ORIGINAL RESEARCH

Inter-Rater Reliability and Agreement Among Mass-Casualty Incident Algorithms Using a Pediatric Trauma Dataset: A Pilot Study

Emilia H. Fisher, MD, MPH;[1] Ilene Claudius, MD;[2,3,4] Amy H. Kaji, MD, PhD;[2,3,4] Alaa Shaban, MD;[5] Nicholas McGlynn, MD;[6] Mark X. Cicero, MD;[7] Genevieve Santillanes, MD;[8] Marianne Gausche-Hill, MD;[2,3,4,9] Todd P. Chang, MD, MAcM;[10] J. Joelle Donofrio-Odmann, DO[11,12]

1. Department of Emergency Medicine, Children's Minnesota, Minneapolis, Minnesota USA
2. Harbor-UCLA Medical Center, Emergency Medicine, Torrance, California USA
3. David Geffen School of Medicine at UCLA, Los Angeles, California USA
4. The Lundquist Institute at Harbor-UCLA, Los Angeles, California USA
5. Community Health Centers, Pediatrics, Santa Maria, California USA
6. Keck School of Medicine of University of Southern California, Los Angeles, California USA
7. Yale University School of Medicine, New Haven, Connecticut USA
8. Department of Emergency Medicine, Keck School of Medicine of University of Southern California, Los Angeles, California USA
9. Los Angeles County Emergency Medical Services Agency, Santa Fe Springs, California USA
10. Children's Hospital Los Angeles, Los Angeles, California USA
11. University of California San Diego Departments of Pediatrics and Emergency Medicine, San Diego, California USA
12. Rady Children's Hospital of San Diego, San Diego, California USA

Correspondence:
Emilia H. Fisher
Children's Minnesota - Department of Emergency Medicine
2525 Chicago Ave South
Mailstop 32-1488
Minneapolis, Minnesota 55404 USA
E-mail: Emilia.fisher@childrensmn.org

Conflicts of interest: The authors declare none.

Abstract

Introduction: Many triage algorithms exist for use in mass-casualty incidents (MCIs) involving pediatric patients. Most of these algorithms have not been validated for reliability across users.
Study Objective: Investigators sought to compare inter-rater reliability (IRR) and agreement among five MCI algorithms used in the pediatric population.
Methods: A dataset of 253 pediatric (<14 years of age) trauma activations from a Level I trauma center was used to obtain prehospital information and demographics. Three raters were trained on five MCI triage algorithms: Simple Triage and Rapid Treatment (START) and JumpSTART, as appropriate for age (combined as J-START); Sort Assess Life-Saving Intervention Treatment (SALT); Pediatric Triage Tape (PTT); CareFlight (CF); and Sacco Triage Method (STM). Patient outcomes were collected but not available to raters. Each rater triaged the full set of patients into Green, Yellow, Red, or Black categories with each of the five MCI algorithms. The IRR was reported as weighted kappa scores with 95% confidence intervals (CI). Descriptive statistics were used to describe inter-rater and inter-MCI algorithm agreement.
Results: Of the 253 patients, 247 had complete triage assignments among the five algorithms and were included in the study. The IRR was excellent for a majority of the algorithms; however, J-START and CF had the highest reliability with a kappa 0.94 or higher (0.9-1.0, 95% CI for overall weighted kappa). The greatest variability was in SALT among Green and Yellow patients. Overall, J-START and CF had the highest inter-rater and inter-MCI algorithm agreements.
Conclusion: The IRR was excellent for a majority of the algorithms. The SALT algorithm, which contains subjective components, had the lowest IRR when applied to this dataset of pediatric trauma patients. Both J-START and CF demonstrated the best overall reliability and agreement.

Fisher EH, Claudius I, Kaji AH, Shaban A, McGlynn N, Cicero MX, Santillanes G, Gausche-Hill M, Chang TP, Donofrio-Odmann JJ. Inter-rater reliability and agreement among mass-casualty incident algorithms using a pediatric trauma dataset: a pilot study. ***Prehosp Disaster Med.*** **2022;37(3):306–313.**

Keywords: disaster; inter-rater reliability; mass casualty; pediatric; triage

Abbreviations:
AVPU: alert, verbal, responsive to pain, unresponsive (responsiveness score)
CF: CareFlight
ED: emergency department
EMS: Emergency Medical Services
EMT: emergency medical technician
GCS: Glasgow Coma Score
IRR: inter-rater reliability
ISS: Injury Severity Score
J-START: combination of START and JumpSTART
JumpSTART: pediatric modification of START
MCI: mass-casualty incident
PTT: Pediatric Triage Tape
SALT: Sort, Assess, Life-Saving Intervention, Treatment
START: Simple Triage and Rapid Treatment
STM: Sacco Triage Method

Received: October 13, 2021
Revised: December 4, 2021
Accepted: December 11, 2021

doi:10.1017/S1049023X22000632
© The Author(s), 2022. Published by Cambridge University Press on behalf of the World Association for Disaster and Emergency Medicine.

