Prehospital Disaster Triage Does Not Predict Pediatric Outcomes: Comparing the Criteria Outcomes Tool to Three Mass-Casualty Incident Triage Algorithms

Mark X. Cicero, MD;[1] Genevieve R. Santillanes, MD;[2] Keith P. Cross, MD;[3] Amy H. Kaji, MD;[4] J. Joelle Donofrio, DO[5]

1. Yale School of Medicine, New Haven, Connecticut USA
2. Keck School of Medicine, University of Southern California, Los Angeles, California USA
3. University at Buffalo Jacobs School of Medicine, Buffalo, New York USA
4. Harbor-UCLA Medical Center, Torrance, California USA
5. Rady Children's Hospital of San Diego and University of California at San Diego, San Diego, California USA

Correspondence:
    Keith P. Cross, MD
    285 Middlesex Rd
    Buffalo, New York 14216 USA
    E-mail: kcross@upa.chob.edu

Abstract

Introduction: It remains unclear which mass-casualty incident (MCI) triage tool best predicts outcomes for child disaster victims.

Study Objectives: The primary objective of this study was to compare triage outcomes of Simple Triage and Rapid Treatment (START), modified START, and CareFlight in pediatric patients to an outcomes-based gold standard using the Criteria Outcomes Tool (COT). The secondary outcomes were sensitivity, specificity, under-triage, over-triage, and overall accuracy at each level for each MCI triage algorithm.

Methods: Singleton trauma patients under 16 years of age with complete prehospital, emergency department (ED), and in-patient data were identified in the 2007-2009 National Trauma Data Bank (NTDB). The COT outcomes and procedures were translated into ICD-9 procedure codes with added timing criteria. Gold standard triage levels were assigned using the COT based on outcomes, including mortality, injury type, admission to the hospital, and surgical procedures. Comparison triage levels were determined based on algorithmic depictions of the three MCI triage tools.

Results: A total of 31,093 patients with complete data were identified from the NTDB. The COT was applied to these patients, and the breakdown of gold standard triage levels, based on their actual clinical outcomes, was: 17,333 (55.7%) GREEN; 11,587 (37.3%) YELLOW; 1,572 (5.1%) RED; and 601 (1.9%) BLACK. CareFlight had the best sensitivity for predicting COT outcomes for BLACK (83% [95% confidence interval, 80%-86%]) and GREEN patients (79% [95% CI, 79%-80%]) and the best specificity for RED patients (89% [95% CI, 89%-90%]).

Conclusion: Among three prehospital MCI triage tools, CareFlight had the best performance for correlating with outcomes in the COT. Overall, none of three tools had good test characteristics for predicting pediatric patient needs for surgical procedures or hospital admission.

Introduction

Across the United States and the world, several mass-casualty incident (MCI) triage algorithms and strategies are in use.[1–6] Previous studies have attempted to determine which triage method is best for determining the prognosis of individual victims and the best use of resources when children are victims of disasters.[7–9] Despite these efforts, there remains no consensus regarding which algorithm is best in the pediatric population.

The inherent accuracy of MCI algorithms in the pediatric population is not well-studied. In 1990, Baxt and Upenieks[10] developed interventional criteria to test the Injury Severity Score's (ISS) ability to predict resource needs. These criteria were modified in 2000 by Garner and have been the most frequently published method of testing an MCI algorithm's ability to distinguish immediate priority patients.[11] The modified Garner criteria include: non-orthopedic operative intervention within six hours, fluid resuscitation of 1000ml or more to maintain blood pressure >89mmHg, invasive central nervous system monitoring or a positive head computed tomography scan, a procedure to maintain the airway or assisted ventilation, and decompression of a tension pneumothorax. The modified Garner criteria have been adapted to pediatrics with the fluid resuscitation changed to 20ml/kg in prior pediatric MCI studies. Other literature has used ISS, ventilator use, mortality, and admission length-of-stays as outcome proxies to test MCI algorithm accuracy.[9,12,13]

Given the lack of a tool to test the ability of MCI algorithms to triage correctly, in the last five years, the first two consensus-based gold standards to evaluate MCIs were developed.[14,15] These two standards have not been widely tested.

One of the consensus-based gold standards, the Criteria Outcomes Tool (COT), was developed specifically to test the inherent capability of MCI triage algorithms to correctly triage the pediatric population using outcomes data.[14] The COT is agnostic to the decision processes of any specific MCI triage method. Additionally, the COT is not another triage strategy in an already crowded field of contenders. Rather, the COT provides an exhaustive index based on illness and injury criteria, the interventions patients are likely to require, and the timeframe in which the procedure must be completed to test an MCI algorithm. Based on these criteria, the COT yields RED (required life-saving intervention), YELLOW (victim required operative repair/reduction during stay to preserve function; eg, vision and mobility), GREEN (no disability nor need of treatment to preserve normal function), or BLACK (did not survive) gold standard triage levels. The COT was derived using a modified Delphi Method with experts in pediatric disaster medicine as Delphi participants. An excerpt from the COT appears in Table 1.

