


METASTART: A Systematic Review and Meta-Analysis of the Diagnostic Accuracy of the Simple Triage and Rapid Treatment (START) Algorithm for Disaster Triage

Jeffrey M. Franc, MD, MSc, MS, FCFP.EM, Dip Sport Med;¹  Scott W. Kirkland, MS;¹ Uirá D. Wisnesky, PhD;¹ Sandra Campbell, MLS;² Brian H. Rowe, MD, MSc, FCFP(EM), FCCP, FCAHS^{1,3}

1. Department of Emergency Medicine, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada
2. J.W. Scott Health Sciences Library, University of Alberta, Edmonton, Alberta, Canada
3. School of Public Health, University of Alberta, Edmonton, Alberta, Canada

Correspondence:

Jeffrey M. Franc, MD
Department of Emergency Medicine
University of Alberta
Room 736c, University Terrace 8203-112
Street NW
Edmonton, AB, Canada, T6G 2T4
Email: jeffrey.franc@ualberta.ca

Conflicts of interest/funding: Scoping and Systematic Review Grant from the Emergency Strategic Clinical Network (ESCN) at Alberta Health Services and the Emergency Medicine Research Group (EMeRG) in the Department of Emergency Medicine at the University of Alberta. Dr. Rowe's research is supported by a Scientific Director's Grant (SOP 168483) from the Canadian Institutes of Health Research (CIHR; Ottawa, ON). The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the funding agencies. The funders had no role in the design, implementation, analysis, and write-up of the study. Dr. Franc is CEO and Founder of Stat59. The remaining authors have no conflicts to declare.

Keywords: disaster medicine; mass-casualty; incidents; meta-analysis; START; triage

Abbreviations:

MCI: mass-casualty incident
MMAT: Mixed Methods Appraisal Tool
RCT: randomized controlled trial

Abstract

Introduction: The goal of disaster triage at both the prehospital and in-hospital level is to maximize resources and optimize patient outcomes. Of the disaster-specific triage methods developed to guide health care providers, the Simple Triage and Rapid Treatment (START) algorithm has become the most popular system world-wide. Despite its appeal and global application, the accuracy and effectiveness of the START protocol is not well-known.

Objectives: The purpose of this meta-analysis was two-fold: (1) to estimate overall accuracy, under-triage, and over-triage of the START method when used by providers across a variety of backgrounds; and (2) to obtain specific accuracy for each of the four START categories: red, yellow, green, and black.

Methods: A systematic review and meta-analysis was conducted that searched Medline (OVID), Embase (OVID), Global Health (OVID), CINAHL (EBSCO), Compendex (Engineering Village), SCOPUS, ProQuest Dissertations and Theses Global, Cochrane Library, and PROSPERO. The results were expanded by hand searching of journals, reference lists, and the grey literature. The search was executed in March 2020. The review considered the participants, interventions, context, and outcome (PICO) framework and followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Accuracy outcomes are presented as means with 95% confidence intervals (CI) as calculated using the binomial method. Pooled meta-analyses of accuracy outcomes using fixed and random effects models were calculated and the heterogeneity was assessed using the I^2 statistic.

Results: Thirty-two studies were included in the review, most of which utilized a non-randomized study design (84%). Proportion of victims correctly triaged using START ranged from 0.27 to 0.99 with an overall triage accuracy of 0.73 (95% CI, 0.67 to 0.78). Proportion of over-triage was 0.14 (95% CI, 0.11 to 0.17) while the proportion of under-triage was 0.10 (95% CI, 0.072 to 0.14). There was significant heterogeneity of the studies for all outcomes ($P < .0001$).

Conclusion: This meta-analysis suggests that START is not accurate enough to serve as a reliable disaster triage tool. Although the accuracy of START may be similar to other models of disaster triage, development of a more accurate triage method should be urgently pursued.

Franc JM, Kirkland SW, Wisnesky UD, Campbell S, Rowe BH. METASTART: a systematic review and meta-analysis of the diagnostic accuracy of the Simple Triage and Rapid Treatment (START) algorithm for disaster triage. *Prehosp Disaster Med.* 2022;37(1):106–116.