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press

Introduction

Background

Mass-casualty incident (MCI) triage algorithms are intended to allow prehospital providers to quickly and accurately identify patients who require immediate life-saving care and to appropriately allocate limited resources. Emotional obstacles, triage rationale and efficiency, and unfamiliarity with pediatric physiology have been identified as impediments to appropriate triage in pediatric MCIs.[1] While previous studies have compared test characteristics and over- and under-triage rates, the inter-rater reliability (IRR) of most disaster algorithms in the pediatric population has not been established.[2–5] This deficit has been highlighted by both Jenkins, et al in their 2008 paper addressing the need for an evidence-based approach with mass-casualty triage as well as by Heffernan, et al in their 2018 publication on MCI algorithm accuracy in the pediatric population.[4,6] Currently, several MCI algorithms, which guide allocation of responders' time and resources, are utilized for pediatric patients in MCIs. A comparison of the clinical factors used by five of these MCI triage algorithms to designate patients into categories ranging from deceased to minor injuries are presented in Tables 1 and 2.

Simple Triage and Rapid Treatment (START), and its pediatric modification, JumpSTART (initially created for eight years of age and under), combined as J-START, are widely used in the United States.[6] One national survey found that it is used by 71% of Emergency Medical Service (EMS) systems employing a pediatric disaster triage algorithm.[7] As such, START is one of the most well-studied disaster algorithms in the adult and pediatric population and is the only algorithm with literature on the IRR of provider application in the pediatric population.[5,8]

The Sort, Assess, Life-Saving Interventions, Treatment/Transport (SALT) algorithm was proposed in 2008 as a new national standard. Currently, SALT is the only disaster algorithm meeting the key components in the Model Uniform Core Criteria,[9,10] a consensus guideline by the Federal Interagency Committee on EMS (Washington, DC USA) that recommends 24 criteria for all MCI systems with four general categories: general considerations, global sorting, life-saving interventions, and individual assessment. The Centers for Disease Control and Prevention (Atlanta, Georgia USA) funded a workgroup that developed the Model Uniform Core Criteria and created the SALT triage algorithm. The SALT algorithm is unique in its sequence, with global sorting followed by life-saving intervention and then categorization of triage. It is the only commonly used algorithm utilizing providers' subjective assessments–whether the patient is "likely to survive given resources" and whether there are "minor injuries only."[10]

The Pediatric Triage Tape (PTT) is the pediatric adaptation of the Trauma Sieve, a disaster algorithm commonly used in the United Kingdom and parts of Australia which evaluates ability to walk, airway patency, respiratory rate, and pulse rate.[6] The PTT incorporates patient height and/or weight, relying upon measurements obtained on the backboard to determine appropriate vital sign ranges for pediatric patients.

CareFlight (CF), developed by Nocera and Garner in 2001 and widely used in Australia, is arguably the simplest of the algorithms with no need for measurements or determination of heart rate or respiratory rate.[11] It is intended for use in all ages and consists of simple yes/no questions with no adjustment for differences in developmental level or physiology of pediatric patients.

The Sacco Triage Method (STM) is also designed for use in all ages. This triage algorithm requires the most calculations,[3,12] a respiratory rate, heart rate, and motor response, followed by an adjustment for age. The STM yields a score of zero to 14 which can be translated into five more traditional categories of Likely Expectant (score of 0-1), Critical (2-4), Compromised/Salvageable (5-8), Delayed/Slow (9-10), and Minor (11+).[6]

Importance and Goals

The goal of this pilot study was to measure IRR of five MCI algorithms in the pediatric population. It was hypothesized that algorithms with the most objective criteria would have the highest IRR and that SALT, the only algorithm incorporating subjective findings, would have the lowest IRR. The secondary aims were to evaluate the inter-rater agreement and inter-MCI algorithm agreement.

Methods

This was a comparative analysis of MCI scoring systems using a pre-existing research database. This study was approved by the Institutional Review Board at the University of California, San Diego (San Diego, California USA; protocol #170467).

Dataset Identification

An existing, de-identified dataset of 253 patients less than 14 years of age who were transported by ambulance and had a trauma team activation to a single Level 1 Pediatric Trauma Center in Los Angeles (California USA) from July 2010 through November 2013 was used. The selected patient dataset was purposely weighted to include a high proportion of those severely injured. Prehospital data abstracted included: age; size including weight, height in centimeters, and/or color designation as measured using a Broselow Tape (2017 Edition; Armstrong Medical; Lincolnshire, Illinois USA); all EMS free-text documentation and EMS interventions performed; ability to walk; level of responsiveness using the AVPU scoring system (A = alert, V = verbal, P = responsive to pain, U = unresponsive); field Glasgow Coma Score (GCS); motor component of the GCS; strength of pulse; heart rate in the field; presence of respiratory distress such as apnea, retractions, grunting, tachypnea, or wheezing; field respiratory rate; and perfusion quality. Poor perfusion was defined as documentation of cool or mottled extremities, capillary refill over three seconds, or explicit EMS documentation of poor perfusion. Given the inability to distinguish whether the patient began breathing with repositioning and five breaths, if bag-mask-ventilation was required during EMS care, apnea and a respiratory rate of zero was assumed. Information collected but not provided to the raters included: gender, injury mechanism, Injury Severity Score (ISS), length-of-stay in days, and disposition from emergency department (ED).