The primary objective of this study was to compare triage outcomes of Simple Triage and Rapid Treatment (START[2,16]), modified START as used by the Fire Department of New York (FDNY; New York USA),[5,9] and CareFlight[9,12] in patients under 16 years of age (Figure 1) to an outcomes-based gold standard using the COT. The secondary outcomes were to determine sensitivity, specificity, under-triage, and over-triage at each level for each MCI triage algorithm along with the overall accuracy of the MCI algorithm. The study sought to determine whether each of the triage methods, if applied prospectively, would correctly predict patient outcomes after trauma.

Methods
Selection of Participants
The study utilized the 2007-2009 National Trauma Data Bank (NTDB), which is compiled by the American College of Surgeons (Chicago, Illinois USA).[17] The NTDB includes individual patient demographics, prehospital information including vital signs (if recorded), destination hospital, destination hospital trauma level designation, and patient interventions and outcomes. The Institutional Review Board of the University of California at

Red Category
Intubated for Intracranial Hemorrhage. Not apneic in field.
Mannitol, intubation, and Operating Room <90 min required for head injury with mass effect. Not apneic in field.
Received blood in Emergency Department for hemorrhagic shock from liver laceration.
Yellow Category
Chest tube placed after 2h, Pediatric Intensive Care Unit stay >48h.
Open Reduction Internal Fixation pf femur fracture, Pediatric Intensive Care Unit stay >48h.
Pediatric Intensive Care Unit stay >48h, admitted for monitoring of splenic laceration, oxygen given in Pediatric Intensive Care Unit for pneumothoraxes.
Green Category
No treatments needed to preserve limb or function.
Black Category
Apneic and pulseless despite bag-valve-mask, declared dead in Emergency Department.
Apneic and pulseless in field, defibrillated in Operating Room. Died within 10min of arrival to operating room and within 2h of arrival.
Head injury, cardiac arrest in Emergency Department (pulseless ventricular tachycardia), required chest compressions, defibrillation x 3.

Cicero © 2021 Prehospital and Disaster Medicine

Table 1. Examples of Criterion Outcome Tool Expected Triage Level Designations

San Diego School of Medicine (San Diego, California USA) approved this study (protocol # 160067).

Patients less than 16 years of age were selected from the full NTDB dataset from reporting years 2007, 2008, and 2009 and their data were extracted from the databank. These data included COT outcomes (apnea, cardiac arrest, and death in the emergency department [ED]; disposition to monitored bed for less than 48 hours; and time to procedure) which were classified according to ICD-9 codes.

There were 530,695 individual patients with trauma incidents for consideration in the resulting data set, as shown Figure 1. For this study, the sub-set of these individual patients who were aged 0-15 years at the time of the trauma, transported from the scene of injury with internally consistent data, was considered. This set of trauma patients has been used in prior studies.[9,18] This previously extracted and formatted dataset was chosen to save time and simplify comparison to results of prior work.

Determination of Triage Levels
Using the NTDB Procedures data, the study assigned COT triage levels (BLACK/RED/YELLOW/GREEN) to each patient using the following steps:

1. Modification of the COT using the NTDB procedure code list. Two experts in pediatric disaster independently reviewed 3,241 procedure codes with a third expert to review discrepancies. A COT triage code was assigned based on each procedure code and the following time constraints for the procedure: any, time ≤90 minutes, time >90 minutes, and ≤24 hours). All else were defaulted to GREEN.
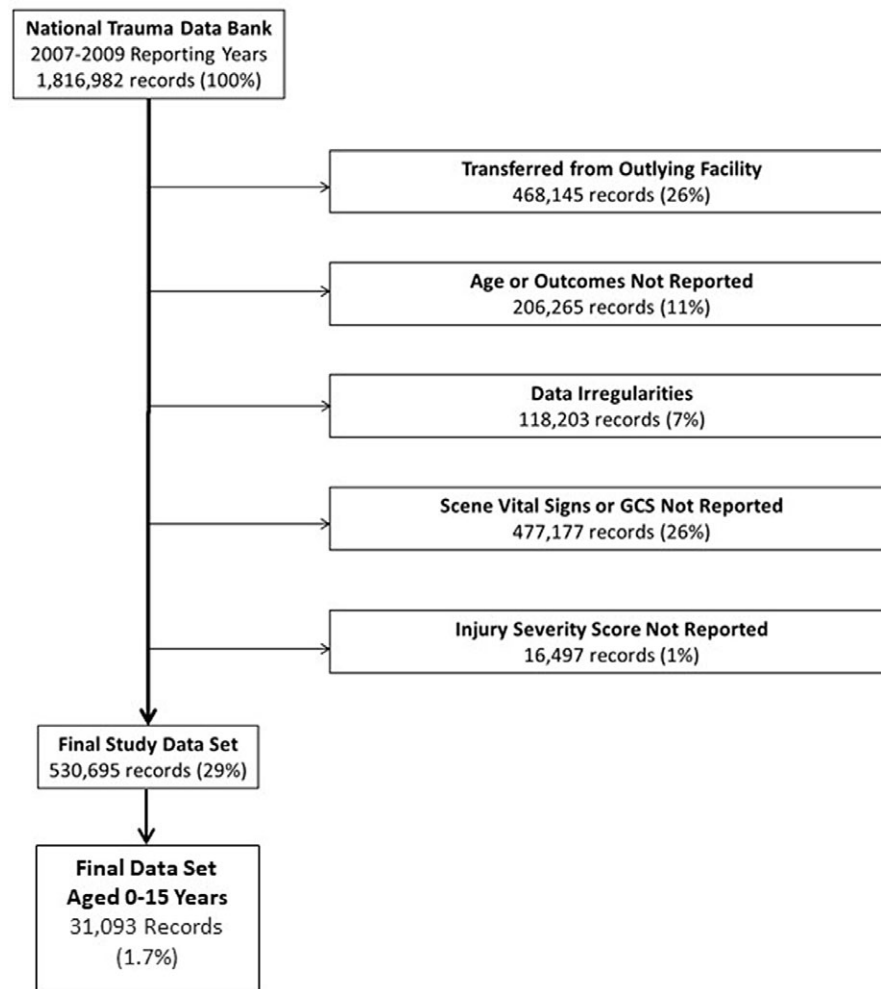
Figure 1. Selection of Patients from the National Trauma Data Bank.
Abbreviation: GCS, Glasgow Coma Scale.

2. Sort procedures performed in the first 72 hours for each patient and select the procedure (if there were multiple procedures) with the highest COT triage code.
3. Assign the highest COT triage code to the patient.
4. Match the list of COT triage code assignments to the previously generated subset of 530,695 incidents with complete data (see #2 above).
5. If needed, further modify COT assignments for each incident with the following COT non-procedure-based criteria:

   a. Initial pulse = 0 at scene à BLACK;
   b. Initial respirations = 0 at scene à BLACK;
   c. Death in the ED à BLACK; or
   d. ED disposition to intensive care unit/telemetry/step down unit, **and** hospital length-of-stay >48 hours à minimum YELLOW.

Based on the outcomes and procedures for each patient, an expected gold standard prehospital triage level was determined retrospectively for each patient with the presupposition that the patient was being triaged in an MCI.

The three triage methods, START, FDNY, and CareFlight, were applied according to their published algorithms, as depicted in Figure 2a, Figure 2b, and Figure 2c. The method of approximating each MCI triage method using the NTDB was previously published.[9] JumpSTART was utilized in children eight years and under and START for children over eight years of age. Ability to walk was assumed for patients who did not meet other triage level criteria and had an ISS of 10 or less. These methods only varied in that infants <12 months of age were not automatically redtagged in the modified START algorithm. Patient triage levels were determined based on the algorithmic depiction of the triage strategy.

Analysis
Using the COT outcomes as the criterion standard, the study determined the overall accuracy, sensitivity, specificity, and occurrences of over- and under-triage of the START, FDNY, and CareFlight MCI triage algorithms. The ORANGE and YELLOW triage categories of FDNY triage were compared to the YELLOW expected triage level yielded by the COT.