START: Simple Triage and Rapid Treatment
VR: virtual reality

Received: August 23, 2021

Accepted: September 20, 2021

doi:[10.1017/S1049023X2100131X](https://doi.org/10.1017/S1049023X2100131X)

© The Author(s), 2021. Published by Cambridge University Press on behalf of the World Association for Disaster and Emergency Medicine.

Introduction

Resources to assist in a surge of critically ill or injured individuals during a disaster or mass-causality incident (MCI) may be limited, creating a discrepancy between needs and available resources.¹ Unlike triage within the hospital, disaster triage must perform the delicate task of balancing the need of each patient for medical care and the overarching needs of the health care system to avoid overwhelming health care capacity. This can be an unusual and frightening event for health care providers who often respond to disasters working in unfamiliar conditions and are forced to make uncomfortable, morally challenging, and critical clinical care decisions with minimal information.

To address these concerns, numerous disaster triage algorithms have been developed. Their goals are to provide an easily learnable and consistent tool to aid in these difficult decisions. Many are based on simple flow charts with the goal of facilitating difficult clinical decisions during times of provider stress while ensuring that victims of a disaster or MCI are prioritized based on their clinical needs. The ability of these disaster triage tools to accurately triage victims is important. Under-triage (poor sensitivity) can result in a failure to recognize victims who could benefit from urgent medical intervention. Conversely, over-triage (poor specificity) results in valuable resources being used prematurely or unnecessarily.

To assist in the triage of victims of a disaster or MCI, the Simple Triage and Rapid Treatment (START) tool was introduced in 1983 by the Newport Beach Fire and Marine Department (Newport Beach, California USA) and the Hoag Hospital (Orange County, California USA) using the now familiar groups of black (expectant), red (immediate), yellow (delayed), and green (minor) to prioritize care of disaster victims.² As the tool is based on a simple flowchart, it has been targeted to providers of all groups (ie, physicians, nurses, trainees, and prehospital providers) as a simple way to promote consistent and reproducible triage.

Despite being one of the most commonly used and studied disaster triage systems world-wide, there are no published syntheses on the accuracy of the START tool. The purpose of this meta-analysis was to assess overall accuracy, as well as the proportion of under- and over-triage, for the START method when used by providers across a variety of backgrounds. In addition, specific estimates of accuracy were obtained for each of the four START categories: red, yellow, green, and black.

Methods

A study protocol was developed “a priori” and registered on PROSPERO (registration # CRD42020175457) to define the objectives, selection criteria, data collection, and analysis. This review conforms to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.³

Differences between Protocol and Review

Several amendments were made to the initial protocol. First, the initial plan was to employ the following tools to access the risk of bias: (1) Cochrane risk of bias tool for studies enrolling patients in a true disaster, randomized controlled trials (RCTs); (2) before-after quality assessment checklist for before-after studies; (3) Newcastle-Ottawa Scale (NOS) for observational cohort/case-control studies; and (4) the Simulation Research Evaluation Rubric for simulation-based MCI studies.^{4–8} However, all collaborators agreed before the implementation that it was more appropriate to use the Mixed Methods Appraisal Tool (MMAT; Montreal, Canada) as it is a single tool that allows for the

assessment of multiple study designs.^{9,10} Second, several sub-group and sensitivity analyses proposed in the protocol will now be reported elsewhere. Third, due to infrequent and incomplete reporting, other proposed accuracy outcomes including specificity, sensitivity, likelihood ratios, as well as reliability/validity were not assessed as planned; the review instead focused on assessing triage accuracy as well as over- and under-triage outcomes.

Inclusion/Exclusion

To identify all relevant studies on the topic, the following inclusion criteria were used.

Population—Adults (≥ 17 years of age) being triaged in or out of hospital during a true or simulated disaster or MCI.

Intervention—Use of START triage assessment, either in a true or simulated disaster or MCI.

Comparator—Studies consisting of a single or multiple groups were eligible for inclusion, including those comparing the accuracy of START to a reference standard (eg, case examples that are defined through consensus), other triage tools, different triage assessors, or different disaster scenarios.

