rate of 0.3% (1%–10%—negative predictive value = 98.9% [95% CI = 98.3% to 99.3%]; >10%—sensitivity = 95.2% [95% CI = 92.3%–97.2%] and specificity = 47.8% [95% CI = 46.0%–49.6%] for diagnosis of appendicitis). However, gestalt demonstrated poor calibration due to overestimation of risk at the higher end of the spectrum (Hosmer-Lemeshow p = 0.001; Figure S2): appendicitis incidences were 7.6% in gestalt 11% to 49% range, 26.9% in gestalt 50% to 89% range, and 48.5% in gestalt 90% to 100% range. Distribution of physician gestalt by appendicitis outcome is shown in Figure 2. There was no evidence of temporal trends in gestalt ROC performance (quarterly comparisons) across the study time period.

Physician-level characteristics are presented in Table 1. Analysis of physician gestalt performance showed variation associated with years of physician experience. Physicians with $\geq 5$ years since medical school graduation had improved c-statistics compared to those with <5 years since medical school graduation (c-statistic = 0.84 vs. 0.74, p = 0.007; Table 2). Other physician-level characteristics were not significantly associated with gestalt performance: years with the medical group (p = 0.06), sex (p = 0.10), and age (p = 0.11; Table 2). Facility pediatric inpatient unit availability (p = 1.00) and teaching hospital designation (p = 0.49) were not significantly associated with physician gestalt performance.

All clinical variables tested, except for duration of pain <24 hours, were significantly associated with physician gestalt assessment (p < 0.05; Table 3).
There were notable increases in prevalence between the low- and high-gestalt strata for anorexia (increased by 55.1%), guarding (57.8%), migration of pain to right lower quadrant (RLQ; 72.2%), pain with coughing/hopping/walking (66.3%), and maximal tenderness in the RLQ (88.8%; Table 3). The highest percentage of ED imaging was for gestalt category 50% to 89% ($p < 0.001$; Table 3). Sensitivity analysis indicated insignificant variation in gestalt performance between those with no WBC count determined before gestalt entry ($n = 3,043$) and the overall cohort ($c$-statistic = 0.84 vs. 0.83, 95% CI = 0.82 to 0.87 vs. 0.81 to 0.85).

Safety and secondary outcomes are presented in Table 4. Of the 334 patients with appendicitis, 56 (16.8%) had a perforation. The negative appendectomy rate was 6.2% (22/356) and the missed appendicitis rate was 0.4% (15/3,426). Chart review analysis of low-gestalt (1%-10%) appendicitis cases ($n = 16$) revealed that 13 (81.0%) of these cases were early presentations of appendicitis (pain <24 hours). Chart review determined characteristics of low-gestalt appendicitis (1.1%), negative appendectomy (15.8%), and missed appendicitis (0.3%) patients are presented in Table 5. The three patients in the 1% to 10% gestalt category with perforated missed appendicitis all had pain <24 hours, no migration of pain, no pain with walking, no RLQ tenderness, and no guarding at the time of gestalt entry. These three patients returned to the ED between 7 and 72 hours following the index ED visit.

The appendicitis rate of nonenrolled patients was 1.1% (252/22,902) and audits assessing patient characteristics confirmed only a limited number of nonenrolled patients were truly eligible for the study. In a separate analysis by Cotton et al. examining a subset...
of this population, enrolled and nonenrolled cohorts did not differ significantly by age, sex, or race.

DISCUSSION

In this prospective study, we describe the diagnostic performance of emergency physician gestalt for the diagnosis of acute pediatric appendicitis and the association of physician gestalt with patient, physician, and facility characteristics.

Emergency physician gestalt in our community setting was found to have excellent ROC curve characteristics (c-statistic = 0.83), although with poorer discrimination at the higher end of the spectrum. Figure 2 demonstrates the especially good performance in the low-gestalt strata. This performance is notably better than that reported in a prior study of patients age 11 years and older (not restricted to pediatrics) who underwent CTs in the ED for possible appendicitis and used a dichotomous gestalt cutoff of 60%. The variation in performance between our study and theirs is multifactorial. Most prominently, our study focused on pediatric patients and treated gestalt as both a categorical and continuous variable. A recent study by Lee et al. found comparable physician gestalt performance to ours (c-statistic = 0.84), although this study was conducted at four EDs (two pediatric only) in Australia, where training pathways and clinical practices (e.g., CT is rarely used in pediatric abdominal pain evaluation) are significantly different from those in the U.S. community ED setting.

Emergency physician gestalt had good discriminatory ability in assigning patients to the low-risk (1%–

<p>| Table 2 |
| Provider and Facility Area Under the Receiver Operating Characteristics Curve Comparisons for Gestalt Performance |</p>
<table>
<thead>
<tr>
<th>Provider characteristics</th>
<th>AUC (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.83 (0.81–0.85)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤40</td>
<td>0.82 (0.79–0.85)</td>
<td>0.11</td>
</tr>
<tr>
<td>&gt;40</td>
<td>0.85 (0.82–0.88)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.81 (0.77–0.84)</td>
<td>0.10</td>
</tr>
<tr>
<td>Male</td>
<td>0.85 (0.82–0.87)</td>
<td></td>
</tr>
<tr>
<td>Years since medical school graduation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–4</td>
<td>0.74 (0.66–0.81)</td>
<td>0.007</td>
</tr>
<tr>
<td>≥5</td>
<td>0.84 (0.82–0.86)</td>
<td></td>
</tr>
<tr>
<td>Years with the medical group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–4</td>
<td>0.81 (0.77–0.84)</td>
<td>0.06</td>
</tr>
<tr>
<td>≥5</td>
<td>0.85 (0.82–0.87)</td>
<td></td>
</tr>
<tr>
<td>Pediatric inpatient unit available</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.83 (0.80–0.86)</td>
<td>1.00</td>
</tr>
<tr>
<td>No</td>
<td>0.83 (0.81–0.86)</td>
<td></td>
</tr>
<tr>
<td>Teaching hospital designation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major</td>
<td>0.81 (0.77–0.85)</td>
<td>0.49</td>
</tr>
<tr>
<td>Not major</td>
<td>0.83 (0.80–0.86)</td>
<td></td>
</tr>
</tbody>
</table>

AUC = area under the curve.
The low appendicitis rate in the low-gestalt category (1.1%) provides confidence in gestalt performance at the low end of the spectrum. Even in cases where initial gestalt was 1% to 10% and the patient had a final diagnosis of appendicitis, including those with perforations, chart review of the ED notes often revealed a progression of disease symptoms throughout the ED visit. However, emergency physicians often acted conservatively, even when their gestalt was low—as evidenced by the high imaging rate

### Table 3
Patient Characteristics, ED Laboratory Values, ED Imaging, and Appendicitis Diagnosis by Physician Gestalt Category (*N* = 3,426)

<table>
<thead>
<tr>
<th>Physician Gestalt Category</th>
<th>n (%)</th>
<th>1%–10%</th>
<th>11%–49%</th>
<th>50%–89%</th>
<th>90%–100%</th>
<th>n</th>
<th>1%–10%</th>
<th>11%–49%</th>
<th>50%–89%</th>
<th>90%–100%</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical characteristics (% yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age (years), mean (±SD)</strong></td>
<td>11.0 (±4.2)</td>
<td>10.5 (±4.3)</td>
<td>11.0 (±4.1)</td>
<td>11.8 (±4.1)</td>
<td>12.0 (±4.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>1,590 (46.4)</td>
<td>656 (43.9)</td>
<td>514 (45.9)</td>
<td>371 (49.9)</td>
<td>49 (72.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Temperature &gt; 38°C</strong></td>
<td>270 (7.9)</td>
<td>95 (6.4)</td>
<td>100 (8.9)</td>
<td>62 (8.3)</td>
<td>13 (19.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nausea/vomiting</strong></td>
<td>2,271 (66.3)</td>
<td>940 (63.0)</td>
<td>738 (65.8)</td>
<td>542 (72.9)</td>
<td>51 (75.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anorexia</strong></td>
<td>1,695 (49.5)</td>
<td>539 (36.1)</td>
<td>600 (53.5)</td>
<td>494 (66.4)</td>
<td>62 (91.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pain &lt;24 hours</strong></td>
<td>1,974 (57.6)</td>
<td>860 (57.6)</td>
<td>645 (57.5)</td>
<td>431 (57.9)</td>
<td>38 (55.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Guarding</strong></td>
<td>609 (17.8)</td>
<td>59 (4.0)</td>
<td>204 (18.2)</td>
<td>304 (40.9)</td>
<td>42 (61.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pain migrating to RLQ</strong></td>
<td>804 (23.5)</td>
<td>85 (5.7)</td>
<td>259 (23.1)</td>
<td>407 (54.7)</td>
<td>53 (77.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pain with coughing/hopping/walking</strong></td>
<td>1,255 (36.6)</td>
<td>241 (16.1)</td>
<td>446 (39.8)</td>
<td>512 (68.8)</td>
<td>56 (82.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Maximal tenderness to RLQ</strong></td>
<td>1,214 (35.4)</td>
<td>144 (9.7)</td>
<td>399 (35.6)</td>
<td>604 (81.2)</td>
<td>67 (98.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Diagnostic test utilization (% yes)**

| **Labs** | WBC count determined | 1,938 (56.6) | 517 (34.6) | 748 (66.7) | 617 (82.9) | 56 (82.4) |   |        |         |         |          |         |
| World Blood Count (WBC) | 923 (47.6) | 206 (39.9) | 330 (44.1) | 347 (56.2) | 40 (71.4) |   |        |         |         |          |         |
| **PMN count done** | 1,836 (53.6) | 484 (32.4) | 715 (63.8) | 585 (78.6) | 52 (76.5) |   |        |         |         |          |         |
| Polymorphonuclear Leukocytes (PMN) | 806 (43.9) | 175 (36.2) | 286 (40.0) | 311 (53.2) | 34 (65.4) |   |        |         |         |          |         |
| **Imaging** | Any ED abdominal imaging | 1,774 (51.8) | 341 (22.8) | 739 (65.9) | 642 (86.3) | 52 (76.5) |   |        |         |         |          |         |
| **Ultrasound only** | 1,338 (75.4) | 301 (88.3) | 586 (40.0) | 311 (53.2) | 34 (65.4) |   |        |         |         |          |         |
| **CT only** | 135 (12.3) | 20 (7.6) | 82 (11.1) | 105 (16.4) | 6 (11.5) |   |        |         |         |          |         |
| **Both** | 167 (12.2) | 25 (17.5) | 71 (50.7) | 119 (82.4) | 13 (91.1) |   |        |         |         |          |         |

**Appendicitis diagnosis (% yes)**

| **Confirmed appendicitis** | 334 (9.8) | 16 (1.1) | 85 (7.6) | 200 (26.9) | 33 (48.5) |   |        |         |         |          |         |

**PMN** = polymorphonuclear leukocytes; **RLQ** = right lower quadrant; **WBC** = white blood cell.

*Statistically significant variation between gestalt categories (p < 0.001).
†p < 0.05 for all categoric comparisons except 11% to 49% vs. 90% to 100% and 50% to 89% vs. 90% to 100%.
‡Percentage of those with a WBC done per gestalt category.
§Percentage of those with a PMN done per gestalt category.
¶Percentage of those with any imaging performed in the ED per gestalt category.

### Table 4
Secondary Outcome Events by Physician Gestalt Category

<table>
<thead>
<tr>
<th>Physician Gestalt Category</th>
<th>n (%)</th>
<th>1%–10%</th>
<th>11%–49%</th>
<th>50%–89%</th>
<th>90%–100%</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negative appendectomy</strong></td>
<td>22/356 (6.2)</td>
<td>3/19 (15.8)</td>
<td>6/91 (6.6)</td>
<td>12/212 (5.7)</td>
<td>1/34 (2.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Perforation</strong></td>
<td>56/334 (16.8)</td>
<td>4/16 (25.0)</td>
<td>14/85 (16.5)</td>
<td>35/200 (17.5)</td>
<td>3/33 (9.1)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Missed appendicitis</strong></td>
<td>15/3426 (0.4)</td>
<td>5/1493 (0.3)</td>
<td>5/1121 (0.5)</td>
<td>5/744 (0.7)</td>
<td>0/68 (0)</td>
<td>0.61</td>
</tr>
</tbody>
</table>

*Negative appendectomy related to index ED visit, percentage of those with an appendectomy in each gestalt cohort.
†Perforation within 7 days of index ED visit, percentage of those with appendicitis in each gestalt cohort.
‡Chart reviewed confirmed appendicitis within 7 days not as part of index visit or immediate transfer.

10%) category. The low appendicitis rate in the low-gestalt category (1.1%) provides confidence in gestalt performance at the low end of the spectrum. Even in cases where initial gestalt was 1% to 10% and the patient had a final diagnosis of appendicitis, including cases where initial gestalt was 1% to 10% and the patient had a final diagnosis of appendicitis, including
(22.8%) in the low-gestalt cohort. Reducing imaging for those deemed to be at low risk of appendicitis has the potential to decrease ED length of stay and resource utilization and, in the case of CT, mitigate a child’s exposure to radiation. Of note, our integrated health care system, with its good follow-up capability, is conducive to this care model. In select care settings with higher prevalence of appendicitis or other surgical diagnoses, for example, tertiary pediatric EDs, an US to magnetic resonance imaging (MRI) algorithm may be appropriate. However, during our study period, abdominal MRI was not readily, rapidly, and consistently available at our community EDs for the pediatric abdominal pain diagnostic algorithm. Notably, the gestalt category 50% to 89% had the highest imaging rate (86.3%), demonstrating a high level of concern regarding an appendicitis diagnosis in this patient strata. The somewhat lower imaging rates in the gestalt 90% to 100% category (76.5%) suggest that in this highest estimated risk decile, physicians may have been somewhat more confident in their diagnosis and the low negative appendectomy rate (2.9%) supports this contention. We were underpowered to robustly evaluate gestalt in this highest risk decile, but our results suggest that it may perform well as an adjunct to existing decision aids for this patient population.

Risk overestimation, especially in the intermediate gestalt categories, likely contributes to the overutilization of imaging. Overestimation may be due to concern for the ramifications of a missed diagnosis, both legal and adverse patient outcomes, and the relatively low-risk tolerance often prevalent in emergency physicians. Risk-minimizing behavior by emergency physicians may also contribute to the overutilization of advanced imaging due to the perceived risk of missing a high-consequence diagnosis.

We did not design this study to compare emergency physician gestalt performance to the pARC and PAS, which would not be a fair comparison because not all physicians who entered a gestalt ordered a WBC count in the ED, and we could not verify if those with a WBC count viewed the result prior to entering gestalt. However, recent work from our study team has reported on the performance of pARC and PAS in the same setting with distinct inclusion criteria (requiring the presence of a determined ED WBC count). The reported c-statistics range from 0.85 to 0.89 for pARC and 0.77 to 0.80 for PAS. While, comparatively, gestalt performed slightly better than the PAS and slightly worse than the pARC, we remind the reader that gestalt overestimated risk in the intermediate ranges (in which imaging rates were high) and as such is likely most useful in identifying low-risk (1%–10%) patients for whom no further ED workup is necessary. As such, the incorporation of gestalt for low-risk patients into CDS tools may facilitate provider buy-in and integration into provider workflow, thus increasing uptake in clinical practice. However, for cases falling in higher gestalt categories further evaluation may be necessary, including surgical consultation and/or imaging. CDS tools may help correct for the overestimation of risk at the higher end of the spectrum and provide reassurance to the provider when deciding if imaging is necessary.

Assessment of emergency physician characteristics and gestalt performance showed no significant variation by physician age, sex, or years with the medical group. Gestalt performance improved for physicians with >5 years since medical school graduation in all risk strata (Figure S3). This finding of enhanced gestalt performance with physician experience aligns with other studies on the performance of gestalt for pulmonary embolism diagnosis. This finding

### Table 5
Clinical Characteristics of Patients in Gestalt Category 1% to 10% With Confirmed Appendicitis or Confirmed Secondary Outcomes Based on Chart Review

<table>
<thead>
<tr>
<th>Outcome</th>
<th>n</th>
<th>Pain &lt;24 Hours</th>
<th>WBC Count Obtained in ED</th>
<th>Imaging During Index ED Visit</th>
<th>Perforation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendicitis</td>
<td>16</td>
<td>13</td>
<td>13</td>
<td>CT Only 1, US Only 5, CT and US 1*</td>
<td>4</td>
</tr>
<tr>
<td>Missed appendicitis</td>
<td>5†</td>
<td>3</td>
<td>3</td>
<td>0, 0, 0</td>
<td>3</td>
</tr>
<tr>
<td>Negative appendectomy</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3, 3, —</td>
<td>—</td>
</tr>
</tbody>
</table>

US = ultrasound; WBC = white blood cell.
*Patient also had intestinal malrotation.
†These patients are a subset of the appendicitis cases.
‡One case had chronic abdominal pain but was not excluded so as not to introduce bias due to selected chart review.
§All had equivocal/nondiagnostic imaging.
supports targeting the use of CDS tools toward more junior clinicians, who have also been reported to be more accepting of prediction rules than more experienced providers. Our evaluation also demonstrated that physician gestalt performance was not associated with specific facility variables.

Our results also provide insight into how physicians formulate their gestalt. For example, the presence of RLQ maximal tenderness was dramatically higher in the 90% to 100% gestalt category compared with the 1% to 10% gestalt category (p < 0.001), while pain <24 hours was not significantly associated with increased gestalt (p = 0.99). Performance of physician gestalt is known to vary in a condition-specific manner, and it is possible that pediatric appendicitis is associated with better performance due to the presence of trademark physical examination findings such as RLQ tenderness. Interestingly, 57.6% of patients in the lowest gestalt subgroup had pain <24 hours, accounting for 81% of appendicitis cases in the low-gestalt cohort. Our finding that pain <24 hours has poor correlation with gestalt demonstrates the difficulty of appendicitis diagnosis in patients with a brief duration of pain, potentially due to a lower likelihood of pain concentration in the RLQ within a short pain duration period.

LIMITATIONS

Several study limitations deserve mention. First, this analysis was undertaken as a component of a larger study on pediatric abdominal pain. The presence of this parent study may have increased physician awareness around the diagnostic evaluation and management of appendicitis, which may have, over time, impacted gestalt estimates. However, the publication of the pARC validation study was in April 2018, near the end of our study period, and at no time during the study was the pARC calculator available on publicly available Web-based platforms (i.e., MDCalc, New York, NY).

Enrollment for this study was initiated by the emergency physician and consequently did not capture all providers at the 11 KPNC EDs and only a sample of the total eligible patient population is represented. Study enrollment for the parent study and this subanalysis was performed on an opt-in basis by the treating physicians to capture an appropriate patient population at risk for appendicitis and meeting all eligibility criteria as defined above. Audits of missed eligible patients demonstrated that less than a quarter of potentially eligible patients were actually eligible for the larger study, and the low rate of appendicitis in this population suggests the we captured a representative risk pool. It is also unclear how the inability to compare physicians who enrolled patients in our study versus those who did not, as well as our specialized practice setting, affect study generalizability. Additionally, due to the necessary data collection design, physicians were asked about the presence of the patient’s clinical variables immediately prior to entering their gestalt. Theoretically, this may have increased the association between clinical variables and gestalt; however, this effect is likely mitigated since the assessed clinical variables are standard components of acute appendicitis evaluation in the ED. Since we could not control for physician gestalt being entered before or after attaining relevant clinical data, we did not consider a “gestalt-only” model and, instead, the availability of these clinical data and determination of gestalt were treated as a single step. Also, physicians could enter their gestalt before or after ordering a WBC count, but only 11% of enrollments had WBC counts determined at the time of gestalt entry. We were also unable to discern if imaging was requested by a consultant, such as a surgeon. Finally, there was the potential for providers to calculate the PAS or other risk scores on their own prior to completing the gestalt form; however, these scores require a WBC count and we are unaware of their regular use by KPNC emergency physicians.

CONCLUSION

Emergency physician gestalt for possible pediatric appendicitis presenting to the ED had excellent receiver operating characteristic curve characteristics. Emergency physicians with less experience showed decreased c-statistics. The very low rate of appendicitis in the low-gestalt risk category (1%–10%) provides support for providers’ decisions to forgo imaging in these patients. In higher-risk gestalt categories, the overestimation of risk suggests a possible benefit of utilizing prediction algorithms to mitigate imaging studies of limited value.

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.13931/full

Data Supplement S1. Supplemental material.
The Effect of Patient Observation on Cranial Computed Tomography Rates in Children With Minor Head Trauma

Sonia Singh, MD, MPH, MBA1,2,3, Stephen J. C. Hearps, MBiostat1, Meredith L. Borland, MBBS4,5, Stuart R. Dalziel, PhD6,7, Jocelyn Neutze, MBChB8, Susan Donath, MA1,2, John A. Cheek, MBBS1,2,9, Amit Kochar, MD10, Yuri Gilhotra, MBBS11, Natalie Phillips, MBBS11,12, Amanda Williams, BN1, Mark D. Lyttle, MBChB1,13,14, Silvia Bressan, PhD1,15, Jeffrey S. Hoch, PhD16,17, Ed Oakley, MBBS1,2,9, James F. Holmes, MD, MPH16,18, Nathan Kuppermann, MD, MPH18,19, and Franz E. Babl, MD, MPH1,2,9, on behalf of the Paediatric Research in Emergency Departments International Collaborative (PREDICT)

ABSTRACT

Background: Management of children with minor blunt head trauma often includes a period of observation to determine the need for cranial computed tomography (CT). Our objective was to estimate the effect of planned observation on CT use for each Pediatric Emergency Care Applied Research Network (PECARN) traumatic brain injury (TBI) risk group among children with minor head trauma.

From the 1Murdoch Children's Research Institute, Melbourne, Australia; and the 2Department of Paediatrics, University of Melbourne, Melbourne, Australia; the 3University of California Davis Medical Center, Sacramento, CA; the 4Perth Children’s Hospital, Perth, Australia; the 5Divisions of Paediatrics and Emergency Medicine, School of Medicine, University of Western Australia, Perth, Australia; the 6Starship Children’s Health, Auckland, New Zealand; the 7Departments of Surgery and Paediatrics: Child and Youth Health, University of Auckland, Auckland, New Zealand; the 8Kidzfirst Middlemore Hospital, Auckland, New Zealand; the 9Royal Children's Hospital, Melbourne, Australia; the 10Women’s & Children’s Hospital, Adelaide, Australia; the 11Queensland Children’s Hospital, Brisbane, Australia; the 12Child Health Research Centre, University of Queensland, Brisbane, Australia; the 13Bristol Royal Hospital for Children, Bristol, UK; the 14Academic Department of Emergency Care, University of the West of England, Bristol, UK; the 15Department of Women’s and Children’s Health, University of Padova, Padova, Italy; the 16Division of Health Policy and Management, Department of Public Health Sciences, University of California at Davis, Davis, CA; the 17Center for Healthcare Policy and Research, University of California at Davis, Sacramento, CA; and the 18Department of Emergency Medicine, University of California Davis School of Medicine, Sacramento, CA; and the 19Department of Pediatrics, University of California Davis School of Medicine, Sacramento, CA.

Received October 29, 2019; revision received February 3, 2020; accepted February 14, 2020.


The study was funded by grants from the National Health and Medical Research Council (project grant GNT1046727, Centre of Research Excellence for Pediatric Emergency Medicine GNT1058860), Canberra, Australia; the Murdoch Children’s Research Institute, Melbourne, Australia; the Emergency Medicine Foundation (EMPJ-11162), Brisbane, Australia; Perpetual Philanthropic Services (2012/1140), Australia; Auckland Medical Research Foundation (No. 3112011) and the A+ Trust (Auckland District Health Board), Auckland, New Zealand; WA Health Targeted Research Funds 2013, Perth, Australia; the Townsville Hospital and Health Service Private Practice Research and Education Trust Fund, Townsville, Australia; and supported by the Victorian Government’s Infrastructure Support Program, Melbourne, Australia. SS is supported by an Australian Government Research Training Program scholarship, and a PREDICT CRE Research Higher Degree scholarship. FEB’s time was partly funded by a grant from the Royal Children's Hospital Foundation and the Melbourne Campus Clinician Scientist Fellowship, Melbourne, Australia, and an NHMRC Practitioner Fellowship, Canberra, Australia. SDR’s time was partly funded by the Health Research Council of New Zealand (HRC13/556). The authors have no potential conflicts to disclose.

Clinical trial registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12614000463673.

Author contributions: SS conceived the study, interpreted the data, and wrote the initial draft of the article; all authors designed the study, approved publication, and agreed to be accountable for all aspects of the work; MLB, SRD, JN, JAC, AK, YG, NP, AW, MDL, SB, EO, and FEB obtained the data; SJCH and FEB had full access to all the data in the study and take responsibility for the integrity of the data; SJCH takes responsibility for the accuracy of the data analysis; SS, SJCH, JCH, SD, JFH, NK, and FEB supervised the analysis of the data, contributed to the interpretation of the data, and revised the article critically; MLB, SRD, JN, JAC, AK, YG, NP, AW, MDL, SB, EO, JFH, NK, and FEB interpreted the data, provided supervision, and drafted or revised the paper critically; SS takes responsibility for the paper as a whole.

Supervising Editor: Robert Cloutier, MD.

Address for correspondence and reprints: Sonia Singh, MD; e-mail: singhsd@student.unimelb.edu.au.


@related_article {title: "Cranial Computed Tomography Rates in Children With Minor Head Trauma", authors: "Sonia Singh, MD, MPH, MBA1,2,3, Stephen J. C. Hearps, MBiostat1, Meredith L. Borland, MBBS4,5, Stuart R. Dalziel, PhD6,7, Jocelyn Neutze, MBChB8, Susan Donath, MA1,2, John A. Cheek, MBBS1,2,9, Amit Kochar, MD10, Yuri Gilhotra, MBBS11, Natalie Phillips, MBBS11,12, Amanda Williams, BN1, Mark D. Lyttle, MBChB1,13,14, Silvia Bressan, PhD1,15, Jeffrey S. Hoch, PhD16,17, Ed Oakley, MBBS1,2,9, James F. Holmes, MD, MPH16,18, Nathan Kuppermann, MD, MPH18,19, and Franz E. Babl, MD, MPH1,2,9, on behalf of the Paediatric Research in Emergency Departments International Collaborative (PREDICT)", journal: "ACADEMIC EMERGENCY MEDICINE", volume: "2020", issue: "27", pages: "832–843", issn: "1553-2712", doi: "10.1111/acem.13942"}
Methods: This was a secondary analysis of a prospective observational study at 10 emergency departments (EDs) in Australia and New Zealand, including 18,471 children < 18 years old, presenting within 24 hours of blunt head trauma, with Glasgow Coma Scale scores of 14 to 15. The planned observation cohort was defined by those with planned observation and no immediate plan for cranial CT. The comparison cohort included the rest of the patients who were either not observed or for whom a decision to obtain a cranial CT was made immediately after ED assessment. The outcome clinically important TBI (c iTBI) was defined as death due to head trauma, neurosurgery, intubation for > 24 hours for head trauma, or hospitalization for ≥ 2 nights in association with a positive cranial CT scan. We estimated the odds of cranial CT use with planned observation, adjusting for patient characteristics, PECARN TBI risk group, history of seizure, time from injury, and hospital clustering, using a generalized linear model with mixed effects.

Results: The cranial CT rate in the total cohort was 8.6%, and 0.8% had c iTBI. The planned observation group had 4,945 (27%) children compared to 13,526 (73%) in the no planned observation group. Cranial CT use was significantly lower with planned observation (adjusted odds ratio \([OR] = 0.2\), 95% confidence interval \([CI]\) = 0.1 to 0.1), with no difference in missed c iTBI rates. There was no difference in the odds of cranial CT use with planned observation for the group at very low risk for c iTBI (adjusted OR = 0.9, 95% CI = 0.5 to 1.4). Planned observation was associated with significantly lower cranial CT use in patients at intermediate risk (adjusted OR = 0.2, 95% CI = 0.2 to 0.3) and high risk (adjusted OR = 0.1, 95% CI = 0.0 to 0.1) for c iTBI.

Conclusions: Even in a setting with low overall cranial CT rates in children with minor head trauma, planned observation was associated with decreased cranial CT use. This strategy can be safely implemented on selected patients in the PECARN intermediate- and higher-risk groups for c iTBI.

Minor head trauma in children is a common emergency department (ED) presentation and is frequently defined by a Glasgow Coma Scale (GCS) score of 14 to 15 following head trauma. Cranial computed tomography (CT) to diagnose traumatic brain injury (TBI) after trauma is the current gold standard investigation; however, the ionizing radiation exposure is associated with an increased risk of lethal malignancies. There has been concern about increasing CT rates in children presenting to EDs in the United States from 1998 to 2008. Retrospective studies on cranial CT trends in pediatric EDs have noted considerably higher rates in the United States (27.6% in 2007 to 30.2% in 2015) compared with Australia (9.5% in 2001 to 9.6% in 2010). This difference has been confirmed with prospective studies in children reporting lower non-risk-adjusted cranial CT rates in Australia (8.3%) compared to the United States (35.3%), with similar rates of clinically important traumatic brain injuries (c TBI). Clinical decision rules (CDRs) have been developed and validated, to aid clinical decision making, improve the sensitivity of identifying patients with TBI, and optimize cranial CT rates by decreasing unnecessary CT scans. With the Choosing Wisely campaign, the American Academy of Pediatrics and the American College of Emergency Physicians have endorsed the benefit of an observation period for the management of pediatric head injury. Observation reduces cranial CT scan rates for children with minor blunt head trauma, with no significant impact on the delayed diagnosis of c TBI. One investigation of 40,113 children reported that observation was associated with lower cranial CT use when compared with those not observed (adjusted odds ratio \([OR] = 0.53\)). In a subsequent study of 1,605 patients, every hour increase in ED observation was associated with decreased cranial CT rates across all three risk groups of children with minor head trauma at risk for c iTBI.

We explored the relationship between planned ED observation and cranial CT use in children with minor blunt head trauma in a clinical setting with low baseline cranial CT rates. Additionally, we stratified patients according to their risk for c iTBI and estimated the effect of ED observation on CT rates for children in the different PECARN-defined TBI risk groups (Table 1) and the impact of observation on missed c iTBI rates.

METHODS

Study Design
This was a planned secondary analysis of a prospective cohort study of children younger than 18 years of age with minor blunt head trauma, enrolled between April
2011 and November 2014 in 10 pediatric EDs in Australia and New Zealand which are members of the Pediatric Research in Emergency Departments International Collaborative (PREDICT) network. The primary study was designed to validate 3 clinical head injury decision rules: PECARN (USA), CATCH (Canada), and CHALICE (UK). The individual hospital ethics committee for each participating site approved the study. The details of the study protocol have been previously published.\(^{20}\)

### Study Setting and Population

The participating EDs were a combination of two mixed and eight free-standing children’s hospital EDs across Australia and New Zealand. The treating clinicians enrolled patients presenting to the ED with blunt head trauma. A standardized case report form was completed recording demographic, epidemiologic, and clinical information. After the initial ED assessment, clinicians documented responses to two questions that were not mutually exclusive: if they planned to observe the patient (“Do you intend to observe the patient in ED for head injury? yes/no”) and if they planned to obtain immediate neuroimaging (“Do you intend to perform neuroimaging for this patient? yes/no”)—if yes, (“What neuroimaging do you intend to perform? CT, MRI, US, skull X-ray”). Based on the responses, the planned observation cohort included those with planned observation and no immediate plan for cranial CT. The comparison cohort included the rest of the patients either who were not observed or for whom a decision to obtain a cranial CT was made immediately after ED assessment. Planned ED observation was defined as “yes” to the question, “Do you intend to observe the patient in ED for head injury?”

### Study Protocol

#### Inclusion Criteria.

All children younger than 18 years of age presenting to the ED after sustaining blunt head trauma of any severity were enrolled in the primary study.\(^{11}\) For this subanalysis, we selected the cohort of children with GCS scores of 14 to 15 presenting within 24 hours of head trauma, to enable direct comparison with prior published results. Children with multiple ED visits, with different episodes of head trauma, were eligible for enrollment at each visit.

#### Exclusion Criteria.

Exclusion criteria were published in the primary study,\(^{11}\) which included patients with trivial facial trauma, those who presented to the ED with neuroimaging for head trauma before study enrollment, and those with history of blunt head trauma who left the ED without being seen by a physician.

### Study Measures

Physicians recorded ED data, and research assistants recorded the data regarding hospital management and conducted structured telephone follow-up on all patients who did not get neuroimaging in the ED 14 to 90 days after the initial ED visit. Results of cranial CTs performed on patients who re-presented to the ED were reviewed by the site investigators. Neuroimaging and neurosurgery reports were reviewed for any patients identified on the follow-up telephone call to have had medical visits outside the original study hospital. Patients who were admitted to the hospital or discharged from ED after a cranial CT was obtained did not receive a follow-up phone call.

All site investigators and research coordinators were trained at the central site prior to commencing the study. Site clinicians and research assistants were trained by the site investigators and research coordinators using standardized teaching materials in 1-hour
sessions. This was repeated as needed for new staff throughout the study period. Site investigators and research coordinators participated in annual updates. The central study coordinator conducted annual site visits for oversight, compliance, and ongoing updates of site personnel.

Timing of events was calculated in hours using information collected on case report forms, including time from injury to ED arrival, time from ED arrival to physician evaluation, ED length of stay (LOS), and total LOS. We defined ED LOS as the interval of time from ED arrival until ED discharge. Data on admission rates included those admitted to short-stay units for observation, hospital wards, or pediatric intensive care units (PICUs). The short-stay unit admitted patients with anticipated LOS < 24 hours. The location of the short-stay units varied across the participating hospitals (part of ED, separate unit in ED, or a hospital ward). As a result, for those patients admitted to observation units, but not the hospital wards or PICU, the ED LOS was the interval of time from ED arrival until hospital discharge. Total LOS was defined as the time from ED arrival to hospital discharge.

To minimize any potential selection bias between the two study cohorts, we stratified the population by their risk for ciTBI using the age-based PECARN TBI risk groups: very low, intermediate, and high (Table 1). Children with no predictors for ciTBI were classified as very low risk. Children with one or more risk factors were classified as intermediate or high risk, depending on the specific risk factor, according to the original PECARN paper. We analyzed CT use for each of the PECARN TBI risk groups.

The primary outcome measures were CT and ciTBI rates. Secondary outcome measures were ED and total LOS, rates of TBI on cranial CT, and hospital admission. The ED clinician’s initial assessment of the patient’s GCS score, attending radiologist reports of the CT scans, and operative reports of neurosurgical interventions were used for each site.

The outcome, cranial CT, was defined as present if the patient underwent the imaging study during ED or short-stay evaluation. The outcome, ciTBI, was defined as death as a result of intracranial injury, neurosurgical intervention, intubation for longer than 24 hours for TBI, or hospital admission for 2 nights or longer in association with TBI on CT. Neurosurgical intervention for TBI included intracranial pressure monitoring, elevation of depressed skull fracture, ventriculostomy, hematoma evacuation, lobectomy, tissue debridement, duro repair, and other procedures. Missed ciTBI was defined as the diagnosis of ciTBI on a follow-up ED visit or identified on the follow-up phone call, with no cranial CT performed on the initial ED visit. TBI on CT was defined as intracranial hemorrhage or contusion, cerebral edema, traumatic infarction, diffuse axonal injury, shearing injury, sigmoid sinus thrombosis, midline shift of intracranial contents or signs of brain herniation, diastasis of the skill, pneumocephalus, and skull fracture depressed by at least the width of the table of the skull. The attending radiologists at each hospital site were blinded to the study hypothesis and patient cohorts.

Data Analysis

The focus of this analysis was a comparison of the cohort with planned observation and the cohort who were either not observed or for whom a decision to obtain a CT was made immediately after ED evaluation. We described these cohorts with population proportions and 95% confidence intervals (CIs). Additionally, we compared the distribution of patients within each PECARN ciTBI risk group between the two cohorts. We performed a bivariate analysis of rates of CT scans, ciTBI, missed ciTBI, TBI on CT, and hospital admission between the two cohorts and for each PECARN TBI risk group. The proportion of patients with return ED visits was calculated for those with follow-up phone calls.

We compared the time in hours as a categorical variable using median with interquartile range, and rate difference using the Hodges-Lehmann method, which accounts for different population distributions in the two cohorts. The total LOS (hours) was compared, as well as ED LOS, LOS for each PECARN TBI risk group, and LOS in patients who had cranial CT scans.

Using a generalized estimating equation model, we estimated cranial CT use with multivariable logistic regression adjusting for patient demographics, PECARN TBI risk group, history of seizure, and time from injury. History of seizure was not included as a risk factor for ciTBI in the PECARN head injury rules. The investigators subsequently described the importance of recurrent seizures as a risk factor associated with TBI on CT. In the current study, history of seizure had a significant bivariate relationship with CT use. Therefore, it was included in the multivariate analysis. Because planned observation and cranial CT rates could be correlated with clinician practice within
a hospital site, we controlled for hospital site as a random effect (with random intercept). Furthermore, because clinician practice could vary across the study time-period (2011–2014), we adjusted for the calendar year as a fixed effect, with 2011 as a reference. We investigated variance inflation factor values to check for multicollinearity between the variables and model residuals and leverage to ensure the absence of outliers and influential cases. Overall model performance was evaluated via assessment of classification and calibration (prediction accuracy). Additionally, we performed a sensitivity analysis of the regression model by adjusting for the individual signs and symptoms and mechanism of injury, instead of the individual PECARN TBI risk groups. Finally, we estimated the adjusted odds (OR) of CT use with planned observation for each PECARN TBI risk group.

Sample size calculations were performed and published in the primary study, and no additional computations were performed for this planned sub-analysis. Additionally, there were sufficient events in the outcome measure of cranial CT use (8.6%, 1,579 patients), such that including the seven covariates ensured a robust regression model. Data analysis was performed with Stata (version 13) for the analysis. This article is reported following the STROBE statement.

