

# Design, Validity, and Reliability of a Pediatric Resident JumpSTART Disaster Triage Scoring Instrument

Mark X. Cicero, MD; Antonio Riera, MD; Veronika Northrup, MPH; Marc Auerbach, MD, MSci; Kevin Pearson; Carl R. Baum, MD

From the Department of Pediatrics, Yale University School of Medicine (Drs Cicero, Riera, Auerbach, and Baum; and Mr Pearson), and Yale Center for Clinical Investigation (Ms Northrup), New Haven, Conn  
Address correspondence to Mark X. Cicero, MD, Pediatric Emergency Medicine, 100 York Street Suite 1F, New Haven, CT 06511 (e-mail: mark.cicero@yale.edu).

Received for publication December 17, 2011; accepted September 30, 2012.

## ABSTRACT

**OBJECTIVE:** To design an instrument for scoring residents learning pediatric disaster triage (PDT), and to test the validity and reliability of the instrument.

**METHODS:** We designed a checklist-based scoring instrument including PDT knowledge and skills and performance, as well as a global assessment. Learners' performance in a 10-patient school bus crash simulation was video recorded and scored with the instrument. Learners triaged the patients with a color-coded algorithm (JumpSTART, Simple Triage and Rapid Treatment). Three evaluators observed the recordings and scored triage performance for each learner. Internal and construct validity of the instrument were established via comparison of resident performance by postgraduate year (PGY) and correlating instrument items with overall score. Validity was assessed with analysis of variance and the *D* statistic. We calculated evaluators' intraclass correlation coefficient (ICC) for each patient, skill, triage decision, and global assessment.

**RESULTS:** There were 37 learners and 111 observations. There was no difference in total scores by PGY ( $P = .77$ ), establishing

internal validity. Regarding construct validity, most instrument items had a *D* statistic of  $>0.5$ . The overall ICC among scores was 0.83 (95% confidence interval [CI] 0.74–0.89). Individual patient score reliability was high and was greatest among patients with head injury (ICC 0.86; 95% CI 0.79–0.91). Reliability was low for an ambulatory patient (ICC 0.29; 95% CI 0.07–0.48). Triage skills evaluation showed excellent reliability, including airway management (ICC 0.91; 95% CI 0.86–0.94) and triage speed (ICC 0.81; 95% CI 0.72–0.88). The global assessment had moderate reliability for skills (ICC 0.63; 95% CI 0.47–0.75) and knowledge (ICC 0.64; 95% CI 0.49–0.76).

**CONCLUSIONS:** We report the validity and reliability testing of a PDT-scoring instrument. Validity was confirmed with no performance differential by PGY. Reliability of the scoring instrument for most patient-level triage, knowledge, and specific skills was high.

**KEYWORDS:** disaster; evaluation; medical education; pediatrics; reliability; triage; validity

**ACADEMIC PEDIATRICS** 2013;13:48–54

## WHAT'S NEW

There is a need for both well-designed curricula in pediatric disaster medicine and for validated, reliable scoring instruments to assess the performance of those learning pediatric disaster triage. Testing of the scoring instrument shows it has validity and reliably measures pediatrics residents' disaster triage performance in a simulated environment.

DISASTERS ARE HIGH-STAKES, low-frequency incidents that overwhelm available health care resources. Infants and children are particularly vulnerable in disasters.<sup>1,2</sup> When disasters strike, triage of child disaster patients is a crucial skill for pediatrics residents and other health care workers.<sup>3</sup>

Several disaster triage strategies are in use internationally. Recognized pediatric disaster triage (PDT) strategies include JumpSTART (Simple Triage and Rapid Treatment),<sup>4</sup> the most commonly used strategy, Smart,<sup>5</sup> SALT (Sort–Assess–Life-Saving Treatment),<sup>6,7</sup> and Sacco

Triage.<sup>8–10</sup> Other strategies allow first responders and prehospital care professionals to triage using their clinical judgment.<sup>11,12</sup>

Triage decisions are rapid and have important consequences. Triage decisions may affect the timeliness of care and the likelihood the patient will survive. Despite these consequences, it is unclear whether PDT should be performed by senior and experienced personnel, or by more junior disaster responders.<sup>13</sup> Further, there is little evidence that educational interventions, including didactic courses, tabletop exercises, and live-action simulations, improve PDT performance.<sup>7,14,15</sup>

There remains a need for reliable and valid pediatric triage evaluation instruments that measure PDT performance.<sup>14,16–18</sup> Characteristics of an ideal PDT-scoring instrument include applicability to many learners, reproducible results, and clear language.<sup>17–19</sup> Further, the instrument should be adaptable to multiple PDT strategies.