Study Design and Measurements

Training—An attending physician who had completed fellowships in both Pediatric Emergency Medicine and EMS trained three independent raters. Raters were volunteer research assistants: one pediatric resident, one medical student who was also an emergency medical technician (EMT), and one pediatric resident with prior EMT experience. Each rater was separately trained in the five triage algorithms with review and discussion of each algorithm. The attending physician provided a sample set of 25 patients from the full data set. The sample dataset consisted of five randomly selected patients from each triage designation (five Green, five

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press

Triage Method	Able to Walk	Breathes if Airway Opened [a]	Respiratory Rate	Respiratory Quality	Palpable Pulse	Heart Rate	Height or Weight	Age
J-START	X	X	X		X			X
SALT	X	X		X				
PTT	X	X	X		X	X	X	
CF	X	X			X			
STM			X			X		

Fisher © 2022 Prehospital and Disaster Medicine

Table 1. Factors Utilized for Triage in each MCI Algorithm - Part 1
Abbreviations: MCI, mass-casualty incident; J-START, Combined Simple Triage and Rapid Treatment and JumpSTART; SALT, Sort Assess Life-Saving Intervention Treatment; PTT, Pediatric Triage Tape; CF, CareFlight; STM, Sacco Triage Method.
[a] Raters instructed to assume apnea if patient required bag valve mask ventilation.

Triage Method	Obeys Commands	AVPU Score	GCS Motor Response	Major Hemorrhage [a]	Minor Only [a]	Likely to Survive [a]
J-START		X				
SALT	X			X	X	X
PTT						
CF	X					
STM			X			

Fisher © 2022 Prehospital and Disaster Medicine

Table 2. Factors Utilized for Triage in each MCI Algorithm - Part 2
Abbreviations: MCI, mass-casualty incident; J-START, Combined Simple Triage and Rapid Treatment and JumpSTART; SALT, Sort Assess Life-Saving Intervention Treatment; PTT, Pediatric Triage Tape; CF, CareFlight; STM, Sacco Triage Method; AVPU, alert/verbal/responsive to pain/unresponsive; GCS, Glasgow Coma Score.
[a] Raters inferred these subjective factors from other information provided.

Yellow, five Red, and five Black), plus an additional five randomly selected patients. Raters used the sample set for baseline evaluation. A 90% concordance with the attending's triage of the 25 patients was determined to be adequate IRR to proceed to triage of the full dataset. All three raters achieved greater than 90% concordance on each algorithm on the first attempt.

Triage—A Microsoft Excel (Microsoft Office 2011; Microsoft Corp.; Redmond, Washington USA) sheet containing the 253 patients' prehospital information followed by six columns (J-START, SALT, PTT, CF, STM color, and STM number) was given to each rater; patients were in the same order for each rater. Raters were instructed to assign each patient a triage designation in each column. For children eight years of age and younger, JumpSTART was utilized, and for patients over eight years of age, the adult START was utilized for the combined triage algorithm (J-START). Both SALT and STM normally sort patients into five triage categories. For consistency, all five algorithms were collapsed into four possible triage designations: Black, Red, Yellow, or Green. For the SALT triage method, the expectant and dead categories were combined into a Black designation. For STM, likely minor (a score of 11-12) was considered a Green designation; delayed/slow (a score of 9-10) Yellow; compromised/salvageable (a score of 5-8) Red; and critical and likely expectant (0-4) were combined into the Black category. Raters could reference a PDF document that included the five algorithms, including the combined J-START algorithm, instructions on converting STM number into a color category, and a list of vital sign parameters based on height/weight for the PTT. The order in which the triage should be performed was not specified and there was no time limit. Time to complete triage of each patient or the complete dataset was not recorded. Raters did not discuss the process with each other.

Characteristics of the Dataset

Patient age ranged from 25 days to 13 years with a mean age of nine years. There was a wide distribution of age, with the majority in the 9-13 (34%) and 2-to-4-year-old (27%) ranges. The male to female ratio of patients was 2:1. The mechanism was vehicle versus pedestrian or bicycle in over one-third of patients and fall was the mechanism in one-quarter of patients. Almost one-half were discharged home after ED evaluation, 18% were admitted to an intensive care unit without going to the operating suite first, 8.5% went directly to the operating suite, and 10% died. The median ISS was 3.5 with a range of one to 75.