RESULTS

Characteristics of Study Subjects

Of the 20,137 enrolled patients in the primary study, 18,781 (93.3%) presented within 24 hours of sustaining head trauma, with GCS scores of 14 to 15 (Figure 1). Documented planned observation and cranial CT status were available for 18,471 (91.7%) of enrolled patients, and all subsequent analyses were performed on this cohort. There were no differences in demographic characteristics, cranial CT, or ciTBI rates between the 310 (1.5%) patients with no documented observation plan and the analytic cohorts.

![Patient flow diagram](image)

**Figure 1.** Patient flow diagram. ciTBI = clinically important traumatic brain injury; GCS = Glasgow Coma Scale.
**Descriptive Data**

The overall cranial CT rate was 8.6% (range = 2% to 16% across the different hospitals), with a ciTBI rate of 0.8 and 0.1% of patients receiving a neurosurgical intervention. Cranial CT rates for each of the PECARN TBI risk groups varied significantly as those at higher risk were more likely to undergo CT: very-low-risk group (1.8% CT rate), intermediate-risk group (10.3% CT rate), and high-risk group (43.1% CT rate). Similarly, the ciTBI rates varied significantly, with 0.01% in the PECARN very-low-risk group, 0.6% in the intermediate-risk group, and 6.3% in the high-risk group.

The strategy of planned observation was made for 4,945 (26.8%) patients (Table 2), with a range of 21% to 49% across the participating hospitals. Those in the planned observation cohort were slightly older and included 33.0% at very low risk, 54.9% at intermediate risk, and 12.2% at high risk for ciTBI.

The median time from injury to ED arrival in the cohort with planned observation was not statistically different from those with no observation (Table 3). The increase in ED LOS with cranial CT use was significant with planned observation and for each of the PECARN risk groups. The total LOS was greater with planned observation for patients at very low risk and intermediate risk and lower for those at high risk for ciTBI than those with no planned observation.

**Main Results**

The unadjusted cranial CT rate was significantly lower in those with planned observation (4.4%) than those in the no planned observation cohort (10.1%), as was the ciTBI rate (0.4% vs. 0.9%; Table 2). There was a significant difference in the ward and PICU admission rates for the patients with planned observation (6.4 and 0.3%) compared to those not observed (4.0 and 0.1%). There was no difference in the rates of missed ciTBI between the cohorts. Two patients in the intermediate PECARN risk group, both with planned observation, were admitted for ciTBI at the follow-up ED visit. Neither patient required neurosurgery. The details of these patients have been published. There were no deaths in either cohort.

**PECARN TBI Risk Groups.** There was no difference in the unadjusted cranial CT, ciTBI, and TBI on CT rates for the cohorts with and without planned observation, for the group at very low risk for ciTBI (Table 2). The planned observation group at intermediate risk for ciTBI had significantly lower CT use (4.4%) than the group without planned observation (13.3%), and there were no differences in rates of ciTBI (0.5% vs. 0.6%) or TBI on CT (0.7% vs. 1.1%). The planned observation group at high risk for ciTBI had significantly lower cranial CT use (11.7%) than those without planned observation (62.5%), with significantly lower rates of ciTBI (1.2% vs. 9.4%) and TBI on CT (2.0% vs. 14.1%).

**Multivariable Analysis.** After age, sex, PECARN TBI risk group, history of seizure, calendar year, and time from injury were adjusted for, the cohort with planned observation had significantly lower cranial CT use (adjusted OR = 0.2, 95% CI = 0.1 to 0.2; Table 4). Additionally, there was no evidence of multicollinearity, with variance inflation factor of less than 1.5 for all variables in the model (except the nonindependent calendar years variable). A sensitivity analysis of the generalized linear model, adjusting for the mechanism of injury and individual signs and symptoms of TBI rather than PECARN risk group, produced similar results (adjusted OR = 0.2, 95% CI = 0.1 to 0.2; 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.13942/full). Good model fit was found, and the classification rate was high (94%; sensitivity = 38%, specificity = 99%). There was no difference in cranial CT use across calendar years for the total cohort and each individual PECARN TBI risk group.

Multivariate logistic regression for each of the PECARN TBI risk groups demonstrated that planned observation was significantly associated with lower CT use in patients at intermediate risk (adjusted OR = 0.2, 95% CI = 0.2 to 0.3) and high risk (adjusted OR = 0.1, 95% CI = 0.0 to 0.1) for ciTBI (Data Supplement S1). Good model fit was found, with high classification rates: high risk 83% (sensitivity = 78%, specificity = 86%); intermediate risk 90% (sensitivity = 7.5%, specificity = 99%). The difference in cranial CT use between those with and without planned observation at very low risk for ciTBI was not significant (adjusted OR = 0.9, 95% CI = 0.5 to 1.4), with a poor model fit.

**DISCUSSION**

This multicenter prospective study evaluated the strategy of planned ED observation and cranial CT use in children with minor head trauma. Despite the low
baseline cranial CT rate, planned observation of patients was associated with 80% lower adjusted odds of CT use. Although the cohorts with and without planned observation were dissimilar in individual patient characteristics, after patient characteristics, signs, symptoms, and risk factors for ciTBI were adjusted for, planned observation was independently associated with lower cranial CT use in all but those at very low risk of ciTBI.

The primary study protocol defined ED observation as an ongoing assessment for less than 6 hours after the initial clinical evaluation. Admission was defined as assessment requiring more than 6 hours to either an observation or short-stay, inpatient, or intensive care unit.

Table 2
Patient Demographics, PECARN TBI Risk Groups, and Patient Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Planned Observation, n (%)</th>
<th>No Planned Observation, n (%)</th>
<th>Rate Difference, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4,945 (26.8)</td>
<td>13,526 (73.2)</td>
<td>−46.4 (−47.8 to −44.9)</td>
</tr>
<tr>
<td>Demographic characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>Mean (±SD)</td>
<td>6.2 (±5.1)</td>
<td>5.5 (±4.4)</td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>4.5 (1.8–10.7)</td>
<td>4.0 (1.9–8.1)</td>
</tr>
<tr>
<td>Age &lt; 2 years</td>
<td></td>
<td>1,393 (28.2)</td>
<td>3,566 (26.4)</td>
</tr>
<tr>
<td>Sex (male)</td>
<td></td>
<td>3,102 (62.7)</td>
<td>8,671 (64.1)</td>
</tr>
<tr>
<td>PECARN TBI risk group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td></td>
<td>1,630 (33.0)</td>
<td>8,201 (60.6)</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td>2,714 (54.9)</td>
<td>4,273 (31.6)</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>601 (12.2)</td>
<td>1052 (7.8)</td>
</tr>
<tr>
<td>CT performed</td>
<td></td>
<td>215 (4.4)</td>
<td>1,364 (10.1)</td>
</tr>
<tr>
<td>PECARN TBI risk group, n (within group %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td></td>
<td>25 (1.5)</td>
<td>138 (1.7)</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td>120 (4.4)</td>
<td>569 (13.3)</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>70 (11.7)</td>
<td>657 (62.5)</td>
</tr>
<tr>
<td>Patient outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ciTBI</td>
<td></td>
<td>20 (0.4)</td>
<td>126 (0.9)</td>
</tr>
<tr>
<td>Neurosurgical intervention</td>
<td></td>
<td>2 (0.0)</td>
<td>21 (0.2)</td>
</tr>
<tr>
<td>Intubation for &gt; 24 hours</td>
<td></td>
<td>0 (0.0)</td>
<td>2 (0.0)</td>
</tr>
<tr>
<td>Hospital admission ≥ 2 nights with + CT</td>
<td></td>
<td>20 (0.4)</td>
<td>125 (0.9)</td>
</tr>
<tr>
<td>Death from head injury</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PECARN TBI risk group, n (within group %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td></td>
<td>0 (0.0)</td>
<td>1 (0.01)</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td>13 (0.5)</td>
<td>26 (0.6)</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>7 (1.2)</td>
<td>99 (9.4)</td>
</tr>
<tr>
<td>TBI on CT</td>
<td></td>
<td>32 (0.7)</td>
<td>195 (1.4)</td>
</tr>
<tr>
<td>PECARN TBI risk group, n (within group %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td></td>
<td>0 (0.0)</td>
<td>1 (0.01)</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td>20 (0.7)</td>
<td>46 (1.1)</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>12 (2.0)</td>
<td>148 (14.1)</td>
</tr>
<tr>
<td>Hospital admission rate</td>
<td></td>
<td>2,238 (45.3)</td>
<td>1,811 (13.4)</td>
</tr>
<tr>
<td>Short-stay admission</td>
<td></td>
<td>2,056 (41.6)</td>
<td>934 (6.9)</td>
</tr>
<tr>
<td>Ward admission</td>
<td></td>
<td>199 (4.0)</td>
<td>864 (6.4)</td>
</tr>
<tr>
<td>PICU</td>
<td></td>
<td>3 (0.1)</td>
<td>46 (0.3)</td>
</tr>
<tr>
<td>Return ED visit‡</td>
<td></td>
<td>703 (14.9)</td>
<td>1,457 (12.0)</td>
</tr>
<tr>
<td>Missed ciTBI</td>
<td></td>
<td>2 (0.04)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

ciTBI = clinically important traumatic brain injury; GCS = Glasgow Coma Scale; IQR = interquartile range; PECARN = Pediatric Emergency Care Applied Research Network; PICU = pediatric intensive care unit; TBI = traumatic brain injury.

*Difference between means (95% CI).
†Difference between medians (95% CI).
‡If had follow-up interview: planned observation n = 4,773; no planned observation n = 12,208.
Although there was no defined period of observation in the study protocol, in a subsequent analysis, a hospital LOS greater than 4 hours was used as a proxy to evaluate the effect of observation. ED LOS is affected by many factors that are often out of the clinician’s control. Therefore, the impact of planned ED observation on CT use in patients for whom there was no plan for an immediate cranial CT provides the most accurate effect of time on clinical decision-making. The total LOS was greater for the cohort with no planned observation at high risk for ciTBI than those with planned observation. This was likely as a result of more children at high risk for ciTBI getting immediate cranial CTs (62.5%), having TBI on CT (14.1%), or ciTBI (9.4%), either of which may have been associated with hospital admission.

Although the primary study for which this was a subanalysis was designed to validate the existing head injury decision rules, clinicians did not report using any particular guideline for their decision making. Additionally, with a stable cranial CT rate in the decade prior to the study, and no difference in cranial CT use across calendar years in the current study, this suggests that the clinicians did not alter their practice.

### Table 3
Median Intervals: Planned Observation Versus No Planned Observation

<table>
<thead>
<tr>
<th>Interval (hours)</th>
<th>Planned Observation</th>
<th>No Planned Observation</th>
<th>Time Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Median (IQR)</td>
<td>n</td>
</tr>
<tr>
<td>Injury to ED arrival</td>
<td>4,892</td>
<td>1.3 (0.8-2.2)</td>
<td>13,293</td>
</tr>
<tr>
<td>ED arrival to physician evaluation</td>
<td>4,936</td>
<td>0.4 (0.2-0.7)</td>
<td>13,502</td>
</tr>
<tr>
<td>ED LOS (hours)</td>
<td>Overall</td>
<td>4,935 3.7 (2.7-5.2)</td>
<td>13,495</td>
</tr>
<tr>
<td></td>
<td>PECARN TBI risk group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low risk</td>
<td>1,628</td>
<td>3.4 (2.4-4.5)</td>
<td>8,186</td>
</tr>
<tr>
<td>Intermediate risk</td>
<td>2,707</td>
<td>3.7 (2.8-5.3)</td>
<td>4,261</td>
</tr>
<tr>
<td>High risk</td>
<td>600</td>
<td>4.4 (3.3-7.3)</td>
<td>1,048</td>
</tr>
<tr>
<td>Cranial CT</td>
<td>Obtained</td>
<td>213 6.0 (4.0-13.0)</td>
<td>1,360</td>
</tr>
<tr>
<td></td>
<td>Not obtained</td>
<td></td>
<td>4,722</td>
</tr>
<tr>
<td>Total LOS (hours)</td>
<td>Overall</td>
<td>4,934 3.7 (2.7-5.5)</td>
<td>13,484</td>
</tr>
<tr>
<td></td>
<td>PECARN TBI risk group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low risk</td>
<td>1,628</td>
<td>3.4 (2.5-4.7)</td>
<td>8,184</td>
</tr>
<tr>
<td>Intermediate risk</td>
<td>2,706</td>
<td>3.4 (2.8-5.6)</td>
<td>4,259</td>
</tr>
<tr>
<td>High risk</td>
<td>600</td>
<td>4.6 (3.4-10.5)</td>
<td>1,041</td>
</tr>
<tr>
<td>Cranial CT</td>
<td>Obtained</td>
<td>213 11.8 (4.6-22.0)</td>
<td>1,351</td>
</tr>
<tr>
<td></td>
<td>Not obtained</td>
<td></td>
<td>4,721</td>
</tr>
</tbody>
</table>

LOS = length of stay; IQR = interquartile range; PECARN = Pediatric Emergency Care Applied Research Network; TBI = traumatic brain injury.

### Table 4
Adjusted Likelihood of Cranial CT Use

<table>
<thead>
<tr>
<th></th>
<th>Total Cohort</th>
<th>N = 17,969</th>
<th>AOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≥ 2 years</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 years</td>
<td>0.4 (0.4-0.5)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (male)</td>
<td>1.11 (0.97-1.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PECARN TBI risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>8.1 (6.7-9.7)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>77.0 (62.8-94.5)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of seizure</td>
<td>4.6 (3.2-6.5)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time from injury (hours)</td>
<td>1.03 (1.02-1.04)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not observed</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planned observation</td>
<td>0.2 (0.1-0.2)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>0.9 (0.6-1.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>0.8 (0.6-1.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>0.8 (0.6-1.0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clustered by hospital site; *p ≤ 0.001.

AOR = adjusted odds ratio; PECARN = Pediatric Emergency Care Applied Research Network; TBI = traumatic brain injury.
during the study period. The comparison cohort was composed of patients with different risk profiles for ciTBI, consisting of those who were either not observed (e.g., asymptomatic children with minor mechanisms of injury) or for whom a decision to obtain a cranial CT was made right after ED evaluation (e.g., symptomatic children at higher risk for TBI). This is similar to a prior study in which 72% of children at very high risk for ciTBI were in the not-observed group. They reported a cranial CT rate of 5% in the patients who were observed prior to the decision of imaging compared with 30% in the not-observed cohort.

Stratifying the cohorts by the PECARN TBI risk groups allowed for a comparison of cranial CT rates in patients at similar risk for ciTBI. There were no differences in the rates ciTBI and TBI on CT for the very-low-risk and intermediate-risk patients between the two cohorts, indicating a similar risk profile. After risk adjustment, however, planned ED observation was associated with significantly lower cranial CT use in the groups at intermediate and high risk for ciTBI. The group at very low risk for ciTBI who were observed did not have a statistically lower rate of cranial CT use compared to those not observed. This is not surprising as cranial CT is not recommended in this low-risk group, and therefore it would be unlikely that observation would have much impact.

The results of this study validate the published literature from North America, where ED observation is associated with lower cranial CT rates in children after minor blunt head trauma. Prior research on the CT rates of each of the PECARN TBI risk predictors in the United States was reported as 13% of low-risk, 54% of intermediate-risk, and 82.5% of high-risk groups. Published evidence of the clinical practice in Australia and New Zealand for pediatric minor head trauma termed as “usual care” report lower cranial CT rates than with the use of decision rules. Those publications report much lower CT rates for each of the PECARN TBI risk groups (1.7% of very low risk, 9.9% of intermediate risk, and 44% of high risk) with similar ciTBI rates.

In addition to differences in clinical practice, this study highlights the use of short-stay units in EDs in Australia and New Zealand. The Australian National Emergency Target (NEAT) or the 4-hour rule was introduced in 2011, and a similar 6-hour target in New Zealand was implemented in 2009, in response to ED overcrowding. These targets/rules require a certain proportion of patients based on their triage acuity to be discharged, admitted, or transferred within 4 hours of ED presentation, with a potential loss of financial incentives when 90% of targets are not met. Because admissions to short-stay units stop the NEAT target clock, they are commonly used to decompress the ED. Perhaps developing similar short-stay units in pediatric EDs in North America might aid with longer observation periods of children with minor head trauma and result in fewer CT scans.

The increase in lifetime attributable risk of cancer associated with ionizing radiation of cranial CT scans has been reported worldwide. The cumulative iatrogenic radiation in children might triple the risk of leukemia and brain tumors, a large retrospective study reports. With increasing awareness of these risks, it has become important to educate health care providers and parents that CT decisions can often be undertaken as a shared decision-making process. Furthermore, implementation of clinical decision tools has resulted in decreased CT rates for minor head trauma at pediatric and general EDs. A recent study of seven health systems in the United States and Ontario, Canada, report a decreasing trend in cranial CT rates for children <18 years from 2006 to 2016. However, this has not been the case at the national level in the United States, with no apparent decrease in cranial CT rates in children between 2007 and 2015.

The PECARN age-based CDRs identify those patients at very low risk for ciTBI and for whom cranial CT is usually not recommended. Incorporating decision support in the electronic medical record based on the PECARN rules reduces cranial CT rates in the ED for children at very low risk for ciTBI. The average cranial CT rates across all the intervention EDs decreased from 5.3% to 4.2%. Furthermore, it is possible to achieve overall cranial CT rates of ~15% in all children with minor blunt head trauma with GCS scores of 14 to 15. We report no benefit from planned observation in patients at very low risk for ciTBI, and cranial CT rate of 1.7%, without any missed ciTBI in this group. Following this analysis, we intend to evaluate the cost-effectiveness of observation as a strategy to reduce CT use while minimizing missed ciTBI.

To our knowledge, this is the first study to validate the impact of planned ED observation as a strategy to lower cranial CT use for each PECARN TBI risk group, in a large cohort of children with minor blunt
head trauma. Despite the lower overall use of CT in this study, our findings are in agreement with the previous study of planned observation conducted on a different continent. Selective observation of patients with minor head trauma at intermediate risk of ciTBI could provide time for resolution of symptoms and/or reassurance of the lack of ciTBI and decrease the need for cranial CT. We found no increased risk of missed ciTBI in the observation without a CT cohort, even among selected patients at higher risk of ciTBI. Therefore, this strategy can be safely implemented on selected patients in the PECARN intermediate- and higher-risk groups for ciTBI.

LIMITATIONS

Our study has several limitations. First, the questions asked of the clinicians on the case report forms regarding planned observation and planned immediate cranial CT were not mutually exclusive. Additionally, the questions did not clarify the reasons for planned observation and, in particular, the patients for whom the decision to obtain an immediate cranial CT was already made. However, there was no difference in the cranial CT rates between these patients and those who received an immediate cranial CT scan and were not observed. As a result, we believe that the effect of observation on the cranial CT rate is accurately reflected in the cohort with planned observation and no plan for immediate cranial CT. Second, there were significant differences between the patients with planned observation and those without planned observation. The patients who received an immediate CT were more symptomatic, with frequent severe mechanisms of injury and at a higher risk for ciTBI than the cohort with planned observation. Although we used multivariable logistic regression to adjust for risk factors of ciTBI to estimate the independent effect of planned observation on cranial CT use, there could be some unaccounted-for residual confounding or bias.

Third, we did not collect data on the time between physician evaluation to the time of the cranial CT order. Therefore, we cannot measure the incremental effect of observation time on cranial CT use. Further research in this area will guide the recommended length of ED observation. Finally, because we did not have time stamps regarding the time of completion of case report forms, it is not clear if the decision for neuroimaging was made because of the progression of symptoms or other factors not recorded at the time the case report form was completed. As the data collected were static, we do not know if the data accurately represent the dynamic nature of the evolution of disease and clinicians’ decision making.

CONCLUSION

Planned observation, as a strategy for the management of minor blunt head trauma, is associated with lower cranial computed tomography use, even in a setting of low baseline computed tomography rates. Children at very low risk for clinically important traumatic brain injury do not benefit from observation as a strategy to lower cranial computed tomography use. Our results validate planned observation as an appropriate strategy for selected patients at intermediate and higher risk for clinically important traumatic brain injury, with lower adjusted odds of cranial computed tomography use and no significant difference in missed clinically important traumatic brain injuries.

We thank the participating families and emergency department staff at participating sites. We thank Sarah Dalton and Mary McCaskill (The Children’s Hospital at Westmead, Sydney); Jeremy Furyk (The Townsville Hospital, Townsville); and Louise Crowe (Murdoch Children’s Research Institute, Melbourne) for their involvement with obtaining the data and prior data analysis. We thank Daniel Tancredi, PhD (Departments of Pediatrics and Emergency Medicine at the University of California Davis School of Medicine, Sacramento, CA) for his review of the statistical analysis; and Amrita Singh (Brown University, Providence, RI) for critical review of the manuscript.

REFERENCES


and outcomes among children with isolated head trauma. Pediatr Emerg Care [2018; Epub ahead of print].


**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.13942/full

**Data Supplement S1.** Adjusted likelihood of cranial CT use by symptoms and signs, and for each PECARN TBI risk group.
Centers for Disease Control and Prevention Recommendations for Hepatitis C Testing: The Need to Adopt Universal Screening in an Appalachian Emergency Department

Elena M. Wojcik, MPH¹, Melinda J. Sharon, MPH¹, Stephen M. Davis, PhD¹,², Owen M. Lander, MD¹, and Carmen N. Burrell, DO¹,³

ABSTRACT

Background: The Centers for Disease Control and Prevention recommends screening baby boomers and high-risk patients for hepatitis C virus (HCV); however, the incidence of HCV is rapidly increasing among younger populations, and screening is limited by access to care and risk factor assessment. The purpose of this study was to evaluate characteristics of HCV antibody-positive (Ab⁺) and ribonucleic acid (RNA)-confirmed-positive patients identified via two screening models in an Appalachian emergency department (ED).

Methods: This was a retrospective cohort study of patients who screened HCV Ab⁺ in the ED from January 1 to October 31, 2018. Data were extracted, and comparative analyses were conducted between the risk-based and the universal screening models.

Results: Overall, 444 patients screened HCV Ab⁺, with a median age of 39 years. From January to May 2018, the risk factor model identified 126 HCV Ab⁺ patients out of 3,014 screened (4%), whereas from June to October 2018, the universal model identified 318 HCV Ab⁺ patients out of 5,407 screened (6%; p < 0.001). A consistently large proportion of diagnoses were new (71%). There was no statistically significant decrease between the RNA-confirmed-positive patients during the risk factor model (76, 60%) and universal model (186, 58%) time periods (p = 0.72). The models had high rates of reported intravenous drug use, and the universal screening adoption was modest at 33%.

Conclusion: This study was the first to present characteristics of HCV Ab⁺ and RNA-confirmed-positive patients identified during the transition to a universal screening model in an Appalachian ED. Most diagnoses were new regardless of screening model, but more patients screened HCV Ab⁺, and a similar proportion were RNA-confirmed-positive, under the universal model. Given that adoption of universal screening was modest, and risk factors remained similar, future research should investigate how to more effectively implement a universal screening model on a wider scale to identify early infections.
Hepatitis C virus (HCV) is the most common blood-borne infectious disease in the United States. The disease was listed as a contributing cause of death in over 19,000 individuals in the United States in 2014 and has outnumbered human immunodeficiency virus (HIV) as a cause of death since 2007. Currently, the two leading risk factors for HCV infection are intravenous drug use (IDU) and a birthdate between 1945 and 1965 (the “baby boomer” age cohort). With the increase in IDU nationally, the number of HCV infections has dramatically increased. Acute HCV infections have surged since 2009 in patients aged 20 to 39 years, with approximately 70% attributed to IDU. In 2015, there were approximately three million people living with HCV in the United States, and the HCV incidence has since continued to increase.

Surges of HCV incidence have been noted in the United States, primarily related to IDU. Regions east of the Mississippi river have been most heavily impacted, with high rates of new infections in the Central Appalachian region. Opioid abuse is often the gateway to injection drug use, and because many Central Appalachians are historically physical laborers (e.g., coal miners and loggers), pain relief is crucial to maintain daily activity. Many of these individuals begin to use prescriptions outside of physician recommendations and then turn to something more widely available, such as heroin. As a result, there are high rates of infectious diseases, such as HCV and HIV, transmitted among users. Central Appalachia’s drug use trends can be attributed to various factors, including low education levels, high unemployment levels, and high rates of job-related injuries. These factors can also be applicable in other regions.

With Central Appalachia at the center of the current opioid epidemic, West Virginia, Virginia, Kentucky, and Tennessee have a combined 364% increase in HCV in people under 30 years of age; IDU was the most common risk factor in 73% of these cases. Additionally, West Virginia has the third-highest incidence rate of HCV in the nation. With the concurrent increases in IDU and HCV infection, the two appear closely related. Current HCV screening recommendations, according to the Centers for Disease Control and Prevention (CDC), are listed in Table 1. However, solely utilizing risk factor screening can be inadequate due to many factors. Due to the number of preventative health services that primary care physicians must provide, some screenings may be overlooked due to assumed lack of patient risk factors. Failure of the patient to disclose medically relevant information to clinicians can also impair care or even lead to harm. Uncomfortable conversations may often be avoided, and patients may not be honest with their physicians. In fact, 60% to 80% of people in a recent study reported that they had withheld or lied to their health care provider about medically relevant information. In many cases, hepatitis infections are not identified until testing is warranted from advanced disease signs and symptoms.

It has been shown that individuals with risk factors for HCV are less likely to have access to health care screenings, as they are less likely to have a primary care provider and/or are uninsured or underinsured. Because one major risk factor of acquiring HCV is past or current IDU, there is also a stigma attached, which impacts an individual’s likelihood of being screened. Not only are individuals blamed for acquiring HCV because of drug use, but also it is strongly associated with HIV. However, these patients still often utilize emergency departments (EDs) for their health care needs, and universal screening in acute care may be needed to adequately capture all at risk individuals. HCV screening in the ED can increase the likelihood of patients receiving appropriate linkage to care and treatment. Since ED patients generally have longer lengths of stay, compared to ambulatory clinics, the HCV Ab+ screen is likely to be resulted prior to the patient discharge. This creates a one-stop opportunity for the emergency physician to identify HCV Ab+ patients and initiate the linkage to care process during a single patient visit. In addition, the electronic medical record has been shown to successfully facilitate emergency physicians conducting these screenings. One study found that an EMR best-practice alert (BPA) led to ED providers...
initiating HCV screenings more often. Further, upon receiving notification of an HCV Ab+ result, the EMR would trigger an automatic reflex to HCV confirmatory testing, thereby removing the task of additional order placement from the provider. Therefore, the ED could be a crucial location for infectious disease–related screenings.

The purpose of this study was to evaluate characteristics of HCV Ab+ and RNA-confirmed-positive patients identified via two different screening models that were implemented in an Appalachian ED: risk-based and universal screenings. Our primary hypothesis was that an increased proportion of Ab+ patients would be identified via the universal method of screening compared to the risk-based model. Additional hypotheses were 1) an increased proportion of RNA-confirmed-positive patients would be identified via the universal method of screening compared to the risk-based model and 2) an increased number of IDU would be identified via the universal model compared to the risk-based model. As these hypotheses were exploratory in nature, formal power calculations were not performed.

**METHODS**

**Study Design**

This ED was the recipient of a FOCUS program award from Gilead Sciences, Inc., which assists in the implementation and establishment of routine infectious disease–related screenings, such as HCV and HIV. Through this grant, the EMR was programmed to identify patients for both screening models by triggering a BPA for emergency physicians and advanced practice providers (APPs) to address during the ED visit. The HCV screening was provided at no cost to the patient through the program. Gilead Sciences, Inc., did not have an influence on generation of the manuscript or the design of this study. This study was reviewed and approved by the affiliated university’s institutional review board.

This study was performed in a cross-sectional manner. A retrospective chart review was conducted on all patients screening HCV Ab+ during two separate screening model implementation time periods by patient navigators (PNs), and data abstraction was a routine aspect of tracking patient screening efforts.

**Study Setting and Population**

This study was conducted at an academic-affiliated, Level I trauma, tertiary care ED in West Virginia that has approximately 50,000 patient visits per year and utilizes a large enterprise electronic health record system (Epic® 2015, Epic Systems Corporation). Participants in this study were all 18 years of age and over and must have presented to the ED at least once within the two screening model time frames.

**Screening Methods**

Two different screening models were utilized in this study. First, from January 1, 2018, to May 31, 2018, patients were identified either based on 1) CDC-established risk factors or 2) a list of triage complaints related to IDU upon presentation. A complete list of risk factors and triage complaints can be found in Table 2. From June 1, 2018, to October 31, 2018, a universal screening model was implemented, replacing the solely risk-based model. Patients who were

---

**Table 2**

<table>
<thead>
<tr>
<th>Triage Chief Complaints Captured by the EMR, Which Then Triggers an On-screen BPA, Prompting Emergency Physicians to Screen Eligible Patients for HCV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abscess</td>
</tr>
<tr>
<td>Epidural</td>
</tr>
<tr>
<td>Intracranial</td>
</tr>
<tr>
<td>Brain</td>
</tr>
<tr>
<td>Subdural</td>
</tr>
<tr>
<td>Acute drug overdose</td>
</tr>
<tr>
<td>Boils</td>
</tr>
<tr>
<td>Cellulitis</td>
</tr>
<tr>
<td>Toe</td>
</tr>
<tr>
<td>Foot</td>
</tr>
<tr>
<td>Chlamydia</td>
</tr>
<tr>
<td>Endocarditis</td>
</tr>
<tr>
<td>Gonorrhea</td>
</tr>
<tr>
<td>Infection</td>
</tr>
<tr>
<td>Bacterial</td>
</tr>
<tr>
<td>Skin</td>
</tr>
<tr>
<td>Blood</td>
</tr>
<tr>
<td>Neck</td>
</tr>
<tr>
<td>Staphylococcus</td>
</tr>
<tr>
<td>Wound</td>
</tr>
<tr>
<td>LSD overdose</td>
</tr>
<tr>
<td>Morphine overdose</td>
</tr>
<tr>
<td>Osteomyelitis</td>
</tr>
<tr>
<td>Forearm</td>
</tr>
<tr>
<td>Multiple sites</td>
</tr>
<tr>
<td>Opioid withdrawal</td>
</tr>
<tr>
<td>Opium overdose</td>
</tr>
<tr>
<td>Tranquilizer/antipsychotic overdose</td>
</tr>
</tbody>
</table>

BPA = best-practice alert; EMR = electronic medical record; HCV = hepatitis C virus.
18 years of age and older and had not been screened in the past year would trigger a BPA for the emergency provider, regardless of whether or not risk factors were presently documented in the EMR at the time of the visit. Patients with risk factors would trigger the BPA every 3 months to assess need for additional testing. Both screening models were opt-out for patients, allowing them to choose not to participate in the screening during their ED visit but making it a routine clinical practice. With both screening models, emergency physicians and APPs had the opportunity to decline testing, by selecting “not clinically appropriate” should they feel that the test was not relevant to the patient’s acute care needs or if additional laboratory orders were not being placed during the ED visit. Some patients were excluded from the screening at physician or APP discretion, which could include time-sensitive conditions such as trauma, myocardial infarction, or stroke.

In both screening models, patients with either an initial Ab+ screen or a confirmed active HCV infection were contacted for follow up by patient navigators or an APP. Should the patient have screened RNA confirmed positive, the patient’s provider would initiate the order for an infectious disease referral and inform the patient of the positive result. The BPA was triggered once yearly on all patients and quarterly for any patients deemed high risk. If a patient had an HCV Ab+ screen, repeat testing was often recommended in 6 to 8 weeks if not confirmed positive with RNA screening at the time. After all screening was completed, PNs automatically received results through the EMR, which were subsequently added to a master log for tracking purposes. Upon an HCV Ab+ result, results were entered into an HCV+ spreadsheet until confirmatory results were available, approximately 1 to 3 days, on average. Upon a positive confirmatory result, PNs followed up with patients via phone and provided the appropriate linkage to care, defined as follow-up contact with a primary care or specialty provider through consultation or medical appointment. In the event that a patient was HCV Ab+, but not RNA-confirmatory-positive, PNs would ask that patients be consulted by an APP within the infectious diseases department. This allowed the provider to assess risk factors, if any; order repeat screening if necessary; and educate patients on safe behavior to avoid potential reinfection in case of a previous exposure. APP consultation was considered to be successful linkage to care.

**Variables**

Patient navigators were trained on the EMR itself upon hire via the institution. Further training on extracting the appropriate data points was conducted by the investigators. Manually extracted variables from the EMR included age, gender, ethnicity, race, and risk factors (IDU, baby boomer cohort, and HIV coinfection). Manually extracted variables were not double coded.

**Statistical Methods**

Descriptive statistics were calculated for all variables. Differences between nominal sociodemographic variables (gender, race, and ethnicity) in the risk-based versus universal-based model were tested using the chi-square test. Fisher’s Exact tests were computed in cases where >20% of cells contained expected counts < 5. Differences in the continuous age variable were tested between groups using the t-test or the nonparametric Wilcoxon test. Evidence of normality was tested using the Shapiro-Wilk W-test with p values < 0.05 accepted as evidence of a nonnormal distribution.

Our primary hypothesis (increase in HCV Ab+) and each additional hypothesis (increase in RNA positivity and increase in identified IDU) were tested using the chi-square test. To minimize the threat of multiple comparisons, a Bonferroni correction was made by dividing the customary 0.05 alpha level by the number of planned hypotheses yielding an alpha threshold of 0.017 (0.05/3) for statistical significance. All statistical tests were calculated using JMP Pro® 14.0 (SAS, Inc.).

**Study Size**

For the purposes of this study, all eligible participants were populated automatically through the EMR based on risk factors present in the chart upon presentation, by age, and on patients without recent history of previous testing or identified active disease. Each patient who was screened for HCV was part of the study sample.

**RESULTS**

From January 1 to October 31, 2018, a total of 8,421 adult patients were screened for HCV out of 31,422 unique patient visits (27%). Of those, 444 patients (5%) screened HCV Ab+, which automatically reflexed to RNA confirmatory testing, which resulted in 262 (59%) patients who were RNA-confirmed-
positive. Overall, 264 (59%) of all Ab+ patients were linked to care. Of those not linked, the patients were lost to follow-up care, had complications affecting follow up, or refused the appointment. Most patients were male (275, 62%), white (405, 91%), and non-Hispanic (436; 98%). This is similar to the overall ED patient population, which is almost evenly split by gender and majority white and non-Hispanic. Overall, 323 (73%) of HCV Ab+ patients fell outside of the baby boomer age group, with a median age of 39 years. Additionally, 396 patients (89%) had at least one risk factor documented in the EMR; of those, 266 (67%) were found to be past or current IDU, leaving 48 patients (11%) who were identified outside of the CDC guidelines. Three patients (1%) were found to have an HIV coinfection, and 315 patients (71%) had documentation stating that this screening identified a new diagnosis of HCV, per chart review. The number of patients who had potential cleared infections was unable to be differentiated from possible false-positive testing due to the current testing method (Figure 1).

Risk-based Screening Model
From January 1 to May 31, 2018, a total of 14,968 unique adult patients presented to the ED; 21,043 total adult visits were made. Of those, 3,014 (20%) were screened for HCV and 126 patients (4%) screened HCV Ab+ when the physician or APP accepted the BPA during the risk-based screening model implementation. In this screening model, 59 (47%) of all Ab+ patients were linked to care. Of these, 76 (60%) were RNA-confirmed-positives. Most patients were male (75, 60%), white (113, 90%), and non-Hispanic (124, 98%). Eighty-six patients (68%) had documentation that this was a newly identified HCV Ab+ diagnosis.

Out of the 100 (79%) patients who had documentation of a risk factor in their chart, including age, IDU was the most common in this screening model (73, 73%), and 59 patients (59%) fell into the baby boomer cohort, indicating that 11 patients (11%) had both risk factors. Additional patients screened consisted of a potential risk factor identified through the patient chief complaint at triage (Table 2).

Universal Screening Model
From June 1 to October 31, 2018, a total of 16,454 unique adult patients presented to the ED, out of 22,200 total adult visits. Of those, 5,407 (33%) were screened for HCV when the BPA triggered based on age or lack of recent testing, and 318 patients (6%) screened HCV Ab+ during the universal screening model implementation, representing a 2% increase in HCV Ab+ identification from the previous risk-based screening model. Of these, 186 (58%) were RNA-confirmed-positive, representing a higher number of patients but slightly lower proportion. In this screening model, 205 (64%) of all Ab+ patients were linked to care. Most patients were male (200, 63%), white (292, 92%), and non-Hispanic (312; 98%). Again, this is similar to the ED patient population, which is almost evenly split by gender and majority white and non-Hispanic. A total of 229 patients (72%) were noted to have a newly identified HCV Ab+ diagnosis.

Out of the 283 (89%) patients who had any documentation of a risk factor in their chart, including age, IDU was the most common risk factor in the universal screening model (166, 59%), and 80 patients (28%) were noted to be in the baby boomer cohort.

Model Differences
There were no statistically significant differences observed among demographic variables between
models (Table 3). Overall, there was a statistically significant increase in detection of HCV Ab+ patients after transitioning from the risk-based model to the universal model (4% vs. 6%, \( p < 0.001 \)). Additionally, there was a nonsignificant decrease in RNA-confirmed-positive patients between the risk-based and universal screening models (60% vs. 58%, \( p = 0.72 \)). No statistically significant differences in IDU were observed from the risk-based model to the universal model (73% vs. 59%, \( p = 0.21 \)).