We aimed to design, validate, and test the inter-rater reliability of a scoring instrument for use in evaluating pediatric residents' PDT performance. The instrument was

created for use in a standardized, multiple-scenario, simulation-based pediatric disaster medicine curriculum. We hypothesized the instrument would measure PDT performance in a valid, reproducible manner.

## METHODS

### SUBJECTS

The learners were first- to fourth-year pediatrics and internal medicine–pediatrics residents at an urban, tertiary care children’s hospital. The learners completed the simulation for this study on February 5, 2010. Six months before, the learners had previous training in JumpSTART<sup>4</sup> PDT, including airway, breathing, and circulation assessment skills and knowledge of the JumpSTART algorithm. Learners completed a preparticipation survey that assessed their prior disaster training and experience. The institutional review board approved subject participation in this educational intervention.

### SIMULATION DESCRIPTION

The learners completed a high-fidelity, 10-patient school bus crash simulation. This simulation was the final of 3 simulations in a PDT educational curriculum. The simulations occurred at a large health care simulation laboratory affiliated with our health care system. The first 2 simulations occurred 5 months before the school bus crash simulation. A learning session with scripted debriefings and a standardized didactic intervention occurred after each simulation. The 5-month delay was designed to test retention of PDT skills and knowledge.

In most triage strategies, including JumpSTART, color codes correspond to patients’ triage priority. Patients with critical illness or injury are assigned the Red category; non-ambulatory, Yellow; walking, Green; dead or likely to die, Black. Common features of the PDT strategies include scene assessment, triage, and decisions about transport to receiving facilities.

The 10 simulated patients suffered various injuries. The types of injuries and the domains for learner assessment are shown in [Table 1](#). Expected color-coded triage levels were determined by a consensus of subject matter experts who applied the JumpSTART triage tool. Nine of the patients represented a decision endpoint of the JumpSTART triage algorithm. An additional patient, seated in a wheelchair, represented a nonverbal child with special health care needs (CSHCN).

In the simulation, the learners served as the sole health care worker performing PDT. The school bus patients had been transported to the emergency department for this scenario. The learners performed the simulation individually and triaged the patients independently. Nine simulation manikins and one standardized patient served as the patients. A subset of manikins responded physiologically to airway maneuvers. Manikins used were 3 SimMan, 2 VitalSim Kid, 2 SimBaby, and one SimMan 3G (all products of Laerdal Medical, Stockholm, Sweden), and one Hal child (Gaumard Scientific, Miami, Fla).

A trained faculty facilitator provided information about patients’ mode of transport, the number of patients, and available health care resources. The learners received standardized prompts from the facilitators regarding manikin-specific physiology, such as the location of manikin pulses.

### DESIGN OF THE SCORING INSTRUMENT

We designed a checklist-based evaluation tool that included learners’ PDT knowledge, skills, and performance, which were defined as follows.

#### *PDT KNOWLEDGE*

Included patient-level application of the JumpSTART algorithm and assignment of the accurate triage level (Red, Yellow, Green, or Black).

#### *PDT SKILLS*

Included patient-level assessment of ambulatory status, airway, breathing, and circulation, repositioning of the airway, bag-valve-mask ventilation, and determining mental status.

#### *PDT PERFORMANCE*

Included efficient triage of patients, taking less than a minute per patient assessment, and performance on the global assessment scale, described below.

The instrument included a separate global assessment scale. The global assessment was included as a measure of aggregate performance, to complement the patient-level assessments. The global assessment included 5-point, Likert-style questions with text anchors at 1 and 5. The 5 points on the Likert scale were labeled Novice, Advanced Beginner, Competent, Proficient, and Expert. The global assessment items included professionalism, triage skill, triage knowledge, and overall performance.

