# Accuracy and Reliability of Emergency Department Triage Using the Emergency Severity Index: An International Multicenter Assessment

Binoy Mistry, MD; Sarah Stewart De Ramirez, MD; Gabor Kelen, MD; Paulo S. K. Schmitz, MD; Kamna S. Balhara, MD; Scott Levin, PhD, MS; Diego Martinez, PhD; Kevin Psoter, PhD; Xavier Anton, MD; Jeremiah S. Hinson, MD, PhD*

*Corresponding Author. E-mail: jhinson4@jhmi.edu, Twitter: @Hinson_EM.

**Study objective:** We assess accuracy and variability of triage score assignment by emergency department (ED) nurses using the Emergency Severity Index (ESI) in 3 countries. In accordance with previous reports and clinical observation, we hypothesize low accuracy and high variability across all sites.

**Methods:** This cross-sectional multicenter study enrolled 87 ESI-trained nurses from EDs in Brazil, the United Arab Emirates, and the United States. Standardized triage scenarios published by the Agency for Healthcare Research and Quality (AHRQ) were used. Accuracy was defined by concordance with the AHRQ key and calculated as percentages. Accuracy comparisons were made with one-way ANOVA and paired *t* test. Interrater reliability was measured with Krippendorff's $\alpha$. Subanalyses based on nursing experience and triage scenario type were also performed.

**Results:** Mean accuracy pooled across all sites and scenarios was 59.2% (95% confidence interval [CI] 56.4% to 62.0%) and interrater reliability was modest ($\alpha$=.730; 95% CI .692 to .767). There was no difference in overall accuracy between sites or according to nurse experience. Medium-acuity scenarios were scored with greater accuracy (76.4%; 95% CI 72.6% to 80.3%) than high- or low-acuity cases (44.1%, 95% CI 39.3% to 49.0% and 54%, 95% CI 49.9% to 58.2%), and adult scenarios were scored with greater accuracy than pediatric ones (66.2%, 95% CI 62.9% to 69.7% versus 46.9%, 95% CI 43.4% to 50.3%).

**Conclusion:** In this multinational study, concordance of nurse-assigned ESI score with reference standard was universally poor and variability was high. Although the ESI is the most popular ED triage tool in the United States and is increasingly used worldwide, our findings point to a need for more reliable ED triage tools. [Ann Emerg Med. 2018;71:581-587.]

Please see page 582 for the Editor's Capsule Summary of this article.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2017.09.036

## INTRODUCTION

Accurate and reliable triage is needed to safely prioritize care and achieve optimal resource allocation in busy emergency departments (EDs). The 5-level Emergency Severity Index (ESI) is among the most popular ED triage algorithms, currently in use by a majority of EDs in the United States and increasingly adopted internationally.[1,2] The ESI facilitates rapid decisionmaking through 3 primary questions: (1) Does this patient require an immediate lifesaving intervention? (2) Is this a high-risk situation? and (3) How many resources are required to care for this patient?[1]

The ESI is unique in its explicit consideration of resource use and use of a simple algorithm that is subjective and relies heavily on nursing intuition to facilitate rapid decisionmaking. In an era in which ED crowding is commonplace, with time and resources universally constrained, the utility of the ESI is obvious and its popularity is not surprising. However, it has not been well validated against time-sensitive or critical outcomes, and the majority of studies establishing ESI validity and reliability were performed soon after its implementation and in the United States, where the tool was developed.[1] Several lines of evidence, including a wide range of interrater reliability,[2] lack of true discrimination, with nearly half of patients nationwide triaged to the middle-tier ESI level 3,[2,3] and reports of suboptimal performance with low accuracy and high variability outside the United

**Editor's Capsule Summary**

*What is already known on this topic*

The Emergency Severity Index (ESI) is a simple, commonly used tool for emergency department (ED) triage categorization.

*What question this study addressed*

Does ESI use by nurses create accurate and consistent acuity designations in similar patients?

*What this study adds to our knowledge*

Using standard triage scenarios presented to 87 ESI-trained nurses in 3 countries, the mean accuracy was 59.2%, with only modest interrater reliability. Medium-acuity and adult patient scenarios had greater but still truncated accuracy.

*How this is relevant to clinical practice*

The ESI has limits when used to guide triage. Opportunity for a better initial ED triage tool exists.

States,[4,5] suggest that the ESI may perform less well in true clinical context than expected.[3,6]

In this internationally collaborative study, we sought to assess accuracy and variability of triage score assignment by ESI-trained nurses using standardized triage cases in 3 countries. We hypothesized suboptimal concordance with reference standard (ie, low accuracy) and high variability between individual raters across all sites. We also hypothesized higher accuracy among more experienced nurses and explored relationships between triage scenario type (ie, trauma versus nontrauma and adult versus pediatric) and accuracy and variability.

## MATERIALS AND METHODS

### Study Design, Setting, and Selection of Participants

This cross-sectional multicenter study was conducted between December 2015 and December 2016 in the EDs of 3 community hospitals located in 3 countries: the United Arab Emirates, Brazil, and the United States. All study sites had used the ESI for ED triage for at least 2 years before study initiation, and participants were recruited from the pool of active triage nurses at each site. Responses were collected anonymously from volunteer participants, and there was no reward or penalty associated with participation. All participating nurses had undergone formalized training in the ESI and been using the tool in clinical practice for at least 1 year before participation (Table 1). This study was approved by the institutional review board of all participating institutions.