**DISCUSSION**

This study was the first to examine characteristics of HCV Ab+ patients identified during the transition from a risk factor-based screening model to a universal screening model in an Appalachian ED. As a result of the transition to universal screening, with similar overall number of ED patient volumes, an additional 192 HCV Ab+ diagnoses and 110 RNA-confirmed-positive patients were identified. The transition not only saw an increase of the total HCV Ab+ patients but also a statistically significant percentage increase (4% vs. 6%) and no statistically significant decrease in patients who were confirmed to have active infection (60% vs. 58%). Noting these numbers, it is evident that screening solely based on CDC risk factors is not currently sufficient for the Appalachian region. A universal model was more successful in identifying HCV Ab+ patients and had almost identical confirmatory rates (a broader screening approach would typically be expected to lead to a lower confirmatory rate). A similar study from an urban counterpart found that conducting a panel test consisting of HIV, hepatitis B virus, and HCV testing with a universal screening method was similarly a successful strategy for identifying HCV-positive patients on a wider scale.18

Many previously undiagnosed cases (315, 71%) were identified during the ED visits in this study. This number is higher than what has been previously documented in the literature and was consistent regardless of the screening model used.19 This is a particularly concerning finding, demonstrating the potential impact and public health risk of the opioid epidemic on the state of West Virginia and Appalachia. Approximately 75% to 85% of acute HCV infections develop into chronic HCV.2 Based on these results, it is possible that implementing a universal screening model, not only for HCV, but also other blood-borne diseases, has the potential to more effectively identify positive patients beyond the current CDC guidelines and decrease the public health impact by identifying early infections before complications arise. Furthermore, screening for HCV and receiving an Ab+/RNA-negative confirmatory result is important in identifying those who do not have a current active infection, but a potential previous infection and exposure. Patients who screened HCV Ab+ but were not confirmed positive for active infection present a public health opportunity to engage in assessing for risk factors,

| Table 3 Demographic Characteristics of HCV Ab+ Patients, Separated by Specific Model |
|-------------------------------|---------------------|---------------------|---------------------|-------|
| Characteristic               | Overall            | Risk-factor Cohort  | Universal Cohort    | \( p \)-value |
| Total number                 | 444                | 126 (28)            | 318 (72)            | <0.0001 |
| Gender                       |                    |                     |                     |       |
| Male                         | 275 (62)           | 75 (60)             | 200 (63)            | 0.5098 |
| Female                       | 169 (38)           | 51 (40)             | 118 (37)            |       |
| Age (years)                  | 39 (19–81)         | 39 (21–73)          | 40 (19–81)          | 0.8928 |
| Race                         |                    |                     |                     |       |
| White                        | 405 (91)           | 113 (90)            | 292 (92)            | 0.5093 |
| African American             | 20 (5)             | 5 (4)               | 15 (5)              |       |
| More than one                | 6 (1)              | 2 (2)               | 4 (1)               |       |
| Unknown                      | 13 (3)             | 6 (5)               | 7 (2)               |       |
| Ethnicity                    |                    |                     |                     |       |
| Non-Hispanic                 | 436 (98)           | 124 (98)            | 312 (98)            | 0.3353 |
| Hispanic                     | 1 (0.2)            | 1 (1)               | 0 (0)               |       |
| Unknown                      | 7 (2)              | 1 (1)               | 6 (2)               |       |
| RNA positive                 | 262 (59)           | 76 (60)             | 186 (58)            | 0.7242 |

Data are reported as \( n \) (%) or median (range). Ab+ = antibody positive.
completing repeat screening if necessary, and educating patients on safe behavior and harm reduction practices to avoid potential reinfection.

The ED is an ideal setting to increase access to routine screening services for high-risk populations that have poor access to primary care services. HCV screening in an ED, when bundled with other infectious disease-related screenings such as HIV, has been shown to be successful in multiple studies. Many urban locations have successfully implemented screenings into their respective ED settings, demonstrating the feasibility of this process. ED-based screening programs have also demonstrated that utilizing an “opt-out” format has been successful, where patients are screened unless they inform a provider they would not like to be tested.

Given that universal screening was not attained in this study despite opt-out, there may be additional barriers such as: 1) blood testing was not always needed during the ED visit and therefore the HCV screening tests could not be added and 2) provider compliance was not consistent. However, our rates of universal screening are comparable to the best rates of universal screening published to date. Continued emphasis on stigma-free screening in the ED has value if it can be made a routine part of the patients who have laboratory tests performed during their visit. Leveraging the EMR, HCV screening can be easily added to the order list without affecting standard of care or increasing the patients’ length of stay.

This study found that 89% of patients had at least one risk factor documented in their ED notes in contrast to another study conducted in an urban ED, which found that 51% of HCV Ab+ patients had a risk factor documented. We note that a larger proportion of patients with a documented risk factor was identified during the universal screening period, which potentially suggests that emergency providers could be sensitized to improving documentation after receiving repeated BPAs for HCV screening. In some cases, the provider may have added more thorough documentation after receiving the positive result and asking the patient additional questions concerning risk factors. Since the universal screening period followed the risk-based screening, it is likely that physician and APP documentation could have significantly improved.

Patient documentation of pertinent risk factors may be inaccurate or lacking altogether for multiple reasons. It is unknown if emergency physicians and APPs commonly pose adequate history questions to determine if a patient is high risk for HCV, as those risk factors may not be pertinent to the acute ED visit complaint. This may represent a missed opportunity to provide public health services through preventative health screenings during the patient visit. Moreover, if the patient is questioned about personal history of IDU or other risky behaviors, he or she may not feel comfortable giving open and honest responses to the physician, especially if he or she feels stigmatized for disclosure. One study found that over 80% of participants reported having ever avoided telling a clinician medically relevant information. Therefore, with the implementation of a universal screening protocol, unless the patient opts out of testing, all eligible patients can be screened with the potential of early identification of infection.

There were high rates of IDU present within both screening models; however, the study failed to identify increased IDU within the universal model. IDU has been established as the primary mode of transmission for HCV, with the highest seroprevalence among middle-aged people. Although IDU has been shown to account for approximately 80% of current infections, only 266 patients (67%) from our overall study sample were noted to have IDU as a risk factor. However, the median age of HCV Ab+ patients was 39 years, matching prior literature. With the current epidemic in Appalachia, it was expected that IDU would have higher rates than what was identified in this study, but it is unknown if stigma is a barrier to disclosure even after an HCV diagnosis is made. Therefore, further in-depth research is needed beyond this single-center retrospective study, to truly capture the interplay of the epidemics at large in the Appalachian region.

Multiple studies have found that universal screening is beneficial and cost-effective. A study by Assoumou et al. found that screening patients starting at age 15 was beneficial, especially in an IDU-prevalent population. They recommended using rapid testing for quicker turnaround times, but did not include a confirmatory result. Additionally, the study by Assoumou et al. found that screening, diagnosing, and treating previously unknown cases of HCV was more cost-effective than treating liver diseases caused by chronic HCV. Although there was reinfection in some patients, there were many cases in which reinfection was eliminated due to harm reduction approaches. Another study by Terrault concluded that universal screening is beneficial because providers could potentially be missing out on an entire group of people that do not meet risk factor recommendations. One-time screening of all adults aged 18 and over was recommended as well as repeat screening in high-risk groups such as IDU.
The United States Preventative Services Task Force (USPSTF) is currently considering updating its HCV screening recommendations because more people are becoming infected with HCV at a younger age.\textsuperscript{25} The USPSTF’s tentative recommendation is universal screening of individuals aged 18 to 79, and those younger than 18 or older than 79 if they are at high risk for infection.\textsuperscript{25} This more closely mirrors current HIV screening guidelines. These draft recommendations from the USPSTF strengthen what healthcare providers already know—universal screening of stigmatized diseases in adults is needed and warranted.

\section*{LIMITATIONS}

There are a number of limitations to this study. First, documentation of risk factors in the EMR may not be accurate and/or updated during the patient visit; therefore, a patient’s history may not accurately reflect the full discussion during their ED visit. It is possible that patients may not feel comfortable admitting to certain behaviors for fear of law enforcement or stigma. Second, there may be issues with the BPA. The physician or APP could choose to defer the alert and return to it later in the visit, order the screening at the time of the alert, or decline to order the test. If not ordering the screening test, the provider could choose either “not clinically appropriate” or “patient refused.” Providers suggested that “not clinically appropriate” was often chosen if the BPA appeared prior to the medical decision making when they were not yet placing routine orders or during times of acute stress, such as with a trauma patient. It is possible that the EMR did not correctly trigger the BPA consistently as this was reported at times by some providers. Additionally, residents involved in patient care may have gotten the BPA alert and may not always appropriately select screening. Data are currently being analyzed as to when each option was chosen to improve the BPA alert process. Third, laboratory tests are not able to distinguish between true HCV Ab\textsuperscript{+} and false-positive rates. False-positive results could decrease accurate identification of potential infections.

Due to the study implementation at a single ED site, the results may not be generalizable to other ED or acute care locations in the Appalachian region or nationwide. Specifically, this study is based in an ED that serves a largely rural population; therefore, recreating this model in urban or metropolitan areas may result in different findings. In addition, the two screening models were performed during different months of the calendar year, which may have introduced temporal bias, such as fluctuations in the number of patient visits due to holidays or seasonality changes. It is possible that these results are not representative of our entire ED population due to the inability to capture all eligible patients. Additional research is needed to make more accurate year-to-year comparisons while controlling for these factors. Finally, if a similar study was done on a wider scale, trained abstractors who are blinded to the study’s hypothesis should be employed in order avoid possible chart review bias or abstraction errors.

\section*{CONCLUSIONS}

This study was the first to present characteristics of hepatitis C virus antibody-positive and RNA-confirmed-positive patients identified during the transition from a risk factor–based screening model to a universal screening model in an Appalachian ED. The universal screening model identified 2\% more hepatitis C virus antibody-positive patients and identified a comparable proportion of RNA-confirmed-positive patients compared to the risk factor–based model. New diagnosis rates were consistently high, and rates of disclosure of intravenous drug use were high in both screening models. Future research should confirm the benefit of transitioning to a universal screening model in diverse ED settings, as well as identify effective implementation methods to a universal screening model in EDs.

The authors thank Justin Burns for his contributions to the completion of the manuscript. The authors also thank Dr. Kimberly Quedado for her contributions to the manuscript.

\section*{REFERENCES}


Shared Decision Making for Syncope in the Emergency Department: A Randomized Controlled Feasibility Trial

Marc A. Probst, MD, MS1, Michelle P. Lin, MD, MS1, Jeremy J. Sze, MA1, Erik P. Hess, MD, MPH2, Maggie Breslin, MDes3, Dominick L. Frosch, PhD4, Benjamin C. Sun, MD, MPP5, Marie-Noelle Langan, MD6, Venkatesh Thiruganasambandamoorthy, MBBS7, and Lynne D. Richardson, MD1

ABSTRACT

Objectives: Significant practice variation is seen in the management of syncope in the emergency department (ED). We sought to evaluate the feasibility of performing a randomized controlled trial of a shared decision making (SDM) tool for low-to-intermediate risk syncope patients presenting to the ED.

Methods: We performed a randomized controlled trial of adults (≥30 years) with unexplained syncope who presented to an academic ED in the United States. Patients with a serious diagnosis identified in the ED were excluded. Patients were randomized, 1:1, to receive either usual care or a personalized syncope decision aid (SynDA) meant to facilitate SDM. Our primary outcome was feasibility, i.e., ability to enroll 50 patients in 24 months. Secondary outcomes included patient knowledge, involvement (measured with OPTION-5), rating of care, and clinical outcomes at 30 days post-ED visit.

Results: After screening 351 patients, we enrolled 50 participants with unexplained syncope from January 2017 to January 2019. The most common reason for exclusion was lack of clinical equipoise to justify SDM (n = 124). Patients in the SynDA arm tended to have greater patient involvement, as shown by higher OPTION-5 scores: 52/100 versus 27/100 (between-group difference = 25.4, 95% confidence interval = 13.5 to 37.3). Both groups had similar levels of clinical knowledge, ratings of care, and serious clinical outcomes at 30 days.

Conclusions: Among ED patients with unexplained syncope, a randomized controlled trial of a shared decision-making tool is feasible. Although this study was not powered to detect differences in clinical outcomes, it demonstrates feasibility, while providing key lessons and effect sizes that could inform the design of future SDM trials.

1From the Department of Emergency Medicine; 2the Department of Emergency Medicine, University of Alabama at Birmingham, Birmingham, AL; 3the School of Visual Arts, New York, NY; 4the Palo Alto Medical Foundation Research Institute, Palo Alto, CA; 5the Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA; 6and the Division of Cardiology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 7and the Department of Emergency Medicine, The Ottawa Hospital, Ottawa, Ontario, Canada.

Received December 16, 2019; revision received February 25, 2020; accepted March 5, 2020.

Presented at the ACEP Research Forum, Denver, CO, October 2019.

Dr. Probst is supported by a Career Development Award from the National Heart, Lung, and Blood Institute of the National Institutes of Health (K23HL132052). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The sponsoring organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

The authors have no potential conflicts of interest to disclose.


Author contributions: MP designed the study and obtained funding; DF, EH, BS, and VT contributed to study design; MP, LR, and ML were responsible for data collection and management; JS and MP analyzed the data; MP, EH, and ML drafted the manuscript; all authors contributed substantially to manuscript revisions; MP takes responsibility for the paper as a whole; MP and JS had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis; and all authors approved the final report for submission.

Supervising Editor: Stephen P. Wall, MD.

Address for correspondence and reprints: Marc A. Probst, MD, MS; e-mail: mprobst@gmail.com.


© 2020 by the Society for Academic Emergency Medicine
doi: 10.1111/acem.13955

ISSN 1553-2712
Syncope is a chief complaint often encountered by clinicians working in the emergency department (ED), representing over 1.4 million visits per year and rising. The health care costs and resource utilization associated with syncope care are substantial. Even after extensive ED evaluation, involving laboratory, electrocardiogram, and radiologic testing, the actual cause of the syncope often remains unknown. Most causes of syncope are benign. Serious causes, including cardiac arrhythmias, acute coronary syndrome, and structural heart disease, are infrequent and difficult to diagnose in the ED. As a result, there is substantial variation in ED management and a tendency to admit intermediate-to-high-risk patients for further testing and monitoring. Continued evaluation, either in an observation or in an inpatient unit, are often low value and may not reflect the values and preferences of the patient.

Shared decision making (SDM) has been proposed as an approach to improve care for ED patients, including those with intermediate-risk syncope. SDM has been defined as a collaborative process in which patients and providers make health care decisions together, taking into account scientific evidence, the clinician’s experience, and the patient’s values and preferences. Although the benefits of SDM have been demonstrated in randomized trials involving patients with low-risk chest pain in the ED, a clinical scenario with some parallels to syncope, no studies have directly evaluated SDM for ED syncope care. We recently developed a personalized, paper-based decision aid to facilitate SDM for patients with unexplained syncope after ED evaluation. In 2016, the Academic Emergency Medicine (AEM) Consensus Conference on SDM in the ED called for further research on the development and testing of SDM interventions and identified syncope as a relevant clinical scenario. Priority research questions identified during this conference included how to communicate risk and uncertainty with patients, what tools can facilitate SDM, and what outcomes are suitable for SDM research in the ED.

Given the recent emphasis on patient-centered care in the ED, the frequency with which unexplained syncope is encountered, and the multiple reasonable management options to consider for this condition, it is important to study the effects of SDM for intermediate-risk syncope patients. Potential benefits include increased patient knowledge and satisfaction and decreased low-value health services, while the risks include unanticipated serious clinical outcomes. A decrease in low-value health services could potentially decrease health care costs and ED crowding, two important challenges facing the American health care system today.

Our objective was to perform a pilot, randomized clinical trial to test the feasibility and preliminary efficacy of using a paper-based patient decision aid to facilitate SDM around the disposition decision for low- to intermediate-risk syncope patients in a busy ED. We sought to determine if execution of such a study protocol was feasible given the frequency of syncope visits in the ED and our inclusion/exclusion criteria as well as to estimate the effect size of the intervention on various study outcomes to inform the design of future studies.

**METHODS**

**Study Design**

We performed a single-center, parallel randomized controlled trial of syncope patients. Patients in the intervention arm were randomized to SDM facilitated by a paper-based, personalized, syncope patient decision aid (SynDA), while those in the control group received usual care. Given the nature of the intervention (decision aid to facilitate SDM), patients and providers were not able to be blinded. However, outcome assessors were blinded to study arm assignment. Full-time, paid research coordinators and a project manager were trained on all aspects of the research protocol, including screening, obtaining consent, randomization, and follow-up prior to participating in study activities. This study was approved by the institutional review board at our medical center and registered at ClinicalTrials.gov (NCT02971163). This research was funded by a Career Development Grant from the National Institutes of Health, which had no role in the conduct or reporting of the study. As much as feasible, we adhered to the CONSORT criteria for randomized controlled feasibility trials. Post hoc, we evaluated where on the pragmatic-explanatory continuum our trial lay using the PRECIS-2 guidelines.

**Study Setting and Population**

This trial was performed at an urban, academic ED in a tertiary care medical center in the United States with an annual volume of over 100,000 visits, which serves a socioeconomically and ethnically diverse patient population.
population. We enrolled a convenience sample of adult patients (≥30 years) who presented with syncope and were deemed appropriate for SDM by the treating attending physician. Patients were recruited for 12 hours/day Monday through Friday and 8 hours/day on Saturday and Sunday over a 24-month period. The minimum age of inclusion was lowered from 40 to 30 years midway through the trial to increase enrollment. There was no upper limit of age for participants. Exclusion criteria included inability to read in English, major communication barrier (e.g., significant visual or auditory impairment, altered mental status), lack of fixed address and telephone number, and a serious diagnosis identified in the ED, as determined by the treating clinician. Clinicians were also able to exclude patients from the trial if they deemed them to be inappropriate for observation care based on their risk profile (e.g., too low risk to be offered observation care or too high risk to be offered direct discharge from ED) or not appropriate for SDM based clinical variables (e.g., dementia, altered mental status, intoxication, other disease process requiring admission). All clinicians (physician assistants [PAs], resident and attending physicians) were eligible to participate, signed informed consent, and were trained on the use of the decision aid to facilitate SDM during a 30-minute training course prior to the start of the trial.

Trained clinical research coordinators screened the electronic ED trackboard for eligible patients (looking for chief complaints of syncope, loss of consciousness, fall, dizziness, weakness, fainting, or passing out) and approached the treating clinician to verify eligibility before approaching the patients to explain the study and obtain written informed consent. After confirming the chief complaint (i.e., syncope), that the diagnostic evaluation had not revealed a serious diagnosis, and that the treating attending physician deemed the patient appropriate for SDM (i.e., has medical decision-making capacity and was neither too high nor too low risk), research staff randomized patients to the SynDA arm versus usual care. The was done after all of the results of ED testing had returned, with the exception of a second troponin test, if ordered. In that scenario, patients were informed that their continued eligibility was contingent on the second troponin results being negative. Randomization was achieved using sealed, opaque envelopes assigning patients on a 1:1 basis to SynDA or to usual care, using a variable block sequences of 4 to 6, and was stratified by age (less than or greater than 60 years) to ensure a balance of older patients in both groups. The random allocation sequence was created by a statistician at our institution.

**Interventions**

Patients were randomized to receive either usual care or a personalized syncope decision aid, named SynDA, designed to facilitate SDM (see Figure 1) around the disposition decision, i.e., direct ED discharge with primary care or cardiology follow-up, versus observation care. This paper-based decision aid was iteratively developed over a 1-year period with broad stakeholder input from patients, emergency physicians, a PA, a health care designer, and experts in the field of SDM research. This development process has been published previously.

Briefly, the tool uses simple language to explain what syncope is, what potential underlying conditions are still possible, what the patient’s risk of a 30-day serious medical event is, and what the possible disposition options are (i.e., direct discharge with primary care provider follow-up, direct discharge with cardiology follow-up, or admit to the observation unit).

Differing versions of the SynDA tool were used based on the patient’s individualized Canadian Syncope Risk Score and the corresponding estimated probability of a serious medical event within 30 days.11 “We chose the Canadian Syncope Risk Score based on the following factors: 1) rigorous developmental methodology, 2) largest sample size of any ED syncope risk score, and 3) stratified outcome with graduated levels of risk.” The components of the Canadian Syncope Risk Score were electronically entered into a database by the research assistant with the guidance of the treating clinician, generating the personalized risk estimate. Versions of the tool differed only in their displayed estimated risk, which ranged from 0.5% to 9%, communicated using percent, natural frequency, and a 100-person colored pictogram to maximize patient comprehension. If the calculated 30-day risk exceeded the comfort level of the attending physician, he or she had the option to exclude the patient from the study.

Participants in the usual care arm received management as per standard care for ED syncope patients, without any study intervention or decision aid. This may or may not have included some form of SDM with the patient, depending on the usual practice of the clinician involved. ED disposition for all patients was determined by the treating clinician and not by study protocol. This disposition discussion, with or without the SynDA tool, was performed either by the
1 Why did you pass out today? You were seen today in the Emergency Department because you passed out or fainted. Something also known as “syncope.” Passing out is usually caused by not enough blood flowing to your head.

All your test results are so far are normal. You do NOT have any evidence of a stroke, heart attack, or major blood loss. This is good news! Over half of the time, we do not figure out what caused you to pass out—even after lots of testing.

2 What is your future risk? It is still possible that you have an underlying condition such as:

- Abnormal heart rhythm (heart beat too fast or too slow)
- Problems with your heart muscle or valves
- Other problem with your blood circulation

Are you worried that passing out was a sign of something else?

3 Your Personal Risk Estimate Based on recent research of patients who came to the Emergency Department after passing out, there is in your risk of a serious medical event within the next 30 days.

What is your reaction to this risk?

4 What would you like to do? In making this decision with your doctor, consider the following questions. Do you have access to a doctor? Do you feel comfortable going home now? You have 4 options.

A. I would like to go home now and follow up with my doctor who will do further testing if needed.
B. I would like to see a Mount Sinai Cardiologist in 2-3 days.
C. I would like to stay in the hospital for heart monitoring and further testing if needed.
D. I would like my Emergency Department doctor to make this decision for me.

Based on the Canadian Syncope Risk Score, 2016

- Past Medical History
- Blood Pressure
- Specific Symptoms
- Blood tests
- Right side of electrocardiogram (electronic tracing of the heart)

Figure 1. SynDA: options for care after passing out. SynDA = syncope patient decision aid.
scale is a brief, 5-item instrument used by a trained observer to measure the degree of patient involvement in clinical decision-making and has demonstrated validity and reliability.32 The principal investigator (PI) trained two research coordinators on OPTION-5 scoring using an online course from the Preference Laboratory at The Dartmouth Institute and demonstration audio-recordings of patient and clinician actors. The demonstration audio-recordings were created by the PI using two volunteer patient actors not aware of the study hypothesis. Six separate audio-recordings were created to illustrate minimal, moderate, and maximal clinician effort toward patient involvement, with and without the decision aid. The two research coordinators listened to the demonstration recordings, while blinded to the intended effort, with the PI and rated and discussed each one until consensus was achieved. Disagreements were resolved by discussion with the PI. Both coordinators were blinded to study assignment when evaluating the study recordings. Intraclass correlation (ICC 3,1) for OPTION-5 scores on a subsample of eight recordings was calculated to measure inter-rater agreement. ICC is a means to measure the reliability of quantitative measurements made by different observers33 recommended by the creators of the OPTION-5 score.32 We used a two-way mixed-effects ICC model, where each subject is assessed once by each rater, and the selected raters are the only raters of interest.

To assess 30-day outcomes, research staff completed electronic chart review and contacted patients by telephone to confirm major adverse cardiac events, hospital admissions, ED visits, physician office visits, and further testing since discharge within 30 days of the index visit. Staff collecting this outcome were blinded to study assignment. Patients received one $25 gift card at the time of enrollment and a second $25 gift card upon completion of the 30-day follow-up telephone survey. All screened patients were reminded that participation in the study was entirely optional and that refusal would not affect the medical care they received.

Key Outcomes Measures
Our primary outcome was feasibility, defined as ability to reach target enrollment: 50 patients in 24 months. This sample size was based on the time and resources available for study completion and is consistent with prior ED feasibility studies.34-36 Secondary outcomes included patient knowledge, satisfaction, engagement (as measured using the validated OPTION-5 scale), decisional conflict (as measured using the validated Decision Conflict Scale)37, admission rates, and clinical outcomes at 30 days postindex visit. We also measured provider satisfaction with the SynDA tool.

Data Analysis
Descriptive statistics for baseline demographic and clinical variables were conducted. Secondary outcomes with continuous distributions were evaluated using independent sample t-tests and a two-sample test of proportions for dichotomous variables. For continuous variables, the Shapiro-Wilk normality and Anderson-Darling tests were used to determine if the continuous variables were normally distributed. For the continuous data that were not normally distributed, we used the Wilcoxon rank-sum test by groups. Point estimates and 95% confidence intervals (CIs) were calculated. Numerical and categorical data were collected with the Research Electronic Data Capture (REDCap) Web-based program (Vanderbilt University). Audio-recordings of the disposition discussion were collected using a tablet device with an embedded digital microphone. Analyses were conducted using STATA version 15.2 (StataCorp LP) with an intention-to-treat approach.

RESULTS
Characteristics of Study Subjects
After screening 351 patients with syncope, we enrolled 50 (14.3%) from January 2017 to January 2019 (see Figure 2). One patient was excluded after randomization since the etiology of their loss of consciousness was actually hypoglycemia, not syncope. The most common reason for exclusion (n = 124, 35.4%) was lack of clinical equipoise to justify SDM (i.e., patient too high/low risk). Among enrolled patients, 58% were female, and the median age was 58.5 years (range = 33–89 years). Thirty-four percent identified as white (Table 1). The two study arms were well balanced with regard to numeracy, health literacy, and education, but not with regard to race and sex.

Main Results
We were able to randomize 50 syncope patients from the ED, demonstrating feasibility. We stopped the trial after reaching our minimum target enrollment, n = 50. Our telephone follow-up rate at 30 days was 88% (44/50). Patients and providers both expressed satisfaction with respect to amount, clarity, and
helpfulness of the information shared (see Table 2). Overall rating of care by patients was similar in the two groups, 7.6/10 in the control group versus 8.3/10 in the intervention group (between-group difference = 0.7, 95% CI = −1.9 to 0.5). Patient knowledge was also similar in the two groups: 5.3/10 versus 5.2/10 in the control versus intervention group (between-group difference = 0.1, 95% CI = −0.9 to 1.1). Patient engagement tended to be higher in the SynDA arm as reflected by a higher mean OPTION-5 scores: 52/100 versus 27/100 (between-group difference = −25.4, 95% CI = −37.3 to −13.5), as shown in Table 3. OPTION-5 scores were found to be normally distributed using the Anderson-Darling test. Inter-rater agreement for OPTION-5 scores was found to be moderately high (ICC = 0.73) on the subsample of audio-recordings. The median duration (not normally distributed) for each the disposition discussion was 3:29 minutes in the control arm and 5:26 minutes in the SynDA arm (Wilcoxon two-sample test, p = 0.003).

Clinical Outcomes
Based on data from the telephone follow-up and the electronic chart review, the rate of new clinically important diagnoses at 30 days was 2/20 (10%) in the SynDA group and 2/24 (8.4%) in the control group (see Table 3). The two clinically important diagnoses in the SynDA group were an acute stroke and heart failure, while the two in the control group were a cardiac arrhythmia and a large pericardial effusion. There were no deaths in either group.

---

**Figure 2.** SynDA trial patient flow diagram. SynDA = syncope patient decision aid.
Resource Utilization

Resource utilization was similar in the two groups with no patients in either group being admitted to an inpatient service during the index visit. The rate of observation unit care was 26.9% in the control group and 16.7% in the SynDA group (risk difference = 10.2%, 95% CI = −12.4% to 32.9%). One patient in each group was admitted to the hospital during the 30-day follow-up period—one for acute stroke and one for a large pericardial effusion.

DISCUSSION

Our single-center, randomized controlled pilot trial evaluating a decision aid for patients with unexplained syncope demonstrated that such a study is feasible in the ED. Our study directly address research gaps identified during the 2016 AEM Consensus Conference on “Shared Decision Making,” including how best to communicate risk and uncertainty with patients, what tools can facilitate SDM, and what outcomes can and

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All, N = 50</th>
<th>SynDA Group, n = 24</th>
<th>Control Group, n = 26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>29 (58%)</td>
<td>12 (50%)</td>
<td>17 (65%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>58.5 (46.3–71.5)</td>
<td>54.5 (45.3–70)</td>
<td>60 (47.8–73)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black/African American</td>
<td>18 (36)</td>
<td>7 (29)</td>
<td>11 (42)</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>10 (20)</td>
<td>5 (21)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>White</td>
<td>17 (34)</td>
<td>10 (42)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (8)</td>
<td>2 (8)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Control Preferences Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 (6)</td>
<td>3 (13)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>2</td>
<td>15 (30)</td>
<td>8 (33)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>3</td>
<td>23 (46)</td>
<td>10 (42)</td>
<td>13 (50)</td>
</tr>
<tr>
<td>4</td>
<td>4 (8)</td>
<td>2 (8)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>5</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Numeracy score, mean (±SD)</td>
<td>12.9 (±2.6)</td>
<td>13.2 (±2.6)</td>
<td>12.6 (±2.6)</td>
</tr>
<tr>
<td>Health literacy score, mean (±SD)</td>
<td>5.3 (±2.7)</td>
<td>5.0 (±2.6)</td>
<td>5.5 (±2.8)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some high school</td>
<td>2 (4)</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>High school graduate/GED</td>
<td>7 (14)</td>
<td>3 (13)</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Some college</td>
<td>15 (30)</td>
<td>8 (33)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>College graduate</td>
<td>10 (20)</td>
<td>4 (17)</td>
<td>6 (23)</td>
</tr>
<tr>
<td>Graduate school</td>
<td>16 (32)</td>
<td>8 (33)</td>
<td>8 (31)</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;$20,000</td>
<td>9 (18)</td>
<td>2 (8)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>$20,000 to $29,999</td>
<td>5 (10)</td>
<td>1 (4)</td>
<td>4 (15)</td>
</tr>
<tr>
<td>$30,000 to $39,999</td>
<td>4 (8)</td>
<td>2 (8)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>$40,000 to $59,999</td>
<td>5 (10)</td>
<td>3 (13)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>$60,000 to $79,999</td>
<td>3 (6)</td>
<td>3 (13)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>$80,000 to $99,999</td>
<td>3 (6)</td>
<td>2 (8)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>$100,000 or more</td>
<td>9 (18)</td>
<td>4 (17%)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>Rather not say</td>
<td>12 (24)</td>
<td>7 (29)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>Canadian Syncope Risk Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low: 0 (≤2% risk)</td>
<td>9 (18)</td>
<td>3 (13)</td>
<td>6 (23)</td>
</tr>
<tr>
<td>Medium: 1 (≥3% risk)</td>
<td>28 (56)</td>
<td>13 (54)</td>
<td>15 (58)</td>
</tr>
<tr>
<td>Medium: 2 (≥5% risk)</td>
<td>13 (26)</td>
<td>8 (33)</td>
<td>5 (19)</td>
</tr>
</tbody>
</table>

Data are reported as n (%), median (IQR), or mean (±SD).
GED = general educational development; IQR = interquartile range; SynDA = syncope patient decision aid.
Table 2
Outcome Comparisons Between Participants Randomized to SynDA Tool Versus Control

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>SynDA Group (n = 24)</th>
<th>Control Group (n = 26)</th>
<th>Difference Between Groups, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient reported outcomes (index visit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amount of information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = too little information</td>
<td>3 (12.5)</td>
<td>1 (3.8)</td>
<td>–8.7 (–23.8 to 6.5)</td>
</tr>
<tr>
<td>2</td>
<td>0 (0.0)</td>
<td>2 (7.7)</td>
<td>7.7 (–2.6 to 17.9)</td>
</tr>
<tr>
<td>3 = just right*</td>
<td>15 (62.5)</td>
<td>21 (80.8)</td>
<td>18.3 (–6.3 to 43)</td>
</tr>
<tr>
<td>4</td>
<td>5 (20.8)</td>
<td>1 (3.8)</td>
<td>–17.0 (–34.8 to 0.9)</td>
</tr>
<tr>
<td>5 = too much information</td>
<td>1 (4.2)</td>
<td>0 (0.0)</td>
<td>–4.2 (–12.2 to 3.8)</td>
</tr>
<tr>
<td><strong>Clarity of information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = not clear at all</td>
<td>1 (4.2)</td>
<td>1 (3.8)</td>
<td>–0.3 (–11.2 to 10.6)</td>
</tr>
<tr>
<td>2</td>
<td>1 (4.2)</td>
<td>1 (3.8)</td>
<td>–0.3 (–11.2 to 10.6)</td>
</tr>
<tr>
<td>3 = somewhat clear</td>
<td>4 (16.7)</td>
<td>4 (15.4)</td>
<td>–1.3 (–21.6 to 19.1)</td>
</tr>
<tr>
<td>4</td>
<td>8 (33.3)</td>
<td>11 (42.3)</td>
<td>9.0 (–17.8 to 35.7)</td>
</tr>
<tr>
<td>5 = extremely clear*</td>
<td>10 (41.7)</td>
<td>9 (34.6)</td>
<td>–7.1 (–33.9 to 19.8)</td>
</tr>
<tr>
<td><strong>Helpfulness of information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = not helpful at all</td>
<td>0 (0)</td>
<td>2 (7.7)</td>
<td>7.7 (–2.6 to 17.9)</td>
</tr>
<tr>
<td>2</td>
<td>0 (0)</td>
<td>1 (3.8)</td>
<td>3.8 (–3.5 to 11.2)</td>
</tr>
<tr>
<td>3 = somewhat helpful</td>
<td>6 (25.0)</td>
<td>5 (19.2)</td>
<td>–5.8 (–28.8 to 17.2)</td>
</tr>
<tr>
<td>4</td>
<td>5 (20.8)</td>
<td>8 (30.8)</td>
<td>9.9 (–14.1 to 34.0)</td>
</tr>
<tr>
<td>5 = extremely helpful*</td>
<td>13 (54.2)</td>
<td>10 (38.5)</td>
<td>–15.7 (–43.0 to 11.6)</td>
</tr>
<tr>
<td><strong>Decisional conflict score (range = 0–10)</strong></td>
<td>8.9</td>
<td>9</td>
<td>0.1 (–0.8 to 1.0)</td>
</tr>
<tr>
<td><strong>Patient knowledge, mean (range = 0–9)</strong></td>
<td>5.2</td>
<td>5.3</td>
<td>0.1 (–0.9 to 1.1)</td>
</tr>
<tr>
<td><strong>Recommend the way of sharing information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = strongly recommend*</td>
<td>19 (79.2)</td>
<td>13 (50.0)</td>
<td>–29.2 (–54.3 to –4.0)</td>
</tr>
<tr>
<td>2</td>
<td>4 (16.7)</td>
<td>5 (19.2)</td>
<td>2.6 (–18.7 to 23.8)</td>
</tr>
<tr>
<td>3 = not sure</td>
<td>0 (0)</td>
<td>2 (7.7)</td>
<td>7.7 (–2.6 to 17.9)</td>
</tr>
<tr>
<td>4</td>
<td>1 (4.2)</td>
<td>2 (7.7)</td>
<td>3.5 (–9.5 to 16.5)</td>
</tr>
<tr>
<td>5 = strongly recommend against</td>
<td>0 (0)</td>
<td>3 (11.5)</td>
<td>11.5 (–0.7 to 23.8)</td>
</tr>
<tr>
<td>Refused</td>
<td>0 (0)</td>
<td>1 (3.8)</td>
<td>3.8 (–3.5 to 11.2)</td>
</tr>
<tr>
<td><strong>Overall rating of care, mean (range = 0–10)</strong></td>
<td>8.3</td>
<td>7.6</td>
<td>–0.7 (–1.9 to 0.5)</td>
</tr>
<tr>
<td><strong>Likelihood to recommend this ED</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definitely no</td>
<td>1 (4.2)</td>
<td>1 (3.8)</td>
<td>–0.3 (–11.2 to 10.6)</td>
</tr>
<tr>
<td>Probably no</td>
<td>1 (4.2)</td>
<td>1 (3.8)</td>
<td>–0.3 (–11.2 to 10.6)</td>
</tr>
<tr>
<td>Probably yes</td>
<td>8 (33.3)</td>
<td>8 (30.8)</td>
<td>–2.6 (–28.5 to 23.3)</td>
</tr>
<tr>
<td>Definitely yes*</td>
<td>14 (58.3)</td>
<td>15 (57.7)</td>
<td>–0.6 (–28.0 to 26.7)</td>
</tr>
<tr>
<td><strong>Provider-reported outcomes (index visit)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amount of information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = too little information</td>
<td>1 (3.8)</td>
<td>0 (0.0)</td>
<td>3.8 (–3.5 to 11.2)</td>
</tr>
<tr>
<td>2</td>
<td>3 (11.5)</td>
<td>1 (4.2)</td>
<td>7.4 (–7.3 to 22.0)</td>
</tr>
<tr>
<td>3 = just right*</td>
<td>19 (73.1)</td>
<td>18 (75.0)</td>
<td>–1.9 (–26.2 to 22.4)</td>
</tr>
<tr>
<td>4</td>
<td>3 (11.5)</td>
<td>5 (20.8)</td>
<td>–9.3 (–29.7 to 11.1)</td>
</tr>
<tr>
<td>5 = too much information</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Clarity of information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = not clear at all</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>1 (4.2)</td>
<td>1 (3.8)</td>
<td>–0.3 (–11.2 to 10.6)</td>
</tr>
<tr>
<td>3 = somewhat clear</td>
<td>0 (0.0)</td>
<td>11 (42.3)</td>
<td>42.3 (23.3 to 61.3)</td>
</tr>
<tr>
<td>4</td>
<td>4 (16.7)</td>
<td>9 (34.6)</td>
<td>17.9 (–5.6 to 41.5)</td>
</tr>
<tr>
<td>5 = extremely clear*</td>
<td>1 (4.2)</td>
<td>5 (19.2)</td>
<td>15.1 (–2.1 to 32.2)</td>
</tr>
</tbody>
</table>

(Continued)
should be measured in emergency care SDM research. Although our trial was not powered to detect differences in clinical, utilization, or patient-oriented outcomes, our results suggest a tendency toward increased patient engagement, lower resource utilization, and no difference in clinical outcomes with use of the SynDA tool. The relatively low enrollment rate of two to three patients per month, even after age range expansion, does speak to the substantial effort that is needed to complete SDM trials in a busy, academic ED with medically complex patients. Since the clinicians were the ones asked to perform SDM with the patients, we needed their approval and were not able to mandate inclusion in the trial using standardized criteria. Our results provide an estimate of the effect size of the intervention on various study outcomes and can be used to inform power calculations for future studies. The primary outcome of interest in a future trial of this nature would be difference in observation unit rates across the two study arms. A minimally clinically important difference would be approximately 5%. Key secondary outcomes would include differences in OPTION-5, patient knowledge, and patient satisfaction, along with clinical safety outcomes. We believe a MCID for OPTION-5, patient knowledge, and patient satisfaction scores to be 15/100, 1.5/10, and 1/10, respectively. To recruit an adequate sample size for a definitive trial, a significant investment of time and resources with multiple recruitment centers would be needed, preferably with some of these being high-volume community EDs caring for less medically complex patients.