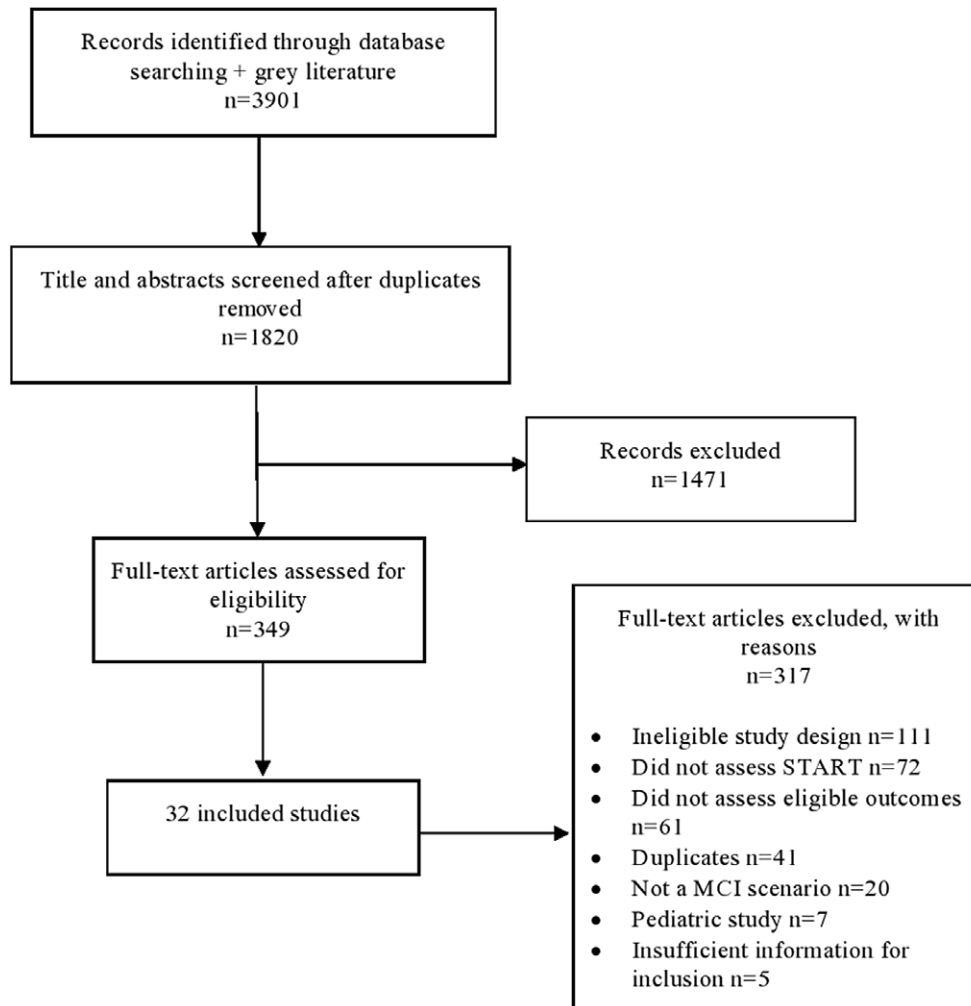
Outcomes—Studies were required to report the diagnostic accuracy of START to be included in the review. The primary outcome of the review was the triage accuracy of START, which was defined as the proportion of cases triaged to the correct category when compared to the reference standard. The overall triage accuracy of START was reported, as well as the accuracy of START based on the triage sub-groups (black, red, yellow, and green). Secondary outcomes of interest included over- and under-triage.

Study Design—Studies described as RCTs or non-RCTs, cohort descriptive studies, or mixed methods studies were eligible for inclusion. Reviews, editorials, and commentaries were not included.

There were no exclusion criteria based on who conducted the triage assessment using START. Studies that strictly reported the accuracy of a modified version of START were not eligible to be included. No other limits regarding publication status or language were applied.