Outcomes and Analysis

The primary outcome measure was the overall IRR for each of the five MCI triage algorithms. Total numbers of Black, Yellow, Red, and Green triage designations assigned by each rater for each algorithm was compared for dyads of raters using Spearman's correlation. The IRR was reported as weighted kappa scores with 95% confidence intervals (CI) for each of the triage algorithms overall. A kappa of 0.81-1.0 was defined as excellent reliability, 0.61-0.8 as

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press

Triage Method	Raters 1 vs 3, 95% CI	Raters 1 vs 2, 95% CI	Raters 2 vs 3, 95% CI	Overall 95% CI Weighted Kappa
J-START	0.98, 0.96-1.0	0.96, 0.93-0.99	0.95, 0.91-0.98	0.91-1.0
SALT	0.57, 0.46-0.68	0.27, 0.19-0.36	0.47, 0.37-0.56	0.19-0.68
PTT	0.96, 0.93-0.99	0.91, 0.86-0.96	0.94, 0.91-0.98	0.86-0.99
CF	0.97, 0.94-1.0	0.94, 0.90-0.98	0.96, 0.93-0.99	0.90-1.0
STM	0.95, 0.88-1.0	0.90, 0.81-1.0	0.89, 0.79-1.0	0.79-1.0

Fisher © 2022 Prehospital and Disaster Medicine

Table 3. Spearman Correlation Inter-Rater Reliability between Triage Methods
Abbreviations: J-START, Combined Simple Triage and Rapid Treatment and JumpSTART; SALT, Sort Assess Life-Saving Intervention Treatment; PTT, Pediatric Triage Tape; CF, CareFlight; STM, Sacco Triage Method.

Triage Method	% Black (mean)	% Red (mean)	% Yellow (mean)	% Green (mean)
J-START	3.2-3.6 (3.3)	5.7-8.5 (7.4)	28.7-31.2 (29.8)	59.1-59.5 (59.2)
SALT	3.2-4.1 (3.5)	8.5-9.3 (9.0)	36.0-75.7 (57.7)	11.7-50.6 (29.7)
PTT	3.2-4.1 (3.5)	4.5-6.1 (5.2)	31.2-33.2 (32.0)	58.7-59.9 (59.4)
CF	3.2-4.1 (3.5)	6.5-8.9 (8.0)	28.7-29.6 (29.2)	58.7-59.9 (59.4)
STM	3.2-3.2 (3.2)	1.6-2.0 (1.9)	3.6-4.5 (4.2)	90.3-91.1 (90.6)
Triage Color	Triage Algorithm	Rater 1	Rater 2	Rater 3
Green	J-START	146 (59.1%)	147 (59.5%)	147 (59.1%)
	SALT	29 (11.7%)	125 (50.6%)	66 (26.7%)
	PTT	145 (58.7%)	148 (59.9%)	147 (59.5%)
	CF	145 (58.7%)	148 (59.9%)	147 (59.5%)
	STM	225 (91.1%)	223 (90.3%)	224 (90.3%)
Yellow	J-START	73 (29.6%)	77 (31.2%)	71 (28.7%)
	SALT	187 (75.7%)	89 (36.0%)	152 (61.5%)
	PTT	82 (33.2%)	78 (31.6%)	77 (31.2%)
	CF	72 (29.2%)	73 (29.6%)	71 (28.7%)
	STM	9 (3.6%)	11 (4.5%)	11 (4.5%)
Red	J-START	20 (8.1%)	14 (5.7%)	21 (8.5%)
	SALT	23 (9.3%)	23 (9.3%)	21 (8.5%)
	PTT	12 (4.9%)	11 (4.5%)	15 (6.1%)
	CF	22 (8.9%)	16 (6.5%)	21 (8.5%)
	STM	5 (2.0%)	5 (2.0%)	4 (1.6%)
Black	J-START	8 (3.2%)	9 (3.6%)	8 (3.2%)
	SALT	8 (3.2%)	10 (4.1%)	8 (3.2%)
	PTT	8 (3.2%)	10 (4.1%)	8 (3.2%)
	CF	8 (3.2%)	10 (4.1%)	8 (3.2%)
	STM	8 (3.2%)	8 (3.2%)	8 (3.2%)

Fisher © 2022 Prehospital and Disaster Medicine

Table 4. Triage Designation by MCI Algorithm and Rater
Abbreviations: MCI, mass-casualty incident; J-START, Combined Simple Triage and Rapid Treatment and JumpSTART; SALT, Sort Assess Life-Saving Intervention Treatment; PTT, Pediatric Triage Tape; CF, CareFlight; STM, Sacco Triage Method.

substantial, 0.41-0.6 as moderate, 0.21-0.4 as fair, and 0-0.2 as poor.