Participating nurses were asked to assign ESI triage acuity (levels 1 to 5) for 25 written standardized triage cases taken verbatim from version 4 of the *ESI Implementation Handbook* (Competency Cases, Set A) (Table E1, available online at http://www.annemergmed.com). Scenarios and the corresponding answer key used were developed by the creators of the ESI and published by the Agency for Healthcare Research and Quality to facilitate ESI competency evaluation.[1] Scenarios included in this study had not been previously used for training or assessment at any study site. Written scenarios were administered in the primary language of clinical practice (English in the United Arab Emirates and United States; Portuguese in Brazil). English ESI scenarios were translated to Portuguese by a physician and nurse with fluency in both languages and were back translated to ensure accuracy of translation before deployment in Brazil. Cases from a broad range of ESI acuity levels were included (Figure). Of the 25 scenarios, 9 were pediatric and 5 were trauma focused. A paper copy of the ESI algorithm was available to all participants, and no time limit for triage assignment was imposed for completion of scoring. Participants were also asked to indicate years of experience as a nurse and years of experience using the ESI.

### Methods of Measurement

Accuracy was defined by concordance with the ESI handbook key for each triage scenario, and mean percentage of scenarios scored correctly was calculated. Under- and overtriage were defined by nurse assignment to a triage level of lower or higher acuity than key, respectively, and represented as percentages. Interrater reliability was defined as the degree of agreement among emergency triage nurses assigning a triage level to the same clinical scenario. To further assess clinical influence of triage errors, we performed supplemental analyses wherein scenarios were designated as high acuity (ESI level 1 or 2), medium acuity (ESI level 3), or low acuity (ESI level 4 or 5) and within-group accuracy (high, medium, and low) was calculated with the same definitions as above. Under the ESI, patients assigned to levels 1 and 2 are considered high risk and time sensitive, those assigned to level 3 may be significantly ill but wait hours to receive definitive care, and many EDs (including all our study sites) track ESI levels 4 and 5 patients to a separate clinical area with a fast-track designation. Thus, out-of-group triage errors would be expected to result in significant changes in care trajectory.

**Table 1.** Demographics of the participating EDs, hospitals, and nurses.

| Location | UAE | Brazil | United States |
|---|---|---|---|
| Number of hospital beds | 165 | 375 | 264 |
| Number of ED treatment spaces | 32 | 45 | 44 |
| Annual ED census, adult | 67,000 | 41,000 | 60,000 |
| Annual ED census, pediatric | 25,000 | 29,000 | 0* |
| Annual ED census, trauma, % | <1 | 0 | 4 |
| Adult ED admission rate, % | 10 | 16 | 26 |
| Introduction of ESI, year | 2014 | 2010 | 2002 |
| **ESI training and competency testing** | | | |
| Initial training | Online ESI training modules and practical in-person course | Intensive 2-day ESI course with classroom and on-site components | Intensive 4-h ESI training course and 12-h on-site training |
| Initial competency testing | Field competency testing | Field competency testing | Written competency examination |
| Ongoing training and assessment | Yes | Yes | Yes |
| Number of ESI-trained triage nurses | 40 | 38 | 35 |
| **Study participants (n)** | 35 | 30 | 22 |
| **ESI triage experience, y** | | | |
| ≤1 | 6 | 4 | 11 |
| 2–4 | 21 | 19 | 7 |
| ≥5 | 7 | 7 | 4 |
| **Nursing experience, y** | | | |
| ≤1 | 0 | 2 | 3 |
| 2–4 | 0 | 7 | 9 |
| ≥5 | 34 | 21 | 10 |

*UAE*, United Arab Emirates.
*Pediatric patients are treated in a separate pediatric ED at the US study site.