We developed the scoring instrument via an iterative process, using a modified Delphi technique. There were 6 Delphi participants, who were local subject matter experts. There were 3 iterations of the Delphi process. Triage Performance was rated via trichotomous scoring (Yes, No, or Unable to Comment). In total, there were 59 points on the scoring instrument. An excerpt from the scoring instrument is shown in [Figure 1](#).

Use of the scoring instrument required instruction for evaluators and learners. The evaluators were oriented to the trichotomous scoring in the scoring instrument. To be scored ‘Yes’ for assessing breathing, for example, evaluators were instructed that learners must verbalize their assessment of the patient’s breathing. Likewise, learners were instructed to verbalize their thought processes and assessments.

### DATA COLLECTION AND PERFORMANCE EVALUATION

The research team video recorded the learners’ triage performances. We standardized the videos by predetermining videography angles and the locations of patients in the simulated emergency department. The video angles were determined during technical rehearsals of the simulations.

**Table 1.** Simulated Bus Crash Victims

Patient No.	Patient Simulation (Expected Triage Level)	Domain for Learner Assessment				Altered Mental Status
		Ambulatory Status	Airway	Breathing	Circulation	
1	Laerdal MegaCode Kid, bar impaled through the chest, no vital signs (Black)	x	x	x	x	
2	Laerdal MegaCode Kid, intubated, no vital signs (Black)	x	x	x	x	
3	Gaumard HAL, CSHCN in wheelchair, nonverbal, normal vital signs (Green)	x	x	x		
4	Laerdal SimMan 3G, apneic, responds to airway maneuvers (Red)	x	x	x		
5	Laerdal SimMan, who is perseverating, with a head injury (Yellow)	x	x	x	x	x
6	Laerdal SimMan, with a bleeding femoral wound, tachypneic, no palpable pulse (Red)	x	x	x	x	
7	Laerdal SimMan, with a head injury and glass imbedded in the scalp (Red)	x	x	x	x	x
8	An actor, ambulatory, with a forearm laceration (Green)	x	x			
9	Laerdal SimBaby, apneic, responds to positive pressure ventilation (Red)	x	x	x		
10	Laerdal SimBaby, with chest bruises, who is tachypneic (Red)	x	x	x		

Black = dead or likely to die; CSHCN = child with special health care needs; green = walking; red = critical; yellow = nonambulatory.

The angles were chosen to optimize visualization of the learners' hands and the patients' airways. The PDT learners consented to videography at the outset of the PDT curriculum. An online, password-protected repository was used for video storage (FlipShare, Cisco Corporation, San Jose, Calif).

We identified 3 evaluators with diverse clinical experience and levels of education. The evaluators were the principal investigator (evaluator A), a collaborating author (evaluator B), and an undergraduate research assistant (evaluator C). There was a deliberate choice to include 3 evaluators of disparate PDT expertise. Including an undergraduate evaluator allowed assessment of the scoring instrument by an evaluator free of clinical experience and biases.

The evaluators independently viewed the recordings and scored PDT performance using the scoring instrument. Each of the 3 evaluators assessed all of the triage videos.

### STATISTICAL ANALYSIS

Scores on all instrument items were summarized using means and standard deviations for all participants, as well as by postgraduate year (PGY). Discrimination (*D*) statistics were obtained with the Pearson correlation coefficient. The *D* statistic was used to determine if participants' performances with individual patients, specific PDT skills, triage decisions, and the global assessment were predictive of overall scores. A *D* statistic >0.5 is considered predictive of the overall score. The *D* statistic was used as a measure of construct validity.

To determine the internal validity of the scoring instrument, we wished to establish that performance during a simulation was not affected by the years of medical training of a participant, as the content of PDT curriculum was not dependent on the years of training and should be equally accessible to and absorbed by all participants. Comparison of interns to senior residents was performed with analysis of variance.

Inter-rater reliability was assessed with intraclass correlation coefficient (ICC). We set the target ICC for the scores at 0.70, a standard for inter-rater reliability coefficient. Reliability analyses were conducted for: (1) the learners' triage accuracy for each of the 10 patients; (2) learners' triage skills, specifically airway maneuvers, assessment of circulatory status, ability to ambulate, airway and breathing assessment, and assessment of mental status; (3) learners' triage knowledge, determined by application of the JumpSTART triage algorithm decision of the patients' triage levels; and (4) for each item included in the global assessment.

Alpha of 0.05 was used in all statistical analyses, and the analyses were conducted using SAS 9.2 (SAS Institute, Cary, NC).