Secondary outcome measures were the inter-rater agreement (unadjusted proportion of patients for which the raters agreed on the triage level for each MCI algorithm) and inter-MCI algorithm agreement (the proportion of patients who were triaged to the same color by each MCI triage algorithm). Descriptive statistics were used to report inter-rater agreement and the inter-MCI algorithm agreement.

Results

Of the 253 patients in the dataset, six were excluded because they lacked one or more triage designations by one of the raters. Two-hundred forty-seven patients were included in the analysis.

The IRR for the five MCI algorithms is presented in Table 3. Overall, J-START and CF had the highest reliability with a kappa of 0.94 or higher (0.9-1.0, 95% CI for overall weighted kappa). Reliability for STM varied from substantial

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press

to excellent (0.79-1.0) and SALT varied from poor to substantial (0.19-0.68).

The least variability was found in patients triaged to the Black category; only two patients who were categorized as Black at least once did not have uniform agreement across all algorithms. Yellow and Green categories had the highest variability among raters. Table 4 details the proportion of patients triaged into each color within each MCI algorithm. In total, STM categorized 90.6% of patients as Green; SALT had the highest number of patients in the Red and Yellow categories, the least in the Green category, and the highest variability between raters in the Yellow and Green categories. There was less than 2.5% variability between all three raters for all algorithms except for SALT (Figure 1). Across algorithms, the Yellow designation had the greatest variability (0.9%-39.7%), followed by Green (0.4%-38.9%; Figure 2). Specific designations per rater for each algorithm are listed in Table 4.

The lowest inter-rater agreement was seen in application of the SALT algorithm with the three raters having 100% agreement in only 53% of patients (Table 5). Inter-rater agreement was high with CF, J-START, STM, and PTT algorithms with 100% agreement between raters in 96%-98% of cases. The three raters had 100% agreement for 93.9% of patients when applying the PTT algorithm. The PTT algorithm was the only algorithm with complete disagreement agreement for any patients; no raters agreed on the triage category for two patients (0.8%).

There were several clusters of triage color cohorts across the algorithms. There was a cohort of eight patients who were triaged Black across all five MCI algorithms by all raters. There was a "Red Cohort"- 20 patients primarily categorized as Red among J-START, SALT, and CF; a "Yellow Cohort"- 73 primarily categorized Yellow by J-START, SALT, PTT, and CF; and a "Green Cohort"- 146 primarily categorized Green among J-START, SALT, PTT, CF, and STM. Within the color cohorts, several patterns of disagreement were identified and these are listed in Figure 3. Both STM and PTT had a high level of disagreement in the "Red Cohort;" SALT had the highest level of disagreement among the "Yellow Cohort" and "Green Cohort."

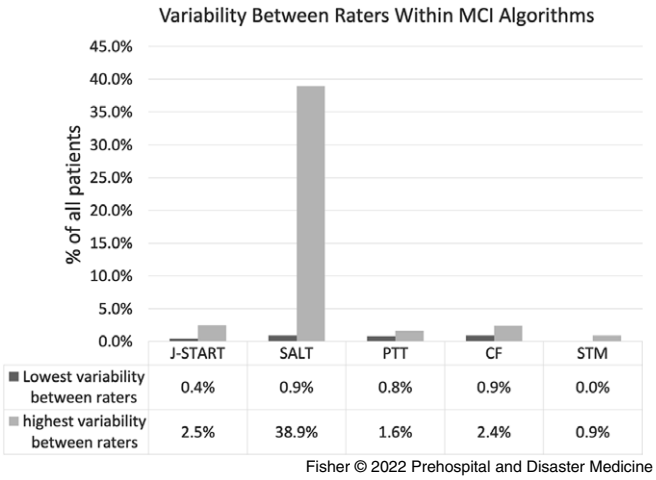


Figure 1. Variability between Raters within MCI Algorithm.
Abbreviations: MCI, mass-casualty incident; J-START, Combined Simple Triage and Rapid Treatment and JumpSTART; SALT, Sort Assess Life-Saving Intervention Treatment; PTT, Pediatric Triage Tape; CF, CareFlight; STM, Sacco Triage Method.

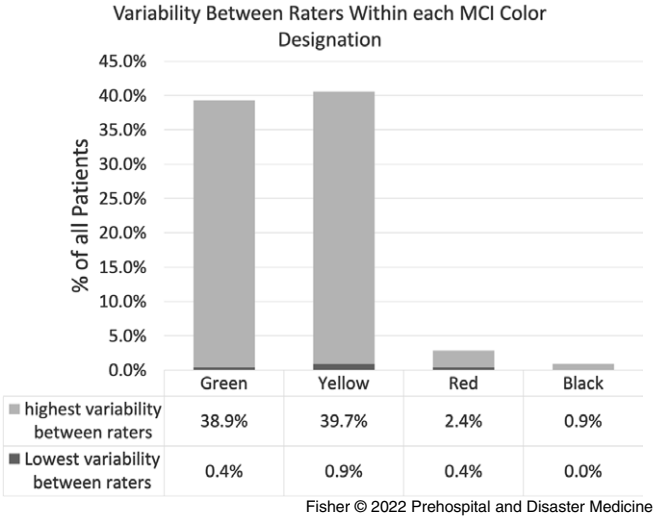


Figure 2. Variability between Raters within Each MCI Color Designation.
Abbreviation: MCI, mass-casualty incident.