The Wilcoxon sign rank test was applied to assess differences between the gold standard triage level choice and the triage level
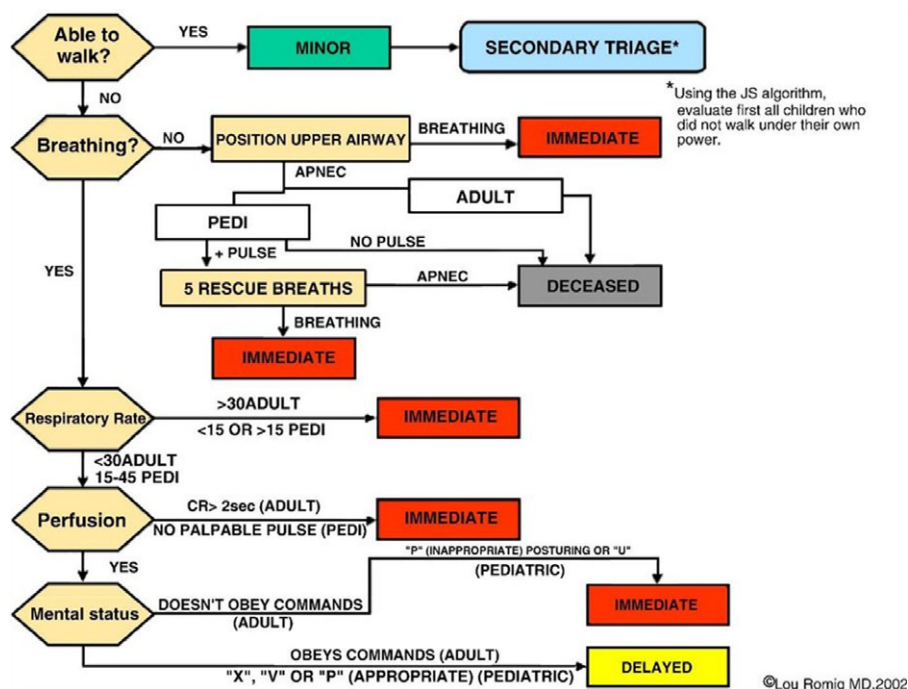
Figure 2A. Algorithmic Depiction of the Combined START/JumpSTART Triage Algorithm. Abbreviation: START, Simple Triage and Rapid Treatment.

that would be applied when each algorithm was used for triage in an MCI. The three triage algorithms were compared regarding sensitivity and specificity for each COT triage level and compared by incidence of under-triage (patients who receive lower priority MCI algorithm triage than the COT gold standard triage) and over-triage (patients who receive higher priority MCI algorithm triage than the COT gold standard triage). For these analysis calculations, the order of priority from lowest to highest was: GREEEN, YELLOW, RED, BLACK. The statistical software package utilized was SPSS (version 25; IBM Corp.; Armonk, New York USA).

Results

In total, 31,093 patients 15 years of age or younger with complete prehospital data for review were identified in the NTDB.

Within those, 3,241 procedures were reviewed with 2,382 codes excluded. The 871 included procedure codes were further categorized into GREEN, YELLOW, RED, and BLACK based on timing (Table 2). A consensus to the procedure codes was made without any discrepancies between the two independent reviewers.

The modified COT (including time adjusted procedure codes, pulselessness/apnea/death in the ED, and admission criteria) was applied to the patients. The resulting breakdown of gold standard triage levels included the following: 17,333 (55.7%) GREEN; 11,587 (37.3%) YELLOW; 1,572 (5.1%) RED; and 601 (1.9%) BLACK patients.

Overall accuracy of the MCI algorithms ranged from 49%-56% with CareFlight having the highest and FDNY the lowest. Of the three MCI triage tools, CareFlight had the best sensitivity for predicting COT outcomes for BLACK (83% [95% confidence interval, 80%-86%]) and GREEN patients (79% [95% CI, 79%-80%]) and the highest specificity for RED patients (89% [95% CI, 89%-90%]). CareFlight also had the lowest under-triage (17%) in the

BLACK category. Table 3 shows the sensitivity, specificity, over-triage, and under-triage percentages of each MCI algorithm compared to the COT gold standard triage levels.

Discussion

A comparison of three commonly used MCI triage tools, START, FDNY triage, and CareFlight, showed many gaps between the performance of these triage tools in children and the gold standard triage level, as determined using the modified COT. CareFlight performed relatively well for the BLACK and GREEN categories, but it demonstrated low sensitivity in identifying RED patients, those with the most immediate need of life-saving interventions. None of the tools demonstrated both high specificity and sensitivity for any COT triage category and no MCI algorithm tested came close to the American College of Surgeons' triage goals.

Using the COT outcomes as the criterion standard, the algorithms tested showed poor sensitivity in predicting RED and YELLOW outcomes and poor specificity predicting GREEN outcomes in pediatric trauma patients.