### RESULTS

There were 40 learners who completed the simulation, of whom 37 consented to videography and to being subjects in the study. Fourteen participants were first-year interns, 10 were in their second year of training, 11 in their third and 2 in fourth year of training. Of note, only one intern had formal disaster medicine training before the curriculum, and one PGY-3 had prior clinical experience in a disaster, a hurricane. Each learner triaged 10 patients, for a total of 370 simulated triaged patients. There were 3 repeat observations of each video, resulting in 110 repeat observations of resident triage.

The mean total score was 44.0 (SD = 5.3), ranging from 30 to 55. Table 2 summarizes the scores for individual patient simulations as well as for the learner performance domains and the global assessment. Discrimination analyses showed that most patient-level scores, PDT skill, knowledge, and performance scores, and global assessment scores had a *D* statistic >0.5, indicating that a learner's performance on a given simulation or in a specific domain was predictive of the total score.

Patient	Expected Actions	Performance Evaluation		
		Yes	No	UTC*
#1. A school-aged child impaled on a pole	1) Determines patient is not ambulating			
	2) Determines the patient is not breathing			
	3) Positions airway (no response)			
	4) Feels for a pulse (No pulse)			
	5) Triage patient as a level Black, or deceased			
	6) Completed triage in <1 minute			
#2. A school-aged child with a head injury	1) Determines patient is not ambulating			
	2) Determines patient is breathing			
	3) Determines patient has a normal respiratory rate			
	4) Checks pulse (femoral or carotid)			
	5) Assesses level of function with AVPU mnemonic (patient is V, verbal)			
	6) Triage patient as level Yellow, delayed			
	7) Completed triage in <1 minute			

\*UTC = Unable to comment

Figure 1. Exemplary excerpt from the scoring instrument.

Exceptions included learners’ scene assessment, pulse assessment, and time to triage patients.

Learners showed no significant variation in PDT performance by PGY. Table 2 shows a comparison of PDT performance for interns and senior residents, including the *D* statistics. There was no difference among resident scores by PGY.

The overall ICC among total scores was 0.83 (95% confidence interval [CI] 0.74–0.89). As a pair, evaluators A and C had the highest overall reliability (ICC 0.87; 95% CI 0.77–0.93), and reliability between evaluators A and B or B and C was the same (ICC 0.71; 95% CI 0.51–0.84).

Inter-rater reliability for individual simulated patients is depicted in Figure 2b. Reliability was generally higher than 0.70, with the exception of patient 8, an ambulatory patient. The highest overall ICCs were for the 2 deceased patients (patients 1 and 2), an adolescent with mental status changes (patient 5), and a tachypneic infant (patient 9).

The ICCs of scores for PDT knowledge, skills and performance are shown in Figure 2a. We observed high levels of inter-rater reliability for most items, with triage decisions via the JumpSTART triage algorithm (ICC 0.92; 95% CI 0.88–0.95) and the skills airway assessment and airway management (ICC 0.91; 95% CI 0.93–0.98) showing the highest reliability. A notable exception was assessment of breathing (ICC 0.53; 95% CI 0.35–0.67). The ICCs for pairs of evaluators showed similar performance to overall ICCs.

Global assessment evaluations showed moderate reliability, with all ICCs <0.7. In ascending order, overall ICCs for global assessments of function were: assessment of professionalism 0.49 (95% CI 0.30–0.64), learners’ overall performance 0.59 (95% CI 0.42–0.72), skills assessment 0.63 (95% CI 0.47–0.75), and knowledge assessment 0.64 (95% CI 0.49–0.76).

**DISCUSSION**

This work has shown the performance of our PDT scoring instrument with a set of evaluators of varying disaster medicine experience. We report the design,

validation, and reliability testing of the first scoring instrument for evaluating residents learning PDT. We have demonstrated the instrument has both construct and internal validity. Our findings show the instrument yields reproducible results in the key domains of PDT knowledge, skills, and performance, as well as global assessment.

The discrimination statistics measure the construct validity of the PDT scoring instrument. The 4 items of the global assessment, particularly overall skill, knowledge, and performance, correlated highly with the overall score on the scoring instrument. At the patient level, simulated patients with more involved assessments (eg, patients 4, 6, and 9) correlated more closely with overall score than the simplest patient (patient 8, an ambulatory patient).