The proportion of patients sent to the observation unit was lower in the SynDA arm by roughly 10%, corresponding to a 10% higher discharge rate, with no difference in hospitalizations. This suggests that, when offered the option, ED patients may be more likely to choose to be discharged home. This is consistent with prior studies showing that ED patients often have greater risk tolerance than expected after being fully informed of the risks, and consistent with systematic review suggesting that SDM tools can reduce resource utilization in non-ED settings. Hess et al. reported that the Chest Pain Choice tool was safe and effective for use in this clinical scenario.

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>SynDA Group (n = 24)</th>
<th>Control Group (n = 26)</th>
<th>Difference Between Groups, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helpfulness of information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = not helpful at all</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2 (8.3)</td>
<td>0 (0.0)</td>
<td>−8.3 (−19.4 to 2.7)</td>
</tr>
<tr>
<td>3 = somewhat helpful</td>
<td>5 (20.8)</td>
<td>16 (61.5)</td>
<td>40.7 (15.9 to 65.5)</td>
</tr>
<tr>
<td>4</td>
<td>10 (41.7)</td>
<td>7 (26.9)</td>
<td>−14.7 (−40.8 to 11.3)</td>
</tr>
<tr>
<td>5 = extremely helpful*</td>
<td>7 (29.2)</td>
<td>3 (11.5)</td>
<td>−17.6 (−39.6 to 4.3)</td>
</tr>
</tbody>
</table>

Would you want to present information about other chief complaints/clinical decisions in the same way?†

<table>
<thead>
<tr>
<th></th>
<th>N/A</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = yes, for sure*</td>
<td>8 (33.3)</td>
<td>8 (33.3)</td>
</tr>
<tr>
<td>2</td>
<td>8 (33.3)</td>
<td>5 (20.8)</td>
</tr>
<tr>
<td>3 = not sure</td>
<td>3 (12.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>4</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

Would you recommend to other providers the way that you presented information today?†

<table>
<thead>
<tr>
<th></th>
<th>N/A</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = yes, strongly recommend*</td>
<td>13 (54.2)</td>
<td>8 (33.3)</td>
</tr>
<tr>
<td>2</td>
<td>8 (33.3)</td>
<td>2 (8.3)</td>
</tr>
<tr>
<td>3 = not sure</td>
<td>1 (4.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>4</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

Data are reported as n (%), unless otherwise specified. N/A = not applicable; SynDA = syncope patient decision aid. *Desired end of the scale. †SynDA group only.
Similar to our study, the OPTION-5 scores were higher in the intervention arm in these two trials. Our study adds to the growing body of literature demonstrating that shared decision making is feasible in emergency care and may decrease resource utilization.\textsuperscript{14,42-45}

Patients and providers appeared to be generally satisfied with use of the decision aid to facilitate shared decision making. Patients in both groups tended to offer high ratings with respect to the amount, clarity, and helpfulness of information shared. On average, use of the decision aid was associated with an additional 2 minutes of discussion time; all potential benefits must be weighed against this increased time. Surprisingly, we did not observe an increase in patient knowledge, as was seen in the two chest pain trials. There also did not appear to be a tendency toward greater patient satisfaction in the SynDA group, a result we had not anticipated. Three possible explanations for this are: 1) selection bias—patients who were generally pleasant and agreeable may have been more likely to participate in the trial; 2) the Hawthorne effect may have caused providers to be particularly kind and attentive when discussing disposition options at the end of the visit; and 3) subjects enrolled in randomized controlled trials often get increased attention from both study staff and clinicians, which may increase satisfaction.

Our enrollment rate of 14%, while relatively low, was roughly as expected given the multiple reasons for exclusion including the patient being too low risk for consideration for the observation unit, and conversely, being so high risk as to merit admission. SDM in this scenario is only appropriate when two disposition options are medically reasonable and the patient has decision-making capacity.\textsuperscript{14} Nonetheless, there is still a substantial proportion of patients for whom direct discharge or observation are both reasonable, and for these, shared decision-making is most likely to have the greatest benefit. Of note, the adverse event rates of 10% and 8% were similar in the intervention and control groups, which confirms that the enrolled subjects were of intermediate risk and suggests that the decision aid is safe for use in the ED. These rates were as expected given our exclusion of low-risk patients. The two clinically important diagnoses in each arm were felt to be potentially related to the initial presentation of syncope by the study team.

### Table 3
Utilization and Clinical Outcomes of Participants Randomized to SynDA Tool Versus Control

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>SynDA Group ( (n = 24) )</th>
<th>Control Group ( (n = 26) )</th>
<th>Difference Between Groups, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utilization outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admitted to hospital</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0</td>
</tr>
<tr>
<td>Sent to observation unit</td>
<td>4 (17)</td>
<td>7 (27)</td>
<td>10.3 (−13 to 32)</td>
</tr>
<tr>
<td>Discharged from ED</td>
<td>20 (83)</td>
<td>19 (73)</td>
<td>−10.3 (−32 to 13)</td>
</tr>
<tr>
<td>CT scan</td>
<td>1 (4.2)</td>
<td>1 (3.8)</td>
<td>−0.3 (−1.6 to 1.5)</td>
</tr>
<tr>
<td>30-day follow-up period</td>
<td>( n = 20 )</td>
<td>( n = 24 )</td>
<td></td>
</tr>
<tr>
<td>Repeat ED visit</td>
<td>1 (5.0)</td>
<td>2 (8.3)</td>
<td>3.3 (−16 to 21)</td>
</tr>
<tr>
<td>Office visit</td>
<td>14 (70)</td>
<td>17 (71)</td>
<td>0.8 (−24 to 27)</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>1 (5.0)</td>
<td>1 (4.2)</td>
<td>−0.8 (−20 to 16)</td>
</tr>
<tr>
<td><strong>Clinical outcomes at 30 days</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New significant diagnosis</td>
<td>2 (10)</td>
<td>2 (8.4)</td>
<td>−1.7 (−23 to 17)</td>
</tr>
<tr>
<td>Acute stroke</td>
<td>1 (5.0)</td>
<td>0 (0)</td>
<td>−5 (−24 to 9.4)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>1 (5.0)</td>
<td>0 (0)</td>
<td>−5 (−24 to 9.4)</td>
</tr>
<tr>
<td>Cardiac arrhythmia</td>
<td>0 (0)</td>
<td>1 (4.2)</td>
<td>4.2 (−12 to 20)</td>
</tr>
<tr>
<td>Large pericardial effusion</td>
<td>0 (0)</td>
<td>1 (4.2)</td>
<td>4.2 (−12 to 20)</td>
</tr>
<tr>
<td>Mortality</td>
<td>0 (0)</td>
<td>0 (0%)</td>
<td>0%w</td>
</tr>
<tr>
<td><strong>Patient involvement (index visit)</strong></td>
<td>( n = 22 )</td>
<td>( n = 22 )</td>
<td></td>
</tr>
<tr>
<td>OPTION-5 Score, mean (±SD)</td>
<td>52.0 (±18)</td>
<td>26.6 (±21)</td>
<td>−25.4 (−13.5 to −37.3)</td>
</tr>
<tr>
<td>[range = 0–100 (Audio recording)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of discussion (minutes:seconds), median</td>
<td>3:40, 6:56</td>
<td>2:25, 5:50</td>
<td>Wilcoxon two-sample test, ( p = 0.003 )</td>
</tr>
</tbody>
</table>

Data are reported as \( n \) (%), unless otherwise specified.

OPTION-5 = Observing Patient Involvement in Decision-Making.
Our study may be helpful to other researchers interested in studying SDM in other ED clinical scenarios, such as transient ischemic attack or minor stroke, imaging for possible appendicitis or acute flank pain, and testing for possible subarachnoid hemorrhage, among others. Researchers would be wise to invest early in training clinicians in SDM, and involve a broad coalition of stakeholders including residents, faculty, physician assistants, and nurses. Financial incentives for providers (as well as patients) may be helpful in increasing enrolment rate, if approved by the institutional review board. To illustrate where our trial lies on the pragmatic-explanatory continuum, we have included a PRECIS-2 table and figure in Data Supplement S1, Appendices S3 and S4, respectively.25

LIMITATIONS

There are a number of important limitations to our study. Our small sample size precluded any definitive statistical comparisons between patients in each study arm. We included only English-speaking patients from a single, busy, urban, academic ED; our results may not apply to other patient populations. Patients and providers were not blinded to the intervention, which may have biased the results. As mentioned above, blinding for a SDM intervention is not feasible. All patient encounters were either directly observed or audio-recorded by research staff, which created the potential for a Hawthorne effect. However, since this direct observation occurred in both study arms, it should not have led to differences between the two. Our enrollment rate of roughly 14% has the potential for selection bias and limits generalizability, which was anticipated given that SDM is only appropriate for select ED patients under certain clinical circumstances. Due to random chance and our small sample, there were more women and more African Americans in the control arm; it is unlikely the imbalance had an effect on our results. Since clinicians in the control arm were not prohibited from using SDM, it is possible that some contamination occurred. However, such contamination would likely bias results toward the null hypothesis. In addition, 12% of patients were lost to follow-up at 30 days (four in the SynDA arm and two in the control arm), which is lower than expected compared to studies of this duration; it is unlikely that higher follow-up rates would change the statistical significance of our findings.

CONCLUSIONS

In summary, our study demonstrates that a randomized controlled trial of a shared decision making intervention for unexplained syncope patients is feasible in the ED. Although our results are promising with regard to improved patient engagement and resource utilization, a larger randomized trial would be necessary to determine the safety and efficacy of such an intervention. Our study provides key lessons with respect to feasibility, enrollment strategies, and effect sizes that could inform the design of future shared decision-making trials conducted in the ED.

The authors acknowledge the research assistants (Diana Gregoriou, Melika Behrooz, Daniela Garcia, Kavey Vidal, Erisa Shehi, Aria Mattias, Luisa Ortiz), faculty, residents, and physician assistants at the Mount Sinai Hospital, Department of Emergency Medicine, for their dedication to this project.

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.13955/full

Data Supplement S1. Supplemental material.
The Utility of Focused Assessment With Sonography for Trauma Enhanced Physical Examination in Children With Blunt Torso Trauma

Aaron E. Kornblith, MD¹, Jahanara Graf, MD², Newton Addo³, Christopher Newton, MD⁴, Rachael Callcut, MD, MSPH⁵, Jacqueline Grupp-Phelan, MD, MPH¹, and David M. Jaffe, MD³

ABSTRACT

Objectives: Computed tomography (CT), the reference standard for diagnosis of intraabdominal injury (IAI), carries risk including ionizing radiation. CT-sparing clinical decision rules for children have relied heavily on physical examination, but they did not include focused assessment with sonography for trauma (FAST), which has emerged into widespread use during the past decade. We sought to determine the independent associations of physical examination, laboratory studies, and FAST with identification of IAI in children and to compare the test characteristics of these diagnostic variables. We hypothesized that FAST may add incremental utility to a physical examination alone to more accurately identify children who could forgo CT scan.

Methods: We reviewed a large trauma database of all children with blunt torso trauma presenting to a freestanding pediatric emergency department during a 20-month period. We used logistic regression to evaluate the association of FAST, physical examination, and selected laboratory data with IAI in children, and we compared the test characteristics of these variables.

Results: Among 354 children, 50 (14%) had IAI. Positive FAST (odds ratio [OR] = 14.8, 95% confidence interval [CI] = 7.5 to 30.8) and positive physical examination (OR = 15.2, 95% CI = 7.7 to 31.7) were identified as independent predictors for IAI. Physical examination and FAST each had sensitivities of 74% (95% CI = 60% to 85%). Combining FAST and physical examination as FAST-enhanced physical examination (exFAST) improved sensitivity and negative predictive value (NPV) over either test alone (sensitivity = 88%, 95% CI = 76% to 96%) and NPV of 97.3% (95% CI = 94.5% to 98.7%).

Conclusions: In children, FAST and physical examinations each predicted the identification of IAI. However, the combination of the two (exFAST) had greater sensitivity and NPV than either physical examination or FAST alone. This supports the use of exFAST in refining clinical predication rules in children with blunt torso trauma.

Trauma is the leading cause of death and disability in children¹ and unrecognized injury from blunt torso trauma is the third most common cause of preventable death in children sustaining blunt trauma.²
Computed tomography (CT) remains the reference standard for diagnosis of intraabdominal injury (IAI), but carries risk, including ionizing radiation. In the past 25 years, there has been a dramatic increase in the number of CT scans performed on children. However, the prevalence of IAI in children undergoing CT scan is only 6% to 15%, making many of these studies potentially avoidable by using safer screening methods.

Evidence-based clinical decision rules have been created to limit CT scans in children with blunt torso trauma. These rules use medical history, clinical findings, and diagnostic testing to identify children with blunt torso trauma who have low risk for IAI and could safely avoid CT scans. However, the two most rigorous and recent prediction rules were developed without the focused assessment with sonography for trauma (FAST) examination, because many study sites were lacking expertise and a standard implementation pathway.

Focused assessment with sonography for trauma is a bedside sonographic evaluation performed within minutes of the patient’s arrival by the treating provider, in conjunction with the physical examination, to rapidly identify intraperitoneal traumatic hemorrhage. FAST has been successfully integrated into the management of injured adult patients where it improves clinical outcomes including reducing CT scan usage. FAST has the potential to be an ideal diagnostic test in the evaluation of IAI in children given its portability, cost efficiency, lack of ionizing radiation, and rapid performance. However, previous studies have shown variability in the test characteristics of FAST in children, with reported sensitivities of 20% to 80% and specificities of 77% to 100%. There are several reasons for this variability in test characteristics including age of patient, injury type, timing of study, provider expertise in image acquisition and interpretation, confidence, and lack of implementation pathways. However, in the past decade there has been substantial improvement in the availability, education, and overall expertise in the use of pediatric-specific point-of-care ultrasound applications including FAST. With the increase in expertise of providers using pediatric emergency ultrasonography during the past decade, we believe FAST may have increasing utility and deserves a reevaluation in the assessment of children with blunt torso trauma as a strategy to reduce the use of CT scans. Therefore, we sought to determine the incremental independent associations of physical examination, laboratory studies, and FAST in predicting IAI in children presenting with blunt torso trauma and to compare the test characteristics of these predictors. We hypothesized that adding FAST would enhance the ability to predict IAI in the diagnostic approach in children with blunt torso trauma.

MATERIALS AND METHODS

Study Design and Population
We performed a retrospective chart review of the trauma database and EMRs of children younger than 18 years of age presenting to an academic, urban, free-standing Level I pediatric trauma center during a 20-month period from November 1, 2013, to July 31, 2015, when FAST was routinely used for children presenting with blunt torso trauma. This study was performed at a tertiary referral center for Northern California with an annual census of approximately 50,000 visits. The pediatric emergency department (ED) is staffed by board-certified pediatric emergency providers and has a full complement of specialists providing pediatric and trauma care. The hospital is an American College of Surgeons verified Level I trauma center. The UCSF Benioff Children’s Hospital Oakland institutional review board approved this study.

Trauma Database. As part of the quality improvement program for trauma services, all patients regardless of disposition requiring a pediatric trauma evaluation in the ED were included in a database. Using this database, we included children younger than 18 years of age who met institutional trauma activation criteria and who also had a FAST performed. Children who had two or more trauma visits during the study period were included only once, and their first visit was selected. Mechanism of injury and International Classification of Disease, Ninth Revision (ICD-9) codes were obtained from the trauma database. Patients with penetrating injury or those who were transferred from a referring institution were excluded. Penetrating trauma was defined as any child with a presenting complaint related to penetrating injury (i.e., stab wound, gunshot wound) or final ICD-9 diagnosis of penetrating injury (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.13959/full). The trauma database was then linked to the electronic medical record (EMR) and subsequently
reviewed by two independent expert reviewers (AK, JF) using a standard data abstraction document. Any disagreement between the reviewers was given to a third independent expert for review (CN). Chart abstraction included patient demographics, physical examination findings, laboratory tests, radiographic findings, and the study’s primary and secondary outcomes.

Predictor and Outcome Variables. Demographic and clinical information was collected, including age, sex, weight, mechanism of injury, Injury Severity Score, ED vital signs (respiratory rate, pulse rate, mean blood pressure), Glasgow Coma Score (GCS), ED disposition, ED length of stay (hours), and hospital length of stay (days). Predictor variables included physical examination findings, ED laboratory results, and FAST. Outcome variables collected included CT scan report, hospital course, procedures performed (ICD-9), and final diagnosis (ICD-9).

All physical examination findings of blunt torso trauma, including thoracic wall trauma, abdominal wall trauma, or abdominal tenderness, as documented by the emergency and trauma providers, were included in the analysis. An abnormal examination was defined a priori as outlined by Holmes et al.10 as evidence of thoracic wall injury (erythema, abrasion, ecchymosis, subcutaneous air, or laceration to the anterior or posterior chest wall; chest auscultation for absent or decreased breath sounds; thoracic tenderness; costal margin tenderness), abdominal injury (erythema, abrasion, ecchymosis, laceration, seat belt sign to the abdominal wall; abdominal tenderness; abdominal dissection; abdominal auscultation for bowel sounds; peritoneal irritation; flank tenderness), or pelvic injury (pelvic bone tenderness or instability on palpation). Data were collected using a standardized data collection sheet. A physical examination was considered positive if any of the above findings were recorded as positive. Missing or omitted physical examinations were considered as negative. If data points were conflicting or ambiguous, they were reviewed by our third blinded author to reach consensus (CN).

Elevated liver enzymes were defined as abnormal if either serum aspartate aminotransferase (AST) was >200 IU/L or serum alanine aminotransferase (ALT) was >125 IU/L. Elevated pancreatic enzymes were defined as abnormal if there was an elevation of the serum lipase ≥100 IU/L.

Focused assessment with sonography for trauma examinations were performed and interpreted by the treating providers. Bedside interpretations by the treating provider were used as the study’s FAST interpretation. All providers participating in this study had undergone formalized didactic training and hands-on instruction. Completeness of FAST was based on the guideline published by the American College of Emergency Physicians.24 Treating providers performed standard FAST including views of the right upper quadrant, the splenorenal fossa, the long and short axis of the pelvis, and cardiac subxyphoid views. All FAST images were uploaded and backed up on a server for routine quality assurance (overread review), by the emergency ultrasound director asynchronously. Agreement between the bedside provider and the overread was measured with a Cohen’s kappa coefficient (κ, 95% CI) for interrater agreement. In addition, we reviewed the CT scan interpretations for children with IAI who had both negative physical examination and FAST.

Outcome Measures
The primary outcome measure was the presence of IAI. The secondary outcome was the presence of IAI requiring acute intervention. The reference standard included CT scan or review of the medical record. CT scan results were those from the final interpretation by attending radiologist. The trauma database and hospital EMR were also reviewed for IAI. IAI has been defined by Holmes et al.10 as any radiographically or surgically apparent injury to the spleen, liver, urinary tract, gastrointestinal tract, pancreas, gallbladder, adrenal gland, intraabdominal vascular structure, or traumatic fascial defect. Similarly, the trauma database and the EMR of hospitalized patients were reviewed to identify patients who had IAI requiring acute intervention, defined as any IAI requiring blood transfusion for anemia resulting from hemorrhage, surgery, angiographic embolization, or intravenous fluids for 2 or more nights in patients with pancreatic or gastrointestinal injury.10

Data Analysis
For reporting the characteristics of the patient population, continuous variables were reported as median and interquartile ranges and categorical variables were presented as frequency and proportions. Predictor variables for univariate regression were chosen based on theoretical and previously demonstrated clinical significance. We determined a priori that 50 positive FAST studies could allow as many as five predictor variables
to be included in our multivariate analysis. Univariate regression was used to evaluate each variable's association with identification of IAI in the present study. Multivariate logistic regression modeling was used to further assess the association of each predictor with IAI. Variables that showed an association with a p-value of <0.2 were considered for multivariate logistic regression and included or eliminated using backward stepwise selection. We used an adjusted alpha level of <0.01 for the multivariate regression analysis. All continuous covariates were reviewed for influential data points and transformed or normalized as appropriate. Patients missing laboratory data were not included in the multivariate analysis. Measurement of goodness of fit was evaluated using Hosmer-Lemeshow test. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and positive and negative likelihood ratios (LR+/LR−) of included examinations were calculated using conventional definitions. All other statistical tests were two-sided with a p-value of <0.05 considered significant. Analyses were performed using SAS 9.4 (SAS Institute).

RESULTS

We identified 1,125 unique initial encounters from November 1, 2013, to July 31, 2015 (Figure 1). We excluded 21 patients (2%) older than 18 years of age and 25 patients (2%) who sustained penetrating injuries. Of the remaining children with blunt torso trauma, 725 patients (67%) did not receive a FAST because of lack of expertise or because FAST images were not available for review. Of the 354 remaining patients who were included in the analysis, more than 70% were younger than 12 years of age (Table 1). Nearly 70% of the sample were boys, and the majority had a mechanism of injury of either vehicle collision (motorized or nonmotorized) or fall from height. The majority of children in this cohort were admitted to the hospital, 287 patients (81%). There were 254 patients (72%) had full results for ALT and AST in addition to the examination data.

Fifty patients (14% of the cohort) had IAI while 304 patients (86%) did not. The age, sex, weight and GCS were similar between those groups with and without IAI (Table 1).

We identified the following variables as associated with the identification of IAI in children with blunt torso trauma: physical examination, FAST, liver enzymes, and lipase. Univariate analysis (Table 2) showed that FAST, physical examination, and abnormal liver enzymes were each associated with the identification of IAI in children with blunt torso trauma. Lipase, however, was not considered a good candidate variable for this model as it demonstrated multicollinearity with liver enzymes and its values were missing for 37% of the population. In the multivariate regression model incorporating FAST, physical examination, and AST and ALT test results, only a positive FAST and physical examination continued to be significantly associated with identification of IAI in children with blunt torso trauma (Table 2). The overall model performance showed Hosmer and Lemeshow of χ² = 1.52, p-value of 0.68, which indicates acceptable goodness of fit. We did not find significant outliers in any of the model covariates. Based on this analysis, we created a composite variable, FAST-enhanced physical examination (exFAST). If both FAST and physical examination were negative, the exFAST was negative. If at least one of the two variables was positive, the exFAST was positive.
We then determined the test characteristics of the exFAST for identifying IAI in children. Of those patients with an IAI, 44 (88%) were detected by the exFAST and six (12%) were missed (Table 3). Of the patients without IAI, 87 (29%) had positive and 217 (71%) had negative exFAST. The sensitivity of the exFAST was 88% (95% CI = 75.7 to 95.5), specificity was 71% (95% CI = 65.9 to 76.4), PPV was 33.6% (95% CI = 10.7 to 18.2), and NPV was 97.3% (95% CI = 94.5 to 98.7). The LR+ was 3.07 and the LR− was 0.17. Of the 50 patients with IAI, 30 (60%) had a physical examination and FAST that were positive. Seven patients (14%) were only positive on FAST and an additional seven patients (14%) were only positive on physical examination.

Six patients (12%) had IAI not identified by the exFAST, of whom none had a IAI requiring acute intervention, but all had some amount of intraperitoneal fluid on CT (Table 4). Four of these patients had erroneously negative bedside FAST, which were subsequently overread FAST as positive. These four children also had intraperitoneal fluid identified on CT scan. When recalculating the exFAST test characteristics using the overread FAST, the sensitivity of the test improved to 96% (95% CI = 86.3 to 99.5), the specificity was 78% (95% CI = 73.2 to 82.8), the PPV was 42.1% (95% CI = 36.8 to 47.6), and the NPV was 99.2% (95% CI = 96.8 to 99.8). The interrater reliability for provider and overread FAST was fair (κ = 0.40, 95% CI = 0.29 to 0.51; Table 4).

One of the two patients with a negative exFAST and negative overread was a 7-year-old girl with a pancreatic tail contusion seen on CT scan. CT scan revealed scant intraperitoneal fluid though the overread FAST was also negative. She remained clinically

---

**Table 1**

Demographics and Characteristics of All Eligible Patients Without and With IAI

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N = 354)</th>
<th>No IAI (n = 304)</th>
<th>IAI (n = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>8 (4–12)</td>
<td>8 (4–12)</td>
<td>8 (6–11)</td>
</tr>
<tr>
<td>Age category (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–2</td>
<td>50 (14.1)</td>
<td>47 (15.5)</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>3–6</td>
<td>93 (26.3)</td>
<td>83 (27.3)</td>
<td>10 (20.0)</td>
</tr>
<tr>
<td>7–11</td>
<td>107 (30.2)</td>
<td>81 (26.6)</td>
<td>26 (52.0)</td>
</tr>
<tr>
<td>12–15</td>
<td>102 (28.8)</td>
<td>91 (29.9)</td>
<td>11 (22.0)</td>
</tr>
<tr>
<td>16–17</td>
<td>2 (0.6)</td>
<td>2 (0.7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Male</td>
<td>245 (69.2)</td>
<td>207 (68.1)</td>
<td>38 (76.0)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>30 (18–49)</td>
<td>30 (17–50)</td>
<td>30 (20–43)</td>
</tr>
<tr>
<td>ISS</td>
<td>4 (1–9)</td>
<td>4 (1–9)</td>
<td>11 (6–17)</td>
</tr>
<tr>
<td>Cause of injury</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle collision (motorized and nonmotorized)</td>
<td>182 (51.4)</td>
<td>161 (53.0)</td>
<td>21 (42.0)</td>
</tr>
<tr>
<td>Fall</td>
<td>143 (40.4)</td>
<td>118 (38.8)</td>
<td>25 (50.0)</td>
</tr>
<tr>
<td>Hit</td>
<td>14 (4.0)</td>
<td>12 (4.0)</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>Other</td>
<td>15 (4.2)</td>
<td>13 (4.3)</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>ED disposition (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICU</td>
<td>78 (22)</td>
<td>51 (16.8)</td>
<td>27 (54.0)</td>
</tr>
<tr>
<td>Operating room</td>
<td>20 (5.6)</td>
<td>17 (5.6)</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>Ward</td>
<td>189 (53.4)</td>
<td>169 (55.6)</td>
<td>20 (40.0)</td>
</tr>
<tr>
<td>Home</td>
<td>66 (18.6)</td>
<td>66 (21.7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Transfer</td>
<td>1 (0.3)</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

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

IAI = intraabdominal injury; ICU = intensive care unit; IQR, interquartile range; ISS = Injury Severity Score; GCS = Glasgow coma score.

**Table 2**

Univariate and Multivariate Analysis for IAI in Children With Blunt Torso Trauma

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>OR (95% CI)</th>
<th>AOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST</td>
<td>14.8 (7.5–30.8)</td>
<td>10.2 (3.9–29.3)</td>
</tr>
<tr>
<td>Physical examination</td>
<td>15.2 (7.7–31.7)</td>
<td>16.4 (6.3–48.3)</td>
</tr>
<tr>
<td>Abnormal AST</td>
<td>15.3 (5.8–43.8)</td>
<td>2.1 (0.1–32.9)</td>
</tr>
<tr>
<td>Abnormal ALT</td>
<td>18.6 (7.5–49.8)</td>
<td>5.4 (0.4–81.7)</td>
</tr>
</tbody>
</table>

ALT = alanine enzymes; AOR = adjusted odds ratio; AST = aspartate enzymes; FAST = focused assessment with sonography for trauma.
stable and began eating without difficulty within 24 hours of arrival to the ED. The second patient was a 13-year-old boy with a grade 1 liver laceration who did not require acute intervention. He stayed in the hospital for 4 days for surgical repair of a displaced radius fracture. Neither his FAST nor his overread FAST showed intraperitoneal fluid; however, CT scan revealed intraperitoneal fluid.

Thirteen of the 50 children with IAI (26%) had IAI that required acute intervention. The exFAST did not miss any children with an IAI requiring acute intervention, resulting in a sensitivity of 100% (95% CI = 75.3 to 100), specificity of 65.4% (95% CI = 60.1 to 70.4), PPV of 9.92 (95% CI = 8.69-11.31), NPV of 100, and LR+ and LR- of 2.89 (2.5 to 3.34) and 0, respectively (Table 3).

### DISCUSSION

We found that physical examination and FAST were both independent predictors of IAI in children with blunt torso trauma. However, in combination, the exFAST outperformed either diagnostic approach alone. A positive exFAST had a positive likelihood ratio of 3.1. The pretest probability of our study cohort for having an IAI was 14.1%. If the result of an exFAST were positive in this study cohort, it would increase the posttest probability for IAI from 14.1% to 43%. Similarly, the negative likelihood ratio of the exFAST was 0.2. If the results of exFAST were negative, the probability of IAI would decrease from 14.1% to 2.4%. Using the risk criteria describe by Menaker et al., a negative exFAST may reduce a provider’s level of suspicion from a high to low likelihood of IAI.

Investigators have long sought CT-sparing risk stratification strategies for children with blunt torso trauma and have focused on patient history, physical examination criteria, and laboratory values. There are two large multicenter studies that derived clinical decision rules for risk stratifying children with possible IAI, the Pediatric Emergency Care Applied Research Network (PECARN, 2013) and the Pediatric Surgery Research Collaborative (PSRC, 2017). The

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Test Characteristics of Independent Predictors for Identification of IAI in Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>exFAST</td>
<td>Physical Examination Alone</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>88% (76%-96%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>71% (66%-76%)</td>
</tr>
<tr>
<td>PPV</td>
<td>33% (29%-38%)</td>
</tr>
<tr>
<td>NPV</td>
<td>97% (95%-99%)</td>
</tr>
<tr>
<td>LR+</td>
<td>3.1 (2.5-3.8)</td>
</tr>
<tr>
<td>LR-</td>
<td>0.2 (0.1-0.4)</td>
</tr>
</tbody>
</table>

exFAST = FAST-enhanced physical examination; FAST = focused assessment with sonography for trauma; IAI = intraabdominal injury; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; NPV = negative predictive value; PPV = positive predictive value.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>List of Characteristics and Findings of the Children With IAI That Would Be Missed by the Compound Variable Physical Examination and Focused Assessment With Sonography for Trauma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Sex</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
</tr>
</tbody>
</table>

FAST = focused assessment with sonography for trauma; IAI-I = intraabdominal injury requiring acute intervention; ICU = intensive care unit; MVC = motor vehicle collision; NAT = nonaccidental trauma.
*Cohen kappa coefficient: (κ = 0.40, 95% CI = 0.29 to 0.51).
†ICU admission for traumatic brain injury.
PECARN study included 12,044 children from which it derived a prediction rule of seven variables to identify children at very low risk of IAI requiring acute intervention. The variables in descending order of importance were evidence of abdominal wall trauma, GCS less than 14, abdominal tenderness, evidence of thoracic wall trauma, complaints of abdominal pain, decreased breath sounds, and vomiting. The PECARN prediction rule has a NPV of 98.9% (95% CI = 98.5% to 99.1%) and a sensitivity of 92.5% (95% CI = 90.4% to 94.3%) for the identification of IAI. The PSRC study included 2,188 children and derived a prediction rule of five variables to identify children at very low risk of IAI, including complaints of abdominal pain, abdominal wall trauma/tenderness/distention, abnormal chest radiography, elevated AST, and abnormal pancreatic enzymes. The PSRC rule has a NPV of 99.4% (95% CI = 99.2% to 99.6%), sensitivity of 98.4% (95% CI = 97.8 to 98.8), and LR- of 0.04 (0.03-0.06) for the identification of IAI. However, it is important to note that FAST was not included in these studies because of low utilization and lack of standardization across research sites. In fact, only a minority of children in both the PECARN and the PSRC studies received a FAST, 935 children (8.6%) in PECARN and 829 children (37.9%) in PSRC, because FAST expertise during this era was still emerging. However, with the growth of point-of-care ultrasound within the discipline of pediatric emergency medicine, FAST has become widespread in usage with consequent improved user quality of image acquisition and interpretation. For these reasons, these prediction algorithms may need to be reconsidered.

Historically, FAST had limited utility for children because of the mixed reported results. In a systematic review and meta-analysis that included 2,135 patients from eight studies, there was a pooled sensitivity of 35% (95% CI = 29% to 40%) and pooled specificity of 96% (95% CI = 95% to 97%). However, in a retrospective secondary analysis of PECARN study children who received FAST, patients with low to moderate clinical suspicion for IAI were less likely to receive a CT scan if they had a negatively interpreted FAST.26 This study laid the groundwork for a single-center randomized clinical trial of 925 children by Holmes et al.34 of FAST in children with blunt torso trauma. The authors found that a negative FAST successfully and safely reduced physician suspicion for IAI, but did not translate into lower rates of CT scans. There were multiple editorials in response to this single-center study, suggesting that the FAST should not be used alone, but instead combined with other variables in the context of a clinical decision rule.22,23,35

In contrast to injured adults, the use of FAST for injured children remains controversial and there is no standard method of application for its use. FAST has been successfully implemented into a consistent pathway for the management of injured adult patients in which a hemodynamically unstable adult patient with a positive FAST can forgo advanced diagnostic imaging and be directed for emergent surgical exploration. This adult pathway has improved patient and hospital-centered outcomes, including decreasing CT scan use.16 In contrast, there has not been a consistent pathway for the implementation of FAST in children with blunt torso trauma. For this reason, we have sought to evaluate the ability of FAST to rule out IAI in combination with other readily available predictor variables in children.

Our results suggest an opportunity to improve the accuracy of real-time provider FAST at the bedside. Four of the six patients who had an erroneously read study (Table 4) had a positive FAST overread on expert review and on CT scan, suggesting opportunity to improve the sensitivity of the test. These erroneous studies most likely represent errors made early in the FAST learning curve, because they all occurred within the first year that FAST was introduced to the study site. If we instead presume that these four erroneously read FAST were correctly identified, we would have improved the test characteristics of the exFAST from a sensitivity of 88% to 96% and the NPV from 97.3% to 99.2%. Similarly, we would have improved the specificity, from 71% to 78%, and PPV, from 34% to 42%, using the overread for exFAST. The agreement between provider and overread exFAST accounts for the fair interrater agreement (κ = 0.4, 95% CI = 0.29 to 0.51), which was similar to a single-center randomized controlled trial of pediatric FAST (κ = 0.45, 95% CI = 0.30 to 0.60). Therefore, through educational intervention, experience, and technologic advancement, we would expect the sensitivity of FAST to improve.

**LIMITATIONS**

There are intrinsic limitations for FAST in children. In this study, two children in whom FAST and overread FAST were negative, scant intraperitoneal fluid
was noted on CT scan interpretation (Table 4). The first possible reason for the discrepancy in FAST and CT scan interpretation was the passage of time. FAST and CT scan are performed in succession. FAST may have been performed too early to detect intraperitoneal fluid. The second possible reason was that CT scan was more sensitive than FAST for detecting small volumes of intraperitoneal fluid. Therefore, like adult FAST, there may be a minimal fluid threshold below which pediatric FAST cannot detect intraperitoneal fluid. Third, there are IAIs that may not cause intraabdominal hemorrhage and, therefore, cannot be detected on FAST.

Our study has several important limitations that predominantly stem from the retrospective design. First, our inclusion sampling of only children who received a FAST may have led to sampling bias. There were 725 children who did not receive a FAST, which could bias the results of the study. However, the cohort of patients included in our study appears to be representative of previous reports. The incidence of IAI in this cohort was 14%, which was similar to the 10% to 15% rate of IAI in other studies, but different than the rate reported by Holmes et al.,. Importantly, in this study, the choice to perform a FAST was left to the discretion of the treating provider, and we believe that children who did not receive a FAST may have had a lower suspicion for injury than those who did. Assuming that other factors remain constant, the NPV of a test improves with a lower prevalence. Therefore, we believe that this potential sampling bias could have led to an underestimate of the NPV. Second, the study’s retrospective design may have led to classification errors. We sought to minimize errors in classification by defining variables a priori and resolving discrepancies between two reviewers. In addition, the data set did not include a category of indeterminate for FAST. Third, although we report demographic, frequency, and accuracy data using all available patients, in computing the ORs for IAI we included only records that had complete data for physical examination and liver enzymes. This could lead to overestimation bias. Fourth, contemporary publications evaluating children with blunt torso trauma focus on identifying children with IAI requiring acute intervention. Our small sample size was not powered to evaluate children with IAI requiring acute intervention. However, the incidence of IAI requiring intervention was 26% of those with IAI, which was similar to previous studies. Furthermore, the exFAST did not have any false-negative studies when considering IAI requiring acute intervention. Although this study was underpowered to truly evaluate IAI requiring acute intervention, we believe that this is hypothesis generating and may lead to future evaluation of this approach as a CT-sparing technique for children with blunt torso trauma. Our study found that the physical examination and the FAST were each significantly and independently associated with the IAI outcome. However, it is important to note that these variables may be interdependent clinically because the person who performed the physical examination also performed the FAST. Knowledge of other elements of the examination, including the ultrasound transducer on the patient’s abdomen, may have influenced the performance and provider interpretation of FAST. However, our analysis indicates that addition of FAST to the physical examination significantly improved the accuracy in detection and exclusion of IAI.

The results of study are significant because the exFAST offers an advantageous CT-sparing strategy, the capability of the clinician to deploy and interpret the test without relying on a specialty service, the ability to obtain results within minutes of the patient’s arrival, and the ease of conducting serial examinations without the patient leaving the resuscitation bay. Similarly, FAST has become ubiquitous in pediatric EDs. Therefore, we believe that the exFAST has desirable test characteristics, including its NPV, which may help clinicians risk stratify patients and identify those who may not require CT scan. Furthermore, the exFAST may offer promising refinement to new or existing clinical decision rules.