This is the first publication to test MCI algorithms against the COT and the first report to test the sensitivity, specificity, over-, and under-triage of MCI algorithms in each of the four triage levels. Prior literature comparing MCI algorithms to the modified Garner criteria found higher rates of sensitivity and specificity in the RED triage cohorts with both START and CareFlight. In 2000, Garner, et al tested START, Manchester Sieve, and CareFlight against the modified Garner criteria using 1,144 retrospective adult trauma patients.[11] They found START to perform better than it did when applied to the COT, having 85% sensitivity and 86% specificity to identify RED patients using the modified Garner criteria versus the current study findings of 54% and 85% when applied against the COT. CareFlight's RED cohort
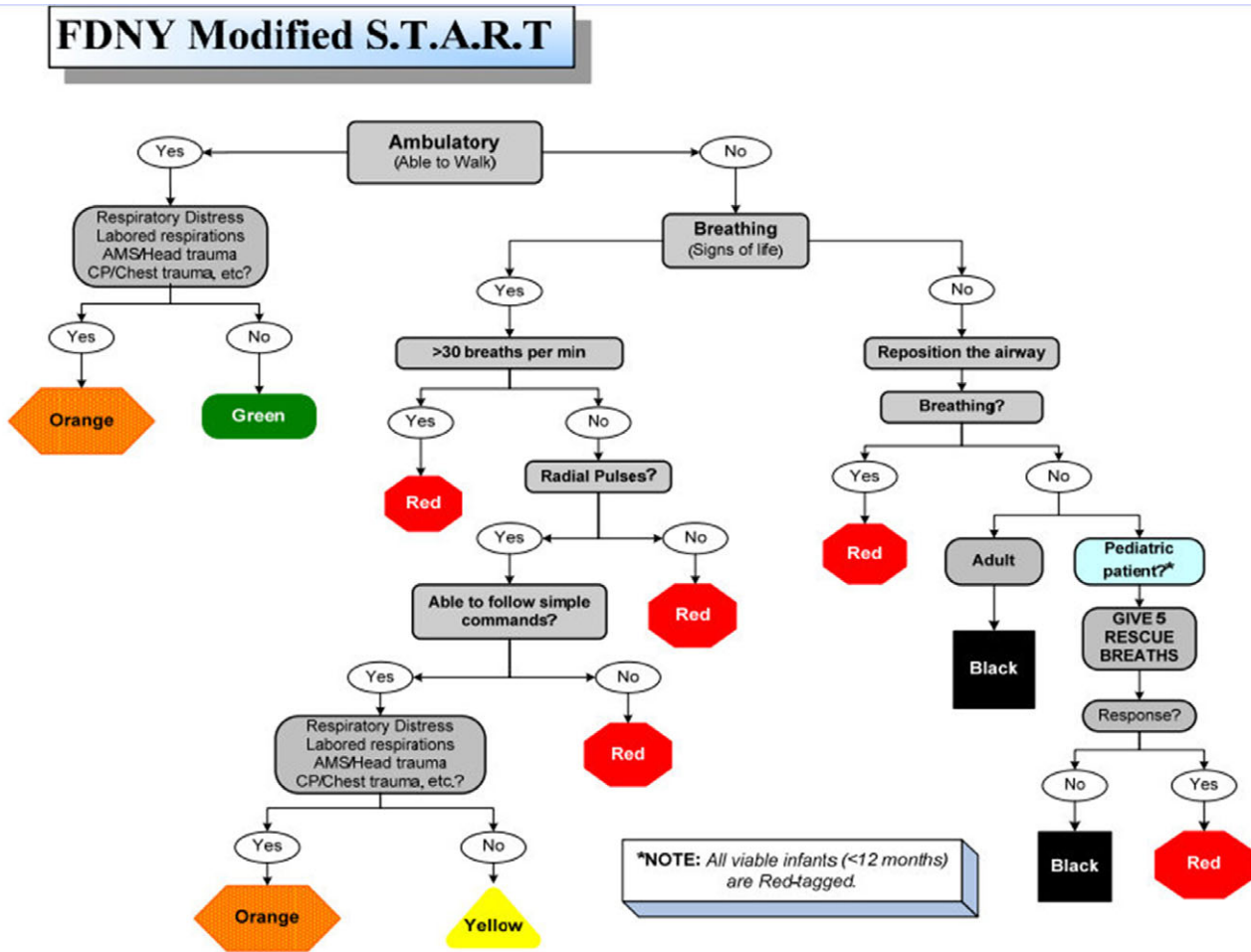
FDNY Modified S.T.A.R.T

Figure 2B. Algorithmic Depiction of the FDNY Modified START Triage Algorithm.
Abbreviations: FDNY, Fire Department of New York; START, Simple Tirage and Rapid Treatment; AMS, altered mental status; CP, chest pain.
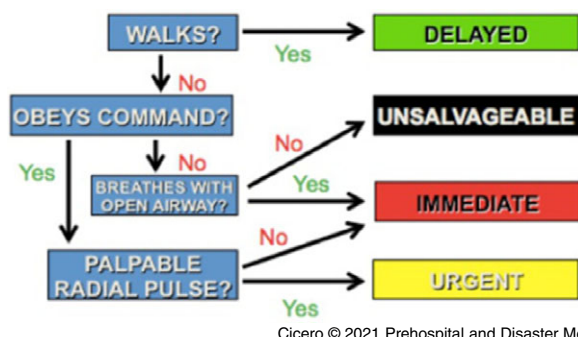
Figure 2C. Algorithmic Depiction of the CareFlight Triage Algorithm.

was found to have a sensitivity of 82% and specificity of 96% with the modified Garner compared to 50% and 89% with the COT.