Discussion

An ideal MCI algorithm would have a high IRR, being consistent regardless of who is applying the tool. It had been hypothesized that an increase in subjective components of a triage algorithm would lead to a decrease in the IRR of that algorithm. In this pilot study applying MCI algorithms to a retrospective cohort of pediatric trauma patients, the algorithm with the most subjective components, SALT, had the lowest IRR. In contrast, J-START, PTT, CF, and STM all had excellent IRR.

Both CF and J-START had similar inter-rater agreement and inter-MCI algorithm agreement across all triage colors. Overall, PTT had the greatest variance in the Red Cohort but performed similarly across groups in the Black, Yellow, and Green Cohorts; SALT performed similarly within Black and Red Cohorts, but stood out as highly variable within Yellow and Green Cohorts. Lastly, STM performed well among Black and Green Cohorts, but had substantial differences among the Yellow Cohort and had a high variability in the Red Cohort.

In comparison to these findings of high IRR in the J-START algorithm, a research abstract published in *Annals of Emergency Medicine* found moderate IRR (kappa = 0.44) in emergency physicians versus fair in pediatricians (0.38 kappa) triaging 11 pediatric patients using JumpSTART.[8] A 2013 study of 37 resident trainees applying JumpSTART to a hands-on mannequin simulation of a 10-patient MCI found IRR to be generally higher than 0.70 with exception of one ambulatory patient. Additionally, there was no difference in triage by post-graduate year.[9] This tabletop study found that J-START has an excellent IRR with minimal variance and strong agreement between the three raters, but results may have differed in a hands-on simulation. One possibility is that variability was introduced by inconsistency in the assessment and interpretation of physical

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press

	START (n, % of all patients)		SALT (n, % of all patients)		PTT (n, % of all patients)		CF (n, % of all patients)		STM (n, % of all patients)	
All Same Across Raters	237	96.0%	131	53.0%	232	93.9%	237	96.0%	242	98.0%
Two Raters Agree	10	4.0%	116	47.0%	15	6.1%	10	4.0%	5	2.0%
No Raters Agree	0	0.0%	0	0.0%	2	0.8%	0	0.0%	0	0.0%
Total Patients	247	100.0%	247	100.0%	247	100.0%	247	100.0%	247	100.0%

Fisher © 2022 Prehospital and Disaster Medicine

Table 5. Inter-Rater Agreement
Abbreviations: START, Simple Triage and Rapid Treatment; SALT, Sort Assess Life-Saving Intervention Treatment; PTT, Pediatric Triage Tape; CF, CareFlight; STM, Sacco Triage Method.

findings. Of mention, CF has not previously been evaluated but performed just as well as J-START in this study.

The SALT algorithm was found to have the highest variability, particularly among the Green and Yellow categorizations. This finding was not surprising as the branch point between the Green and Yellow triage classification was raters' response to the question: "minor injuries only?" As SALT is the algorithm most adherent with the Model Uniform Core Criteria for MCI triage, and is the algorithm being promoted at a national level in the United States, determining how users apply the algorithm is vital.[10] A provider performing triage who classifies only a few injuries as minor will lead to more transports while a provider who classifies more injuries as minor may lead to more Green categorizations, potentially resulting in under-triage. Regardless, the inconsistency between raters seen in this pilot study is concerning.

As a proprietary system with a numerical output intended to prioritize resources, STM has been found to predict mortality well in prior studies.[3,12] The age-adjusted STM was used even though that version has not been demonstrated to improve mortality prediction in the pediatric population.[13] No prior studies have evaluated how STM compares to other disaster algorithms when the STM number is split into triage color categories. Zero-to-four was used as a cut off for Black in the STM and found that cut point to have a high level of agreement with Black categorization in all four of the other algorithms. In comparing STM categorization to the other algorithms, an overwhelming number of patients in STM were categorized as Green, including all of the patients triaged as Yellow in J-START, SALT, and CF and some who were triaged as Red by a majority of the other algorithms. While this study does not look at the inherent accuracy of an algorithm, or the accuracy of the rater applying the algorithm, the fact that STM has a remarkably higher proportion of Green patients compared to other algorithms suggests the possibility of under-triage. The clinical significance of this under-triage when STM is applied to a pediatric population includes non-detection of significant illness and injuries.