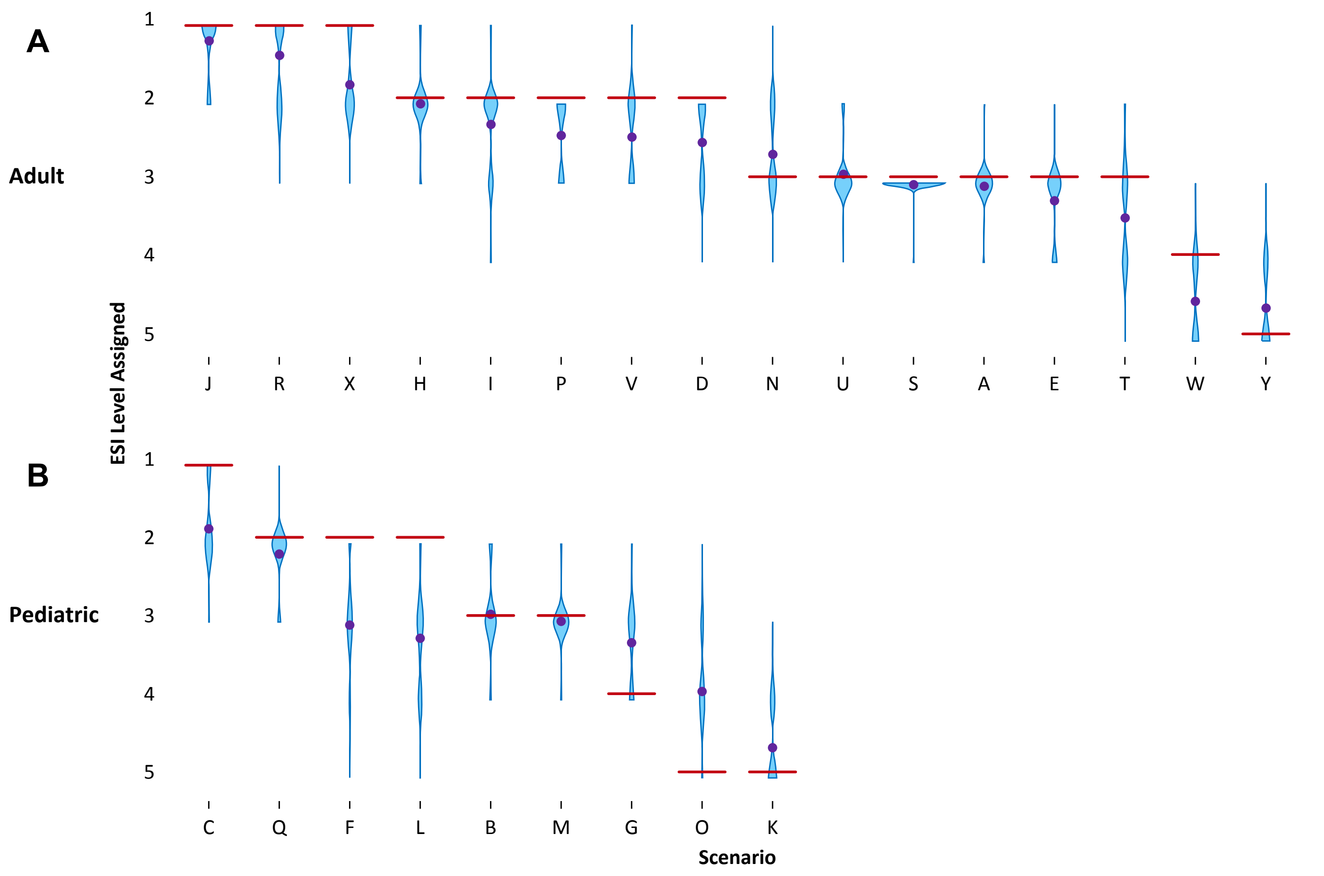


**Figure.** Violin plots showing range and distribution of nursing score assignments for each standardized triage scenario. Scenarios are grouped by patient population (*A*, adult; *B*, pediatric) and presented in descending order of *ESI Implementation Handbook* answer key assigned triage acuity. Letters correspond to scenario identification and values in parentheses to the triage level designated by the ESI handbook key. Black dots correspond to mean value of nursing responses for each scenario.

**Table 2.** Accuracy of ESI-level score assignment by acuity and scenario type across sites.*

| Site and Scenario by Hospital | N | Mean % Correct, (95% CI) | Undertriage, No. (%) | Overtriage, No. (%) |
|---|---|---|---|---|
| **Overall scoring** | 2,175 | 59.2 (56.4–62.0) | 600 (27.6) | 287 (13.2) |
| UAE | 875 | 58.7 (54.5–63.0) | 253 (28.9) | 108 (12.3) |
| Brazil | 750 | 58.3 (54.1–62.4) | 211 (28.1) | 102 (13.6) |
| United States | 550 | 61.3 (53.6–68.9) | 136 (24.7) | 77 (14.0) |
| **High triage acuity** | 1,044 | 54.0 (49.9–58.2) | 468 (44.8) | 12 (1.1) |
| UAE | 420 | 53.8 (46.6–61.0) | 192 (45.7) | 2 (0.5) |
| Brazil | 360 | 52.8 (47.0–58.6) | 165 (45.8) | 5 (1.4) |
| United States | 264 | 56.1 (46.2–65.9) | 111 (42.0) | 5 (1.9) |
| **Medium triage acuity** | 696 | 76.4 (72.6–80.3) | 86 (12.4) | 78 (11.2) |
| UAE | 280 | 72.1 (67.5–76.8) | 46 (16.4) | 32 (11.4) |
| Brazil | 240 | 77.1 (69.1–85.0) | 23 (9.6) | 32 (13.3) |
| United States | 176 | 82.4 (74.4–90.3) | 17 (9.7) | 14 (8.0) |
| **Low triage acuity** | 435 | 44.1 (39.3–49.0) | 46 (10.6) | 197 (45.3) |
| UAE | 175 | 49.1 (40.9–57.4) | 15 (8.6) | 74 (42.3) |
| Brazil | 150 | 41.3 (36.2–46.5) | 23 (15.3) | 65 (43.3) |
| United States | 110 | 40.0 (26.9–53.1) | 8 (7.3) | 58 (52.7) |
| **Adult scenarios** | 1,392 | 66.2 (62.9–69.7) | 368 (26.4) | 103 (7.4) |
| UAE | 560 | 64.3 (58.9–69.7) | 154 (27.5) | 46 (8.2) |
| Brazil | 480 | 67.5 (62.1–72.9) | 123 (25.6) | 33 (6.9) |
| United States | 352 | 67.3 (60.3–74.4) | 91 (25.9) | 24 (6.8) |
| **Pediatric scenarios** | 783 | 46.9 (43.4–50.3) | 232 (29.6) | 184 (23.5) |
| UAE | 315 | 48.9 (44.3–53.4) | 99 (31.4) | 62 (19.7) |
| Brazil | 270 | 41.9 (37.8–45.9) | 88 (32.6) | 69 (25.6) |
| United States | 198 | 50.5 (39.9–61.1) | 45 (22.7) | 53 (26.8) |
| **Trauma scenarios** | 435 | 60.9 (57.0–64.9) | 63 (14.5) | 107 (24.6) |
| UAE | 175 | 60.0 (54.2–65.8) | 25 (14.3) | 45 (25.7) |
| Brazil | 150 | 54.7 (48.2–61.2) | 19 (12.7) | 49 (32.7) |
| United States | 110 | 70.9 (62.4–79.4) | 19 (17.3) | 13 (11.8) |
| **Nontrauma scenarios** | 1,740 | 58.8 (55.6–62.0) | 537 (30.9) | 180 (10.3) |
| UAE | 700 | 58.4 (53.5–63.4) | 228 (32.6) | 63 (9.0) |
| Brazil | 600 | 59.2 (54.6–63.7) | 192 (32.0) | 53 (8.8) |
| United States | 440 | 58.9 (50.3–67.4) | 117 (26.6) | 64 (14.5) |