In reviews of real-life disaster incidents, using the modified Garner criteria and retrospective MCI algorithm application, sensitivity and specificity in the RED triage level were higher than reported in this study. Challen, et al retrospectively assessed a cohort of patients treated at the Royal London Hospital (United

Kingdom) from the July 7, 2005 transport bombing.[19] In the 204/404 bombing victims, when the algorithms were applied retrospectively, both CareFlight and START had 75% sensitivity and 100% specificity with the four critical patients. In 2008, Kahn, et al assessed 148 patients involved in a 2003 train crash who were triaged with START in the field. Using the modified Garner outcomes, they found START to have 100% sensitivity and 77% specificity in the RED triage. Patients not meeting the modified Garner criteria but admitted to hospital for at least 24 hours were considered YELLOW and START was found to have a sensitivity of 39% and specificity of 12%.[20] This varies considerably from the 21% and 88%, respectively, for START's YELLOW triage level in the current study results. Lastly, in Kahn's study, START was noted to have 46% sensitivity and 89% specificity in the GREEN outcome whereas the current study found 77% sensitivity and 44% specificity. The variabilities between application of the MCI algorithm to the modified Garner and the COT point to a need for future research. The current study patients' pediatric age may have contributed to the differences in triage.

A majority of the literature, including all studies mentioned in the above paragraph, have been reported on the adult population. Given children have differing physiology and responses to injury,

	Any Time	Took Place in ED	Occurred ≤90 Minutes	Occurred >90 Minutes	Occurred <24 Hours	Totals
GREEN	78	11	0	0	0	89
YELLOW	276	23	7	348	112	766
RED	0	55	349	0	1	405
BLACK	8	11	1	0	0	20
Totals	362	100	357	348	113	1280

Cicero © 2021 Prehospital and Disaster Medicine

Table 2. Categorization of COT based on Timing of NTDB Procedure Code
Abbreviations: COT, Clinical Outcomes Tool; NTDB, National Trauma Data Bank; ED, emergency department.

COT Level	% Sensitivity (95% CI)	% Specificity (95% CI)	Patients Over-Triaged	Patients Under-Triaged
		START		
Black	39 (35 – 43)	100 (99 – 100)	n/a	61%
Red	54 (51 – 56)	85 (84 – 86)	0%	46%
Yellow	21 (20 – 22)	88 (87 – 89)	16%	63%
Green	77 (76 – 77)	44 (43 – 45)	23%	n/a
		FDNY		
Black	39 (35 – 43)	100 (99 – 100)	n/a	61%
Red	56 (54 – 59)	80 (79 – 80)	0%	56%
Yellow	20 (19 – 21)	88 (87 – 88)	19%	60%
Green	69 (68 – 69)	47 (46 – 48)	31%	n/a
		CareFlight		
Black	**83** (80 – 86)	100 (99 – 100)	n/a	17%
Red	50 (48 – 53)	**89** (89 – 90)	0%	50%
Yellow	22 (21 – 23)	87 (87 – 88)	14%	65%
Green	**79** (79 – 80)	43 (42 – 44)	21%	n/a

Cicero © 2021 Prehospital and Disaster Medicine

Table 3. Comparison of Three MCI Algorithms' Sensitivity, Specificity, Under-Triage, and Over-Triage
Abbreviations: COT, Clinical Outcomes Tool; START, Simple Triage and Rapid Treatment; FDNY, Fire Department of New York; MCI, mass-casualty incident.

specific evaluation of how well an MCI algorithm performs in the pediatric population is important. Classically, START has been used for patients over eight years of age and its pediatric variate, JumpSTART, for patients under eight years. The Pediatric Triage Tape varies the triage parameters based on height. The Sacco Triage Method has an age adjusted score. No other algorithm has adaptations for children.