An area of particular interest that deserves further study is the 20 patients who were found by J-START, SALT, and CF to be primarily triaged as Red (ie, the Red Cohort). In this cohort, the algorithms with the most objective criteria, STM and PTT, had the least amount of inter-rater and inter-algorithm agreement. This may be in part attributed to the limitations of only using the documented first set of prehospital vital signs as the objective data for MCI algorithm triage. For example, in a five-year-old patient with a gunshot wound to the head with flexural posturing requiring bag-mask-ventilation, the patient had a documented heart rate of 120 and respiratory rate of 33. Given these normal for age vital signs, despite a catastrophic injury, both the STM and PTT triage the patient as a Yellow. Additionally, the PTT algorithm does not take mental status into account, which may account for a large proportion of the patients triaged as Yellow by PTT and Red by other algorithms.

Limitations

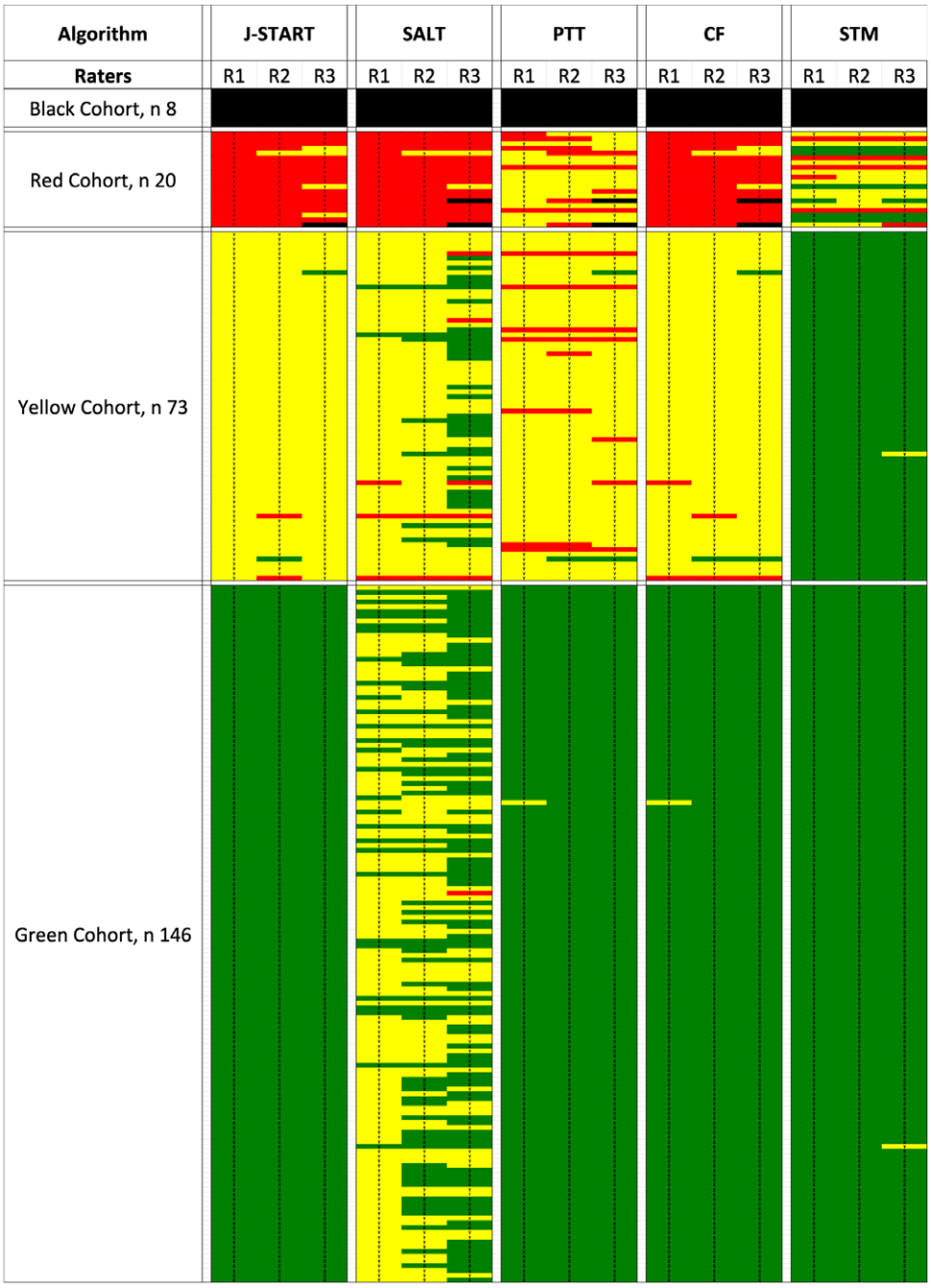
This is a pilot study to allow comparison of five triage systems in a known dataset of trauma patients. There are a number of limitations that can be addressed with further study of selected triage methods in a more realistic simulation. First and foremost, the reliability of the algorithms in a high-stress environment and the timeliness of application were not analyzed. This study evaluated disaster algorithms utilizing a retrospective dataset in a tabletop drill. Raters did not assess the patients themselves and were not limited in time. This created an artificial environment without the pressure of physical triage.