*N is the total number of scored scenarios. High triage acuity: ESI level 1 or 2; medium triage acuity: ESI level 3; low triage acuity: ESI level 4 or 5. Accuracy was defined by exact concordance of nurse-assigned ESI acuity with the *ESI Implementation Handbook* answer key. Undertriage was defined as assignment to a lower level of acuity than designated by key. Overtriage was defined as assignment to a higher level of acuity than designated by key.

### Primary Data Analysis

Mean percentage of accurate score assignment for all scenarios was calculated for participants in aggregate and for each site separately. Subanalyses were performed for triage scenarios divided according to acuity (high, medium, or low) and scenario type (adult or pediatric and trauma or nontrauma). One-way ANOVA was used to compare pooled accuracy between sites. If a significant difference was found, a pairwise $t$ test with unequal variance was performed to quantify differences between individual sites.

To allow measurement of interrater reliability among the large number of raters included, we used graphic plotting and Krippendorff's $\alpha$, a method that uses bootstrapping to generate reliability estimates ($\alpha$) with 95% confidence intervals (CIs). Data sets with $\alpha$ greater than or equal to .800 are considered reliable, whereas those with $\alpha$ less than .800 but greater than or equal to .667 are adequate for tentative conclusions only, and those with $\alpha$ less than .667 are unreliable.[7]

Relationships between years of nursing experience and years of experience using the ESI with accuracy of triage score assignment were assessed with linear regression analysis.

## RESULTS

### Characteristics of Study Subjects

Site characteristics are shown in Table 1. A total of 87 nurses from the 3 study sites participated, with highest rates of participation achieved in the United Arab Emirates (87.5%) and Brazil (78.9%) and lowest in the United States (62.9%). Overall, participating nurses from the United Arab Emirates reported the most years of experience in nursing and ESI triage, whereas nurses from the United States reported the least (Table 1).

Mean accuracy pooled across all sites and scenarios was 59.2% (95% CI 56.4% to 62.0%); no difference in overall accuracy between sites was observed (Table 2). A larger proportion of scenarios was undertriaged than overtriaged

(27.6% versus 13.2% overall). There was no relationship between accuracy of score assignment and participants' years of nursing experience or years of experience using the ESI ($R^2$=0.0082 and 0.0416, respectively). Subanalysis based on acuity of standardized cases revealed lower accuracy for low-acuity cases (ESI levels 4 to 5) (44.1%; 95% CI 39.3% to 49.0%) and high-acuity cases (ESI levels 1 to 2) (54.0%; 95% CI 49.9% to 58.2%) compared with medium-acuity cases (ESI level 3) (76.4%; 95% CI 72.6% to 80.3%). There were no differences in accuracy between sites within any acuity-based grouping (Table 2). There were also no differences in accuracy between sites for adult or pediatric subpopulations (Table 2), but accuracy of score assignment was higher for adult scenarios than pediatric ones (66.2%, 95% CI 62.9% to 69.7% versus 46.9%, 95% CI 43.4% to 50.3%). Trauma and nontrauma scenarios were scored with similar accuracy at all sites except the US site, where trauma scenarios were scored more accurately (70.9%, 95% CI 62.4% to 79.4% versus 58.9%, 95% CI 50.3% to 67.4%). Accuracy of trauma scenario scoring was also higher at this site than in Brazil (54.7%; 95% CI 48.2% to 61.2%), with an intermediate level of accuracy observed in the United Arab Emirates (60.0%; 95% CI 54.2% to 65.8%) (Table 2).

Mean within-group accuracy (high, medium, or low acuity) pooled across sites was only 74.2% (95% CI 72.1% to 76.2%). As reported for unadjusted accuracy above, undertriage (18.2%) was more common than overtriage (7.7%) using within-group criteria, and a higher level of accuracy was achieved for adult scenarios (80.0%; 95% CI 77.5% to 82.4%) than pediatric ones (63.9%; 95% CI 61.0% to 66.7%). Overall, higher within-group accuracy was also achieved for nontrauma scenarios than trauma scenarios (76.4%, 95% CI 74.0% to 78.7% versus 65.3%, 95% CI 61.4% to 69.2%) and for low-acuity scenarios than high-acuity ones (79.5%, 95% CI 76.2% to 82.9% versus 70.4%, 95% CI 67.3% to 73.5%) (Table E2, available online at http://www.annemergmed.com).