Search Methods for Identification of Studies

A search was executed by an expert health librarian (SC) on the following databases: OVID Medline (US National Library of Medicine, National Institutes of Health; Bethesda, Maryland USA); OVID EMBASE (Elsevier; Amsterdam, Netherlands); OVID Global Health (EBSCO Information Services; Ipswich, Massachusetts USA); EBSCO CINAHL (EBSCO Information Services; Ipswich, Massachusetts USA); Compendex (Engineering Village [Elsevier; Amsterdam, Netherlands]); SCOPUS (Elsevier; Amsterdam, Netherlands); ProQuest Dissertations and Theses Global (Ann Arbor, Michigan USA); Cochrane Library (The Cochrane Collaboration; London, United Kingdom); and PROSPERO (University of York; York, United Kingdom) using controlled vocabulary (eg, MeSH [Medical Subject Headings] and Emtree [Embase]) and keywords representing the concepts “START” and “triage” and “mass casualties.” The searches were complete up through March 2020 and search strategies were adjusted appropriately for different databases. For primary databases, searches were limited to 1983 through March 2020. Results of the searches were exported to the citation management system RefWorks (Version 2.1.0.1; ProQuest, LLC; Ann Arbor,



Franc © 2022 Prehospital and Disaster Medicine

Figure 1. PRISMA Flow Diagram.

Abbreviations: START, Simple Triage and Rapid Treatment; MCI, mass-casualty incident.

Michigan USA) and also exported to the COVIDENCE systematic review program (Veritas Health Innovation Ltd; Melbourne, Australia). Detailed search strategies are available in Appendix 1 (available online only).

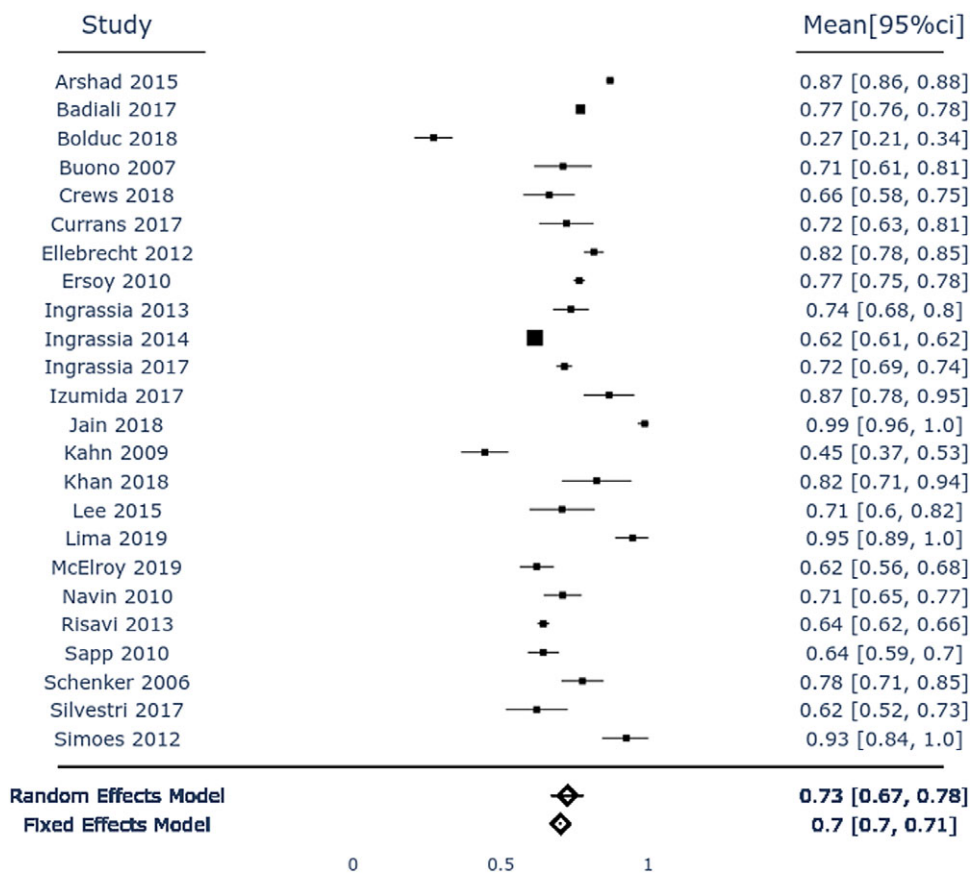
Additional searches of the grey literature were conducted to identify any studies missed from the search of the databases. Sources of the grey literature that were searched included Google Scholar (Google Inc.; Mountain View, California USA); clinical trial registries (clinicaltrials.gov, Cochrane Central Register of controlled trials [The Cochrane Collaboration; London, United Kingdom], and controlled-trials.com); Web of Science (Thomson Reuters; New York, New York USA); backward and forward SCOPUS searches of included studies; bibliographies from included studies and known reviews; as well as hand-searching of abstracts from emergency medicine conferences, including Canadian Association of Emergency Medicine (2018-2020; Ottawa, Ontario Canada), Society of Academic Emergency Medicine (2018-2020; Des Plaines, Illinois USA), and American College of Emergency Physicians (2018-2020; Irving, Texas USA). When possible, search results were imported into EndNote (Clarivate Analytics; Philadelphia, Pennsylvania