Previous work has focused on senior residents outperforming interns as a marker of scoring instrument validity.<sup>20</sup> Our work shows no difference between intern and senior resident performance. This finding is not unexpected. Few residents have opportunities to care for children in disasters, and simulation-based disaster training has not been integrated into our learners’ residency curriculum. Therefore, the similarity of performance across all training levels supports the internal validity of the instrument. The PDT scoring instrument does not detect experience in pediatrics residency; rather, it measures PDT performance.

Regarding the instrument’s reliability, correlation of scores at the individual patient level was high, meeting our target ICC of 0.70. Among patients with high ICC, patient 3, a CSHCN, deserves special discussion. The high overall ICC of 0.73 is an incomplete representation of the evaluator’s findings. Evaluators noted that learners often struggled with this nonverbal disaster patient. Learners verbalized a desire to assign a higher than necessary priority triage code for patient 3, effectively overtriaging that patient. Reasons for overtriage included uncertainty about whether the child’s deficits were the result of the bus crash or were typical for the child. Other learners noted a CSHCN is particularly vulnerable during disaster, a well-established concept.<sup>21</sup> In its current iteration, the evaluation instrument does not capture learners’ thought processes when they triage CSHCN. Future iterations of the instrument should include more evaluation of learners’ triage of CSHCN.

Evaluations with patient 8, an ambulatory disaster patient, represent the poorest reliability performance of the scoring instrument. Learners did not ignore the patient, nor did they fail to assign the expected Green triage tag to this patient with minor injuries. The reason for the poor ICC findings was evaluator interpretation of the scoring instrument. It was not clear to evaluators whether learners must verbalize their assessment of whether patient 8 could walk. Consequently, there was systematic disagreement between the reviewers. Subsequent versions of the scoring instrument explicitly require the learner to verbalize assessment of the patient’s ability to walk.

There was generally excellent correlation for objective skills and discrete knowledge assessed in the tool. Among PDT skills, speed of triage, and assessment of circulatory

**Table 2.** Descriptive Statistics of Learner Scores and Comparison of Scores by PGY

Scored Item	Discrimination ( <i>D</i> ) Statistic for Prediction of Total Score*	Mean Points (SD)			Overall <i>P</i> Value†
		n = 14	n = 10	n = 13	
Scene assessment	-0.19	3.5 (0.5)	3.4 (0.4)	3.5 (0.5)	.96
Patient total	0.99	40.6 (4.9)	41.4 (5.2)	39.7 (6.2)	.77
Patient 1	0.56	5.0 (0.8)	5.0 (0.6)	4.8 (1.0)	.82
Patient 2	0.58	4.6 (0.8)	4.5 (0.9)	4.0 (0.9)	.13
Patient 3	0.44	3.0 (0.9)	3.7 (0.4)	3.3 (0.7)	.14
Patient 4	0.62	3.5 (0.8)	3.7 (1.0)	3.7 (1.0)	.87
Patient 5	0.57	4.6 (1.5)	4.9 (1.1)	4.5 (1.3)	.81
Patient 6	0.73	4.3 (0.9)	3.9 (1.2)	4.1 (1.1)	.56
Patient 7	0.60	4.9 (1.1)	4.4 (1.6)	4.8 (1.1)	.55
Patient 8	0.37	2.7 (0.5)	2.8 (0.3)	2.7 (0.4)	.59
Patient 9	0.67	4.5 (0.8)	5.4 (0.5)	5.0 (1.0)	.0512
Patient 10	0.59	3.2 (0.9)	3.2 (1.0)	2.9 (1.2)	.65
Total score‡	...	44.1 (4.7)	44.8 (4.9)	43.2 (6.3)	.77
PDT Knowledge, Skill, and Performance					
Ambulation	0.76	3.7 (2.4)	4.4 (2.8)	3.4 (3.1)	.72
Breathing	0.56	8.7 (0.4)	8.6 (0.8)	8.6 (0.7)	.91
Airway	0.66	4.3 (2.1)	4.3 (1.7)	3.7 (2.3)	.71
Pulse	0.07	3.9 (0.9)	3.1 (1.1)	3.9 (0.9)	.10
Triage Level	0.67	7.9 (1.4)	8.1 (1.3)	7.9 (1.7)	.92
Time to Triage	0.22	9.1 (1.5)	9.7 (0.4)	9.2 (1.3)	.43
Mental Status	0.43	1.4 (0.5)	1.4 (0.5)	1.4 (0.6)	.97
Global Assessment of Learners					
Skills	0.74	2.6 (0.8)	3.0 (0.5)	2.4 (0.8)	.12
Knowledge	0.79	2.8 (0.8)	3.1 (0.7)	2.9 (0.8)	.67
Professionalism	0.61	2.8 (0.6)	3.2 (0.5)	2.8 (0.5)	.11
Overall performance	0.79	2.6 (0.5)	2.9 (0.5)	2.5 (0.7)	.21