Variability of rater response is represented graphically in the Figure. Interrater reliability across sites was modest ($\alpha$=.730; 95% CI .692 to .767). Similar reliability was observed within sites (United Arab Emirates $\alpha$=.777, 95% CI .746 to .807; Brazil $\alpha$=.763, 95% CI .727 to .798; US $\alpha$=.735, 95% CI .691 to .776).

## LIMITATIONS

This study has several important limitations. A thorough assessment of ESI validity should be performed in true clinical context and include analysis of the relationship between triage classification and actual ED resource utilization and clinical outcomes (eg, predictive accuracy). However, such an approach is resource intensive and beyond the scope of the current study, which aimed to measure consistency in application of the ESI algorithm by ED nurses trained and practicing in various clinical contexts. In addition, we relied on a convenience sample of nurses from a relatively small number of sites, potentially limiting the external and internal validity of our findings. This latter weakness was minimized by inclusion of a majority of triage nurses from each site (Table 1). Finally, although our findings were consistent across institutions on 3 continents, it is possible that there are clinical sites where accuracy and variability of score assignment are significantly different from that presented here. For example, our scenarios included a relatively high proportion of pediatric cases (9 of 25). Although this proportion is similar to that of pediatric patients treated at our study sites in the United Arab Emirates and Brazil, no pediatric patients are treated at our US site, and it is conceivable that nurses who work in a US ED where pediatric patients are treated exhibit higher accuracy and lower variability when scoring these cases.

## DISCUSSION

In what is to our knowledge the first study designed to assess and compare accuracy and reliability of the ESI score assignment across multiple countries, we found that concordance of nurse-assigned ESI score with reference standard was universally poor and variability of score assignment was high within and across sites. Furthermore, supplemental within-groups accuracy analysis suggested consequential changes in care trajectory for 1 in 4 patients, with nearly 30% of high-acuity patients (ESI level 1 or 2) undertriaged to ESI level 3 or below, events that may result in substantial delays to time-sensitive interventions. The universality of our findings argues strongly against nurse- or institution-specific deficiencies in training or implementation. Our study of the ESI differs from those performed within the United States previously in that we included a large number of nurses from each site, all of whom had been trained in the ESI at least 1 year before and routinely used it in clinical practice. This approach increases the external validity of our findings and suggests high variability of ESI score assignment in clinical practice. Our finding that years of experience as a nurse and using the ESI in particular had no effect on accuracy of score assignment was somewhat surprising, but again argues against nurse-specific deficiencies and points toward systemic weaknesses of the triage tool itself.

We also found that pediatric scenarios were scored less accurately than adult scenarios, as has been previously reported,[8] and that trauma scenarios were scored more accurately at a site where a higher number of trauma patients are treated. Although the latter is not surprising, the finding that less than half of all pediatric scenarios were scored correctly by the ESI should be alarming. We did not find any association between nursing experience and accuracy; performance on pediatric cases was comparable at sites where pediatric patients are treated (United Arab Emirates and Brazil) to that at a site where they are not (United States). This suggests that lower accuracy for pediatric cases is not explained by experience alone. A more likely explanation is that triage of pediatric patients under the ESI is more challenging because the ESI algorithm does not account for the increased complexity of decisionmaking inherent in the pediatric population (eg, variable interpretation of vital sign derangements depending on age, limited reliability of clinical history).

The ESI was developed nearly 20 years ago to address a demand for rapid needs assessment and resource distribution in the face of significantly increasing ED volumes in the United States.[1] Since then, ED crowding has become commonplace across the world and is recognized as a significant threat to patient safety, making reliable identification of critically ill patients at ED triage and effective distribution of limited ED resources more important than ever.[9] During the past decade, breadth and severity of pathology encountered in the ED have also increased, as has resource use,[6] potentially undermining a major component of the ESI decision algorithm. Although direct comparison of ESI performance with that of other widely adopted triage systems (eg, Australasian Triage Scale, Canadian Triage and Acuity Scale, Manchester Triage Scale) may be useful according to our findings, these scales rely heavily on triage provider judgment, with little systematic evidence of improved reliability compared with the ESI to date.[2] Recent advancements in electronic health record systems, decision support, and data science offer new opportunities to strengthen evidence and control untoward variation in triage decisionmaking. Several reports describe the development and use of information technology–based solutions to improve current triage practice (eg, ESI)[10] or provide new data-driven approaches adaptive to local ED populations and systems.[3,11] In light of data presented here, consideration of these approaches is well warranted.

*Supervising editor:* Donald M. Yealy, MD

*Author affiliations:* From the Departments of Emergency Medicine (Mistry, Stewart De Ramirez, Kelen, Levin, Martinez, Hinson) and Pediatrics (Psoter), Johns Hopkins University School of Medicine, Baltimore, MD; the Emergency Department, Hospital Moinhos de Vento, Porto Alegre, Brazil (Schmitz); the Department of Emergency Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX (Balhara); and the Emergency Department, Al-Rahba Hospital, Abu Dhabi, United Arab Emirates (Anton).

*Author contributions:* BM, SSDR, KSB, and JSH conceived and designed the study. BM, SSDR, PSKS, KSB, and XA recruited participants. GK, DM, and KP provided statistical advice on the study design. DM and KP helped analyze the data. SL provided content expertise. BM and JSH drafted the article, and all authors contributed to its revision. JSH takes responsibility for the paper as a whole.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

*Funding and support:* By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). This work was funded by a Synergy Award from Johns Hopkins University and grant R21HS023641 from the Agency for Healthcare Research and Quality. Dr. Levin is an engineer and the owner of a start-up company that has developed several machine-learning and electronic health record–based tools to improve patient safety and throughput throughout the hospital, including a tool that aims to improve reliability of ED triage.