In 2006, Wallis, et al applied a series of MCI algorithms, Pediatric Triage Tape, CareFlight, START, and JumpSTART, to 3,461 children less than 12 years of age brought to a Red Cross Children's Hospital trauma unit.[13] They applied each MCI algorithm and assessed the sensitivity and specificity of the RED triage using ISS >15, the NEW ISS >15, and the modified Garner Criteria. They found CareFlight to perform best with sensitivity 31.5%-48.4% and specificity 98.8%-99.0%. This was the only study to separate JumpSTART and START and found them to have lowest level of performance. JumpSTART had 0.8%-3.2% sensitivity and 97.7%-97.8% specificity while START had 22.3%-39.2% sensitivity and 77.3%-78.7% specificity.

When Cross and Cicero[9] evaluated the Sacco Triage Method Algorithm, applied to a large trauma registry using pediatric patients under 15 years, they found that it had a high ability to

predict mortality but was not as reliable at predicting the secondary outcomes of major injury (ISS >15) or a combined endpoint of serious outcomes (all deaths, all transfers for acute care, intensive care and operating room admissions, and admissions over two days). They also found that the Sacco age adjustment for children adds little to its predictive accuracy.

The current study found poor predictive values of the tested MCI algorithms to correctly sort pediatric patients. All these studies illustrate the need for further data assessing the best method of accurately sorting high volumes of pediatric disaster patients quickly.

Past work has forwarded the Model Uniform Core Criteria (MUCC),[7,21] which serves as a standard for the ideal MCI triage tool. The criteria include general triage considerations, global sorting, life-saving interventions, and assignment of triage categories. The criteria are only meant to be used by providers in incidents with a discrete geographic location or locations when they are organizing multiple victims, regardless of the size of the incident. The MUCC consider primary triage, the prehospital sorting of MCI victims.

The current study builds on past work investigating the test characteristics, under-triage, and over-triage expected when various MCI tools are used. It provides a unique contribution to triage

science because it is the first study to use an established outcomes-based tool to compare the predictive value of commonly used triage algorithms. This work differs from the aforementioned studies in that it uses data from 31,093 actual pediatric trauma patients with information from the scene of the injury, interventions needed for each patient, and short- and long-term patient outcomes. Prior studies have used simulated patients and included far fewer cases. The current study established methods for the application of the COT for evaluating MCI triage tools based on patient outcomes after ED evaluation. Based on this work, START, FDNY, and CareFlight do not appear to accurately predict pediatric patient care needs at a receiving facility. These findings underscore the need for secondary triage at the scene and further triage when patients reach definitive care at the hospital.

Study Limitations
This investigation has several limitations. First, the patients were trauma victims who had limited, discrete information entered in the NTDB. These data are prone to the limitations of all database research, including misclassification bias, secondary to incorrect and incomplete entries. Because of these shortcomings, the study omitted patients with incomplete and/or unreliable data.

Second, most patients in the NTDB were singleton trauma victims, rather than victims of an MCI. It is possible that outcomes for these trauma patients would have been less favorable had they presented as part of an MCI cohort.

Third, START, FDNY triage, and CareFlight were applied only to trauma victims in this study, thus limiting the applicability of the findings presented here to victims of other kinds of disasters, such as natural catastrophes, chemical, biological, or radiation events.

Fourth, the use of ISS cutoff values to approximate such difficult-to-measure assessments like ability to walk in the process of determining triage assignments likely is not completely accurate. Similarly, certain response-to-intervention elements of MCI triage algorithms (eg, "Give two breaths and see if the patient starts breathing on their own") could not be modeled with the available data.

Regarding limitations of FDNY triage specifically, the ORANGE and YELLOW categories of FDNY were collapsed into one category for this analysis. Patients who would be triaged as ORANGE in practice were under-triaged in this analysis. Also, FDNY triage in practice assigns RED triage to all infants less than one year old, a factor not modeled in this analysis to facilitate comparisons with the other triage strategies.

Finally, this work did not consider other prehospital factors, such as traits known to improve readiness for pediatric emergencies:[22] level of prehospital care (Basic or Advanced Life Support), transport time, receiving facility characteristics (including American College of Surgeon's trauma designation), and the presence of a pediatric intensive care unit or children's hospital.

Future Directions
Future directions include additional applications of the COT for prehospital triage. These methods can be used to evaluate other existing MCI triage tools, or to derive novel triage algorithms with high sensitivity and specificity for predicting patient outcomes and the complexity of patient health care needs. Additionally, the COT could be used in concert with the MUCC as the basis for a single, evidence-based MCI triage tool with high predictive value for patient outcomes and broad applicability across a range of disaster events. Compared to present strategies, a primary triage tool with these characteristics would have greater utility in MCIs and would promote more efficient use of resources.