**CONCLUSION**

We conclude that physical examination and focused assessment with sonography for trauma are both independent predictors associated with the identification of intraabdominal injury in children with blunt torso trauma. Furthermore, the combination of physical examination combined with focused assessment with sonography for trauma may have adequate negative predictive value to decrease the need for computed tomography scan in certain low- to moderate-risk populations of children with blunt torso trauma and may be used in conjunction with new or existing clinical decision rules. Further testing is warranted to determine extent of computed tomography sparing that can be achieved using the focused assessment with sonography for trauma–enhanced physical examination.
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.13959/full](http://onlinelibrary.wiley.com/doi/10.1111/acem.13959/full)

**Data Supplement S1.** Final International Classification of Disease, Ninth Revision (ICD-9) diagnosis of penetrating injury.
Care Transition Decisions After a Fall-related Emergency Department Visit: A Qualitative Study of Patients’ and Caregivers’ Experiences

Cameron J. Gettel, MD1,2, Kelsey Hayes3, Renee R. Shield, PhD4, Kate M. Guthrie, PhD5, and Elizabeth M. Goldberg, MD, ScM4,6

ABSTRACT

Objective: Falls are a leading cause of injury-related emergency department (ED) visits and may serve as a sentinel event for older adults, leading to physical and psychological injury. Our primary objective was to characterize patient- and caregiver-specific perspectives about care transitions after a fall.

Methods: Using a semistructured interview guide, we conducted in-depth, qualitative interviews using grounded theory methodology. We included patients enrolled in the Geriatric Acute and Post-acute Fall Prevention Intervention (GAPcare) trial aged 65 years and older who had an ED visit for a fall and their caregivers. Patients with cognitive impairment (CI) were interviewed in patient-caregiver dyads. Domains assessed included the postfall recovery period, the skilled nursing facility (SNF) placement decision-making process, and the ease of obtaining outpatient follow-up. Interviews were audio-recorded, transcribed verbatim, and coded and analyzed for a priori and emergent themes.

Results: A total of 22 interviews were completed with 10 patients, eight caregivers, and four patient-caregiver dyads within the 6-month period after initial ED visits. Patients were on average 83 years old, nine of 14 were female, and two of 14 had CI. Six of 12 caregivers were interviewed in reference to a patient with CI. We identified four overarching themes: 1) the fall as a trigger for psychological and physiological change, 2) SNF placement decision-making process, 3) direct effect of fall on caregivers, and 4) barriers to receipt of recommended follow-up.

Conclusions: Older adults presenting to the ED after a fall report physical limitations and a prominent fear of falling after their injury. Caregivers play a vital role in securing the home environment; the SNF placement decision-making process; and navigating the transition of care between the ED, SNF, and outpatient visits after a fall. Clinicians should anticipate and address feelings of isolation, changes in mobility, and fear of falling in older adults seeking ED care after a fall.
Falls are the leading cause of fatal and nonfatal injuries among older adults over age 65. About 30% of older adults fall every year, often resulting in serious injuries, loss of function, and fear of falling again. Estimated at $31.3 billion, the annual Medicare costs for older adult falls is expected to increase given the aging population. Aside from the high associated morbidity and cost, falls serve as a sentinel event in predicting future functional decline, emergency department (ED) recidivism, and mortality. Within 6 months of a fall-related ED visit, 22.6% of older adults have at least one recurrent fall, 42.6% revisit the ED, and 2.6% died. Furthermore, the risk of admission to skilled nursing facilities (SNFs) increases with injurious and noninjurious falls. Indeed, fall-related injuries account for up to one-third of SNF admissions.

The Society for Academic Emergency Medicine has specified “falls” as one of four high-yield research opportunities. Despite an increasing focus on fall detection using sensor technology, there is limited emergency medicine research on falls and little is known about the difficulties older adults and their caregivers face during recovery following a fall-related ED visit. This research addresses this question. We asked patients and caregivers to describe the lived experience of the ED visit, how they made decisions about their postinjury health care needs (e.g., whether to go to a SNF, what to change about their home environment), and what barriers they encountered accessing care in follow-up. We aimed to elucidate the physical and psychological trajectory after an ED visit for a fall, elucidate patient- and caregiver-specific perspectives in determining the appropriate living setting for older adults after a fall, and identify barriers to receiving necessary health care after the event. Understanding the unique experiences of patients after a fall could allow emergency medicine clinicians to provide more patient-centric care and anticipatory guidance to patients and caregivers after this sentinel event.

METHODS

Study Design
In this qualitative study, we used grounded theory methodology and included participants who were a part of the Geriatric Acute and Post-acute Fall Prevention Intervention (GAPcare) study. The GAPcare study aimed to examine the feasibility and acceptability of the GAPcare intervention to determine if ED-based pharmacy and physical therapy (PT) consultations reduced subsequent falls and health care utilization. The trial was registered at www.clinicaltrials.gov (ClinicalTrials.gov identifier NCT03360305). In this study, we analyzed data from semistructured interviews of patients and/or their caregivers to assess their experiences after a fall-related ED visit. We also evaluated the transition of care after the ED visit including the decision to consider/not consider SNF placement for the care recipient. The qualitative research methods used in this study provided an open-ended inquiry to focus on discovery and interpretation, allowing investigators to understand patient and caregiver experiences of their care after an ED-related fall visit. The hospital institutional review board approved the study. Study methods and results are presented according to the consolidated criteria for reporting qualitative research (COREQ).

Study Setting and Population
Recruitment for GAPcare occurred at two different EDs: an academic community hospital and a Level I trauma center and tertiary referral ED within the same health system. Patients 65 years and older were eligible to participate in the main study if they presented to the ED within 7 days of a fall, if they primarily spoke English or Spanish, and if their ED clinician determined they were likely to be discharged from the ED (i.e., not admitted). As they represent an important understudied subgroup of people who experience falls, patients with cognitive impairment (CI) were eligible for participation. The cognitive status of the patient was assessed using the validated Six-Item Screener. A score of <4 on the 6-point questionnaire indicated high risk for CI. If a patient identified with CI remained interested in participating in the study, a legally authorized representative was asked to provide written consent. Individuals who were altered (e.g., intoxicated), undomiciled, living in a SNF, or could not provide a phone number for follow-up were excluded. After providing written consent, participants were randomized to the usual care (UC) or intervention (INT) arm. Participants in the UC arm received routine care as directed by the ED clinicians. Participants in the INT arm received the pharmacy and PT GAPcare consultations.

Following the ED visits, participants from both study arms were contacted over the phone and asked to participate in a brief interview about their experience in GAPcare. We aimed to interview participants...
until thematic saturation was reached. Interviews took place over the phone or at the home of the participant, depending on the individual needs. We completed interviews with randomly selected patients and caregivers from the original study from June 2018 until January 2019. Interviews occurred within 6 months of the fall-related ED visit. We felt that patients and caregivers would be able to remember details regarding the fall and ED visit within this time frame. The ideal time frame between fall event and interview has not been studied. Dyads (i.e., patient with CI and caregiver) were only interviewed in person. When desired by the patient or caregiver, we interviewed them together.

Study Protocol
We developed an interview guide based on prior qualitative research on falls and the clinical expertise of the study authors. The qualitative interview consisted of queries into the following domains: 1) experiences with ED care, 2) symptom management after ED evaluation, 3) quality of in-ED and outpatient professional and provider communication, 4) views of the care transition, 5) perceptions of barriers to engage with follow-up care, and 6) perceptions of clinical trajectories after the ED visit.

We developed two interview guides—one tailored to the patient and the other to the caregiver. The patient and caregiver interview guides are available as supplementary material accompanying the online article (Data Supplement S1, Appendixes S1 and S2 [, available as supporting information in the online version of this paper, which is available at http://online library.wiley.com/doi/10.1111/acem.13938/full]). The semistructured interview guide included the study rationale, an overview of the qualitative in-depth interview process, and potential queries designed to capture the domains listed above. It contained open-ended questions with follow-up questions and probes specific to study goals. We piloted the planned questions with two older adults prior to recruiting participants into the study to ensure that the questions were understood as intended. The piloted study questions were refined based on feedback from these individuals and members of the study team with expertise in qualitative methods, including incorporating exact language of the older adults from the piloting process.

Prior to starting the interview, participants were asked to verbally consent to the interview and its recording. Study personnel (EG and AM; both female) conducted interviews with patients and caregivers. EG is an emergency physician with formal qualitative research training who performs research on fall-related ED visits in older adults. AM completed graduate studies in qualitative research techniques and was trained to the interview guide by the PI. Interviewers collected basic demographic information and recorded brief field notes immediately after the interview. The interviewers were not part of the care team that recruited participants into the study during the index ED visit and therefore did not have an existing relationship when the interviews commenced.

Interviews were recorded, deidentified, and transcribed verbatim by professional medical transcriptionists. Transcripts were reviewed by the authors and corrected when the transcript passage was incomprehensible or had transcription errors. Recordings were destroyed after certification of the transcription to ensure confidentiality.

Data Analysis
We used applied thematic analysis to guide the data analysis, which included the following steps: 1) familiarization with the data through reading and rereading the transcripts and noting initial observations and 2) developing a set of codes based on our interview questions and interview debriefs to identify and sort textual data. Major topics and subtopics were coded in pairs by four study personnel and then reconciled through team discussion. NVivo software (version 12) was used to organize the coded data. Transcripts and final codes were entered into the NVivo project. 3) We used a team-based approach and two authors iteratively reduced the data, identifying patterns, themes, and subthemes that emerged across participants and interviews. 4) We reviewed themes in relation to the coded extracts and entire data set and selected representative quotes from the interviews to illustrate the themes. 5) We recorded coding definitions and decisions as well as ideas about emerging themes in an ongoing audit trail. 6) Two study authors (CG and EG) then prepared the analytic narrative and contextualized it using the existing literature.

RESULTS
The GAPcare intervention study recruited 110 patients from January 2018 until March 2019. None of the patients or caregivers called refused to participate or had hearing impairment. We conducted interviews
until thematic saturation was achieved with a total of 22 interviews with 14 patients and 12 caregivers. Four interviews were patient–caregiver dyads. No repeat interviews were performed. Interviews lasted a mean of 43 minutes, with a range of 14 to 109 minutes. Interviews were conducted a mean (±SD) of 95 (±41) days after the initial ED encounter, with a range of 18 to 173 days. Patients were on average 83 years old; nine of 14 patients were women, and two of 14 had CI (Table 1). Six of the 12 caregiver interviews were with reference to a patient with CI. Of the 12 caregivers, four were daughters, one was a son, three were wives, one was a husband, one was a sister, one was a niece, and one was a female cousin. We identified four main themes from the qualitative data: 1) the fall as a trigger for psychological and physiological changes, 2) SNF placement decision-making process, 3) direct effects of fall on caregivers, and 4) barriers to receipt of recommended follow-up. Subthemes and representative quotes for each theme are identified in Table 2.

Illustrative quotes below are noted with a transcript ID number and participant characteristics: C = caregiver; CD = caregiver of person with dementia; P = person without dementia; PwD = person with dementia).

**Theme 1—The Fall as Trigger for Psychological and Physiological Changes: Patients and Caregivers Noted Psychological and Physiological Changes After a Fall, Including the Development of Fear of Falling and Changes in Gait and Mobility**

Patients were noted to have varying levels of physical recovery in the postfall period. Several patients appeared to be unfazed by the fall, stating that they had not appreciated a decline in functional ability after the fall. In caregivers and patients who did note a change in function after the fall, those who noted a quicker recovery to prefall level of function attributed the response to various visiting in-home services as well as staff at SNFs. Several other older adults bolstered their postfall level of independence by engagement in classes.

In addition to the potential physical sequelae of the fall, patients recounted a significant psychological burden in the postfall time period. One participant in the UC arm of the study stated,

I felt as if I was under house arrest being in the house for like six weeks and not going out because, of course, the weather was awful. Very isolated. Isolated. There was a feeling of really being isolated. [220P_woman]

Participants reported experiencing the fear of falling again after their initial fall. Despite family member’s installation of safety-supporting instruments in the house and the continued use of assistive devices, one caregiver stated, “She’s nervous about a bunch of other things too, like falling up the stairs. She won’t go up the stairs” [130C_niece]. After the fall, two patients also noted giving up hobbies (e.g., gardening) that they enjoyed and found meaningful due to the fear of falling.

---

**Table 1** Participant and Interview Characteristics

| Patient sex | 9 (64.2) |
| Age (years) | 83 (±7) |
| Primary injury at ED visit |  |
| Abrasion/laceration | 6 (27.2) |
| Fracture | 1 (4.5) |
| Contusion | 11 (50.0) |
| Weakness/decreased function | 2 (9.1) |
| Back pain | 2 (9.1) |
| ED discharge disposition |  |
| Home | 6 (27.2) |
| Home with services | 11 (50.0) |
| Facility | 5 (22.8) |
| Any SNF stay during study time frame, yes | 7 (31.8) |
| Insurance status |  |
| Medicare | 5 (22.8) |
| Private | 1 (4.5) |
| Medicare and private | 16 (72.7) |
| Interview types |  |
| Patient | 10 (45.5) |
| Caregiver | 8 (36.3) |
| Patient–caregiver dyads | 4 (18.2) |
| GAPcare study arm |  |
| Intervention | 20 (90.9) |
| Usual care | 2 (9.1) |
| In reference to patient with cognitive impairment* |  |
| Yes | 6 (27.3) |
| Time between fall ED-visit and interview (days) |  |
| Mean (±SD) | 95 (±41) |
| Range | 18-173 |
| Primary language |  |
| English | 22 (100.0) |

Data are reported as n (%) unless otherwise specified.

SNF = skilled nursing facility.

*Identified by Six-Item Screener.
### Summary of Themes and Illustrative Quotes

<table>
<thead>
<tr>
<th>Themes/Subthemes</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theme 1—The fall as trigger for psychological and physiological changes: patients and caregivers noted psychological and physiological changes after a fall, including the development of fear of falling and changes in gait and mobility.</strong></td>
<td></td>
</tr>
<tr>
<td>1A. Gait speed and walking habits</td>
<td>“Yeah, well [I’m] more careful, I take a little bit more time because I used to speed around. Now I got two moves, slow and stop.”</td>
</tr>
<tr>
<td>1B. Social isolation</td>
<td>“I took one shower a week because getting undressed, getting out of the sling was hard because my arm was just hanging down useless…I think that was the hardest thing in a way because if you can’t dress yourself. That’s a really special feature I think of being human, I slept in my clothes for weeks because it was so difficult to, such an operation for me to change what I was wearing.”</td>
</tr>
<tr>
<td>1C. Giving up prior hobbies and difficulties within the home environment</td>
<td>“I felt as if I was under house arrest being in the house for like six weeks and not going out…very isolated. Isolated. There was a feeling of really being isolated.”</td>
</tr>
<tr>
<td>1D. Reluctance to use assistive devices after a fall</td>
<td>“I actually was told that by one of the doctors I see. If you fall, you could end up in the nursing home and it really might be the end of your life as you know it. So that was terrifying.”</td>
</tr>
<tr>
<td><strong>Theme 2—SNF placement decision-making process: several factors influence the decision to agree to SNF placement including the amount of community and home caregiver support as well as concerns over the loss of independence.</strong></td>
<td></td>
</tr>
<tr>
<td>2A. Family and community support in discharge to home</td>
<td>“It would be a lot of personal factors that would go into that kind of decision [to be discharged to SNF], and I had everything going for me here but I could see how many people would not and would have to go to rehab. My situation at home was so good it was better for me to be here than in a rehab place dependent on nurses or whatever that goes on there. If I didn’t have a husband who I knew would do everything for me I would have needed the rehab.”</td>
</tr>
<tr>
<td>2B. Perceptions of SNFs</td>
<td>“Why don’t you go and live in one of them?” I says, “No! I don’t like ‘em.” I don’t. It’s too clustered. It’s alright for people in their 80s and 90s, but not for me! I’m used to being on the go. But, I gotta take it easy.”</td>
</tr>
<tr>
<td>2C. Caregivers discern between “good” and “bad” rehabilitation facilities</td>
<td>“I mean, a lot of times it is the family members that do the care, but when it’s the hand-off to the homes, I feel sometimes…I mean, she’s at a good facility, it’s not like, you know, we stuck her some place, but…”</td>
</tr>
<tr>
<td>2D. Costs of long-term care</td>
<td>“I’m very grateful and I do have concerns of I really want to keep him at home. We have long-term care insurance. So when the time comes, I’m hoping that I have people in place that I have reliable help. But I do want to keep him at home. I want him to be comfortable and feeling he’s in a safe environment that he knows. But if the day comes that I really can’t do that, then we will go into skilled living or dementia care and I’ll be there every day.”</td>
</tr>
<tr>
<td>2E. Safety at SNFs</td>
<td>“My mother has exhausted all her income. Now, we got to pay for everything. It would be a lot cheaper for them to keep her here.”</td>
</tr>
<tr>
<td><strong>Theme 3—Direct effects of fall on caregivers: caregivers make changes to their lives to accommodate their family member after a fall occurs.</strong></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
After a fall, the physical and psychological recoveries were inextricably linked for several interviewees. One participant sustained a humerus fracture and required the use of a sling. In describing her new mobility limitations and the subsequent psychological toll, she described her arm as “hanging down useless” and her inability to get dressed as particularly dehumanizing.

Additionally, the fear of falling noted by many older adults resulted in its own form of physical limitations, articulated by one patient, “I take a little bit more time...” [150P_woman]. Despite the varying physical needs surrounding recovery, patients and caregivers noted a reluctance to use instrumental and assistive devices after the fall, as it seemingly made the patient appear “old.”

**Theme 2—SNF Placement Decision-making Process: Several Factors Influence the Decision to Agree to SNF Placement Including the Amount of Community and Home Caregiver Support as Well as Concerns Over the Loss of Independence**

Patients and caregivers noted many things that influenced their decision to go (or not) to a SNF after the ED visit. The level of community support was frequently shared as a positive benefit to continued community dwelling. Patients stated that dedicated and effective caregiver support also helped them decide against SNF placement. One patient captured the sentiments of several others by acknowledging that patients lacking significant support in the home environment may require a SNF.

It would be a lot of personal factors that would go into that kind of decision [to be discharged to SNF], and I had everything going for me here but I could see how many people would not and would have to go to rehab. My situation at home was so good it was better for me to be here than in a rehab place dependent on nurses or whatever that goes on there. If I didn’t have a husband who I knew would do everything for me I would have needed the rehab. [130P_woman]

Caregivers noted that they played a significant role in the SNF placement decision, particularly when taking into account their own functional abilities. The decision to place their loved one in a SNF was made clear when the assistance needs became greater than their own functional abilities. Several additional caregivers described their attempt to discern between...
“good” and “bad” SNF facilities, not wanting to be viewed as abandoning their loved ones.

In the postfall time period, caregivers also noticed a difference in their loved one’s acceptance of assistance between the home setting and the SNF setting. One daughter, whose mother lived in a SNF, noted that her mother with CI was more open to receiving help at a SNF than at home. In the home environment, the patient was more likely to ignore assistive devices, despite their incorporation into numerous rooms within the house. Alternatively, attentive SNF staff limited the patient from walking without an assistive device, as they “wouldn’t let her get away with it.” The decision regarding SNF placement was also influenced by financial considerations described by caregivers. Another caregiver also cited the importance of long-term care insurance and the comfort that is provided to the family should SNF services be required in the future.

Two related patient-specific barriers were frequently communicated during interviews regarding the decision for SNF placement. Prominent among many participants was the desire for older adults to maintain independence in the community setting. Despite prior falls, patients also did not want to be placed in SNFs, because that step seemed a loss of independence and a prominent sign of aging. Recounting her conversation with health care professionals and caregivers, one participant stated,

‘Why don’t you go and live in one of them?’ I says, ‘No! I don’t like ‘em.’ I don’t. It’s too clustered. It’s alright for people in their 80s and 90s, but not for me! ... I felt like I was an old lady ... I came home, they checked me all out, I was okay. Then I came home, I had one more fall, they took me to the ED to check me out. And they said that I should put me in for rehabilitation. So, I says, ‘Okay. That was fine.’ [200P_woman]

Caregivers also made modifications to their own lives, particularly regarding driving requirements, sleep habits, and daily care in the postfall time period. To avoid the risk associated with driving, one caregiver stated, “I’m going to pick him up. Pick him and my mother up because I just don’t want my dad driving long distances as much” [90C_daughter]. In the time period while trying to gather additional home services, one caregiver shared the direct effects the patient’s fall had on their sleep pattern as well as the uncomfortable need to ask neighbors for support.

I sleep on the sofa now so that if he needs anything, I’m there and that’s fine. That’s no big deal. I’m trying, you know, to find something that can accommodate me that I’m comfortable with leaving him with whoever it is. My neighbor next door will stay with him if I want to go grocery shopping. The other night I needed to go to the drug store. My neighbor across the street came and sat with him, but I hate to ask too often. [30CD_wife]

Given the implementation of new safety modifications, the need for frequent monitoring of their loved ones, and the coordination of care in the outpatient setting, caregivers were noted to undertake a significant burden in the postfall time period. One caregiver felt as though she were part of the health care team, stating, “I call myself a healthcare worker while I had him home, ‘cause it was difficult” [100CD_wife]. Another caregiver relayed the psychological burden frequently evident when managing a loved one without assistance, “I can’t do it by myself. I can’t be trapped in the house 24 hours a day because there’s nobody coming in here to watch her” [40CD_daughter].

Theme 4—Barriers to Receipt of Recommended Follow-up: Patients and Families Feel Overwhelmed and Uncertain and Are Often Choosing Not to Obtain Recommended Follow-up

After a fall, older adults noted discussions with their PCP regarding fall contributors and the effects of prescribed medications. Aside from interviewees reporting shortcomings with ED to PCP communication on ED discharge and the time-consuming nature of organizing follow-up care, participants noted barriers regarding their ability to understand the complex outpatient setting and the time-consuming nature of arranging

Theme 3—Direct Effects of Fall on Caregivers: Caregivers Make Changes to Their Lives to Accommodate Their Family Member After a Fall Occurs

Caregivers noted many home modifications that were made after the fall to accommodate the return home of their loved one. Modifications implemented by caregivers included additional handrails, handicapped-accessible home entrances, and the removal of rugs.
follow-up. One caregiver described a situation of discordant recommendations between her mother’s PCP and specialist, resulting in subsequent confusion as to the appropriate diagnostic evaluation and management.

We go to the face-to-face, and now he tells us this, this, this, this. The hospital wanted him to have one of those [Holters] on to monitor his heart. He felt, let’s do a stress test, not [the Holter]... We go over the hospital visit, all that. He said, basically, the hospital wanted him to do one thing, he decided something else. And then, my mother the boss decides well, we’re not doing that either. We’re going to see the heart doctor, which is good because the heart doctor told us that he didn’t want him for the stress test.” [90C_daughter]

Many narratives highlighted the numerous providers seen after a fall-related ED visit, prompting patients to exhibit ambivalence. When asked if a patient had her lip repaired after the fall, the interviewee responded, “Probably. Probably. I’ve been going to so many doctors and dentists. This tooth was coming out, too” [70P_woman]. Similarly, a separate caregiver noted the prolonged wait to see the PCP gave the impression that follow-up care was not important. “I called the primary care doctor and the first appointment I could get was right before Christmas [2 months after ED visit]. So, I just didn’t think it was that important…” [40CD_daughter]. When the importance of follow-up was evident to caregivers, they were still noted to struggle with the psychological burden that patients experience in the postfall period. A significant fear of falling prevented one patient from attending a PCP visit: her caregiver stated, “Only hurdle now is to get her out of the house to see her primary care physician. He has put his foot down and said he’s not renewing anymore of her prescriptions until he sees her … She doesn’t want to go out of the house because she’s afraid to fall” [140CD_cousin]. A developed care transition conceptual model (Figure 1) identifies four areas considered by older adults in selecting between home and SNF care settings: social support versus isolation, safety, prior SNF perceptions, and caregiver burden.

**DISCUSSION**

This study is the first to characterize perspectives of older adults who presented to the ED after a fall and their caregivers regarding post-fall recovery and the SNF placement decision-making process. Through in-depth qualitative interviews we identified four central themes: 1) the fall as a trigger for psychological and physiological changes; 2) SNF placement decision-making process; 3) direct effects of fall on caregivers; and 4) barriers to receipt of recommended follow-up. The lived experiences of patients and caregivers highlight opportunities to improve the care of older adults presenting to the ED after a fall. Clinicians should be aware of the concerns that patients and families have about SNF placement and what their anticipated experiences are after ED discharge. This is particularly important given the aging of population.

Our study revealed that patients experience varying trajectories of physiological recovery after a fall. Many of the interviewees described a major psychological burden that was brought on by the fall, with prominent feelings of isolation and further fear of falling. The important link between the physical recovery and the psychological outcomes that older adults experience after a fall is a finding not to be overlooked. Prior qualitative studies, including Yardley et al., noted the views of older adults towards fall prevention interventions. While interventions (e.g. strength and balance training) were not ED-initiated, participants
noted that they experienced many unanticipated benefits aside from solely future fall reduction. Benefits of exercise for fall prevention can be discussed with older adults in the ED and include: interest and enjoyment, improved health, mood, and independence. Similar to the findings of our study, barriers to participation in fall prevention strategies identified by this prior literature include: denial of falling risk, the belief that no additional falls-prevention measures were necessary, and practical barriers to participation (e.g. transport, cost). Given the noted reluctance, more recent technology-based (e.g. sensors, cameras) interventions have been used in a myriad of ways for older adults including injury prevention \cite{19, 20} and fall detection.\cite{21, 22} Older adults seeking care in the ED after a fall offer healthcare providers a critical ‘sentinel moment’ to assess fall risk, discuss prevention techniques and perceived barriers, and set expectations regarding the possibility of future fear of falling.

Given the complexity of the post-fall time period and the SNF placement decision-making process, qualitative research is well suited to assess generated themes. Prior qualitative studies have assessed community caregivers’ concerns of care recipients’ risk of falling and management at home,\cite{23} as well as families’ experiences with relationships and the quality of care following SNF placement of an older relative.\cite{24} Ang et al.\cite{23} identified four themes in their interviews with carers: carers’ perception of fall and fall risk, care recipient’s behavior and attitude towards fall risk, care recipient’s health and function, and care recipient’s living environment. The findings of our study highlight the many factors in the post-fall period that patients and caregivers consider in determining if SNF placement is indicated. The decision to accept SNF care after a fall represents a branch point, where older adults confront the sense of losing independence, being away from family on a daily basis, and also the perception of appearing ‘old’. With the known and observed desire for older adults to remain independent for as long as able, the SNF setting is viewed as a ‘last resort’. Our conceptual model (Figure 1) identifies key considerations of the patient in the SNF placement decision-making process: social support vs. isolation, prior SNF perceptions, safety, and caregiver burden. We found that patients prefer the home environment if they have a robust support system and the financial means to pay for long-term care costs. Patients stated they would consider SNF placement if they had physical and environmental limitations or if the caregiver identified home safety concerns. Prior SNF research has focused on predictors for institutionalization,\cite{25} with less known regarding the placement decision-making process from the patient and caregiver’s perspective. Program development and future research studies are needed to better incorporate support persons of older adults as they provide unique perspectives to fall prevention intervention strategies and the SNF placement decision-making process.

Patient deterioration has long been considered a driving factor for the decision to end home care,\cite{26} with this study adding qualitative caregiver fatigue themes such as sleep changes, feeling ‘trapped’, and reluctantly requiring neighborhood assistance to care for their loved one after a fall-related ED visit. Aside from the direct effects on caregivers, interviews often highlighted the financial considerations that families experience in determining the optimal care setting for their loved ones. Furthermore, the caregivers in this study were noted to frequently balance promoting independence and autonomy in their loved one, with attempting to maintain their safety in the community setting. Notably, caregivers sometimes cited the familiarity of the home environment as a reason their loved ones were more likely to be resistant to assistance in the home setting in comparison to the SNF environment.

Future research should focus on the relationships between caregivers and patients and explore tools that could be used to aid the decision-making process surrounding SNF placement. Prior research highlights that caregivers appreciated a healthcare professional acting as a decision-making coach,\cite{27–29} yet few interventions exist to improve caregivers’ involvement in the decision-making process.\cite{30} Clinicians should be encouraged to seek the perspectives of both patients and caregivers in the acute setting to promote an open and early conversation regarding an environment that balances autonomy and safety.

Finally, our study determined that older adults recognize many barriers in acquiring post-fall follow-up care in the outpatient setting that should be considered by ED clinicians before discharge. Most evident from caregivers and patients was the confusion regarding conflicting care plans of different providers, as well as the high volume of providers individuals were asked to see in follow-up. Optimal triage of patients into primary care physician (PCP) visits, using technology to help with transitions, and providing patients and caregivers with tools to make health care appointments.
easier would help address some of the barriers found in this study. From this study, we have learned that a disjointed and uncoordinated approach to follow-up care in the outpatient setting often will result in ambivalence and the impression that follow-up care is not important. This may, in part, strengthen the argument that fall prevention strategies should start in the ED before discharge and should provide urgency to policymakers to present solutions that help patients get the care they need expeditiously.

LIMITATIONS

Our study was conducted at two EDs within one health system in the Northeast United States, therefore potentially restricting generalizability. However, we expect that many older adults and caregivers will have similar experiences after a fall. This study mainly included patients from the INT arm of the GAPcare study, because a major objective of this work was to improve the intervention and care delivered to older adults after a fall. These patients were provided with additional support through PT and pharmacy consultations in anticipating their recovery from injury and identifying reasons for the fall and may have had a more favorable experience than the average older adults presenting to the ED for a fall. However, we believe that lessons learned from these interviews can inform ED care of older adults as a whole. Additionally, women were overrepresented in our study population as both patients and caregivers. This reflects that women are more likely to experience nonfatal falls and are more likely to be caregivers. Additionally, our understanding and interpretation of the data may have potentially introduced researcher or confirmation bias. We minimized this bias by using semistructured interview guides and reconciling discrepancies through team discussion. We included team members who are neither physicians nor trained in emergency medicine in this study to independently evaluate analyses and avoid researcher bias. Although many best practices of qualitative research\textsuperscript{12,13} were followed, we did not return transcripts to participants for checking or solicit interviewee feedback after themes were generated.

CONCLUSIONS

In summary, this study qualitatively reports on the perspectives of older adults and caregivers after a fall-related ED visit. The fall was seen as a trigger for postfall physiological and psychological changes including changes in gait and a prominent fear of falling. The multifactorial decision-making process regarding skilled nursing facility placement for older adults after a fall often evoked patient discussions of their desire for independence and not appearing “old” as primary reasons for reluctance to skilled nursing facility placement. The direct effects of their loved one’s fall caused caregivers to cite changes to sleep location, driving habits, and seeking assistance from neighbors. Patients and caregivers also recognized significant barriers in the coordination of care to the outpatient setting after discharge from the ED after a fall. Consideration of these perspectives is essential in developing practices to improve the overall transition of care between the ED, primary care physician, and skilled nursing facility settings in older adults after a fall.

References

10. Carpenter CR, Shah MN, Hustey FM, Heard K, Gerson LW, Miller DK. High yield research opportunities in


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.13938/full

Data Supplement S1. Supplemental material.
Can Emergency Physicians Accurately Rule Out a Central Cause of Vertigo Using the HINTS Examination? A Systematic Review and Meta-analysis

Robert Ohle, MBBCh, MA, MSc1, Renee-Anne Montpellier, MD1, Virginie Marchadier, MD1, Aidan Wharton, MD1, Sarah McIsaac, MBBCh, MEd1, Mackenzie Anderson2, and David Savage, MD, PhD3

ABSTRACT

Introduction: Dizziness is a common complaint presented in the emergency department (ED). A subset of these patients will present with acute vestibular syndrome (AVS). AVS is a clinical syndrome defined by the presence of vertigo, nystagmus, head motion intolerance, ataxia, and nausea/vomiting. These symptoms are most often due to benign vestibular neuritis; however, they can be a sign of a dangerous central cause, i.e., vertebrobasilar stroke. The Head Impulse test, Nystagmus, Test of Skew (HINTS) examination has been proposed as a bedside test for frontline clinicians to rule out stroke in those presenting with AVS. Our objective was to assess the diagnostic accuracy of the HINTS examination to rule out a central cause of vertigo in an adult population presenting to the ED with AVS. Our aim was to assess the diagnostic accuracy when performed by emergency physicians versus neurologists.

Methods: We searched PubMed, Medline, Embase, the Cochrane database, and relevant conference abstracts from 2009 to September 2019 and performed hand searches. No restrictions for language or study type were imposed. Prospective studies with patients presenting with AVS using criterion standard of computed tomography and/or magnetic resonance imaging were selected for review. Two independent reviewers extracted data from relevant studies. Studies were combined if low clinical and statistical heterogeneity was present. Study quality was assessed using the QUADAS-2 tool. Random effects meta-analysis was performed using RevMan 5 and SAS 9.3.

Results: A total of five studies with 617 participants met the inclusion criteria. The mean (±SD) study length was 5.3 (±3.3) years. Prevalence of vertebrobasilar stroke ranged 9.3% to 44% (mean ± SD = 39.1% ± 17.1%). The most common diagnoses were vertebrobasilar stroke (mean ± SD = 34.8% ± 17.1%), peripheral cause (mean ± SD = 30.9% ± 16%), and intracerebral hemorrhage (mean ± SD = 2.2% ± 0.5%). The HINTS examination, when performed by neurologists, had a sensitivity of 96.7% (95% CI = 93.1% to 98.5%, I² = 0%) and specificity of 94.8% (95% CI = 91% to 97.1%, I² = 0%). When performed by a cohort of physicians including both emergency physicians (board certified) and neurologists (fellowship trained in neurootology or vascular neurology) the sensitivity was 83% (95% CI = 63% to 95%) and specificity was 44% (95% CI = 36% to 51%).

Conclusions: The HINTS examination, when used in isolation by emergency physicians, has not been shown to be sufficiently accurate to rule out a stroke in those presenting with AVS.

From the 1Division of Clinical Sciences, Northern Ontario School of Medicine, Health Science Northern Research Institute, Sudbury, Ontario; the 2Faculty of Science, Laurentian University, Sudbury, Ontario; and the 3Division of Clinical Sciences, Northern Ontario School of Medicine, Thunder Bay, Ontario, Canada.

Received November 29, 2019; revision received February 10, 2020; accepted February 19, 2020.

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

Supervising Editor: J. Stephen Huff, MD.

Address for correspondence and reprints: Robert Ohle, MBBCh, MA, MSc; e-mail: robert.ohle@gmail.com.

ACADEMIC EMERGENCY MEDICINE 2020;27:887–896.

© 2020 by the Society for Academic Emergency Medicine
doi: 10.1111/ace.13960
ISSN 1553-2712
887
Dizziness is a common presenting complaint in the emergency department (ED). A subset of these patients present with acute vestibular syndrome (AVS), which is a clinical syndrome defined by vertigo, nystagmus, head motion intolerance, ataxia, and nausea/vomiting. The majority of patients with AVS are suffering from a benign disorder called vestibular neuritis. However, in a small proportion of patients, the presenting symptom of vertigo is secondary to a central cause such as a vertebrobasilar stroke. Computed tomography (CT) of the head is not sensitive enough to rule out the diagnosis of posterior circulation stroke, and magnetic resonance imaging (MRI) is often not available in the acute setting.

The Head Impulse test, Nystagmus, positive Test of Skew (HINTS) examination was developed by neuroophthalmologists as a bedside test to rule out a central cause of vertigo in those presenting with AVS. This combination of physical examinations has been incorporated into clinical practice by frontline ED clinicians. We sought to assess the sensitivity of the HINTS examination for use by emergency medicine physicians compared to neurologists/neurootologists in ruling out a central cause for vertigo (i.e., vertebrobasilar stroke) in those presenting with AVS in the ED.

**METHODS**

**Data Sources and Search Strategy**

An electronic search was performed by a trained librarian on PubMed, EMBASE, and Cochrane library databases (January 2009 to September 2019). The search period was based on original derivation published being published in 2009. A combination of keywords and Medical Subject Heading (MeSH) terms were used, including vertigo, HINTS, vestibular stroke, Bayes theorem, sensitivity, specificity, reproducibility of results, physical examination, clinical exam, and diagnostic tests (Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://online library.wiley.com/doi/10.1111/acem.13960/full). A citation search of included articles was undertaken using Google Scholar. The references of included studies were also hand-searched for relevant papers. No language or age restrictions were placed on the searches. Where data were not clear from the published study materials, authors were contacted for clarification. We searched conference abstracts from the Canadian Association of Emergency Physicians, European Congress on Emergency Medicine, and the International Conference of Emergency Medicine from 2009 to 2019. This study was not registered. The results of this study were reported according to PRISMA guidelines. Ethical approval exemption was sought and obtained from the Health Sciences North (Sudbury, Ontario) Institutional Research Ethics Board.

**Study Selection**

We included studies that met the following criteria: 1) Population—adult patients presenting to an ED with AVS; 2) Test—HINTS examination described in adequate detail to generate 2 × 2 table; 3) Reference standard—CT and/or MRI; and 4) Outcome—central cause, i.e., stroke. We included only articles published in peer-reviewed journals and we did not seek unpublished data. The risk of bias was assessed using the quality assessment tool for diagnostic accuracy studies (QUADAS-2) included in systematic reviews as per Cochrane collaborative recommendations.

Two reviewers (RO and SM) completed the review process with the inclusion criteria defined a priori. RO and SM reviewed titles and abstracts independently and then selected studies that met inclusion criteria were compared and a measure of agreement was calculated using the kappa statistic. The selected articles were then retrieved and reviewed independently by RO and HK for inclusion in the systematic review. Each reviewer extracted data from the selected studies independently. Disagreements between the reviewers concerning which studies met inclusion criteria or differences in data extraction were resolved through discussion. In cases where agreement could not be reached through arbitration, adjudication was performed by a third reviewer (DS).