*Publication dates:* Received for publication June 29, 2017. Revision received August 29, 2017. Accepted for publication September 27, 2017. Available online November 24, 2017.



## REFERENCES

1. Gilboy N, Tanabe T, Travers D, et al. *Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care, Version 4. Implementation Handbook 2012 Edition*. Rockville, MD: Agency for Healthcare Research and Quality; 2011; AHRQ Publication 12-0014.
2. Christ M, Grossmann F, Winter D, et al. Modern triage in the emergency department. *Dtsch Arztebl Int*. 2010;107:892-898.
3. Dugas AF, Kirsch TD, Toerper M, et al. An electronic emergency triage system to improve patient distribution by critical outcomes. *J Emerg Med*. 2016;50:910-918.
4. Jordi K, Grossmann F, Gaddis GM, et al. Nurses' accuracy and self-perceived ability using the Emergency Severity Index triage tool: a cross-sectional study in four Swiss hospitals. *Scand J Trauma Resusc Emerg Med*. 2015;23; 62-62.
5. Bergs J, Verelst S, Gillet J-B, et al. Evaluating implementation of the Emergency Severity Index in a Belgian hospital. *J Emerg Nurs*. 2014;40:592-597.

6. Pitts SR, Pines JM, Handrigan MT, et al. National trends in emergency department occupancy, 2001 to 2008: effect of inpatient admissions versus emergency department practice intensity. *Ann Emerg Med*. 2012;60:679-686.e673.
7. Krippendorff K. Reliability in content analysis: some common misconceptions and recommendations. *Hum Commun Res*. 2004;30:411-433.
8. Travers DA, Waller AE, Katznelson J, et al. Reliability and validity of the Emergency Severity Index for pediatric triage. *Acad Emerg Med*. 2009;16:843-849.
9. Committee on the Future of Emergency Care in the United States Health System. *Hospital-Based Emergency Care: At the Breaking Point*. Washington, DC: The National Academies Press; 2006.
10. Aronsky D, Jones I, Raines B, et al. An integrated computerized triage system in the emergency department. *AMIA Annu Symp Proc*. 2008;8:16-20.
11. Levin S, Hamrock E , Hinson J , et al. Machine learning–based electronic triage more accurately differentiates patients with respect to clinical outcomes compared to the Emergency Severity Index. *Ann Emerg Med*. 2017; https://doi.org/10.1016/j.annemergmed.2017.08.005.

**Table E1.** Full text of triage scenarios with identification, population, scenario, and *ESI Implementation Handbook* answer key.