PGY = postgraduate year; SD = standard deviation.

\**D* statistics denote the predictive value of a scored item for the total score.

†*P* value obtained by analysis of variance.

‡The highest possible score is 59.

status showed good inter-rater reliability. However, correlation was lower for other domains, notably assessment of breathing, ambulatory status, and the global assessment of function. These instances of more moderate correlation bear consideration.

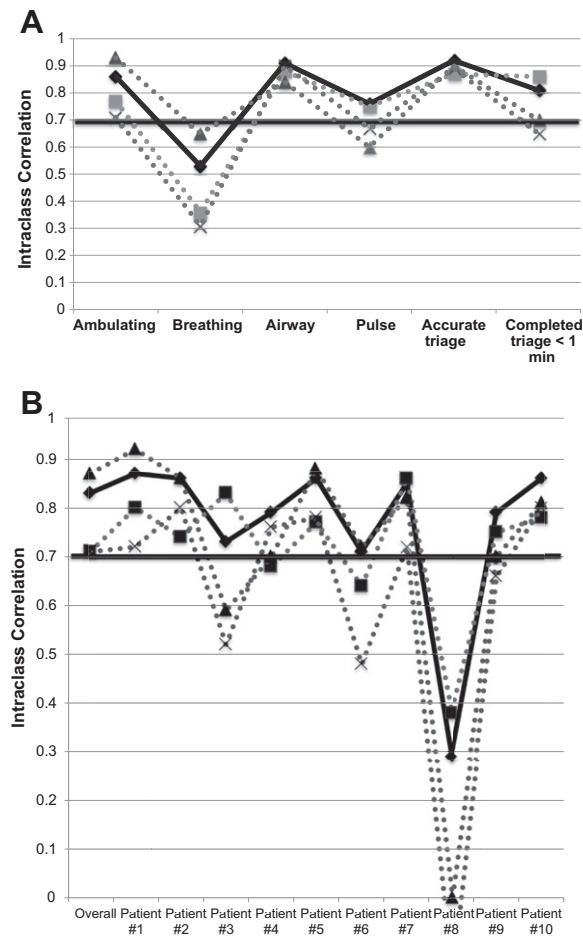
Scoring instrument skill items that did not correlate as well with the total score included scene assessment, the recognition that a disaster existed. Reasons for this might include 1) the small number of points allotted to this section, 2) that scene assessment involves consideration of the overall population of patients, rather than the individuals, and 3) the resources available to provide care. Other skills with poor correlations to the total score included assessment of pulses, and time to triage each patient, suggesting the fastest learners may not have been the best at PDT.

It was surprising that ICCs of assessment for breathing, both presence and rate, was lower than for other PDT skills. Evaluators were instructed to observe for learners' verbalization of breathing assessment and not to positively score learners' actions, such as chest palpation or auscultation. Further review of videos shows our facilitators did not consistently prompt learners to verbalize their actions (eg, "I am checking the patient's breathing" or "the patient is tachypneic"). Video review also showed limitations of the simulation manikins, with failure of the manikins to resume respirations after the learners had performed correct airway repositioning.

We believe the lack of standardized instructions for both learners and facilitators to focus on a verbal announcement of a patient's breathing status resulted in the lower ICCs for breathing. In this context, evaluators may have had disparate interpretations about whether breathing status was verbalized. We think the lower ICCs are not due to a shortcoming of the scoring instrument, and ICCs would have been improved by emphasizing and normalizing the expectations for the learners before the triage scenario. As an example, clarifying that learners must prompt an ambulatory patient, like patient 8, to actually walk, would likely improve the ICC for such patients.