USA) before being exported into the COVIDENCE systematic review program.

Study Selection and Data Extraction

Potentially eligible studies were selected using a two-stage screening process. At the first stage, titles and abstracts of all studies identified in the literature search were screened by two independent reviewers (UDW, SWK) against pre-determined inclusion and exclusion criteria to identify potentially eligible studies. At the second stage, the full-text manuscripts of all studies identified as potentially eligible by at least one of the reviewers were retrieved and reviewed for eligibility by two independent reviewers (UDW, SWK) using the pre-defined eligibility criteria. Disagreements between the reviewers regarding the eligibility of studies was mediated via a third-party adjudicator (JMF).

The data of all included studies were extracted independently by two of three available reviewers (UDW, SWK, JM). Pre-defined outcomes were extracted using standardized extraction forms. The completed data extraction was then verified for accuracy by the third reviewer (either UDW, SWK, or JM) who did not complete the initial assessment. Disagreements that could not be settled



Franc © 2022 Prehospital and Disaster Medicine

Figure 2. Overall Triage Correct.

via discussion between the reviewers were mediated by a third-party adjudicator (JMF). Outcomes regarding study characteristics, characteristics of simulation or MCI, implementation of START, description of the reference standard, and primary/secondary outcomes of interest were extracted onto standardized forms.

Quality Appraisal

The MMAT was employed to assess the quality of the included studies.^{9,10} Two reviewers (SK, UDW) independently evaluated the MMAT level of evidence for each article and completed a data extraction table. Discrepancies were resolved through discussion.

Data Synthesis

Accuracy outcomes including triage accuracy, over-triage, and under-triage are presented as means with 95% confidence interval (CI) as calculated by the binomial method.¹¹ A meta-analysis for the outcomes of triage accuracy, over-triage, and under-triage was performed using Stat59 (Build73a7cf; Stat59 Services Ltd.; Edmonton, Alberta Canada). Pooled estimates were calculated using the DerSimonian and Laird method for the random effects model and the inverse variance method for the fixed effects model.¹² The heterogeneity statistic Q was calculated by the inverse variance method and P values for heterogeneity were based on the chi-square test of Q.¹² A sensitivity analysis of the primary outcome (overall triage accuracy) based on the inclusion of studies in which triage accuracy had to be imputed due to incomplete outcome reporting was also completed.