| Identification | Population | Scenario | Key | Scenario Text |
|---|---|---|---|---|
| A | Adult | Medical | 3 | "I think I picked up a bug overseas," reports a 34-y-old man who presents to the ED, complaining of frequent watery stools and abdominal cramping. "I think I am getting dehydrated." Vital signs: T 98°F, RR 22, HR 112, BP 120/80, $SpO_2$ 100. |
| B | Pediatric | Trauma | 3 | "I think he broke it," reports the mother of a 9-y-old boy. "He was climbing a tree and fell about 5 feet, landing on his arm. I am a nurse, so I put on a splint and applied ice. He has a good pulse." The arm is obviously deformed. Vital signs: T 98°F, RR 26, HR 90, $SpO_2$ 99%, pain 5/10. |
| C | Pediatric | Medical | 1 | "I don't know what's wrong with my baby girl," says a young mother. She reports that her 2-wk-old baby is not acting right and is not interested in eating. As you begin to undress the baby, you notice that she is listless and her skin is mottled. |
| D | Adult | Medical | 2 | "My pain medications are not working anymore. Last night I couldn't sleep because the pain was so bad," reports a 47-y-old woman with metastatic ovarian cancer. "My husband called my oncologist, and he told me to come to the emergency department." The patient rates her pain as 9/10. Vital signs are within normal limits. |
| E | Adult | Medical | 3 | A 48-y-old man tells you that he has a history of kidney stones and thinks he has another one. He has right costovertebral angle pain that radiates around to the front and into his groin. He is nauseated but tells you he took a pain pill, and right now he has minimal pain. He denies vomiting. Vital signs: T 98°F, RR 16, HR 80, BP 136/74, $SpO_2$ 100%, pain 3/10. |
| F | Pediatric | Medical | 2 | "After my pediatrician saw my son's rash, he said I had to bring him to the emergency department immediately. He has this rash on his face and chest that started today. He has little pinpoint purplish spots called petechiae. My son is a healthy kid who has had a cold for a couple of days and a cough. My pediatrician said he had to be sure nothing bad is going on. What do you think?" |
| G | Pediatric | Trauma | 4 or 5 | "Her grandfather pulled her by the wrist up and over a big puddle. Next thing you know, she is crying and refusing to move her left arm," the mother of a healthy 3-y-old tells you. Vital signs are within normal limits. |
| H | Adult | Medical | 2 | A 46-y-old asthmatic patient in significant respiratory distress presents by ambulance. The paramedics report that the patient began wheezing earlier in the day and has been using her inhaler, with no relief. On her last admission for asthma, she was intubated. Vital signs: RR 44, $SpO_2$ 93% on room air, HR 98, BP 154/60. The patient is able to answer your questions about allergies and medications. |
| I | Adult | Medical | 2 | A 56-y-old man with a recent diagnosis of late-stage non-Hodgkin's lymphoma was brought to the ED from the oncology clinic. He told his oncologist that he had facial and bilateral arm swelling and increasing shortness of breath. The patient also reports that his symptoms are worse if he lies down. Vital signs: BP 146/92, HR 122, RR 38, $SpO_2$ 98% on room air, temperature normal. |
| J | Adult | Trauma | 1 | EMS arrives with a 28-y-old man who was stabbed in the left side of his neck during an altercation. You notice a large hematoma around the wound, and the patient is saying he cannot breathe. Vital signs: HR 110, RR 36, $SpO_2$ 89%. |
| K | Pediatric | Medical | 5 | An 11-y-old presents to triage with his mother, who reports that her son has had a cough and runny nose for a week. The child is running around the waiting room and asking his mother for a snack. Vital signs are within normal limits. |
| L | Pediatric | Medical | 2 | "I don't know what is wrong with my son," reports the mother of a normally healthy 8-y-old boy. "He's losing weight and acting so cranky. Last night he was up to the bathroom every hour, and he can't seem to get enough to drink." The child is alert and oriented and answers your questions appropriately. Vital signs: T 98.6°F, RR 30, HR 98, BP 92/78, $SpO_2$ 98%. |
| M | Pediatric | Medical | 3 | "He has had diarrhea for 2 days, and he just started throwing up this morning. This has been going around the family, and he seems to have it the worst. He had been drinking before today, but now he doesn't want anything to drink," reports the mother of a 19-mo-old. The toddler is awake and alert but quiet in the mother's arms, and you notice his lips are dry and cracked. Vital signs: T 99°F, RR 30, HR 130, $SpO_2$ 100%. |
| N | Adult | Trauma | 3 | EMS arrives with an 87-y-old man who slipped on the ice and injured his right hip. His right leg is shortened and externally rotated. The patient's only complaint is hip pain. He rates his pain as 5/10, and his vital signs are within normal limits. |
| O | Pediatric | Medical | 5 | "My baby is a having a hard time drinking his bottle," reports the young mother of a 3-mo-old. The baby is alert and looking around. You notice a large amount of dried mucus around both nares. Vital signs: T 98°F, RR 40, HR 132, $SpO_2$ 99%. |
| P | Adult | Medical | 2 | A 72-year-old female is brought in by ambulance from the nearby nursing home. They report that she has become increasingly confused over the last 24 hours. She is usually awake, alert, and oriented and takes care of her own activities of daily living. At triage she has vital signs of 99.6°F, HR 86, RR 28, BP 136/72, and $SpO_2$ 94% on room air. |
| Q | Pediatric | Medical | 2 | Melissa, a 4-y-old with a ventriculoperitoneal shunt (drains excess cerebrospinal fluid), is brought to the ED by her parents. The mother tells you that she is concerned that the shunt may be blocked because Melissa is not acting right. The child is sleepy but responds to verbal stimuli. When asked what was wrong, she tells you that her head hurts and she is going to throw up. Vital signs: T 98.6°F, RR 22, HR 120, $SpO_2$ 99% on room air, BP 94/76. |

**Table E1.** Continued.

| Identification | Population | Scenario | Key | Scenario Text |
|---|---|---|---|---|
| R | Adult | Medical | 1 | The overhead page announces the arrival of the code STEMI. Paramedics arrive with a 62-y-old man with a history of a myocardial infarction 4 y ago who is complaining of chest pressure that started an hour ago. The field ECG shows anterior lateral ischemic changes. Currently, the patient's pulse rate is 106 beats/min, RR 28, BP 72/53; $SpO_2$ is 95% on a nonrebreather mask. His skin is cool and clammy. |
| S | Adult | Medical | 3 | "I had a knee replacement 3 months ago. Now look at it!" states a 64-y-old man. The knee is red, swollen, and tender to touch. Vital signs: T 99°F, RR 20, HR 74, BP 164/74, $SpO_2$ 97%, pain 6/10. |
| T | Adult | Medical | 3 | "This is so embarrassing," reports a 29-y-old man. "For the last 12 hours, I have had this thing stuck in my rectum. I have tried and tried to get it out with no success. Can someone help me?" The patient denies abdominal pain or tenderness. Vital signs are within normal limits. Pain 4/10. |
| U | Adult | Medical | 3 | EMS arrives with a 67-y-old woman who lives alone. The patient called 911 because she was too sick to take herself to the physician. The patient has had a fever and cough for 3 days. She reports coughing up thick green phlegm and is concerned that she has pneumonia. She denies shortness of breath. Medical history: hypertension. Vital signs: T 102°F, RR 28, HR 86, BP 140/72, $SpO_2$ 94%. |
| V | Adult | Trauma | 2 | EMS arrives with a 14-y-old boy who was snowboarding at a nearby ski area, lost control, and ran into a tree. The patient was wearing a ski helmet; is currently aware, alert, and oriented; and is complaining of left upper quadrant pain and left thigh pain. His left femur appears to be broken. Vital signs: BP 112/80, HR 86, RR 14, $SpO_2$ 98%, temperature normal. |
| W | Adult | Medical | 4 | "I woke up this morning, and there was a bat flying around our bedroom. Scared me half to death, and now I am so worried about rabies," an anxious 48-y-old woman tells you. "My husband opened the window and the bat flew out." Medical history of ovarian cysts, no medications or allergies. Vital signs are within normal limits. |
| X | Adult | Medical | 1 | The family of a 74-y-old man called 911 when he developed severe midabdominal pain. "My husband is not a complainer," reports his wife. "The only medication he takes is for high blood pressure." On arrival in the ED, the patient's HR is 140, RR 28, $SpO_2$ 94%, BP 72/56. |
| Y | Adult | Medical | 5 | "I woke up this morning, and my eyes are all red and crusty," reports a 29-y-old kindergarten teacher. "I think I got it from the kids at school," she tells you. She denies pain or other visual disturbances. Her vital signs are within normal limits. |