**Individual Evidence Quality Appraisal**

Two authors (VM and RAM) independently used the QUADAS-2 for systematic reviews to evaluate the quality of evidence for the studies identified. These are reported in a risk of bias table in Review Manager 5 software from the Cochrane collaboration. These two authors independently performed the quality assessment and any disagreements were resolved by discussion. They used several a priori criteria to evaluate an individual study’s risk of bias and degree of applicability.
Domain 1: Patient Selection.

- Risk of bias—For each study selected, the inclusion and exclusion criteria were evaluated to determine whether suspected stroke patients who were more or less acutely ill than those typically evaluated in ED settings were included in the study. Depending on the patients included in the study, a spectrum bias or spectrum effect and skew observed estimates of sensitivity (with sicker populations) or specificity (with less sick populations) may be present.8 If studies reported inappropriate exclusion criteria they were recorded as high risk of bias, if no exclusion criteria were reported they were recorded as unclear risk of bias, and if appropriate inclusion criteria were reported they were recorded as low risk of bias.

- Applicability concerns—If studies reported extensive testing and risk stratification of patients prior to entering the study, applicability concerns were recorded as high. If patient testing prior to enrolment was unclear, such as those referred or admitted for investigation of stroke, the study was recorded as unclear.

Domain 2: Index Tests.

- Risk of bias—If the symptoms were not described in sufficient detail to allow reproducibility of the index test (e.g., description of what constituted central nystagmus) applicability was assessed as unclear.

- Applicability concerns—If performed by an emergency physician concerns were marked as low, and if performed by a neurologist/neuroophthalmologist the concern was marked as high.

Domain 3: Criterion Standard.

- Risk of bias—Studies where patients did not undergo one of the acceptable reference standards for neurologic imaging (i.e., CT to rule in a stroke or MRI to either rule in or rule out) were marked as high risk. Note that CT was only deemed criterion standard if there was a positive finding, allowing rule in of a central cause. If a patient had a negative CT but did not undergo an MRI the patient was deemed not to have undergone the criterion standard.


- Risk of bias—Misclassification may occur if index tests are assessed at different time intervals. Any delay in diagnosis may lead to progression of a stroke.9 A patient with more advanced disease may exhibit a wider range of signs and symptoms. Studies that recorded signs and symptoms >14 days from presentation were marked as high risk. The reference standard of MRI may provide a false negative within the first 24 hours after symptom onset; therefore, studies that used a negative MRI <24 hours after symptom onset to exclude stroke were recorded as high risk.

Diagnostic Accuracy: Data Extraction and Data Analysis

For the diagnostic accuracy (i.e., discrimination performance) of the HINTS examination, data were extracted and compared to the diagnosis of a central cause as defined by a reference standard. Two reviewers independently extracted the data using a standardized data collection tool.

A bivariate random-effects model was used to compute summary diagnostic sensitivity, specificity, and likelihood ratios, which allowed for an evaluation of heterogeneity beyond chance as a result of clinical and methodologic differences between the studies. Study heterogeneity is displayed through both confidence intervals (CIs) and $I^2$. The broad quantification of heterogeneity using the $I^2$ statistic is 0% to 40% = low/might not be important, 30% to 60% = may represent moderate heterogeneity, 50% to 90% = may represent substantial heterogeneity, and 75% to 100% considerable heterogeneity.10 In addition, heterogeneity was visually assessed with forest plots. Studies were not pooled if significant clinical or statistical ($I^2 > 40\%$) heterogeneity was present. Analyses were carried out using SAS software using the “metadas” command, in addition to RevMan 5 and Meta-Disc.11,12

Results were presented as sensitivity and specificity as well as positive and negative likelihood ratios. We a priori defined the definition of clinically useful likelihood ratios, not useful (LR+ = 1 to 2, LR− = 0.6 to 1), potentially useful (LR+ = 2 to 5, LR− = 0.2 to 0.5), and useful (LR+ = >5, LR− = <0.2).13

RESULTS

Our search yielded 2,695 titles and abstracts for screening (see Data Supplement S1 for the search strategy). A total of 11 articles met the inclusion criteria after full-text review and were retrieved (Figure 1).
The kappa value for the full-text review was 0.89 (95% CI = 0.82 to 0.95). Detailed characteristics of all selected studies are presented in Table 1. Forest plots of selected studies are presented in Figure 2. The risk of bias summary for each selected study is shown in Figure 3. All studies enrolled patients presenting with AVS and a suspicion for a central cause for vertigo.

**Characteristics of Included Studies**

Results of the quality assessment are shown in Figure 3. The overall quality of the selected studies was considered acceptable for the majority of the quality criteria. The lowest rated items were patient selection inclusion and conduct of the index test. Most studies had a high prevalence of stroke, raising concerns for spectrum bias. Adherence to the STARD reporting guidelines was overall poor with only two of five reporting on most items.\(^1\)

The prevalence of stroke among adult patients presenting to the ED with AVS ranged from 9.3% to 44% (mean ± SD = 39.1% ± 17.1%). All studies were prospective in design (n = 5) with participants recruited from a combination of hospital inpatients and the ED (Table 1).

**Accuracy of HINTS Examination**

The sensitivity and specificity of the HINTS examination were 96.7 (93.1–98.5) and 94.8 (91–97.1) for neurologist/neuroophthalmologists. There were no studies including only emergency physicians; however,
<table>
<thead>
<tr>
<th>Study</th>
<th>Batuecas-Caletrio22</th>
<th>Carmona23</th>
<th>Chen19</th>
<th>Kerber16</th>
<th>Newman-Toker20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>91</td>
<td>114</td>
<td>20</td>
<td>202</td>
<td>126</td>
</tr>
<tr>
<td>Study design</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Prospective</td>
<td>Prospective</td>
</tr>
<tr>
<td>Country</td>
<td>Spain</td>
<td>Argentina</td>
<td>Australia</td>
<td>United States</td>
<td>United States</td>
</tr>
<tr>
<td>Setting</td>
<td>Inpatient</td>
<td>Inpatient</td>
<td>ED referred</td>
<td>ED/inpatient</td>
<td>ED/admitted (with diagnosis of cerebellar stroke)</td>
</tr>
<tr>
<td>Age (years), mean ± SD (range)</td>
<td>56 (17–82)</td>
<td>59 ± 11</td>
<td>64 (42–83)</td>
<td>64</td>
<td>—</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>49 (53.8)</td>
<td>11 (61.1)</td>
<td>14 (63)</td>
<td>160 (50)</td>
<td>89 (70.6)</td>
</tr>
<tr>
<td>Stroke, n (%)</td>
<td>8 (8.8%)</td>
<td>46 (37%)</td>
<td>107 (44%)</td>
<td>24 (12%)</td>
<td>52 (41%)</td>
</tr>
<tr>
<td>AVS</td>
<td>Unclear</td>
<td>Sudden onset of vestibular symptoms and signs: vomiting, nystagmus, and postural instability</td>
<td>Those with acute prolonged rotatory vertigo associated with nausea and/or vomiting, without other brainstem signs, were included</td>
<td>Acute ongoing dizziness and nystagmus (spontaneous or gaze evoked) or imbalance when walking (cannot walk 10 steps in tandem for analysis used those with nystagmus)</td>
<td>Acute, persistent vertigo or dizziness with nystagmus, plus nausea or vomiting, head motion intolerance, and new gait unsteadiness</td>
</tr>
<tr>
<td>Study population</td>
<td>AVS/vestibular neuritis and admitted to ENT unit for further work up</td>
<td>Acute isolated vertigo to the ED was identified by referral. The indications for referral were uncertain diagnosis, presence of vascular risk factors and failure of symptom improvement for safe discharge.</td>
<td>AVS and underwent an MRI</td>
<td>AVS presenting within 1 week of symptom onset. Patients were required to have one or more stroke risk factors.</td>
<td></td>
</tr>
<tr>
<td>Exclusion</td>
<td>Hearing loss or symptoms &lt;12 hours, no MRI</td>
<td>Not reported</td>
<td>Tinnitus, antecedent viral illness Prior diagnosis of or attacks suggestive of Meniere’s disease Vestibular migraine (Neuhauser’s criteria) Corticospinal tract dysfunction Appendicular and truncal cerebellar signs Hemianopia or other visual field defect Horner’s syndrome Sensory disturbance (facial or limb) Facial palsy Bulbar dysfunction and dysarthria Dense ocular motor signs—3, 4 or 6th nerve palsy, internuclear ophthalmoplegia, gaze palsy</td>
<td>Age &lt;18 years, prisoners, patients not fluent in English or not able to provide informed consent because of cognitive or psychiatric impairment more than 14 days since onset of continuous dizziness at the time of study examination Chronic recurrent dizziness History of multiple sclerosis Dizziness thought to be the result of trauma, orthostatic hypotension, medication/drug intoxication, or a known medical or neurologic disorder (e.g., hepatic encephalopathy, hydrocephalus) Posterior canal benign paroxysmal positional vertigo Moderate to severe, new, CNS examination abnormalities Patients with a contraindication to MRI</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
one study included both emergency physicians and neurologists and the sensitivity and specificity were 83.3 (63.1–93.6) and 43.8 (36.7–51.2).

### Heterogeneity

#### Clinical Heterogeneity.

There was a wide variation in the prevalence of AVS, the setting of recruitment and differences among the inclusion criteria for the studies selected (Table 1).

#### Statistical Heterogeneity.

The $I^2$ was 0 for both sensitivity and specificity for the HINTS examination when performed by neurologists. There was only one study that included emergency physicians (Table 2).

### DISCUSSION

#### Principal Findings

This systematic review and meta-analysis demonstrated that the HINTS examination performed by emergency physicians cannot in isolation rule out a central cause for vertigo in patients presenting with AVS. The HINTS examination has been proposed as a bedside assessment tool to rule out a central cause for vertigo without advanced imaging and has been incorporated into the practice of emergency physicians.\(^4\),\(^15\) We only found one study that examined the accuracy of the HINTS examination when used by an emergency physician with vascular/neurology fellowship training. In Kerber et al.\(^16\) the HINTS examination was performed by specialist trained neurologists and emergency physicians. The sensitivity and specificity in this study were lower than those in all other studies with the HINTS examination performed by specialist neurologists/neuroophthalmologists. They also found a significant difference in the accuracy of performing the HINTS examination between emergency physicians and neurologists, as demonstrated by a kappa of 0.24 to 0.4. The main outcome in the study by Kerber et al. was restricted to only stroke/intracerebral hemorrhage and ignored other potential central causes of vertigo (e.g., multiple sclerosis or brain tumor). This exclusion of other diagnoses could in part explain the lower specificity; in the remaining studies they classified any central causes for AVS as a true positive. The nearly 50% false-positive rate (and associated low specificity) observed in the study by Kerber et al. could result from the implicit and often necessary ED approach of caution rather than definitive rule outs.
To assure patient safety while maintaining clinical flow, it is possible that some ED physicians use the test primarily to rule out “clear cases” rather than to make a definitive evaluation. This tendency would inherently hurt specificity; however, given the low number of patients that present with AVS every year, a low specificity is not as much of a concern as low sensitivity.

One retrospective study was found with the search strategy that explored the accuracy of the HINTs examination when performed by emergency physicians without any specialist training. This study was excluded due to its retrospective design. It reported that 96% of HINTS examinations were performed on those who did not meet the inclusion criteria of a patient presenting with AVS resulting in a sensitivity of 66.7% (95% CI: 46.3–98.3) and a specificity of 92.7% (95% CI: 89.4–98.8). The data are shown in Table 2.

**Table 2**

<table>
<thead>
<tr>
<th>Test</th>
<th>No. Patients</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HINTS (16, 22–25)</td>
<td>617</td>
<td>87.5 (46.3–98.3)</td>
<td>52.8</td>
<td>94.5</td>
<td>24.2 (7.7–75.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99.8 (82–99.9)</td>
<td>93.8</td>
<td>95.0</td>
<td>16 (6.6–39.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99.4 (69–1)</td>
<td>90.1</td>
<td>95.0</td>
<td>17.2 (3.6–81.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>83.3 (63.1–93.6)</td>
<td>43.8</td>
<td>94.5</td>
<td>1.48 (1.19–1.85)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>96.8 (91.7–98.8)</td>
<td>98.5</td>
<td>92–99.8</td>
<td>63.9 (9.1–446.8)</td>
</tr>
</tbody>
</table>

LR+ = positive likelihood ratio; LR− = negative likelihood ratio.
When only those with a documented history and physical examination that was consistent with AVS were analyzed, the sensitivity of the test was 100% (95% CI = 15.8% to 100%). Unfortunately, not all patients were assessed with criterion standard of MRI; therefore, sensitivity may be falsely elevated. Studies by Vanni et al. examined those presenting with AVS using a partial HINTS examination together with a neurologic examination and the ability to stand/ambulate. These studies were conducted in the ED by the agreement between ED physicians for the assessment of nystagmus ($\kappa$ = 0.93) and the head impulse test ($\kappa$ = 0.73) were excellent. They do not report agreement between ED physicians and neurologists on the individual components of their algorithm but overall for a diagnosis of central versus peripheral vertigo the agreement was excellent with a kappa of 0.73. A limitation of these studies is that ED physicians used Frenzel goggles to assess nystagmus. These goggles use high-powered magnifying glasses with an illumination system. The magnifying lens prevents fixation that can suppress nystagmus. These studies suggest that using components of the HINTS examination together with a neurologic examination may improve the agreement between ED physicians and neurologists when diagnosing central causes of vertigo for patients with AVS.

The quality of the selected studies was low with all showing some risk of bias. A CT or MRI was required as a reference standard for all patients recruited into the selected studies. This may have introduced partial verification workup/referral bias. This occurs when only patients tested with the criterion standard are included in the study and components of the index test (i.e., HINTs examination) make it more likely that a patient undergoes the reference standard. This bias will falsely increase the sensitivity of the index test by lowering the rate of false negatives. Four out of the five studies required an emergency physician to identify the patients by making a referral for the patient to be assessed for admission to hospital.

The studies selected did not report on the sampling methodology; it is likely that they used convenience sampling, which could introduce a spectrum bias by excluding those who were indeterminate or ambiguous cases and thus falsely increase the sensitivity and specificity. In addition, the studies by Chen et al. and Newman-Toker et al. required patients to have at least one vascular risk factor, which could partially account for the high prevalence of stroke in both of these studies, contributing to spectrum bias. In addition there were admitted patients included as part of the study population in all included studies. The studies where the HINTS examination was performed by a neurologist or neurootologists likely led to a detection bias, which will artificially inflate both sensitivity and specificity.

Stiell and Wells proposed criteria for the development of clinical decision aids, including derivation, validation, and impact analysis. The HINTS examination has been derived and validated for use by neurologists and neuroophthalmologists. We only found one validation study for its use by emergency physicians in which the sensitivity was not sufficient to rule out a central cause for AVS. Vanni et al. although not including the full HINTS examination suggests that in combination with a neurologic examination the HINTS examination may be useful in ruling out a stroke. The only study examining emergency physicians accuracy reports a sensitivity that is not sufficient to rule out a central cause for a patients with AVS. In addition, there has been no formal impact analysis. The retrospective study suggests that the unstructured implementation of the HINTS examination has resulted in inappropriate application to a large number of patients significantly affecting its diagnostic accuracy. The lack of evidence supporting the use of the HINTS examination by emergency physicians is not evidence that it should not be performed by emergency physicians. However, care should be taken and physicians should have a low threshold for referral/MRI in those presenting with AVS and one or more stroke risk factors.

**LIMITATIONS**

The low number of studies meeting the inclusion criteria reduced our ability to investigate for publication bias. Four out of five studies had a component that was at high risk of bias according to the QUADAS-2 assessment. We used a librarian-assisted search strategy together with citation searches and conference abstracts for emergency medicine and neurology conferences but it is possible that relevant studies may have been missed. The sensitivity, specificity, and LRs are properties of the test and are not expected to be
different with higher prevalence. The positive and negative predictive values are properties of both the test and the population being tested and thus the predictive value of the HINTS test in groups with different prevalence of stroke will be different.

**Future Research**

Future research should focus on validating the use of the HINTS examination by appropriately trained emergency physicians. After validation the next step is implementation and assessing its impact on patient-oriented outcomes, such as reducing complications from a missed stroke, reducing advanced imaging, and reduced time in the ED. Given the low number of patients that present per year with AVS this would likely need to be a large multicenter study.

**CONCLUSION**

The HINTS examination, when used in isolation by emergency physicians, has not been shown to be sufficiently accurate to rule out a stroke in those presenting with acute vestibular syndrome.

**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.13960/full

Data Supplement S1. Search strategy for MEDLINE.
Confidence at 100%: Characteristics of Likelihood Ratio Confidence Intervals in the Emergency Medicine Diagnostics Literature

Ari B. Friedman, MD, PhD, Aric W. Berning, BS, and Keith A. Marill, MD, MS

ABSTRACT

Objective: We hypothesized that “perfect” 100% sample sensitivity or specificity (PSSS) is common in the emergency medicine (EM) literature. When results yield PSSS, calculating the likelihood ratio (LR) 95% confidence interval (CI) has been challenging. Consequently, we also hypothesized that studies with PSSS would be less likely to report the LR and associated CI, and those that did would use imperfect methods.

Methods: We searched PubMed or Scopus for all articles reporting diagnostic test results in the 20 top EM journals from 2011 to 2016 and randomly sampled 124 articles. Trained researchers coded the articles as having PSSS or not (controls). We separately sampled 100 articles with PSSS and compared them to 100 controls in terms of their reporting of diagnostic tests and associated CIs.

Results: Of the 124 articles, 19.4% (95% CI = 13% to 27.6%) feature a diagnostic test with PSSS. The LR is reported significantly less often in PSSS studies versus control studies: 18 of 100 articles (18% [95% CI = 11.3% to 27.2%]) versus 34 of 100 articles (34% [95% CI = 25% to 44.2%]), with an odds ratio (OR) of 0.43 (95% CI = 0.21 to 0.86). The LR 95% CI is also reported less often in PSSS versus control studies: five of 100 articles (5% [95% CI = 1.9% to 11.8%]) versus 27 of 100 articles (27% [95% CI = 18.8% to 37%]), with an OR of 0.11 (95% CI = 0.02 to 0.44). Five articles with perfect sample sensitivity reported their negative LR CI. The bootstrap method resulted in CIs that were 42.7% smaller on average (range = 16.6% to 63.6%).

Conclusion: This analysis provides systematic evidence of diagnostic test reporting in the EM literature. Sample sensitivity or specificity of 100% is common. LRs and their associated 95% CIs are infrequently reported, particularly for PSSS samples. When the LR CI is reported in this scenario, it is overly wide. Improved reporting and methods can enhance the utility and confidence in diagnostic tests in EM.

To substantially reduce the probability of uncommon but highly morbid and/or mortal events, diagnostic tests and clinical decision rules designed for emergency medicine (EM) literature often aim to have a sensitivity approaching 100%. Given the finite sample sizes used in research, this focus on nearly perfect sensitivity may result in 100% sensitivity within the study sample by chance, even though the test’s true underlying sensitivity is almost never perfect. When 100% sample sensitivity occurs, clinicians using the study results may overestimate the certainty given by the test because human cognition seems to have a discontinuity around 100% (certainty) versus values close to 100% (which carry uncertainty). This interpretation problem is particularly marked when confidence intervals (CIs) are not reported.

For instance, a clinician evaluating a study with 100% sample sensitivity of a clinical decision rule to...
rule out subarachnoid hemorrhage might conclude that since the rule was negative, his or her patient was very unlikely to have subarachnoid hemorrhage as the cause of their headache, yet if informed that the lower bound of the 95% CI for sensitivity in that same study was 63%, that same clinician might pursue computed tomography imaging or lumbar puncture to fully rule out a sentinel bleed.2

Clinicians can also be misled in using statistics that do not take the population prevalence into account. For this reason, likelihood ratios (LRs) are the preferred method for clinical interpretation of the diagnostic tests typically used in EM and are recommended by reporting guidelines for diagnostic accuracy studies.3 Unlike positive/negative predictive values, LRs can be interpreted outside the context of the study’s sample because they are not based on disease prevalence. While sensitivity and specificity also are population-invariant in theory,4 LRs also allow a method of interpretation that more closely aligns with clinical decision making. By multiplying the LR by the pretest odds of disease, LR speaks directly to the patient’s probability of disease once the test result is known.

Under perfect sample sensitivity or specificity (PSSS), LRs become more difficult to calculate, in part because a denominator in commonly used formulas becomes zero, leading to an undefined result. However, CIs can still be calculated by several methods, including available free software using the “bootLR” method of Marill et al.5 For samples with perfect sensitivity, this approach identifies the highest population sensitivity likely to yield a sample sensitivity of 100% and bootstraps the LR 95% CI using this population sensitivity and the sample specificity. Unfortunately, LR 95% CI results obtained may vary across different calculation methods, particularly when applied to samples with extreme (e.g., 100% sensitivity) data.

We hypothesized that sample sensitivities or specificities of 100% are common in EM journals and that studies with PSSS tend to have smaller sample sizes than non-PSSS studies. We further hypothesized that reporting of LRs and their CIs would be less common among papers reporting PSSS, due to the technical challenges of doing so. Finally, we sought to determine the degree to which using the bootLR method to calculate LR CIs produced a change in the test result.

METHODS
Selection of Journals and Articles Reporting Diagnostic Tests
We identified the top 20 journals in EM using the 2014 ISI list, ranked by eigenfactor.6 For each journal, with the assistance of a biomedical librarian, we used PubMed to obtain a list of all articles in the journal with diagnostic tests, as identified by the structured keywords “sensitivity” or “specificity.” This search strategy to identify diagnostic studies was previously validated, and “the single term ‘specificity’ identified essentially all relevant studies compared to an exhaustive manual search.”7 The journal Emergencias was not indexed in PubMed/Medline and therefore was searched via Scopus. Articles from 2011 to 2016 were included regardless of language (by chance no non-English papers were sampled). Because the goal of this investigation was to determine how results were reported, we included only published papers.

This search strategy located 1,415 papers reporting diagnostic tests in the top 20 EM journals in PubMed and Scopus published during 2011 or afterward. Sixty-three of these papers were excluded because they lacked an abstract, typically because they were editorials or other article types unlikely to primarily report the results of diagnostic tests. This left 1,352 papers as our sampling frame of all articles reporting the results of diagnostic studies in the top 20 EM journals during the study period (2011-2016). Figure 1 diagrams the enrollment of papers into the two analyses.

Power Calculations
To identify the prevalence of articles with PSSS in a simple random sample of our sampling frame, we specified an estimated prevalence of PSSS of 15% and a 95% CI of less than 10% on either side of the point estimate; the resulting power analysis recommended sampling 120 papers.

To estimate the number of papers necessary to compare the reporting of the LR CIs between PSSS and non-PSSS papers, we assumed that one-third of studies without PSSS report a LR CI and one-sixth (half of one-third) of these with PSSS report a LR CI. We calculated that 100 studies are necessary in each group to detect a difference in LR CI reporting with a power of 0.8 and two-tailed alpha of 0.05.
Sampling Strategy for Prevalence Estimation and Non-PSSS Studies

We randomly ordered the papers in the sampling frame and then began at the top of the list and manually evaluated the abstract of each to confirm that the paper met inclusion criteria. To minimize the total number of papers required to be manually abstracted, we collected enough papers in the initial prevalence sample to include 100 non-PSSS papers. This required manual review of 309 papers, at which point there were 124 articles in the simple random sample used to determine the prevalence of PSSS among diagnostic studies in the top 20 EM journals.

Sampling Strategy for PSSS Studies

The 100 non-PSSS papers located earlier for the prevalence analysis were reused for the second analysis comparing reporting of diagnostic test statistics among PSSS compared to non-PSSS papers. To obtain 100 PSSS papers, we conducted a free-text search of the abstracts for the phrase “0%” or “100%.” In the overall sampling frame, 234 of the 1,352 papers matched this text. These 234 potential cases were randomly sorted, and beginning at the top of the list, each sampled case was evaluated by an investigator to determine whether it actually contained an analysis with PSSS until 100 cases were reached. This required evaluating 130 potential cases. The sampling frame was papers rather than individual diagnostic tests within papers; where an article contained more than one primary diagnostic test, a single test within the paper was randomly sampled unless one of the values had PSSS, in which case that estimate was chosen.

Data Abstraction

The two abstractors (KAM, who trained AB) were both study authors and therefore not blinded to the study hypotheses. They performed all abstract and manuscript evaluations and data abstraction using a standard form (Google Sheets). Approximately 20% of included manuscripts were randomly chosen for abstraction by both investigators to assess inter-rater reliability of data abstraction.

Investigators also categorized each manuscript according to whether it reported the results of a derivation study (where the data used to derive the
diagnostic test under consideration were the same as the data used to calculate the diagnostic test characteristics or a validation study (where a different data set was used to validate the diagnostic test characteristics of a previously reported diagnostic test).

Data Analysis
We used Fisher’s exact test to determine whether the number of studies reporting the LR CI and other binary factors differed between cases and controls. We utilized a Pearson Phi correlation coefficient for correlations between binary variables. To nonparametrically assess the secondary hypothesis (that studies which report 100% sensitivity or specificity have smaller samples on average), we utilized the Mann-Whitney U-test and Hodges-Lehmann estimator of the median of differences. CIs for percentages were calculated by inverting the score test, and use the NORMAL approximation to the binomial distribution, by the rule of thumb that np and n(1 – p) are both greater than five. We recalculated CIs for every analysis’s sensitivity using Wilson’s method to allow for comparison. We used the Kappa statistic and intraclass correlation coefficient as appropriate to assess interrater reliability of data abstraction. All tests were two-sided and used an alpha of p = 0.05 as the threshold for significance. All analyses were conducted in R 3.4.4 (R Foundation for Statistical Computing). This study was presented to the University of Pittsburgh Medical Center Institutional Review Board and exempted from review.

RESULTS
Prevalence
Of the 124 diagnostic studies in EM journals sampled to ascertain prevalence, there were 24 articles with 100% sensitivity or specificity (19.4% [95% CI = 13% to 27.6%]). Overall, 38 articles (30.6% [95% CI = 22.9% to 39.7%]) reported a positive LR (LR+), and 41 articles (33.1% [95% CI = 25% to 42.2%]) reported a negative LR (LR–), with 37 articles (29.8% [95% CI = 22.1% to 38.8%]) reporting both. As Figure 2 demonstrates, papers with high sensitivity or specificity dominate our sample: 74 articles (59.7% [95% CI = 50.5% to 68.3%]) of 124 have greater than or equal to 90% sensitivity or specificity.

Although more papers report sensitivity and specificity than LRs, for papers that do report each statistic, approximately the same percentage report CIs: 84 articles (69.4% [95% CI = 60.3% to 77.3%]) of 121 for sensitivity/specificity and 27 articles (73% [95% CI = 55.6% to 85.6%]) of 37 for LRs. The probability of reporting LRs did not increase over time during the study period, nor was journal ranking associated with the probability of reporting LRs (results not shown). Reporting of positive/negative predictive values was not correlated with reporting of LRs (correlation = 0.11, 95% CI = −0.06 to 0.28).

Reporting of Diagnostic Statistics According to Sample Sensitivity or Specificity
We next compared 100 PSSS articles to 100 non-PSSS articles (Table 1). Approximately the same proportion were derivation and validation studies in each category. PSSS studies were smaller on average. Figure 3 shows the distribution of sample sizes among PSSS and control articles, scaled so each has the same overall area. The median sample size among PSSS articles was 127, compared to 276 among controls (Hodges-Lehmann median of the differences = 127, 95% CI = 70 to 209).

Diagnostic test statistic reporting followed the patterns hypothesized (Table 2). Papers are less likely to report statistics that are more difficult to calculate and/or interpret (sensitivity/specificity vs. LRs and point estimates vs. CIs). This effect is most pronounced for PSSS studies.

Because they are smaller on average, PSSS papers have wider CIs than non-PSSS papers. For non-PSSS papers with sensitivity above 95% but below 100%, the median lower bound of the 95% CI of sensitivity is 93.6%, compared to 85.1% for PSSS papers.

We randomly selected 40 of the 100 PSSS papers to be reviewed independently by two investigators to assess the reliability of data collection. There was agreement across all 40 reports on the LR+, the LR–, and the sensitivity (1.00 intraclass correlation coefficient). There were three disagreements regarding the specificity: one that reflected a different judgment call by the raters, another that reflected errors in reading or transcribing the data, and a third regarding an intraclass correlation coefficient of 0.99.

Reporting of LRs and Their CIs According to Sample Sensitivity or Specificity
Table 2 lists the reporting of diagnostic test metrics for non-PSSS and PSSS reports. Among the 32 papers reporting a CI around both LR+ and LR–, 24 articles (75% [95% CI = 56.2% to 87.9%]) reported the
method they used to calculate their CIs. Of these, all but five stated the software used. An additional two added an arbitrary value to the LR before calculating CIs.

**PSSS Papers**

Of the 100 PSSS cases, 61 (61% [95% CI = 50.7% to 70.4%]) had sensitivity of 100% (zero LR−), 50 (50% [95% CI = 40.4% to 59.6%]) had specificity of 100%

---

**Figure 2** Reported sensitivity and specificity in a representative sample of diagnostic tests in the EM literature. Each abstracted study’s sample size is presented as the logarithmic size of the circle representing that study.

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th>Non-PSSS</th>
<th>PSSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of articles</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Sample size, median (range)</td>
<td>276 (22–151,152)</td>
<td>127 (10–4,827)</td>
</tr>
<tr>
<td>Journal (no. of articles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Am J Emerg Med</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Resuscitation</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>Acad Emerg Med</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Ann Emerg Med</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Emerg Med J</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Pediatr Emerg Care</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Eur L Emerg Med</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>J Emerg Med</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Injury</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Prehosp Emerg Care</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Scand J Trauma Resusc Emerg Med</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>CJEM</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Emerg Med Australas</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Eur J Trauma Emerg Surg</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>World J Emerg Surg</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Validation, % (±SD)</td>
<td>22 (±41.6)</td>
<td>25 (±43.6)</td>
</tr>
</tbody>
</table>

PSSS = perfect sample sensitivity or specificity.
(infinite LR+), and 11 (11% [95% CI = 5.9% to 19.2%]) had both sensitivity and specificity of 100%.

Among 61 perfect sample sensitivity articles, only five reported a LR− CI.9-13 Of this small, underpowered sample, all of the estimates with sensitivity of 100% had a narrower CI using the bootLR method than the method used in the paper itself (Figure 4), with the upper bound of the CI being 42.7% smaller on average (range = 16.6% to 63.6%). A single article with 100% specificity calculated the lower bound of their CI for the LR+ at 36.5.14 With the bootLR method, the lower bound of the CI would be 101.6.

DISCUSSION

Overall, diagnostic tests published in EM journals substantially underutilize the LR, with only 42 of 124

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Non-PSSS</th>
<th>Non-PSSS 95% CI</th>
<th>PSSS</th>
<th>PSSS 95% CI</th>
<th>OR 95% CI</th>
<th>OR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>100 articles</td>
<td></td>
<td>100 articles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reports +/− PV</td>
<td>46</td>
<td>36.1–56.2</td>
<td>70</td>
<td>59.9–78.5</td>
<td>2.72</td>
<td>1.47–5.11</td>
</tr>
<tr>
<td>Reports Sens CI</td>
<td>97</td>
<td>90.8–99.2</td>
<td>100</td>
<td>95.4–100</td>
<td>Inf</td>
<td>—</td>
</tr>
<tr>
<td>Reports LR+</td>
<td>72</td>
<td>62.0–80.3</td>
<td>69</td>
<td>59.0–77.9</td>
<td>0.87</td>
<td>0.45–1.66</td>
</tr>
<tr>
<td>Reports LR−</td>
<td>34</td>
<td>24.8–44.2</td>
<td>20</td>
<td>12.9–29.4</td>
<td>0.49</td>
<td>0.24–0.96</td>
</tr>
<tr>
<td>Reports LR+ and LR−</td>
<td>36</td>
<td>26.8–46.3</td>
<td>20</td>
<td>12.9–29.4</td>
<td>0.45</td>
<td>0.22–0.88</td>
</tr>
<tr>
<td>Reports LR+ CI</td>
<td>34</td>
<td>24.8–44.2</td>
<td>18</td>
<td>11.0–26.9</td>
<td>0.43</td>
<td>0.21–0.86</td>
</tr>
<tr>
<td>Reports LR− CI</td>
<td>27</td>
<td>18.8–36.8</td>
<td>14</td>
<td>8.1–22.7</td>
<td>0.44</td>
<td>0.20–0.95</td>
</tr>
<tr>
<td>Reports LR+ CI and LR− CI</td>
<td>29</td>
<td>20.6–39.1</td>
<td>9</td>
<td>4.5–16.8</td>
<td>0.24</td>
<td>0.10–0.57</td>
</tr>
<tr>
<td>Reports LR+ CI and LR− CI</td>
<td>27</td>
<td>18.6–36.8</td>
<td>5</td>
<td>1.6–11.3</td>
<td>0.14</td>
<td>0.04–0.40</td>
</tr>
</tbody>
</table>

LR+ = positive likelihood ratio; LR− = negative likelihood ratio; PSSS = perfect sample sensitivity or specificity.

*OR refers to a comparison of PSSS relative to non-PSSS studies.

Figure 3 Sample size distribution of PSSS compared to non-PSSS. PSSS = perfect sample sensitivity or specificity.

Figure 4 Comparison of upper bound of reported negative LR (−LR) CIs to bootLR method.
articles (33.9% [95% CI = 25.8% to 43%]) reporting either a LR+ or LR− in our prevalence sample, despite the fact that this is easily calculated from provided data. This phenomenon is unlikely to be simply due to certain papers failing to report a number of metrics other than sensitivity and specificity, as the correlation between reporting a predictive value and a LR is essentially zero. This raises the possibility of reporting bias, with researchers less likely to report statistics which appear less clinically significant. However, we find that failing to report a LR and its associated 95% CI is significantly more likely among papers with PSSS, compared to papers without the technical challenges that this brings. Fortunately, this is a remediable problem.

An improved CI now can easily be obtained using the free bootLR package for R or via the website https://abfriedman.shinyapps.io/bootLRshiny/. The bootLR technique is found to provide narrower CI results when sample sensitivity or specificity is 100%. Specifically, it tends to yield a lower upper bound of the LR− 95% CI when sample sensitivity is 100%. This lowers the maximal odds of disease after a negative test result. The bootLR result allows the clinician to assess accurately the ability of a diagnostic test to alter the odds of disease. These data are critical to evaluate the risk of serious illness, convey this risk to the patient as appropriate, and consider the need for further testing.

In our sample of the EM literature, with its focus on ruling out highly morbid disease processes, 100% sample sensitivity or specificity is reported about 20% of the time. LRs and their associated CIs are less likely to be reported in these cases. We also find, paradoxically, that on average clinicians and researchers should infer less certainty from studies with 100% sensitivity than from studies with 93% to 99% sensitivity. This occurs because PSSS studies are, on average, smaller than non-PSSS studies, which generally results in a wider CI.

Studies with perfect sensitivity and specificity are not intrinsically untrustworthy. They simply require reporting of accurate 95% CIs to assess properly the impact of their results on clinical decision making. For all studies of diagnostic tests, clinical discussion should focus on the 95% CI, not the point estimate. To facilitate better reporting, journals should require reporting of both point estimates and CIs for sensitivity, specificity, and LR+/LR− for every diagnostic test they publish.

Previous systematic evaluations of the diagnostic test literature with an eye toward identifying underreporting of appropriate statistics are few.\textsuperscript{15,16} There are, however, studies examining sample size that use similar methodologies. A number of previous studies have used a journal-based approach—necessary in this case due to the difficulty of locating EM-relevant articles in general medical journals.\textsuperscript{17–19} Our results add to the literature investigating reporting of diagnostic tests. For instance, Lijmer\textsuperscript{20} found larger effect sizes among lower-quality studies. Whereas Selman et al.\textsuperscript{16} found that only 28.2% of articles reporting diagnostic tests in obstetrics and gynecology reported estimates of variability, 7 years later and in a different specialty’s literature, we found that 84 articles (69.4% [95% CI = 60.3% to 77.3%]) reported CIs for sensitivity/specificity. While they did not discuss reporting of CIs, Honest and Khan\textsuperscript{15} investigated diagnostic statistic reporting in 90 systematic review articles from the general medical literature and found similar reporting of point estimates of LR (20/90, 22%) but with lower reporting of sensitivity or specificity (65/90, 72%) and predictive values (26/90, 28%).

\textbf{LIMITATIONS}

Publication bias may increase the proportion of small studies with perfect sensitivity/specificity and impressive LRs published relative to the proportion of such studies actually performed. Consequently, our results may overestimate the proportion of such studies performed. Furthermore, investigators may adjust the threshold of diagnostic tests to obtain PSSS. These are important threats to the validity of small study results that we do not directly address in this report.

The search strategy did not identify papers with 100% sensitivity or specificity that did not report the term “100%.” For instance, a study that reported that “200 out of 200 true positives were detected by the test” would not have been found by the search. However, such papers are unlikely to be systematically different than papers using the term “100%.” This would not affect prevalence estimates.

We considered the idea of searching for the most highly cited EM-relevant articles across all journals, but rejected this approach on the advice of a biomedical librarian. Since EM is such a broad clinical discipline yet one focused on acute presentations of disease, finding only EM-relevant articles across all journals would be insufficiently targeted. By excluding EM-relevant
articles in general medical journals, we likely biased our results toward smaller trials that are more likely to observe 100% sensitivity or specificity by chance.

CONCLUSION

Clinicians can take several findings from this investigation. First, 100% sensitivity or specificity is not rare in studies reporting diagnostic tests in emergency medicine journals. Second, confidence intervals of diagnostic test metrics are underreported in the emergency medicine literature, as even the sensitivity confidence interval was unreported in three in 10 papers in our sample. Third, in the five instances of 100% sample sensitivity where a confidence interval was reported, a bootstrapped confidence interval calculation led to a narrower negative likelihood ratio confidence interval than the original papers’ calculations. Thus, a negative test implies a lower posttest likelihood of morbidity and mortal disease.