One additional application of the COT is for after-action assessment of actual mass-casualty events and subsequent Emergency Medical Service response. This work would provide greater insight regarding management and outcomes for MCI victims and guidance regarding interventions, such as tourniquets, that improve survival for patients in disasters.

Conclusion
When pediatric trauma patients aged 0-15 years are triaged using START, FDNY triage, or CareFlight, comparison with the COT standard shows that CareFlight has the relative best performance for predicting patient outcomes. All of START, FDNY triage, and CareFlight algorithms do not appear to accurately predict pediatric patient care needs at a receiving facility. The COT can be used to evaluate the correlation between the sorting outcomes of a triage strategy and the short- and long-term outcomes for a population of disaster victims.

References
1. SALT mass casualty triage: concept endorsed by the American College of Emergency Physicians, American College of Surgeons Committee on Trauma, American Trauma Society, National Association of EMS Physicians, National Disaster Life Support Education Consortium, and State and Territorial Injury Prevention Directors Association. *Disaster Med Public Health Prep.* 2008;2(4):245-246.
2. Romig L. Pediatric triage. A system to JumpSTART your triage of young patients at MCIs. *JEMS.* 2002;27(7):52–58, 60-53.
3. Sacco W, Navin D, Waddell Rn, Fiedler K, Long W, Buckman RJ. A new resource-constrained triage method applied to victims of penetrating injury. *J Trauma.* 2007;63(2):316–325.
4. 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(6):314–321.
5. Arshad FH, Williams A, Asaeda G, et al. A modified simple triage and rapid treatment algorithm from the New York City (USA) Fire Department. *Prehosp Disaster Med.* 2015;30(2):199–204.
6. Jenkins J, McCarthy M, Sauer L, et al. Mass-casualty triage: time for an evidence-based approach. *Prehosp Disaster Med.* 2008;23(1):3–8.
7. Lerner EB, Cone DC, Weinstein ES, et al. Mass casualty triage: an evaluation of the science and refinement of a national guideline. *Disaster Med Public Health Prep.* 2011;5(2):129–137.
8. Cross KP, Cicero MX. Independent application of the Sacco Disaster Triage Method to pediatric trauma patients. *Prehosp Disaster Med.* 2012;27(4):306–311.
9. Cross KP, Cicero MX. Head-to-head comparison of disaster triage methods in pediatric, adult, and geriatric patients. *Ann Emerg Med.* 2013;61(6):668-676.e667.
10. Baxt WG, Upenieks V. The lack of full correlation between the Injury Severity Score and the resource needs of injured patients. *Ann Emerg Med.* 1990;19(12):1396–1400.
11. Garner A, Lee A, Harrison K, Schultz CH. Comparative analysis of multiple-casualty incident triage algorithms. *Ann Emerg Med.* 2001;38(5):541–548.
12. Wallis LA, Carley S. Comparison of pediatric major incident primary triage tools. *Emerg Med J.* 2006;23(6):475–478.
13. Wallis LA, Carley S. Validation of the pediatric triage tape. *Emerg Med J.* 2006;23(1):47–50.
14. Donofrio JJ, Kaji AH, Claudius IA, et al. Development of a pediatric mass casualty triage algorithm validation tool. *Prehosp Emerg Care.* 2016;20(3):343–353.
15. Lerner EB, McKee CH, Cady CE, et al. A consensus-based gold standard for the evaluation of mass casualty triage systems. *Prehosp Emerg Care.* 2015;19(2):267–271.
16. Cicero MX, Overly F, Brown L, et al. Comparing the accuracy of three pediatric disaster triage strategies: a simulation-based investigation. *Disaster Med Public Health Prep.* 2016;10(2):253–260.
17. Fantus RJ, Fildes J. NTDB data points. How national is the trauma data bank? *Bull Am Coll Surg.* 2003;88(5):37.

18. Cross KP, Petry MJ, Cicero MX. A better START for low-acuity victims: data-driven refinement of mass casualty triage. *Prehosp Emerg Care.* 2015;19(2):272–278.

19. Challen K, Walter D. Major incident triage: comparative validation using data from 7th July bombings. *Injury.* 2013;44(5):629–633.

20. Kahn C, Schultz C, Miller K, Anderson C. Does START triage work? An outcomes assessment after a disaster. *Ann Emerg Med.* 2009;54(3):424–430, 430.e421.

21. American Academy of Pediatrics; American College of Emergency Physicians; American College of Surgeons - Committee on Trauma; et al. Model uniform core criteria for mass casualty triage. *Disaster Med Public Health Prep.* 2011;5(2):125–128.

22. Remick K, Kaji AH, Olson L, et al. Pediatric readiness and facility verification. *Ann Emerg Med.* 2016;67(3):320–328.e321.