*T*, Temperature; *RR*, respiratory rate; *BP*, blood pressure; *HR*, heart rate; *$SpO_2$*, oxygen saturation; *mo-old*, months old; *EMS*, emergency medical services; *y-old*, years old; *ECG*, electrocardiogram.
All scenarios taken from the *Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care, Version 4. Implementation Handbook 2012 Edition*, Competency Cases, Set A.

**Table E2.** Within-group accuracy of ESI-level score assignment by acuity and scenario type across sites.*

| Site and Scenario by Hospital | N | Mean % Correct, (95% CI) | Undertriage, No. (%) | Overtriage, No. (%) |
|---|---|---|---|---|
| **Overall scoring** | 2,175 | 74.2 (72.1–76.2) | 395 (18.2) | 167 (7.7) |
| UAE | 875 | 71.5 (68.7–74.3) | 185 (21.1) | 64 (7.3) |
| Brazil | 750 | 75.3 (71.9–78.7) | 120 (16.0) | 65 (8.7) |
| United States | 550 | 76.7 (71.6–81.9) | 90 (16.4) | 38 (6.9) |
| **High triage acuity** | 1,044 | 70.4 (67.3–73.5) | 309 (29.6) | NA |
| UAE | 420 | 66.9 (61.2–72.6) | 139 (33.1) | NA |
| Brazil | 360 | 73.1 (68.6–77.5) | 97 (26.9) | NA |
| United States | 264 | 72.3 (66.1–78.6) | 73 (27.7) | NA |
| **Medium triage acuity** | 696 | 76.4 (72.6–80.3) | 86 (12.4) | 78 (11.2) |
| UAE | 280 | 72.1 (67.5–76.8) | 46 (16.4) | 32 (11.4) |
| Brazil | 240 | 77.1 (69.1–85.0) | 23 (9.6) | 32 (13.3) |
| United States | 176 | 82.4 (74.4–90.3) | 17 (9.7) | 14 (8.0) |
| **Low triage acuity** | 435 | 79.5 (76.2–82.9) | NA | 89 (20.5) |
| UAE | 175 | 81.7 (76.3–87.1) | NA | 32 (18.3) |
| Brazil | 150 | 78.0 (74.4–81.6) | NA | 33 (22.0) |
| United States | 110 | 78.2 (68.3–88.0) | NA | 24 (21.8) |
| **Adult scenarios** | 1,392 | 80.0 (77.5–82.4) | 217 (15.6) | 62 (4.5) |
| UAE | 560 | 75.7 (71.6–79.8) | 104 (18.6) | 32 (5.7) |
| Brazil | 480 | 83.3 (79.3–87.4) | 57 (11.9) | 23 (4.8) |
| United States | 352 | 82.1 (78.0–86.2) | 56 (15.9) | 7 (2.0) |
| **Pediatric scenarios** | 783 | 63.9 (61.0–66.7) | 178 (22.7) | 105 (13.4) |
| UAE | 315 | 64.1 (60.8–67.5) | 81 (25.7) | 32 (10.2) |
| Brazil | 270 | 61.1 (57.2–65.0) | 63 (23.3) | 42 (15.6) |
| United States | 198 | 67.2 (58.4–76.0) | 34 (17.2) | 31 (15.7) |
| **Trauma scenarios** | 435 | 65.3 (61.4–69.2) | 46 (10.6) | 105 (24.1) |
| UAE | 175 | 62.9 (56.6–69.1) | 20 (11.4) | 45 (25.7) |
| Brazil | 150 | 59.3 (53.3–65.4) | 13 (8.7) | 48 (32.0) |
| United States | 110 | 77.3 (69.9–84.7) | 13 (11.8) | 12 (10.9) |
| **Nontrauma scenarios** | 1,740 | 76.4 (74.0–78.7) | 349 (20.1) | 62 (3.6) |
| UAE | 700 | 73.7 (70.2–77.2) | 165 (23.6) | 19 (2.7) |
| Brazil | 600 | 79.3 (75.6–83.1) | 107 (17.8) | 17 (2.8) |
| United States | 440 | 76.6 (70.9–82.2) | 77 (17.5) | 26 (5.9) |

*High triage acuity: ESI level 1 or 2; medium triage acuity: ESI level 3; low triage acuity: ESI level 4 or 5. Accuracy was defined by nurse assignment of an ESI triage score within the same 3-tier acuity-based grouping as the level recommended by the *ESI Implementation Handbook* answer key (ie, answers of 1 or 2 and 4 or 5 were considered equivalent). Undertriage was defined as assignment to a lower level of acuity than designated by key. Overtriage was defined as assignment to a higher level of acuity than designated by key.