The authors acknowledge the assistance and expertise of Melanie Cedrone of the University of Pennsylvania Biomedical Sciences Library in constructing structured database searches, Whitney Kramer and Mia Wells of the Lippincott Library of the Wharton School of Business in locating manuscripts, Amy Yang Liu in abstracting articles, Lindsey Ades in providing research assistance, Brian Beeler of Storage Review in providing assistance with data storage, and Eve Ettlinger in copy editing and formatting assistance. The authors thank the discussants at the Society for Academic Emergency Medicine Annual Meeting for their helpful comments.

References

CME Information: Antacid Monotherapy Is More Effective in Relieving Epigastric Pain Than in Combination With Lidocaine: A Randomized Double-blind Clinical Trial

CME Editor: Corey Heitz, MD

Authors: Jaimee Warren, MBBS, Blake Cooper, MBBS, Anton Jermakoff, MBBS, and Jonathan C. Knott, PhD

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 effectiveness of antacid monotherapy for epigastric pain.

Activity Disclosures
This activity received no commercial support.

CME Editor Corey Heitz discloses no relevant financial relationships.

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

Accreditation
John Wiley and Sons, Inc. is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.
Antacid Monotherapy Is More Effective in Relieving Epigastric Pain Than in Combination With Lidocaine: A Randomized Double-blind Clinical Trial

Jaimee Warren, MBBS¹,², Blake Cooper, MBBS¹,², Anton Jermakoff, MBBS¹,², and Jonathan C. Knott, PhD¹,²,³

This was a double-blind, randomized clinical trial comparing three different solutions for the treatment of adults with epigastric pain or dyspepsia presenting to the emergency department (ED). It was conducted in the Royal Melbourne Hospital, a tertiary, adult-only, inner-city center in Melbourne with 75,000 annual ED visits. Data were collected over 3 months, from June to August 2019, between 0800 and 2300, 7 days a week.

Epigastric pain and dyspepsia in EDs around the world are typically treated with an antacid, either alone or combined with other medications. Such medications include viscous lidocaine, an antihistamine, a proton pump inhibitor, or an anticholinergic.¹,²

The aim of this study was to compare antacid monotherapy, antacid/lidocaine 2% solution, and antacid/lidocaine 2% viscous gel in reducing pain at 30 minutes. The primary outcome was change in pain scores 30 minutes after treatment. Thirty minutes was chosen to match previous studies, and it was expected to be sufficient time for the analgesia to take effect.³,⁴ The predetermined minimum clinically important difference was a 13-mm decrease on a 100-mm visual analog scale (VAS) from baseline.⁵ Secondary outcomes were medication palatability (taste, bitterness, texture, and overall acceptability) using a VAS and change in pain score 60 minutes postadministration.

Patients prescribed an antacid/lidocaine mixture by the treating emergency doctor were approached for enrollment. To replicate clinical practice, no standardized criteria for inclusion were used; rather, the study relied on the treating doctor’s clinical discretion and their documented prescription of an antacid/lidocaine mixture. Patients were excluded if they were unable to consent or were under 18 years of age. The study was approved by the Melbourne Health Human Research and Ethics Committee and preregistered at the Australian New Zealand Clinical Trials Registry (ACTRN12619000928112).

Randomization was conducted in a ratio of 1:1:1 in blocks of six, using a random-number table. Opaque envelopes were prepared by a research assistant not involved in recruitment. The envelopes were provided from the ¹Emergency Department, Royal Melbourne Hospital, Melbourne; the ²Melbourne Medical School; and the ³Centre for Integrated Critical Care, University of Melbourne, Parkville, Victoria, Australia.

Received April 16, 2020; revision received June 22, 2020; accepted June 24, 2020.


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

The study was preregistered at the Australian New Zealand Clinical Trials Registry (ACTRN12619000928112).

Author contributions: all authors were involved in the design of, and recruitment to, the study. JW drafted the manuscript, supervised by JK. JK undertook the analysis of the data. The study was unfunded.

Supervising Editor: Kabir Yadav, MDCM, MS, MSHS.

Address for correspondence and reprints: Jonathan C. Knott, PhD; e-mail: jonathan.knott@mh.org.au.

ACADEMIC EMERGENCY MEDICINE 2020;27:906–909

ISSN 1553-2712 © 2020 by the Society for Academic Emergency Medicine doi: 10.1111/acem.14069
in sequence to the attending nurse and contained instructions to give one of three medication mixtures.

Solutions were not made to look identical, because a secondary outcome of this study was palatability. An attempt to make the solutions of equivalent color, appearance, or viscosity would potentially interfere with these assessments. The volume given was identical; investigators and patients remained blinded to the solution they received but not the nursing staff who prepared and administered the mixtures.

Nurses paged investigators immediately after an antacid/lidocaine mixture was prescribed and before it was administered. Investigators were expected to present to the bedside in under 5 minutes to begin enrollment to minimize delays to analgesia administration.

- Arm 1 (viscous): received 10 mL of oral lidocaine 2% viscous gel plus 10 mL of antacid (traditional antacid/lidocaine mixture).
- Arm 2 (solution): received 10 mL of lidocaine 2% solution plus 10 mL of antacid.
- Arm 3 (antacid): received 20 mL of antacid alone.

Lidocaine 2% viscous gel is manufactured in Australia by Perrigo and consists of lidocaine hydrochloride in a 2.13% weight-to-volume gel for oral use. Lidocaine 2% solution is manufactured by Pfizer as 2% lidocaine hydrochloride for injection. The antacid, Gastrogel, is manufactured by Aspen Pharma. Each 10 mL contains dried aluminum hydroxide gel equivalent to 500 mg, 240 mg of magnesium trisilicate, and 240 mg magnesium hydroxide.

All data were collected electronically in REDCap. Baseline data included date, age, sex, pain at time 0, brief past medical history, current medications, and medications taken prior to ED presentation. Pain was recorded along an electronic VAS from 0 to 100 mm, 0 mm being no pain and 100 mm being maximal pain. The patient self-selected their pain score. Immediately after the first pain score was obtained, medication was administered. After 30 minutes, the pain was scored again, with the patient unable to view their previous score. A similar VAS was used to obtain scores for taste, bitterness, texture, and overall acceptability at 30 minutes, with 0 mm being unacceptable and 100 mm being acceptable. At 60 minutes, a final pain score was obtained. Data regarding effects experienced, ED medications given, and ED discharge diagnosis were recorded once the final VAS was obtained.

Analysis was undertaken on an intention to treat. Proportions were tested for significance using the chi-square test. Continuous variables were assumed to be nonparametric and tested using the Kruskal-Wallis test. Stata software was used for all analyses. The null hypothesis was that there would be no difference in pain scores at 30 minutes comparing the addition of either lidocaine viscous or lidocaine solution to an antacid. Assuming a standard deviation of 15 mm and the equivalence limit of 13 mm (power $= 80\%$ and alpha $= 0.05$), 20 patients needed to be recruited into each arm. Allowing for potential differences in patient allocation and heterogeneous recruitment, the target was increased from 60 to 80 patients.

Lidocaine viscous was accessed 219 times in the ED during the recruitment period, 120 patients (55%) were approached for recruitment, 94 were enrolled, and five were excluded because pain scores were not obtained. Eighty-nine (95%) enrolled patients completed the protocol.

Table 1 outlines patient characteristics and findings. There were no statistically significant baseline differences between the treatment groups. Importantly, all three groups started with a similar pain score. In terms of the primary outcome, solution and antacid provided clinically important (>13 mm) analgesia at 30 minutes; viscous did not. Although the traditional mixture of antacid/viscous lidocaine was least effective and antacid monotherapy demonstrated the greatest degree in pain relief, none of the differences between treatments were statistically significant.

Regarding secondary outcomes, at 60 minutes, all treatment groups experienced additional pain relief. The change in median pain scores was clinically significant (>13 mm) for all three arms. Participants found antacid monotherapy to be the most palatable solution, with statistically significant differences in taste, bitterness, and overall acceptability.

The most prominent adverse effect was oral numbness, experienced by treatment groups containing lidocaine, viscous ($n = 6, 20\%$) and solution ($n = 8, 26\%$). Patients in the viscous arm reported dizziness and tiredness ($n = 2, 7\%$), and patients in the solution arm reported cough, nausea, and dizziness ($n = 4, 13\%$). One patient in the antacid arm reported a dry mouth ($n = 1, 4\%$).

The overall finding of this study was the beneficial effect of antacid monotherapy in multiple ways. In addition to no statistical difference in pain relief at 30 and 60 minutes, antacid monotherapy was favored in terms of palatability and acceptability, and there were fewer side effects.
Previous studies of acute dyspepsia management in the ED have been of varying methodologic quality with mixed results. In a 1990 single-blind study comparing 30 mL of antacid with or without 15 mL of viscous lidocaine, Welling and Watson found that the addition of lidocaine significantly increased pain relief (decreased pain score by 40 mm compared to 9 mm with antacid monotherapy). That antacid monotherapy did not produce clinically significant pain relief contrasted with prior studies that demonstrated just that. Another randomized single-blind study comparing antacid plus either benzocaine solution or viscous lidocaine found no difference between the two arms, but there was no antacid monotherapy arm. A larger, more rigorous double-blind randomized clinical trial in 2003 enrolled 113 patients and compared 30 mL of antacid monotherapy, antacid with 10 mL of an anticholinergic, and antacid with anticholinergic and 10 mL of 2% viscous lidocaine. Similar to our study, Berman et al. found that all treatments were clinically effective and there was no difference in pain relief between the three arms. Their conclusion was to recommend antacid monotherapy.

In addition to being a single-center study, this clinical trial had several limitations. Enrollment was determined prospectively by the prescribing of an antacid mixture by ED medical staff, rather than based upon a final diagnosis of acute dyspepsia. A final diagnosis

Table 1
Population Characteristics and Findings (N = 89)

<table>
<thead>
<tr>
<th></th>
<th>Viscous</th>
<th>Solution</th>
<th>Antacid</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>30 (34)</td>
<td>31 (35)</td>
<td>28 (31)</td>
<td></td>
</tr>
<tr>
<td>Sex, female</td>
<td>21 (70)</td>
<td>21 (68)</td>
<td>15 (54)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>43 (32–70)</td>
<td>38 (28–61)</td>
<td>42 (32–72)</td>
<td></td>
</tr>
<tr>
<td>Past medical history</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acid-related (gastroesophageal reflux disease/peptic ulcer disease/gastritis)</td>
<td>7 (23)</td>
<td>8 (26)</td>
<td>4 (14)</td>
<td></td>
</tr>
<tr>
<td>Other gastrointestinal</td>
<td>4 (13)</td>
<td>2 (6)</td>
<td>6 (21)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>17 (57)</td>
<td>16 (52)</td>
<td>15 (54)</td>
<td></td>
</tr>
<tr>
<td>Previous antacid/proton pump inhibitor use</td>
<td>8 (27)</td>
<td>9 (29)</td>
<td>6 (21)</td>
<td></td>
</tr>
<tr>
<td>Medication count</td>
<td>2 (0–4)</td>
<td>1 (0–3)</td>
<td>1 (0–3)</td>
<td></td>
</tr>
<tr>
<td>Prehospital medication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antacid/proton pump inhibitor</td>
<td>6 (20)</td>
<td>3 (10)</td>
<td>4 (14)</td>
<td></td>
</tr>
<tr>
<td>Other analgesia</td>
<td>9 (30)</td>
<td>10 (32)</td>
<td>8 (29)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>7 (23)</td>
<td>6 (19)</td>
<td>8 (29)</td>
<td></td>
</tr>
<tr>
<td>Emergency medication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proton pump inhibitor</td>
<td>9 (30)</td>
<td>11 (35)</td>
<td>12 (43)</td>
<td></td>
</tr>
<tr>
<td>Other analgesia</td>
<td>19 (63)</td>
<td>13 (42)</td>
<td>13 (46)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>11 (37)</td>
<td>7 (23)</td>
<td>12 (43)</td>
<td></td>
</tr>
<tr>
<td>Discharge diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>22 (73)</td>
<td>27 (87)</td>
<td>22 (79)</td>
<td></td>
</tr>
<tr>
<td>Cardiac</td>
<td>5 (17)</td>
<td>2 (7)</td>
<td>3 (11)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3 (10)</td>
<td>2 (6)</td>
<td>3 (10)</td>
<td></td>
</tr>
<tr>
<td>Initial VAS pain score (mm)</td>
<td>64 (36–81)</td>
<td>65 (31–78)</td>
<td>69 (57–80)</td>
<td></td>
</tr>
<tr>
<td>Change in pain score (mm), t = 30 min</td>
<td>9 (3–26)</td>
<td>17 (7–27)</td>
<td>20 (7–36)</td>
<td>0.30</td>
</tr>
<tr>
<td>Change in pain score (mm), t = 60 min</td>
<td>21 (3–31)</td>
<td>26 (9–41)</td>
<td>32 (13–42)</td>
<td>0.18</td>
</tr>
<tr>
<td>Taste VAS score (mm)</td>
<td>37 (12–62)</td>
<td>29 (15–50)</td>
<td>76 (34–88)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Bitterness VAS score (mm)</td>
<td>42 (24–82)</td>
<td>38 (12–55)</td>
<td>82 (66–94)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Texture VAS score (mm)</td>
<td>36 (27–78)</td>
<td>52 (32–80)</td>
<td>64 (27–87)</td>
<td>0.26</td>
</tr>
<tr>
<td>Overall acceptability VAS score (mm)</td>
<td>50 (32–79)</td>
<td>57 (50–73)</td>
<td>75 (50–89)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Data are reported as or n (%) mean (±SD). IQR = interquartile range; VAS = visual analog scale.

*Other non-acid-related gastrointestinal disorders.
†Other past medical history unrelated to the gastrointestinal system.
‡Other analgesia included acetaminophen, ibuprofen, fentanyl, and morphine.
§Other medications included glyceryl trinitrate, hyoscine butyl bromide, and ondansetron.
¶Other discharge diagnoses included psychiatric, kidney stones, prostatitis, T12 fracture, vertigo.
of cardiac pathology was made in 14% of enrolled patients; these were spread evenly across the three arms. This subgroup tended to have an increase in pain scores over time and would dilute the efficacy of the antacid mixtures.

The amount of antacid used in the monotherapy arm was 20 mL, compared to 10 mL in the other arms. This was to ensure that liquid volumes were equivalent and keep patients and physicians blinded. No studies could be found on dosage–response curve for antacids. However 10 to 20 mL of the study antacid is within therapeutic guidelines for treatment of dyspepsia. A future study might consider using a minimum of 20 mL of antacid in each arm.

Nursing staff dispensing the medication were unblinded. However, nursing staff did not collect study data and researchers remained blinded to which medication had been given.

In conclusion, a 20-mL dose of antacid alone is no different in analgesic efficacy than a 20-mL mixture of antacid and lidocaine (viscous or solution). Antacid monotherapy was more palatable and acceptable to patients. A change in practice is therefore recommended to cease adding lidocaine to antacid for management of dyspepsia and epigastric pain in the ED.

The authors thank the doctors and nurses in the Royal Melbourne Hospital for incorporating the study protocol into their medical management of patients, Hayley Zarth for her expertise and protocol implementation as an ED pharmacist, and to Maggie Bock and Celine Yap for their logistical support of the study design.

References

The coronavirus disease 2019 (COVID-19) pandemic has significantly affected health care utilization in the United States. Although reductions in routine outpatient visits and elective procedures were intentional in preparation for increases in COVID-19–related volume, National Syndromic Surveillance Program data indicate that weekly emergency department (ED) visits decreased 42% during the early stages of the pandemic. This reduction may have been driven by a public fear of seeking care, ultimately delaying interventions for time-sensitive serious conditions. A group of U.S. hospitals recently reported a 38% reduction in ST-elevation myocardial infarction (STEMI) activations, and a national neuroimaging database indicated a 39% reduction in patients undergoing stroke imaging. To yield a more complete picture of the COVID-19 pandemic’s effect on emergency care, we sought to describe ED visit trends for other serious diagnoses requiring acute intervention or hospitalization.

We conducted a cross-sectional study of adult visits to seven EDs in a large health system located in greater Chicago, Illinois, with a combined 2019 visit volume of 308,000. These seven EDs include one urban academic hospital, five suburban community hospitals, and one free-standing ED. Additional hospital characteristics are included in Data Supplement S1, Appendix S1 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.14099/full). We measured ED visit frequencies for serious diagnoses during the early stages of the COVID-19 pandemic (March 8 to May 2, 2020), a period of time encompassing the World Health Organization pandemic declaration on March 11, 2020, up to the date of data collection. We then compared this early pandemic period to a pre-pandemic period (December 31, 2019, to March 7, 2020) and a historical control period from the prior year (March 10, 2019, to April 27, 2019). Thus, the total study duration spanned from March 10, 2019, to May 2, 2020. This study was approved by the Northwestern University Institutional Review Board.

We selected a limited set of serious diagnoses associated with the need for emergency care by multiple rounds of consensus-driven discussion. This process was guided by the priority selection of conditions requiring time-sensitive intervention (e.g., acute myocardial infarction) or the need for timely therapy to prevent clinical deterioration (e.g., gastrointestinal hemorrhage). We were additionally guided by the need to limit the overall number of diagnoses to preserve meaningful data visualization and to minimize the risk of achieving statistical significance through multiple comparisons alone. We did not include symptom-based diagnoses (e.g., chest pain) given that these are nonspecific by definition.
We identified International Classification of Diseases, Tenth Revision (ICD-10), codes corresponding to selected diagnoses using the Agency for Healthcare Research and Quality Clinical Classifications Software Refined (AHRQ CCSR) system, and we grouped similar diagnoses into the following major diagnosis classes: cardiac (acute myocardial infarction, dysrhythmias, conduction disorders, coronary atherosclerosis), surgical (appendicitis, biliary tract disease, intestinal obstruction), orthopedic (dislocations, fractures), neurologic (hemorrhagic and ischemic stroke, seizure), gastrointestinal (diverticulitis, gastritis/duodenitis and ulcers, hemorrhage, pancreatitis, enteritis and colitis, hepatitis, hepatic failure), chronic respiratory (asthma, chronic obstructive pulmonary disease, heart failure), and COVID-19 for comparison. Because the AHRQ CCSR system was created prior to the COVID-19 pandemic, we identified these ICD-10 codes based on Centers for Disease Control and Prevention guidance. A full list of ICD-10 codes is included as Data Supplement S1, Appendix S2.

We searched ICD-10 codes of interest during the study period via structured query language of the health system’s electronic data warehouse. We restricted our search criteria to the primary ICD-10 code associated with the ED visit in order to avoid double counting ED visits (e.g., a single visit with both appendicitis and atrial fibrillation with rapid ventricular response).

We compared the rate of ED visits for each major diagnosis class in the early pandemic period to the prepandemic period using incidence rate ratios (IRRs) with corresponding 95% confidence intervals estimated by Poisson regression with robust standard errors in Stata v14.2. We used a two-sided significance threshold of $p < 0.05$.

There were a total of 283,187 visits across the seven EDs during the study period. Mean ED volumes were 3,709 visits per week in the early pandemic period compared to 4,854 visits per week in the prepandemic period and 4,839 visits per week in the historical control period. Figure 1 shows the number of weekly ED visits for each major diagnosis class over the early pandemic and prepandemic periods. A longer time series including the historical control period is included as Data Supplement S1, Appendix S3. The first diagnosis of COVID-19 in the health system occurred on March 4, 2020, with COVID-19 diagnoses increasing weekly thereafter until reaching 325 ED visits in the last week of the study period. In contrast, ED visits for all major diagnosis classes decreased in the early pandemic period compared to the prepandemic period (all $p \leq 0.001$): cardiac (41.0 vs. 78.3 weekly visits, IRR = 0.52 [95% CI = 0.44 to 0.62]), surgical (53.6 vs. 80.9 weekly visits, IRR = 0.66 [95% CI = 0.57-0.77]), neurologic (40.0 vs 57.9 weekly visits, IRR = 0.69 [95% CI = 0.56 to 0.86]), orthopedic (51.5 vs 96.9 weekly visits, IRR = 0.53 [95% CI = 0.41 to 0.69]), gastrointestinal (91.3 vs 131.1 weekly visits, IRR = 0.70 [95% CI = 0.56 to 0.86]), and chronic respiratory (79.4 vs 120.6 weekly visits, IRR = 0.66 [95% CI = 0.52-0.84]). IRRs for the early pandemic period compared to the historical control period were similarly decreased (Data Supplement S1, Appendix S4).

In summary, we noted significant reductions in ED visits for all major diagnosis classes during the early stages of the COVID-19 pandemic, including a 48% reduction in cardiac diagnoses and a 31% reduction in neurologic diagnoses. These findings support previous reports of decreased health care utilization during COVID-19, particularly for STEMI activations and stroke neuroimaging. We extend these previous reports using ED visit data and additionally report reductions in care for other serious surgical and medical diagnoses, including a 34% reduction in surgical diagnoses and a 47% reduction in orthopedic diagnoses.

The state of Illinois and the city of Chicago instituted a number of social distancing measures (e.g., statewide stay-at-home order, closure of city parks) in the early pandemic period to delay community transmission of COVID-19. While these measures might be expected to reduce orthopedic injuries frequently associated with outdoor activities and motor vehicle use, they would not be expected to decrease the incidence of other major diagnoses, such as acute myocardial infarction or acute appendicitis. These diagnoses would be expected to occur at a similar rate in the early pandemic and prepandemic periods, given the random probability of any event (e.g., appendicitis) occurring in any member of the population at a given time. Our study findings therefore support ongoing concerns that the public may be refraining from seeking care for a wide breadth of serious conditions due to fears of COVID-19 exposure in the health care setting.

It is possible that many serious diagnoses, such as acute myocardial infarction or stroke, could be directly related to primary COVID-19 infections given the heterogeneity of disease findings reported and our relative lack of knowledge regarding the pathophysiology of this novel disease. However, this possible association
would be expected to increase the incidence of ED visits for these serious diagnoses (i.e., in the direction of the null hypothesis), rather than reduce the number of ED visits observed in the early pandemic period.

We did not evaluate ambulatory minor conditions (e.g., lacerations) given the focus of this study on serious time-sensitive diagnoses, nor did we evaluate symptom-based diagnoses, such as chest pain or headache, which may ultimately have been caused by more serious conditions undiagnosed during the ED evaluation. However, we anticipate that ED visits for both nonemergent and symptom-based diagnoses might have similarly declined, given that others have reported that encounters for routine preventive care, such as pediatric vaccinations, have decreased during the COVID-19 pandemic.9

This study is limited by its focus on a single health system in the greater Chicago area. Although we compiled data from seven EDs with diverse care settings, our findings may not be generalizable to other settings with different patient population characteristics and/or local government responses to the COVID-19 pandemic. Additionally, this study is limited by its reliance on ICD-10 coding data, which are dependent on the impression of the treating ED physician. We attempted to minimize this limitation by excluding symptom-based diagnoses and including diagnoses that typically require objective findings available during the ED evaluation (e.g., appendicitis). Finally, ICD-10 codes are fundamentally a billing mechanism rather than a clinical organization system. Although we matched ICD-10 codes to clinical diagnoses using a validated, publicly available system, it is possible that not all relevant ICD-10 codes for serious diagnoses were captured. However, this limitation would not be expected to differentially impact our comparison of the early pandemic period to the prepandemic period.

In summary, we found significant reductions in ED visits for many emergent and time-sensitive diagnoses during the early COVID-19 pandemic. These findings raise concern that the public may be delaying or avoiding necessary emergency care, which may ultimately increase the burden of morbidity and mortality associated with COVID-19, especially in light of concomitant reports that out-of-hospital cardiac arrests have significantly increased compared to the prior year.10 Future efforts should focus on public health messaging regarding the necessity of seeking care for serious conditions and ensuring the safety of patients presenting to EDs during the COVID-19 pandemic.

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.14099/full
Data Supplement S1. Supplemental material.
Traditionally, authors will present a manuscript at a conference prior to submitting the subsequent manuscript for consideration in a journal. While peer-reviewed publications are typically considered to be the criterion standard in research, presentations of abstracts at conferences have several distinct advantages. First, presentation of abstracts at a conference offers an opportunity for early feedback and review, which can help identify issues prior to submission. This can allow authors the ability to refine their study question or revise their study design prior to completion of the project. In addition to this prepublication feedback, conference presentation can also provide an opportunity to identify future collaborators and gain early presentation experience for more novice researchers. Moreover, the average time delay between conference presentation and eventual publication is 1 to 3 years, so this may allow for an early route for dissemination. However, this must be balanced with increasing time and funding limitations as well as alternate dissemination avenues through social media. In light of these factors, it is important to determine the degree to which manuscripts are presented prior to publication and which factors influence this decision. The first step in this process is to quantify the degree to which this occurs. To the best of our knowledge, there are no data demonstrating what percentage of manuscripts are actually presented as abstracts prior to being published.

This was a cross-sectional, observational study that sought to determine the percentage of manuscripts that were presented at a conference prior to publication and the distribution by publication year, conference type, and study design. The study was conducted in accordance with best practice recommendations for chart reviews. Two investigators independently reviewed all publications from January 2009 to December 2018 from two major Emergency Medicine (EM) journals (Annals of Emergency Medicine and Academic Emergency Medicine). These journals were selected because they had the highest impact factor among EM journals and explicitly require authors to disclose any presentations of the work prior to publication. All original research articles were included. Narrative reviews, case reports/images, letters to the editor, and consensus conference proceedings were excluded.

A data extraction tool was developed, piloted, and modified in accordance with the pilot data. Two investigators independently dual-extracted the data into the data collection form for all included studies. When less than five manuscripts were published from a given study design, they were placed into the “other” category. Any discrepancies were resolved by
consensus with the addition of a third investigator as needed. Data were grouped by year of publication, study design, and distribution of conference types. Data are presented primarily as descriptive statistics including means and standard deviations (SD). Statistical analyses were performed using Microsoft Excel (version 16.1).

A total of 2,409 articles met our inclusion criteria (comprising 1,414 from *Academic Emergency Medicine* and 995 from *Annals of Emergency Medicine*) with 1,192 (49.5%) having been presented as conference abstracts prior to publication. Of those that were presented, the mean (±SD) number of presentations was 1.4 (±0.7). The rate of conference presentations decreased over time, with 57.0% of manuscripts presented in 2009 versus 38.8% in 2018 (Figure 1). However, the mean number of presentations per manuscript was relatively stable over the 10-year period, ranging from 1.2 to 1.5 (Data Supplement S1, Figure S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/acem.13911/full).

Of the total conference presentations, 89.6% were presented at a national or international conference, 8.3% were presented at a regional conference, and 2.0% were presented at a local conference. Overall, 61.7% of presentations were at an EM conference and 38.3% were at non-EM conferences. Among non-EM conferences, the most common specialties were pediatrics (33.2%), toxicology (5.6%), and cardiology (5.2%; Data Supplement S1, Figure S2). When assessed by study design, the most commonly presented studies were before–after studies (70.8%), prospective observational studies (55.3%), and randomized controlled trials (51.6%), while the least common were consensus/Delphi analyses (31.6%) and systematic reviews (27.6%; Data Supplement S1, Figure S3).

This study identified several interesting findings that warrant further discussion. Overall, only half of all manuscripts were presented prior to publication with a notable decline over time. This may be influenced by the increasingly prominent role that peer-reviewed publications have played within the promotion and tenure process.1 Because the time from abstract presentation to manuscript presentation can be 1 to 3 years,2,3 it is possible that authors may opt to forego abstract presentation in order to shorten the time to publication. Additionally, there can be high costs associated with attending a national or an international conference. A recent survey found that the average funding for physician continuing medical education ranges from $3,500 to $5,000 per year.5 This may limit the ability of physicians to attend more than one or two conferences each year. Moreover, competing time commitments may further restrict their ability to present to specific times of the year. Furthermore, with the rise of social media, there are new avenues of collaboration and dissemination pathways that were previously limited primarily to conference attendance. This may also explain the decreasing number of conference presentations over time, which is inversely correlated with the rise of social media. However, it is important to weigh this against the potential for unintentional consequences, including decreased networking opportunities and reduced ability for junior researchers (both residents and faculty) to gain early presentation exposure.6

Another interesting finding was that over one-third of abstracts published in EM journals were presented...
at non-EM conferences. We found that pediatric conferences represented the most common non-EM conference type occurring nearly six times more than the second most frequent conference. This may be reflective of the increasing number of publications seen within pediatric emergency medicine over the past 10 years. Other fields, such as toxicology, cardiology, and neurology were also represented in 3% to 5% of cases. This may be due to increasing expansion of EM research into areas typically dominated by other fields. Importantly, presenting at non-EM conferences provides an opportunity for sharing research outside of EM with the associated fields, as well as opportunities to expand researchers’ networks by attending these conferences.

Finally, we found that before–after, prospective observational, and randomized controlled trials were the most frequently presented, while retrospective, consensus/Delphi, and systematic reviews were less frequently presented. This is an interesting finding, as those most likely to receive modifiable feedback were least likely to have the opportunity to receive presubmission feedback. With prospectively collected data, it may be more challenging to incorporate feedback after a study has already begun. However, because authors often present preliminary data prior to study completion, presenters may identify major design flaws and redesign (and potentially restart) their study to better address these issues before the full study is completed.

This study also has several limitations to consider. First, it was performed in two EM journals and may not reflect the full spectrum of EM research. It is possible that the true incidence of prepublication presentations may be higher or lower than the present study if assessing all EM journals. However, since most journals do not explicitly require that authors list any prior presentations, this was not feasible to perform with most other journals. It is also possible that some authors may have presented their work without disclosing it to the journals. Additionally, the present study is limited to only EM literature and may not reflect that of other fields. Future studies should determine whether this finding is consistent in other fields. Moreover, this study was retrospective in nature and subject to all the inherent limitations of this approach. Finally, the present study does not explain why many authors did not present their research prior to publication and future studies should evaluate this component.

In conclusion, less than half of all original research manuscripts were presented at a conference prior to publication, with a progressive decline in the number of manuscripts presented over time. Future efforts should identify the underlying reasoning for not presenting at a conference and why this trend has been increasing over time.

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.13911/full

Data Supplement S1. Supplemental material.
Hot Off the Press: Troponin Testing and Coronary Syndrome in Geriatric Patients With Nonspecific Complaints: Are We Overtesting?

Justin Morgenstern, MD<sup>1</sup>, Corey Heitz, MD<sup>2</sup>, Chris Bond, MD<sup>3</sup>, and William K. Milne, MD<sup>4</sup>

**BACKGROUND**

Non-specific complaints, such as “fatigue,” “dizziness,” and “not feeling well” are relatively common among elderly patients presenting to the emergency department (ED).<sup>1,2</sup> The appropriate workup of such complaints is not always evident. When employed in populations with very low risk of disease, our tests can result in many more false positives than true positives, frustrating physicians and potentially harming our patients. However, a significant number of elderly patients presenting with non-specific symptoms are ultimately diagnosed with serious or life-threatening conditions.<sup>3</sup> Elderly patients diagnosed with acute coronary syndrome (ACS) are more likely than younger patients to present without chest pain, including up to 20% who present with weakness as their chief complaint.<sup>4,5</sup>

Therefore, the troponin presumably has some role in the workup of elderly patients presenting with non-specific symptoms. This study by Wang and colleagues uses a chart review to explore the frequency of ACS in elderly patients presenting with non-specific complaints and the utility of troponin testing in this population.

**ARTICLE SUMMARY**

This is a chart review study that identified patients aged 65 years and older presenting to the ED with a non-specific chief complaint in whom the treating physician decided to order a troponin. Of 594 elderly patients with non-specific complaints, troponins were ordered in 416 (69%), of whom 52 (12%) were positive. However, only five patients were determined to have ACS. The ED troponin was 80% sensitive and 88% specific for ACS, with a negative predictive value (NPV) of 99.7%, but a positive predictive value (PPV) of only 7.7%.

**QUALITY ASSESSMENT**

Strategies to improve the accuracy of medical chart reviews include training abstractors before starting the study, explicitly defining inclusion and exclusion criteria, precisely defining variables of interest, using a standardized abstraction form, holding periodic meetings to resolve disputes and review coding rules, monitoring the performance of the abstractors, blinding the abstractors to the study hypothesis, and testing inter-rater agreement between multiple abstractors.<sup>6</sup>
Impressively, this study utilized all of these recommended methodologic techniques.

However, there are inherent limitations to chart review methodology that limit the reliability of the reported findings. Clinicians may have altered the chief complaint after the troponin value was reported. For example, in the context of a positive troponin, the physician might decide to emphasize the fact that the patient mentioned chest pain, despite an initial chief complaint of “fatigue” or “multiple complaints,” resulting in the exclusion of patients in whom troponin testing was truly valuable. Furthermore, the retrospective look at these data necessitated a dichotomous approach to troponin values (in which the test was considered either positive or negative at a specific cutoff). However, even before the introduction of high-sensitivity troponins, the troponin result has always had to be interpreted within the clinical context and often considering the trend of multiple values. In a chart review, we lose this crucial aspect of clinician judgment and therefore do not know how these troponin results would have been interpreted in real practice. For these reasons, although we believe the general conclusion that troponin testing in this population is low yield with many false positives, the precise numbers reported should be interpreted cautiously.

**KEY RESULTS**

They initially identified 1,146 potentially eligible patients. After excluding the patients who had a specific complaint listed and those with documented fever, they were left with a total of 594 patients. Of those, 412 (69%) had troponins ordered. The average age of the cohort was 78 years old, 58% were female, and 75% were admitted to hospital. The most common chief complaints were altered mental status (43%), weakness/fatigue (33%), and dizziness (21%).

The troponin was positive in 52 patients (12.6%), but only five (1.2%) were adjudicated as having ACS at the index visit or within 30 days. Focusing specifically on the first troponin in the ED, this results in a sensitivity of 80%, a specificity of 88%, a NPV of 99.7%, and a PPV of 7.7%. Considering all troponins, the sensitivity was 100% (95% confidence interval [CI] = 48% to 100%), the specificity was 81% (95% CI = 77% to 85%), the NPV was 100%, and the PPV was 6.1%.

**AUTHOR’S COMMENTS**

As a general rule, tests perform poorly in populations with a very low prevalence of disease. ACS is relatively rare in patients without chest pain or dyspnea, and therefore we should expect the troponin to be a low value test in this population. However, ACS is more common in elderly patients, and the elderly are also more like to present with atypical symptoms, such as “weakness.” The limitations of chart review data preclude any definitive rules, so clinicians will have to continue to use their clinical judgment to determine whether the nonspecific symptoms in the patient in front of them warrant testing for possible ACS.

**TOP SOCIAL MEDIA COMMENTARY**

Dr. Art Sanders on the SGEM blog: I am concerned about the 19% mortality of elder patients with no documented ACS (15 deaths of 77 patients). The authors note “There is a wide body of literature demonstrating that troponin elevation predicts worse outcomes in a variety of noncardiac conditions.”

Patients died from a variety of other causes such as sepsis, dehydration, etc. Would these other conditions have been diagnosed and warrant admission if the troponin had not been tested? The lesson might be that, especially with high sensitivity troponin, we may need to think differently about troponin testing – as an ED doctor, I may not diagnose the patient as NSTEMI, but admission to assess for the other serious conditions is warranted. If we discourage troponin testing, we may not be aware of some of these serious conditions, especially with atypical presentations in elder patients.

Dr. Wand responds: We did not specifically look into how troponin would perform as a prognostic factor. However, I did go back into the original data set to look at the original presentations and reasons for admission... None of these 16 patients who had an initial troponin elevation and died within 30 days of the index visit had just an elevated troponin as cause of admission. All of them had other reasons that were found during their ER visit that cause them to be admitted.

Again, I am unable to fully comment on whether the troponin added valuable prognostic information or risk stratification on top of the other tests that were abnormal for these patients as that was not our paper’s intention. But I think your question is a
interesting one as a potential follow up study- why do physicians order troponins- is it for prognostic purposes? risk stratification? to help admit? or concern for ACS?

Christian H. Nickel (@replynickel): Great study, great podcast @TheSGEM – topic merits prospective study. HsTroponin algorithms (such as 1 hr rule out) are derived from patients with “symptoms suggestive of ACS” – should (IMHO) not be extrapolated to NSCs [nonspecific complaints]!

Dr. Ken Milne – EBM and Rural (@TheSGEM) responds: Good history, followed by a directed physician examination and then judicious use of investigations. #SGEMHOP

Christian H. Nickel (@replynickel) responds: We seem to have similar problems: Biomarkeritis (Troponinitis) disseminate

Daniel Jafari (@DanielJafari): Interesting paper. I wish the poll was a bit more nuanced (well appearing, I’ll appearing). Paper’s pop admitted 75% of the time, so likely sicker. Also, worth noting single trop 100% sensitive. Question remains: how many unnecessary procedures for 93% FP trops.

Dr. Ken Milne – EBM and Rural (@TheSGEM) responds: Wait until high-sensitivity is widely available in North America. Combine that with an aging population, indiscriminate testing and a zero-miss culture, I’m concerned.

Marc A. Probst (@probstyMD) responds: I agree, there is definitely a risk of doing harm by over-testing and overreacting to “positive” HS-trop. We may need to rethink the clinical significance of a positive trop with these high-sensitivity assays... more of a prognostic factor than diagnostic factor in many cases.

Rick Body (@richardbody) responds: Definitely. The advance in the assays needs to be matched by an advance in our thinking. We need to understand more about what troponin is telling us. It’s a marker of myocardial injury. Our job is to understand what caused that

Paper in a Pic by Dr. Kristy Challen:

TAKE-TO-WORK POINTS

The yield of troponin testing is low in elderly patients presenting with nonspecific complaints, and there are many more false positives than true positives. However, limitations in this study prevent any strong recommendations for practice change. Physicians will need to continue to apply clinical judgment in deciding which patients require a workup for ACS.