The simultaneous use of simulation for conducting patient-outcomes research and education creates a challenge. In clinical situations, health care workers do not verbalize every action and assessment they perform. Balancing realism and assessment of learner's actions and knowledge can be difficult. In later iterations of this work, we have introduced more formal train-the-trainer interventions, and prompts for learners with unclear thought processes. Previous work by Henry et al has shown verbalization of thought processes does not affect clinical performance.<sup>22</sup>

Decisions about triage category had very high ICC. This likely reflects the highly objective nature of this assessment domain. Here, evaluators are using the instrument to record the learners' choices of the Red, Yellow, Green, or Black categories.



**Figure 2.** (A) Inter-rater reliability for triage performance. The solid line at an ICC of 0.7 indicates the target correlation. (B) Inter-rater reliability for individual simulated patient performance. The solid line at an ICC of 0.7 indicates the target correlation. The ◆ = overall correlation; ■ = reviewers A and B; ▲ = reviewers A and C; and X = reviewers B and C.

As a domain, the global assessment had the lowest ICCs. Although there were anchors to the Likert scores for the global assessment of function, it is by nature a less objective score, and previous authors have reported poor inter-rater reliability.<sup>23</sup> When expert, trained raters use a global assessment, the reliability of the assessment improves.<sup>24</sup> Given that this study used 3 evaluators of varying expertise, low ICCs may have stemmed from disparities in expertise. Further refinement of this domain should include formal training of the evaluators with a user's guide to the scoring instrument. Training of evaluators with videos modeling overall performance, professionalism, and triage skills may have improved correlation. Videographic examples of novice, advanced beginner, competent, proficient, and expert performance could facilitate this aim.

## LIMITATIONS

There are limitations to our evaluation of the PDT scoring instrument. First, there were 37 total learners from a single institution. This limits our ability to draw conclusions about the instrument's use for evaluating other

health care providers who may perform PDT, including emergency medicine physicians and emergency medical technicians. Two of the 3 evaluators were involved in the derivation of the instrument. A separate evaluation by independent users would help support the generalizability of the scoring tool when it is applied to resident learners. An additional analysis with 3 different reviewers, all of whom are pediatric disaster medicine experts, would complement the current study.

A final limitation is in the derivation of the evaluation tool, which is specific to JumpSTART-PDT. At this time, JumpSTART remains the prevalent triage strategy in the United States. Further, the skills evaluated in the instrument, such as airway assessment and maneuvers, assessment of pulses and mental status, are common to all PDT strategies in wide use. However, we have not evaluated the instrument's performance when other PDT strategies are used.

## FUTURE DIRECTIONS

There are several potential future applications of the PDT scoring instrument and its parent curriculum. First, we must revise the instrument, and add a train-the-trainer component. Items with poor ICC, such as patient 8, will drive these revisions and additions. A revised version of the instrument may be validated with different kinds of learners, including prehospital care providers, medical learners from other institutions, and school nurses. Using the instrument to assess performance in larger simulated mass-casualty events, or in actual disasters, would yield data about the generalizability of the instrument. Validation of the instrument with triage strategies other than JumpSTART would support its use in PDT education and evaluation in other settings, including international use.

## CONCLUSION

We report the design, validation, and reliability testing of the first scoring instrument for evaluating PDT learners. We have addressed specific areas for improving the instrument. The instrument has both construct and internal validity, measuring PDT performance, not experience as defined by PGY. The instrument is reliable, with high correlation of evaluator scores for most patients, discrete triage skills, assignment of triage levels, and the global assessment.

## ACKNOWLEDGMENTS

This work was sponsored by a grant from the Yale Pediatric Faculty Scholars Program. Dr Cicero would like to acknowledge Dr Lindsey Lane and the Academic Pediatric Association's Educational Scholars Project for mentorship and guidance during this work. Biostatistical collaboration was provided through CTSA grant UL1 RR024139 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. Information on Re-engineering the Clinical Research Enterprise can be obtained from the NIH website.