The visual presentation of the data may have influenced the IRR within and between some of the systems. In this study, the raters were provided a data sheet with patients' prehospital data; raters filled in six columns for each of the triage algorithms (including STM color and number). This led to an artificial environment and the potential to replicate the same triage category across algorithms. Similarly, those systems with redundant elements would naturally score closer to each other if raters were scoring them consecutively. As the order of triage completion and time to triage were not controlled in this study, it cannot be determined to what extent these elements may have affected the outcomes.

The raters were provided with actual prehospital vitals along with the documented paramedic narrative of findings and treatments. However, response to repositioning of jaw and administration of rescue breaths was not consistently documented. The raters were taught to assume apnea if respiratory rate was recorded as zero or if patient required assisted ventilation. However, in the only two patients with inter-rater differences between Red and Black, both patients received assisted ventilation but a respiratory rate was documented for both. This is likely a systemic data entry issue as presumably paramedics are entering the respiratory rate at which they are bagging the patient. If true, use of large datasets, such as the National Emergency Medicine Services Information System or the National Trauma Data Bank, to evaluate MCI algorithms may have similar limitations.

Finally, forcing the STM from its numerical designations into the four categories required by the other algorithms may have

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press



Fisher © 2022 Prehospital and Disaster Medicine

Figure 3. Inter-MCI Algorithm Agreement by Color Cohort.
Abbreviations: MCI, mass-casualty incident; J-START, Combined Simple Triage and Rapid Treatment and JumpSTART; SALT, Sort Assess Life-Saving Intervention Treatment; PTT, Pediatric Triage Tape; CF, CareFlight; STM, Sacco Triage Method; R1, Rater 1; R2, Rater 2; R3, Rater 3.

affected the IRR of that method, if the score fell above or below the cutoff for color. Comparing individual scores demonstrated the least variability with this triage method. Changing the cutoff points for Yellow or Red designations may have altered the IRR between this method and the others. These may require re-examination for the pediatric population as most of the patients in this data set were triaged Green.

Conclusions

In this study, J-START and CF were found to have the highest IRR, highest inter-rater agreement, and highest level of agreement within MCI algorithms across all triage colors. Overall, J-START, PTT, and CF were found to have excellent IRR. Predictably, SALT was found to have the lowest IRR, particularly in the Yellow and Green designations, but it performed well

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press

in Red and Black categories. As had been hypothesized, the algorithm with the subjective components, SALT, had the highest rate of variability, and the most concrete objective algorithm, STM, had the least variability. Future studies should explore IRR and agreement in a hands-on simulation of these MCI algorithms.

References

1. Koziel JR, Meckler G, Brown L, et al. Barriers to pediatric disaster triage: a qualitative investigation. *Prehosp Emerg Care.* 2014;19(2):279–286.
2. Wallis LA, Carley S. Comparison of pediatric major incident primary triage tools. *Emerg Med J.* 2006;23(6):475–478.
3. Cross KP, Cicero MX. Head-to-head comparison of disaster triage methods in pediatric, adult, and geriatric patients. *YMEM.* 2013;61(6):668–676.
4. Heffernan R, Lerner EB, McKee C, et al. Comparing the accuracy of mass casualty triage systems in a pediatric population. *Prehosp Emerg Care.* 2019;23(3):304–308.
5. 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.
6. Jenkins JL, McCarthy ML, Sauer LM, et al. Mass-casualty triage: time for an evidence-based approach. *Prehosp Disaster Med.* 2008;23(1):3–8.
7. Nadeau NL, Cicero MX. Pediatric disaster triage system utilization across the United States. *Pediatr Emerg Care.* 2017;33(3):152–155.
8. DeGeorge L, Glick R, Ackerman A, Wu Y. Pediatric triage decision making in a large-scale pediatric disaster. *Ann Emerg Med.* 2015;66(4):S143–S144.
9. Cicero MX, Riera A, Northrup V, Auerbach M, Pearson K, Baum CR. Design, validity, and reliability of a pediatric resident JumpSTART disaster triage scoring instrument. *Acad Pediatr.* 2013;13(1):48–54.
10. Lerner EB, Schwartz RB, Coule PL, et al. Mass casualty triage: an evaluation of the data and development of a proposed national guideline. *Disaster Med Pub Health Prep.* 2008;2 Suppl 1;S25–S34.
11. Nocera A, Garner A. An Australian mass casualty incident triage system for the future based upon triage mistakes of the past: the Homebush Triage Standard. *Aust NZ J Surg.* 1999;69:603–608.
12. 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.
13. Sacco WJ, Navin DM, Fiedler KE, Waddell RK II, Long WB, Buckman RF. Precise formulation and evidence-based application of resource-constrained triage. *Acad Emerg Med.* 2005;12(8):759–770.

https://doi.org/10.1017/S1049023X22000632 Published online by Cambridge University Press