References


Accuracy of Physical Examination and Imaging Findings for the Diagnosis of Elevated Intracranial Pressure

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

Summary heading
No single finding is sufficiently sensitive nor specific to diagnose elevated intracranial pressure

Positive LR findings
LR+: 2.0 for pupillary dilation
LR+: 2.20 for compression or absence of the basal cisterns on CT
LR+: 1.92 for midline shift > 10 mm on CT

Negative LR findings
LR-: 0.84 for pupillary dilation
LR-: 0.23 for compression or absence of the basal cisterns on CT
LR-: 0.89 for midline shift > 10 mm on CT

Who was in the studies
40 studies comprising 5,123 patients

NARRATIVE

Elevated intracranial pressure (ICP) is a common but severe complication of several medical and traumatic conditions.1,2 Prolonged increases in ICP are associated with poor patient outcomes in a variety of conditions, including traumatic brain injury (TBI), spontaneous subarachnoid or intracerebral hemorrhage, space-occupying lesion, meningitis, cerebral infarct, and cerebral edema from hepatic encephalopathy.3–5 Invasive ICP monitoring is the reference standard for cases suspected of increased ICP but is not universally available and has several potential complications, including intracranial infection and hemorrhage.5,6 Consequently, clinicians must often rely on noninvasive methods for assessing increased ICP.

The systematic review and meta-analysis discussed here included all retrospective, prospective observational, and randomized controlled trials of adult patients (age ≥ 16 years) in the emergency department or intensive care unit that assessed physical examination findings, brain computed tomography (CT), ocular nerve sheath diameter on ultrasound, or transcranial Doppler indices.7 The reference standard consisted of either an ICP reading 20 mm Hg on invasive ICP monitoring or craniectomy with an operative diagnosis of elevated ICP. The primary outcome was the diagnostic accuracy of the aforementioned tests for diagnosing increased ICP.

The systematic review identified 40 studies (n = 5,123 patients) who met the inclusion criteria. Twenty-four studies were prospective cohort studies, 15 were retrospective, and one was a randomized controlled trial. Twenty studies included patients with TBI, three subarachnoid hemorrhage, two intracerebral hemorrhage, two hepatic failure, one ischemic stroke, and 12 mixed populations of primary brain injury. Only three physical examination findings (pupillary dilatation, motor posturing, and altered mental status) had adequate relevant studies to allow for a meta-analysis. Table 1 demonstrates the findings associated with elevated ICP.
ity may limit the external validity of older studies with regard to current imaging. Furthermore, while 10 studies (1,035 patients) evaluated ocular nerve sheath diameter, the variations in ocular nerve sheath diameter thresholds precluded the ability to perform meta-analyses on this group. Finally, there were limited data on transcranial Doppler with differences in both the parameters assessed between studies.

Based on the existing evidence, most findings are insufficiently sensitive or specific for the diagnosis of increased ICP. Therefore, physical examination findings, CT, ocular nerve sheath diameter, and transcranial Doppler do not appear to be reliable in identifying or excluding increased ICP.

### CAVEATS

This study has several important limitations. Most of the included studies were relatively small, with only 13 studies enrolling more than 100 people. Additionally, the prevalence varied significantly between studies which may lead to spectrum bias, and evidence quality for findings suggestive of elevated ICP was predominantly low or moderate, with only midline shift > 10 mm on CT associated with high evidence quality. Nearly one-third of the studies were retrospective in nature. There was also significant heterogeneity with regard to the etiology and severity of the associated injuries, with no definition of the search strategy pertaining to the specific etiology of elevated ICP. Clinical signs were evaluated independently, which is not typical of clinical practice, and it is unclear how the diagnostic accuracy would change when using combinations of findings. Moreover, the inclusion of studies with either invasive monitoring or intraoperative diagnosis may lead to misclassification of the target condition in the latter case. Advances in imaging quality may limit the external validity of older studies with regard to current imaging.

### References

Medical Expulsive Therapy (Alpha Blockers) for Urologic Stone Disease

John C. Conway and Benjamin W. Friedman, MD

NNT color recommendation
Green (benefits > harms)

Summary Heading
Alpha blockers increase the chance of ureteral stone passing (especially for stones >5 mm)

Benefits in NNT
1 in 4 were helped (stones passed)
1 in 7 were helped (fewer hospitalizations)
No one was helped (no surgical intervention was prevented)

Benefits in percentages
28% higher chance of passing the stone
14% decrease in hospitalization

Harms in NNT (NNH)
No difference in major adverse events

Harms in percentages
No difference in major adverse events

Efficacy endpoints
Passing the stone, hospitalization, surgical intervention

Harm endpoints
Major adverse events (orthostatic hypotension, collapse, syncope, palpitations, or tachycardia)

Who was in the studies
10,509 adult patients with symptomatic ureteral stones less than 1 cm confirmed by imaging

NARRATIVE

Urinary tract stones are common and usually painful. Lifetime prevalence is approximately 10%. Direct health care costs are estimated to be over $10 billion dollars annually. First-line treatment is typically analgesia with nonsteroidal anti-inflammatory drugs until the stone passes. If the stone does not pass spontaneously, urologic intervention may be necessary. Spontaneous passage rates for small stones less than 5 mm is 68% and for stones between 5 and 10 mm is 47%. Certain medications such as alpha blockers are sometimes used to hasten passage of stones and decrease the need for urologic intervention or hospitalization. Alpha blockers act on ureteral alpha-1 receptors and decrease the basal tone and peristalsis, thereby facilitating stone passage. However, conflicting results from randomized controlled trials (RCTs) have limited their use. The systematic review discussed here is an update of a 2014 Cochrane review. It includes several new, large, RCTs.

The purpose of this systematic review was to determine the effectiveness of alpha blockers for adult patients with symptomatic ureteral stones measuring less than 1 cm and confirmed by imaging. The systematic review included 67 trials with 10,509 patients. The included studies compared alpha blockers with placebo or medical therapy with nonsteroidal anti-inflammatory drugs, corticosteroids, or antispasmodics. The primary outcomes were stone clearance (defined as stone free imaging, symptomatic relief, or stone collection by the last day of the trial) and major adverse events (defined as orthostatic hypotension, collapse, syncope, palpitations, or tachycardia). Secondary outcomes included hospitalization and the need for surgical intervention. Subgroup analysis compared stone clearance rates for stones 5 mm or smaller versus stones greater than 5 mm. Further analyses examined...
only high-quality studies, excluding studies at high risk of bias. Overall, the use of alpha blockers was associated with increased stone passage (relative risk [RR] = 1.45, 95% confidence interval [CI] = 1.36 to 1.55, absolute risk difference [ARD] = 30%, number needed to treat [NNT] = 4, low-quality evidence) without increasing the risk of major adverse events. Alpha blockers were also associated with a lower risk of hospitalization (RR = 0.51, 95% CI = 0.34 to 0.77, ARD = 14%, NNT = 7, moderate-quality evidence) and no difference in the risk of surgical intervention (low-quality evidence). The subgroup analysis based on the size of the stone revealed that alpha blockers did not impact passing of stones ≤ 5 mm but did improve passing of stones > 5 mm (RR = 1.45, 95% CI = 1.22 to 1.72, ARD = 30%, NNT = 3, moderate-quality evidence).6 When the analysis was performed using high-quality trials only, alpha blockers increased stone passing (RR = 1.09, 95% CI = 1.06 to 1.13; ARD = 7%, NNT = 15, high-quality evidence, five studies, 4,133 participants) while having no effect on major adverse events, hospitalization, or surgical intervention.6

CAVEATS

This review is limited in several ways. Most importantly, the quality of evidence for most outcomes was low due to several methodologic limitations of the included studies, inconsistency in study results, publication bias, a lack of prospectively stratified subgroups, and clinically important heterogeneity.

The findings of this meta-analysis are consistent with other recently published meta-analyses.7 However, some included RCTs, such as the SUSPEND trial, did not demonstrate a benefit for MET.8–10 The findings of individual RCTs may have been skewed toward no benefit because of limited sample size, a high percentage of smaller stones, and insufficient power to detect group differences between small and large stones. Additionally, a recent, large RCT, the STONE trial, was not included in this meta-analysis. The STONE trial, which included 512 patients found no significant differences in outcomes.11 These findings are unsurprising as this trial has the same limitations as other individual RCTs. Because of the lack of support for MET by several well-designed RCTs, it is important to counsel patients on the potential limitations of the evidence that is being used to recommend MET.

In summary, using alpha blockers appears to be beneficial in increasing ureteral stone passage (especially if stones are >5 mm) and reducing hospitalization. They appear to be safe as they do not increase the risk of major adverse events when compared to placebo, nonsteroidal anti-inflammatory drugs, corticosteroids, or antispasmodics. Because benefit is likely (particularly for stones larger than 5 mm) and there is no apparent harm, we have assigned a color recommendation of green (benefits > harm) to this treatment.

References

Cultivating a Better Understanding of COVID-19 Amidst a Shifting Landscape

As COVID-19 swept across the globe, emergency physicians stood on the front lines of a pandemic unlike any seen in our lifetimes. In areas where cases surged, we rapidly assessed and stabilized large numbers of critically ill patients using the limited data available to us, for that is, fundamentally, what we do. In the beginning, we relied on first principles. Vital signs are vital. Isolate the infected. Intubate ARDS early. Use lung-protective ventilation. Monitor severely ill patients for complications of a diffuse inflammatory state, like coagulopathy and disseminated intravascular coagulation (DIC). These interventions are based on decades of observational studies and clinical trials, and they serve us well in times of uncertainty. But, practicing by analogy, “This disease looks like another disease I know, so I will treat it the same way” is ultimately inadequate. To provide optimal care for patients during this pandemic, we need data specific to COVID-19 itself. We clinician scientists are duty bound to provide these data using sound epidemiologic methods and to communicate our findings to the world.

The first data to emerge during a pandemic are case reports and small case series based on clinical or laboratory observations. Like grass growing in the ashes after a forest fire, these initial data are quick spreading with shallow roots. They provide stability for an eroding ecosystem, but they are invasive, and grassland bears no resemblance to the mature forest of high-quality data that must ultimately inform the practice of emergency medicine. One early case series by Tang et al.\(^1\) described hypercoagulability in 183 patients with COVID-19 and suggested an inverse association between higher D-dimer and fibrin degradation product levels and longer prothrombin and activated partial thromboplastin times and survival. In this study, 71% of nonsurvivors and 0.6% survivors met the criteria of DIC during their hospital stay. A follow-up analysis of 449 patients corroborated these findings and suggested that patients with high sepsis-induced coagulation scores might benefit from anticoagulation.\(^2\) Subsequently, dozens of case reports of pulmonary emboli, deep vein thromboses, and other thromboses were published, and emergency physicians were forced to consider the possibility that a patient’s respiratory distress might be due to both COVID-19 and a pulmonary embolus (PE).

This possibility increased the complexity of the diagnosis and management of patients with COVID-19 exponentially. Because patients with severe COVID-19 almost always have elevated D-dimer levels, the already low specificity of this test fell further. As is the case with intensive care unit patients, postpartum women, and patients with active cancer, the low likelihood of a negative result rendered the D-dimer essentially useless for ruling out PE in emergency department (ED) patients with COVID-19. On the other hand, performing computed tomography pulmonary angiography (CPTA) on every COVID-19 patient who might have a PE was, at best, impractical because of the number of COVID-19 patients who would need scans, and, at worst, dangerous since moving patients to radiology might spread the virus throughout the hospital during patient transport. Finally, without a good understanding of the risk of PE imparted by COVID-19, we could not know how to apply the established diagnostic algorithms or clinical decision rules to COVID-19 patients. Should COVID-19 be added to the PERC rule, Wells score, or revised Geneva score as a high-risk marker? If so, how many points should we assign? How does the risk of PE change with the severity of

---

\(^1\) Tang et al.  
\(^2\) Subsequently, dozens of case reports of pulmonary emboli, deep vein thromboses, and other thromboses were published, and emergency physicians were forced to consider the possibility that a patient’s respiratory distress might be due to both COVID-19 and a pulmonary embolus (PE).

A related article appears on page 811.
the patient’s COVID-19? Is the risk of PE in severe COVID-19 so high that a negative D-dimer is insufficient to rule out PE? Algorithms guiding the diagnostic approach to PE, which had been among the most data-driven in emergency medicine, became entirely open to question. We needed data.

In this issue of Academic Emergency Medicine, Freund et al.³ report the results of the PEPCOV study, a large, international, retrospective analysis of 3,253 patients undergoing CTPA for possible PE before and during the COVID-19 pandemic. In doing so, they attempt to move us out of the grassland of case reports and anecdotes, toward the mature forest of high-quality data that we need to guide the diagnostic workup of PE. Perhaps surprisingly, they found the prevalence of PE in patients with COVID-19 (15%) to be identical to the prevalence of PE in patients without COVID-19 (15%). The findings were robust in subanalyses limited to the pandemic period and when subsegmental PE were excluded. The results of PEPCOV are also in line with a recently published study by Al-Samkari et al.⁴ that found that 7.8% of hospitalized COVID-19 patients receiving standard heparin-based prophylaxis are diagnosed with venous thromboembolism (VTE), and 5.6% suffer bleeding within the first 10 days of hospitalization. These proportions are similar to rates of VTE (7%–8%) and bleeding (5%–6%) in critically ill patients without COVID-19, which raises the question of whether COVID-19 actually represents a pathophysiologically distinct coagulopathy or simply coagulation system activation in the setting of severe inflammation. The clinical importance of this line of inquiry is undeniable, and the results of PEPCOV are enticing since they imply that our well-learned algorithms for the diagnosis of PE still apply.

The coordination of a large, international effort like the PEPCOV study in such a short span of time is remarkable and one for which the authors should be applauded. Achieving a study this size during a global emergency demonstrates a dedication to science and evidence-based medicine that is desperately needed, especially now and especially for those of us in the United States where science is frequently undermined by our elected leaders.

It is this same dedication to high-quality epidemiology that requires us to read the methods of this study critically and to interpret its results with a scientist’s skepticism. PEPCOV represents the largest study of PE and COVID-19 yet performed, but we must remember that bigger is not necessarily better. Large sample sizes can actually magnify bias associated with error resulting from sampling or study design.⁵ Most importantly we must interpret this study in the context in which it was performed, since the potential for bias (ever-present in observational research) can only be amplified by the shifting landscape of an expanding pandemic. The potential for bias was very clearly stated in the publication by Al-Samkari, which noted that “Without a uniform protocol to image all patients with suspected VTE, thrombotic events may have been missed.”

The uniformity of the protocol by which patients are selected for CTPA is critical to our interpretation of PEPCOV as well. Patients were included in PEPCOV when a CTPA was ordered for the suspicion of PE. Prior to the pandemic, we can assume that most CTPA were ordered for patients who had either a high clinical probability of PE or a positive D-dimer. Patients with nonhigh clinical probability and a negative D-dimer typically had PE ruled out without imaging. Before COVID-19, this threshold to order a CTPA was well established and incorporated into numerous guidelines. But what if the threshold to order a CTPA changed during the pandemic? Might this have biased the results of the PEPCOV study? If, during the pandemic, emergency physicians had a higher threshold to order CTPA (e.g., due to concerns about transporting patients with COVID-19 to radiology), then only patients very likely to have PE might be sent for imaging. In this circumstance, the proportion of positive CTPA would appear higher than before the pandemic. Conversely, if, during the pandemic, emergency physicians had a lower threshold to order CTPA, more patients with a low probability of PE would be sent for imaging. A decrease in the (real or perceived) specificity of the D-dimer could cause this. Typically, negative D-dimer tests cull patients with a low probability of PE from the ranks of those requiring CTPA. If, however, the very low specificity of D-dimer testing in COVID-19 patients rendered the test clinically useless, emergency physicians would default to CTPA as the first-line diagnostic test, even for patients with a low probability of PE. In this circumstance, the rate of positive CTPA diagnoses would appear lower than before the pandemic. Might these conflicting biases balance out? Maybe, but only if they were of similar magnitude and they exerted their influences during the same time period, which we cannot know. We must, therefore, acknowledge the possibility that the results of PEPCOV were biased by our changing testing practices during the pandemic. Other potential sources of bias, for example, brought on by
inaccuracy of tests to diagnose COVID-19, must also be considered.

Freund and colleagues recognized the potential for bias. In their discussion, they note that “It is likely that COVID-19 is associated with higher risk of PE in the general population, but our report suggests that this is not the case among ED patients with suspicion of PE.” They point to a similar phenomenon that occurs as a result of the disproportionate testing of pregnant women who present to the ED with symptoms suspicious for PE: “...in the general population [pregnant women] are at increased risk of thromboembolic events. However, in ED patients with suspected PE, pregnancy was not reported to be associated with higher risk of PE.” To put this another way, there is a systematic bias in the diagnostic workup of PE such that the propensity to test pregnant women for PE is greater than increased risk of PE associated with pregnancy. As such, the low prevalence of PE in pregnant ED patients tested for PE tells us more about our testing practices than it does about the epidemiology of venous thrombosis in pregnancy. The same seems true for PEPCOV, which may tell us more about how our testing for PE changed amidst the shifting landscape of the pandemic than about the actual risk of PE in patients with COVID-19.

Clinical epidemiology is difficult to perform well, especially when the ground is moving under our feet. Despite the difficulty, it is our obligation as clinician scientists to generate high-quality data that advance science and saves lives. I applaud Dr. Freund and colleagues for their effort to do just that. It is also our obligation to interpret the results of studies like PEPCOV in light of factors that may have affected the results. This is the process by which the science of emergency care grows.

Christopher Kabrhel, MD, MPH
(ckabrhel@mgh.harvard.edu)
Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA

References
It’s Time to Normalize Observation for Pediatric Minor Head Trauma

The 2009 PECARN minor head injury clinical decision rule (CDR) represented an important advance in pediatric emergency care. For over a decade, it has offered clinicians a proven evidence-based tool to thoughtfully guide neuroimaging for pediatric minor head trauma while dramatically reducing missed clinically important traumatic brain injuries. When appropriately applied it simultaneously limits harmful radiation exposures and better utilizes health care resources. Nonetheless, in spite of the PECARN CDR, neuroimaging of children for minor head trauma in the United States (as studied between the years 2007 and 2015) has not changed appreciably and remains at approximately 30%. In this month’s issue of AEM, Singh et al. describe their experience, with pediatric minor head trauma and the use of a planned observation period prior to making a decision to CT them. This study is similar to a PECARN study published in 2011; this current study, however, was conducted by the PREDICT Collaborative, specifically in New Zealand and Australia, where the CT rates for pediatric minor head trauma are significantly lower than those in the United States (approximately 8%–10%). Planned observation led to an 80% lower adjusted odds of CT use. These findings parallel those of the 2011 PECARN study where observed patients odds of CT use was reduced by nearly half.

While there are clearly numerous factors contributing to the wide differences in CT rates between the United States and other nations, what is clear is the value of observation as an adjunct to medical decision making in minor head trauma. Management of patients in both the high- and the intermediate-risk groups in this study benefitted from planned observation. While all risk groups in the United States could benefit from more pointed clinical decision support tools, near-term research should focus on intermediate-risk groups where clinician comfort levels may vary in how they interpret symptoms. Decisions to perform a medical test, for better or for worse, are steeped in heuristics defined by experiences that may or may not be evidence based. If time can dull our impulses to CT children with minor head trauma, then quality initiatives should support the use of observation to modify current heuristic tendencies. By evaluating numerous patients who improve during their planned observation in the ED, clinicians could potentially learn to shift their medical judgments toward less interventional stances on neuroimaging.

Over a decade has passed since the publication of the PECARN CDR and yet, in spite of sound evidence, we are not making significant progress in reducing overall numbers of CT scans for minor pediatric head injury. Studies, now from both PREDICT and PECARN, have demonstrated the effectiveness of planned observation as a means to trim CT rates. As far medical expenses go, observation time is relatively cheap relative to the costs of unnecessary imaging and cancer. If building planned observation time into the PECARN CDR potentiates the diffusion of the tool to a broader audience, we can begin to chip away at our country’s unnecessarily high CT rate.

Robert L. Cloutier, MD, MCR
(cloutier@ohsu.edu)
Emergency Medicine and Pediatrics, UME Admissions, OHSU and Doernbecher Children’s Hospital, Portland, OR

Supervising Editor: Jeffrey A. Kline, MD

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


Letter to the Editor: D-dimer and Presumptive Diagnosis of COVID-19

To the Editor:

We read with interest, “D-dimer Triage for COVID-19” by Li et al.1 This observational cohort study from Wuhan, China, concludes that a D-dimer performed at admission could be an effective and easily available diagnostic surrogate marker for coronavirus disease-19 (COVID-19). We are concerned that several limitations in the study’s methodology affect the authors’ conclusions.

Our first concern is the use of D-dimer as a rule-in strategy. A negative D-dimer is useful but when positive can be caused by multitude of conditions that are unrelated to COVID-19. The utility of the D-dimer requires age adjustment,2 which this study doesn’t seem to have taken into account as the positive threshold was set as >1.0 mg for all patients.

We are also concerned about the lack of reporting of important outcomes for the population included in this study. No information is provided on the impact of the clinical management on patients with positive D-dimer results. As well, the proportion of patients who subsequently underwent a pulmonary angiogram and the result of imaging studies are not reported. The characteristics of the study population such as age, sex, risk factors, pregnancy, comorbidities, or current medications especially anticoagulation medications are not provided. These data are necessary to be able to understand to which population the conclusions, if valid, could be applied to.

A particular statement caught our attention: “Elevated D-dimer levels could be presumptively diagnosed as COVID-19 and triaged as higher risk.” We feel that this conclusion is erroneously inflating the diagnostic specificity of a positive D-dimer for COVID-19 and we would like the authors to report a 2 x 2 table showing the positive predictive and negative predictive values with a criterion standard PCR test for COVID-19.

Another use for D-dimer might be as a predictor of the severity of the disease rather than its diagnostic value. However, the authors do not report all the relevant outcomes to assess whether or not a positive D-dimer in their cohort was associated with the severity COVID-19.

Finally, because this is a single-center study of 749 patients, we believe that a larger number of patients and reporting of both clinical outcomes and COVID PCR results are necessary to validate the accuracy of D-dimer value as a diagnostic tool in COVID-19 patients.

While we appreciate the authors’ intent to establish diagnostic testing for COVID-19–suspected patients, we feel that this study’s claims are far too bold and require much more transparency of information.

Estabraq Al-Najar, MD
Fahd Alrumaih, MD
Danah Arab, MD
Anne Sophie Drouin, MD
Sophie Gosselin, MD, FRPC
Jake Sawa, MD
(Jacob.sawa@mail.mcgill.ca)
Royal College Emergency Residency Program and Emergency Department, Faculty of Medicine, McGill University, Montréal, Canada

Supervising Editor: John H Burton, MD

References

In Reply:

We thank Dr Al-Najar and colleagues for their letter.1 Our research letter2 from work completed early April 2020 is intended to rapidly communicate our finding of 93% 28-day survival (95% CI = 92% to 95%) for a normal D-dimer on admission and 98% survival (95% CI = 96% to 99%) for a normal D-dimer 2 days later. Because we reported on 749 patients, those estimates had high precision. We chose death as a hard relevant outcome, rather than severity. A total of 100% of patients had criterion standard PCR RNA positivity.

D-dimers are most useful when normal, as with venous thromboembolism evaluation. McGill’s Emergency Department may not see the onslaught of COVID-19 suspects coupled with delayed and limited COVID-19 PCR result availability that plagues so many other hospitals and patients. We suggested that an elevated D-dimer could aid in diagnosing serious COVID-19 for the latter. Regarding reporting outcomes other than death and D-dimer, another manuscript has been under review elsewhere. Regarding a 2 x 2 table for PCR-positive patients, the data are already in our research letter but require facility with subtraction: with a normal D-dimer in 586 of 671 survivors, subtraction shows that it was elevated in 85, and with an elevated D-dimer in 36 of 78 nonsurvivors, subtraction shows that it was normal in 42. The Day 3 results are similarly already presented.

Our findings are for triage. They are not the last word. Replicating findings is always important but wrenching numbers of unattended deaths and morbidity and mortality of our fellow health care workers led us to urgently submit these transparent data and suggest how to use them. We disagree that doing so is "far too bold."

Chenghong Li, MD
Zhenguo Zhai, MD, PhD
Bruce L. Davidson, MD, MPH
(brucedavidson@pobox.com)

1Department of Pulmonary and Critical Care Medicine, Affiliated Hospital of Jianghan University, Wuhan
2Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, China
3Providence Health System, Seattle, WA

Supervising Editor: John H. Burton, MD

References
Recognized as a specialty by the American Board of Medical Specialties over 40 years ago, emergency medicine is unique since its inception—providing care to all patients in all times of unscheduled, time-sensitive need. Members of our specialty—whether working in an academic or a private practice setting—provide this care without regard to a patient’s sex, sexual orientation, gender, creed, physical ability, ability to pay, or race/ethnicity. The last human determinant of health bears emphasis in our current climate of a race-based inequity crisis:

Of its many lessons, the COVID-19 pandemic has certainly taught us the integral importance of public health to the practice of emergency medicine. We, in emergency departments, are often first to detect a public health crisis. In other instances, we find ourselves with a “room with a view” of the scourges of public health crises on the lives of our often most vulnerable patients.

The high-profile, recent killings and assaults by police officers of fellow African-American citizens have highlighted yet another public health crisis—that of systemic racism and unchecked, race-based hatred. As an example of many health effects of systemic racism (to include stress-induced hypertension and its “downstream” effects, lower birth weights, lower rates of renal transplantation, historically unethical experimentation, etc.), African-American men are up to 3.5 times more likely than their White-American counterparts to be slain by police. According to Edwards et al. in a recent study, one in 1,000 African-American men will die at the hands of a law enforcement officer—twice the rate of men in general in the United States. Excessive police force—often disproportionately targeted toward African Americans because of racism—is a public health crisis, too.

Battling the COVID-19 pandemic has afforded an almost unparallel opportunity for the world to witness the life- and limbsaving care we provide to patients every day in emergency departments. The COVID-19 pandemic, a scourge again disproportionately impacting Black and Brown communities—has shown everyone that emergency departments are the place of heroes. Now, it is our time to show the world we are also heroes against systemic racism—a public health crisis literally killing our fellow citizens and our patients from Black and Brown communities.

To rise to the occasion, we in emergency medicine need to be actively antiracist. In doing so, we are being pro-patient—for each and every one of our patients regardless of the color of their skin (or any other social and moral determinant). And in being pro-all-patients, we are being pro—emergency medicine.

It is our duty—regardless of our own backgrounds—to speak up and speak out for our patients and the communities in which they reside. We call on our colleagues in EM to commit to the following pledge to act against racism toward a solution for our patients and communities.

Toward a solution, for our patients and our communities:

We pledge to continuously provide health care equity to the communities we serve regardless of race or ethnicity or any other human factor.
We pledge to leave the bedside and meaningfully and genuinely engage our respective communities.
We pledge to address the social and moral determinants of health that adversely affect our patients and their outcomes.
We pledge to insist on diversity, inclusion, and belonging of all peoples in all matters.
We pledge to speak up and speak out for our patients now and whenever we bear witness to racial injustices and inequities and health disparities.
We pledge to become active antiviolence and proviolence leaders.
We pledge to become antiracists. 
We pledge not to be silent!
Our patients and the communities we serve are counting on it.

Ian B. K. Martin, MD, MBA
(imartin@mcw.edu)
Stephen Hargarten, MD, MPH
Medical College of Wisconsin, Milwaukee, WI
Supervising Editor: Brian J. Zink, MD

References

1. Geronimus AT, Hicken M, Keene D, Bound J. "Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. Am J Public Health 2006;96:826–33. https://doi.org/10.2105/AJPH.2004.060749
The COVID-19 pandemic has created an historic global crisis, resulting in unprecedented stressors on frontline health care systems. The population remains on edge, and hospitals are stretched to capacity. This is not our first pandemic. We have learned from prior outbreaks that, in addition to the sheer physical burden, there are also likely to be important but often unacknowledged behavioral and psychosocial costs.1 Frontline health care workers may be particularly vulnerable to harmful psychological consequences.

The challenges have been monumental. Although acute care providers are no stranger to the extremes of human tragedy, the scale and enormity of the COVID-19 crisis has shaken even the most battle hardened among them. Coupled with the onslaught of extraordinary patient volume and acuity playing out on a daily basis in crowded emergency departments, providers have struggled with nearly relentless physical and psychological exhaustion. Heightened risk of infection, rampant illness, and even fatalities within their own ranks have instigated in many frontline providers previously unheard-of fears for their own personal safety.2 Layered on top of these already formidable stresses are a lurking hopelessness and despair over the ethical struggles that inevitably result when patient survival is weighed against dwindling resources.

Recent data suggest that the most acute phase of the pandemic may be fading or at least leveling off. But determining when the crisis will end, whether it will return, and what the ongoing crush on hospital resources may be is still largely a matter of guess work. What this means for the mental well-being of providers is difficult to know. There are, however, at least two things we can say with some certainty. The news is both good and bad.

First, the good news. Human beings are resilient. We will get through this. It is not uncommon to experience heightened distress in the face of extreme or potentially traumatic events. That response is both normative and adaptive. Moreover, as an abundant body of research has shown, the majority of people exposed to even the most aversive events are able to weather those events with little or no enduring psychological costs. Frontline providers are no exception. Although the intense stresses and strains of acute care can be highly aversive, most health care workers manage those stresses and strains without enduring harm to their mental and physical health.

Now the bad news. There will be psychological casualties. Even at the highest levels of resilience, there are always some casualties. Early evidence has already shown that frontline health care providers are reporting significant increases in symptoms of depression and anxiety.2 We fear that many will inevitably endure longer-term psychological consequences, including posttraumatic stress disorder (PTSD), prolonged grief, or major depression. Even in normal times, health care providers in critical care settings experience PTSD at a considerably higher prevalence than in the general adult population but similar to that seen among disaster survivors.3 All indications suggest that these rates are likely to be even greater for critical care responders in the aftermath of the COVID-19 pandemic. In addition to the enormous toll such psychological costs will extract from the personal lives of providers, they may also compromise treatment and exacerbate the already existing shortage of critical care personnel, especially amongst our health care safety nets, such as emergency departments.

We can improve this situation. Resilience is not static, but rather a matter of flexibility and adaptation.4 With health care systems already taxed to capacity, substantive changes at present may be difficult if not impossible. But there are lessons to be learned and, given the gravity of the current crisis, opportunity to address at least some of these concerns even now.

Consider, for example, the medical culture of stoicism. Frontline providers are surrounded by human pain and tragedy, yet their own suffering is typically...
endured in silence. The experience of talking with and receiving support from others is one of the best predictors of resilience we know.\(^1\) In this digital age pandemic, opportunities to share intimate and personal reactions with colleagues have already blossomed. Health care workers have shared tips on COVID-19 management, while also revealing stories of grief and personal tragedy with their peers throughout the crises via informal chat groups or social media platforms. These efforts have arisen spontaneously and organically, but they highlight an obvious need for the development of more formalized peer support networks (either virtual or in person) that could play a key role in the nurturing of resilience among frontline workers.

Beyond the level of the individual provider, it will be crucial to think of resilience in terms of larger, system-level factors. The term “moral injury” has gained increasing currency among providers as they describe their emotional struggle to reconcile the obligation to provide the best patient care in the face of profound system-level limitations and challenges. This was felt acutely by frontline staff during the crush of the current pandemic, where shortages in essential protective gear, beds, and equipment (e.g., ventilators, oxygen tanks) led to unimaginable despair in emergency departments across the city. Any efforts to foster resilience in providers must occur in concert with a deep, introspective analysis at the health system level to identify factors that can enhance flexibility in resource management during times of crisis.

We are still in midst of the storm of COVID-19, yet already, there are concerning warning signs portending a dramatic mental health fallout among frontline providers. Strategies to support and nurture the resilience of these brave women and men will be essential, as their continued physical and mental health is critical in our efforts for this current pandemic, but also to ensure their readiness for the likely future crises to come.

Bernard P. Chang, MD, PhD\(^1\)  
(bpc2103@cumc.columbia.edu)  
George A. Bonanno, PhD\(^2\)

\(^1\)Department of Emergency Medicine, Columbia University Irving Medical Center, New York, NY, USA  
\(^2\)Department of Counseling and Clinical Psychology, Teachers College, Columbia University, New York, NY, USA

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

Supervising Editor: Brian J. Zink, MD

References

Being the child of an emergency medicine (EM) doctor means a lot of unappetizing dinner conversations. The nauseating remarks are usually casually weaved into common discourse, with a disturbing disregard for the family’s digestion. Perhaps a family member mentions an upcoming dental appointment—an innocent remark—which somehow elicits from my mother the gruesome suctioning noise that occurred when she replaced her patient’s displaced tooth in its socket. The conversations are a great way to make sure no one overeats.

Being the child of an EM doctor means trampolines are out. I am not sure I am even allowed to be friends with someone who has a trampoline.

Being the child of an EM doctor bends time. We never know if Mom is asleep or awake and have been forced to learn the art of swift and silent movements throughout our house. The care and fear with which we approach walking draws comparisons to stepping in a field with land mines or a horror movie with an unnamed beast. You never know which board will creak and awaken the beast from its silent slumber. You do not want to wake the beast. I take very seriously the reports of scientists who discovered that American black bears react as though they are preparing for a sudden attack on them if they are awakened in their winter dens.

Being the child of an EM doctor means my own injuries and illnesses are judged against a warped yardstick. Sometimes, it feels like I am in a silent competition with my mother’s patients to receive validation of my injured-ness. I am not sure if that’s a competition I want to “win.” At a young age, I would become immensely confused as I watched other mothers seem anxious when their child fell during a sporting event. I thought every parent would use obvious limb deformity as the criteria for concern.

Being the child of an EM doctor means never putting nonfood items (like a bottlecap) in my mouth—not even for a second.

Being the child of a pediatric EM doctor means my mother loves interacting with children. Every time I go out into public with my mother, I pray we do not see a baby, for if we do, all hell breaks loose. She begins to talk aloud in high-pitched tones designed to attract and simultaneously soothe the baby. Soon thereafter—my personal least favorite part—she undertakes a conversation with the parents about how special their baby is. I’m sure these are great techniques at work, but they are painful when deployed at Stop-and-Shop.

Being the child of an EM doctor means I have had some sort of medical procedure performed on me in my kitchen.

Being the child of an EM doctor means I am increasingly concerned about Mom’s safety. I acknowledge the risks my mom faces in dealing with COVID-19, but I also acknowledge that she will do her best to protect herself and her family.

Most of all, being the child of an EM doctor means I feel proud. I understand how much work and dedication it takes to be an EM doctor. I accept the loss of nights, weekends, holidays, and normal life—traded for passion and purpose. I feel a sense of pride and reflected glory talking about my mother and what she does for a living. Although I have nothing to do with her profession, I am right there with her in her medical journey. When the mother is called to duty, the son also rises. It’s a family job.

Jameson Mannix
(j.mannix@rivers.org)
The Rivers School, Weston, MA
Supervising Editor: Brian J. Zink, MD
Orange Alert

It was the 3rd of March, 2013, during my internship year as I was headed off to The Aga Khan University Hospital for an evening shift in the ED. After I arrived in the “Back” (noncritical) area, our endorsement was interrupted by an unusual announcement: “The hospital is on orange alert.” We all got extremely worried; orange alert was the color code for a mass casualty disaster! The hospital was anticipating the influx of a lot of patients. A bomb blast had leveled two residential buildings in a probable act of sectarian terrorism. We were all devastated by the news! However, before we could even process this information, the blast victims started flowing in the ED one by one.

I went up to my first patient and performed a focused assessment. As I turned around, I was caught completely off guard: there were close to a hundred people jammed in the small space of the “Back” area—which was supposed to have only 10 official beds. I realized that there was no point looking for my patient's chart or assigned nurse. In fact, there were no MRNs or charts to begin with—it was a state of total chaos!

I grabbed some supplies and went over to my patient to place an IV myself. I felt a little anxious at first; though my hands shook a little, I was able to secure the IV on my first attempt. Shortly thereafter, the ED attending on call asked me to attend to another patient. My second patient had shrapnel injuries to his right groin with possible critical limb ischemia. I secured an IV access and began irrigating his wounds. As soon as I was done, I heard the squeaking sound of a Doppler device. The vascular surgery resident on call seemed to have appeared out of nowhere. After his quick evaluation, he wheeled the patient out of the ED toward the OR himself.

As the night went on, the number of people in the ED kept growing, except that some of them were wearing white coats. These were the residents and attending physicians of different specialties, who—on this particular day—were just doctors and healers attending to patients irrespective of their “specialties.” I could begin to sense that the ED was now in a state of a controlled chaos and there seemed to be some method to the madness.

One by one, I kept attending to various patients. I can’t recall how many IV cannulations, wound dressings and clinical evaluations I did during the night. At last, things settled down around 6 AM. At this point, I got a chance to use the restroom and drink some juice. It almost felt unreal and unbelievable. I had not sat down for a second during the whole night, but yet, I never felt the slightest urge to relax or take a break. I only felt sad for the blast victims and I felt a sense of great responsibility on my shoulders. For the first time in my life, I felt loyal to the oath that I had taken during my medical school graduation; I had realized what being a doctor entailed. I felt that I had walked my talk in true spirit for the first time. After finishing my ED shift, I checked the news on my cellphone again. I saw a news item on Reuters: “Bomb at Shi’ite mosque kills 45 in Pakistan.” For the first time in my life, I could feel the pain and agony summarized in that one news item. For the first time, I realized that human lives are not statistics.

Abdul Rehman, MBBS

(jsmawais@yahoo.com)

1Department of Medicine, Rutgers–New Jersey Medical School, Newark, NJ

2Department of Biological & Biomedical Sciences, Aga Khan University Hospital, Karachi, Sindh, Pakistan

Supervising Editor: Brian Zink, MD
In SAEM Annual Meeting Abstracts, the abstracts below were attributed to Michael Ward, MD, at UW-Madison and they should have been attributed to Michael Ward MD, PhD, MBA, at the Vanderbilt University Medical Center.

#68, S40: “Burn-Related Injuries: A Nationwide Analysis of Interfacility Transfers Over a Six-Year Period”

#266, S116: “Examining Factors Associated With Potentially Avoidable Transfers to a Large Academic Emergency Department”

#379, S155: “ST-Elevation Myocardial Infarction Transfer Instability: Hospital-Level Predictors and Association With Mortality and Readmission”

#380, S155: “Dissolution of Medicare Interhospital ST-Elevation Myocardial Infarction Transfer Networks and Associated Characteristics, 2013-2016”

We apologize for the errors.

Reference