## REFERENCES

1. Brandenburg M, Arneson W. Pediatric disaster response in developed countries: ten guiding principles. *Am J Disaster Med.* 2007;2:151–162.
2. Fendya D. When disaster strikes—care considerations for pediatric patients. *J Trauma Nurs.* 2006;13:161–165.
3. American Academy of Pediatrics, American College of Emergency Physicians, American College of Surgeons Committee on Trauma, American Trauma Society, Children's National Medical Center CHAI, et al. Model uniform core criteria for mass casualty triage. *Disaster Med Public Health Prep.* 2011;5:125–128.
4. Romig L. Pediatric triage. A system to JumpSTART your triage of young patients at MCIs. *JEMS.* 2002;27:52–58. 60–63.
5. Cone DC, Serra J, Kurland L. Comparison of the SALT and Smart triage systems using a virtual reality simulator with paramedic students. *Eur J Emerg Med.* 2011;18:314–321.
6. Lerner E, Schwartz R, Coule P, et al. Mass casualty triage: an evaluation of the data and development of a proposed national guideline. *Disaster Med Public Health Prep.* 2008;2(suppl 1):S25–S34.
7. Cone D, Serra J, Burns K, et al. Pilot test of the SALT mass casualty triage system. *Prehosp Emerg Care.* 2009;13:536–540.
8. Sacco W, Navin D, Waddell RN, et al. A new resource-constrained triage method applied to victims of penetrating injury. *J Trauma.* 2007;63:316–325.
9. Jenkins J, McCarthy M, Sauer L, et al. Mass-casualty triage: time for an evidence-based approach. *Prehosp Disaster Med.* 2008;23:3–8.
10. Navin D, Sacco W, McGill G. Application of a new resource-constrained triage method to military-age victims. *Mil Med.* 2009;174:1247–1255.
11. Qazi K, Kempf JA, Christopher NC, et al. Paramedic judgment of the need for trauma team activation for pediatric patients. *Acad Emerg Med.* 1998;5:1002–1007.
12. Emerman C, Shade B, Kubincanek J. A comparison of EMT judgment and prehospital trauma triage instruments. *J Trauma.* 1991;31:1369–1375.
13. Sapp RF, Brice JH, Myers JB, et al. Triage performance of first-year medical students using a multiple-casualty scenario, paper exercise. *Prehosp Disaster Med.* 2010;25:239–245.
14. Cicero M, Blake E, Gallant N, et al. Impact of an educational intervention on residents' knowledge of pediatric disaster medicine. *Pediatr Emerg Care.* 2009;25:447–451.
15. Ballow S, Behar S, Claudius I, et al. Hospital-based disaster preparedness for pediatric patients: how to design a realistic set of drill victims. *Am J Disaster Med.* 2008;3:171–180.
16. Cicero M, Baum C. Pediatric disaster preparedness: best planning for the worst-case scenario. *Pediatr Emerg Care.* 2008;24:478–481.
17. Adler MD, Vozenilek JA, Trainor JL, et al. Development and evaluation of a simulation-based pediatric emergency medicine curriculum. *Acad Med.* 2009;84:935–941.
18. Murray DJ, Boulet JR, Kras JF, et al. A simulation-based acute skills performance assessment for anesthesia training. *Anesth Analg.* 2005;101:1127–1134, table of contents.
19. Donoghue A, Ventre K, Boulet J, et al. Design, implementation, and psychometric analysis of a scoring instrument for simulated pediatric resuscitation: a report from the EXPRESS pediatric investigators. *Simul Healthc.* 2011;6:71–77.
20. McBride ME, Waldrop WB, Fehr JJ, et al. Simulation in pediatrics: the reliability and validity of a multisenario assessment. *Pediatrics.* 2011;128:335–343.
21. Stalcup SA, Oscherwitz M, Cohen MS, et al. Planning for a pediatric disaster—experience gained from caring for 1600 Vietnamese orphans. *N Engl J Med.* 1975;293:691–695.
22. Henry SB, LeBreck DB, Holzemer WL. The effect of verbalization of cognitive processes on clinical decision making. *Res Nurs Health.* 1989;12:187–193.
23. Chipman JG, Beilman GJ, Schmitz CC, et al. Development and pilot testing of an OSCE for difficult conversations in surgical intensive care. *J Surg Educ.* 2007;64:79–87.
24. Adler MD, Vozenilek JA, Trainor JL, et al. Comparison of checklist and anchored global rating instruments for performance rating of simulated pediatric emergencies. *Simul Healthc.* 2011;6:18–24.