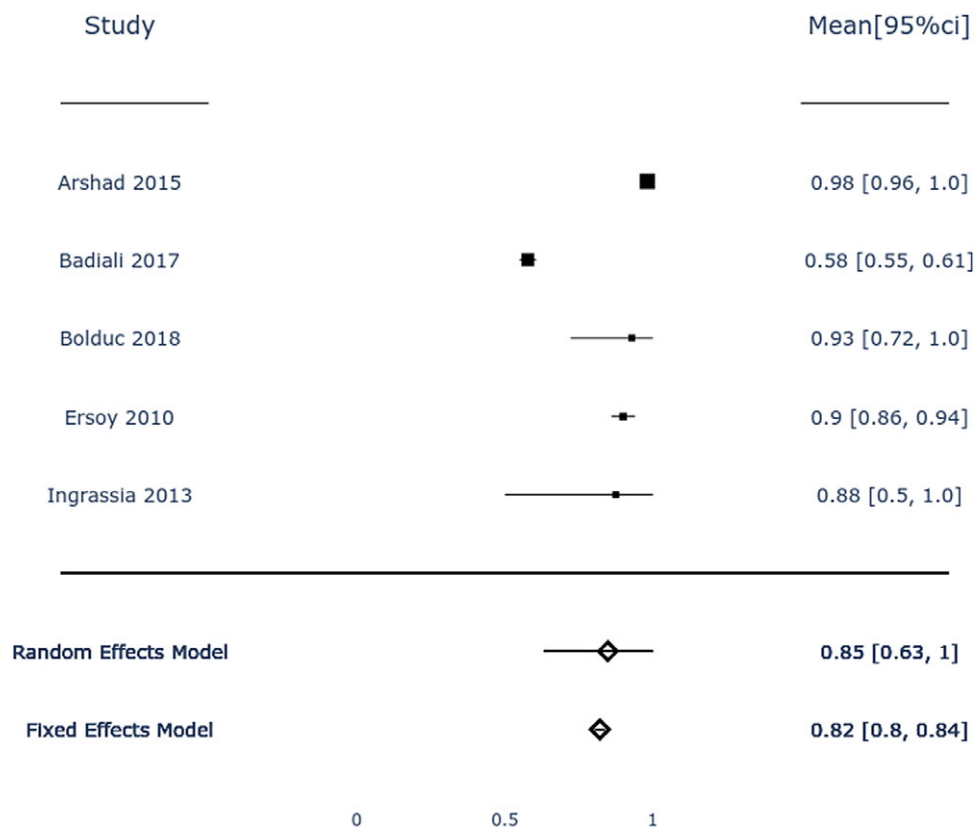
Results

Search Results

A total of 3,901 articles were identified in the search of the electronic databases and grey literature (Figure 1). After duplicates were removed, the title and abstracts of 1,820 studies were reviewed. A total of 1,471 studies were excluded for irrelevance and 349 were identified as potentially eligible. After full-text review, 317 studies were excluded for various reasons including ineligible study design (n = 111), did not assess START (n = 72), did not report eligible outcomes (n = 61), duplicate publications (n = 41), not set in a disaster/MCI (n = 20), included pediatric victims (n = 7), and in some cases, the full text could not be retrieved (n = 5). As a result, a total of 32 studies were included in the review (Table 1¹³⁻⁴⁴; available online only).

Study Characteristics

A descriptive analysis of these 32 studies will be published as a scoping review elsewhere and a summary of the characteristics of the studies are provided in Table 1.⁴⁵ Briefly, one-half of the studies were based in North America, with the remaining studies being conducted in various countries across Europe, Asia, the Middle East, and Oceania. The majority of studies involved triage of simulated patients. Based on the classification of simulation techniques developed by Alinier, the majority of studies used a Level 3 (live) simulation involving actors in moulage playing victims in various MCI scenarios.⁴⁶ The next most common simulation type was paper-based exercises (Level 0) or computer-based simulations including the use of virtual reality (VR) simulations (Level 2).



Franc © 2022 Prehospital and Disaster Medicine

Figure 3. Category BLACK Triage Correct.

The simulations most commonly involved motor vehicle collisions with some studies simulating a collision of trains or airplanes. Eleven studies did not specify the nature of the disaster or MCI simulated. Participants who completed the triage assessments included: paramedics; firefighters and other first responders; physicians; nurses; and students. All but two studies assessed the accuracy of START by comparing to a reference standard.^{38,39} However, the majority of studies ($n = 22$) did not specify the reference standard used when assessing the accuracy of START. When specified, the reference standard was most commonly expert opinion.^{15,16,25,30,32,36,40,41,47}

Quality of Studies

Five studies were assessed using the MMAT quality criteria for quantitative RCTs (Appendix 2; available online only).^{14,18,20–22} Across the studies, the method of randomization was not clear.^{14,18,20–22} In addition, in several studies, blinding was not adequately described.^{14,20–22} Assessments using MMAT, however, did identify complete outcome reporting across all five studies and that participants did adhere to the assigned groups.

Seventeen studies were assessed using the MMAT quality criteria for quantitative non-RCTs which included before-after studies and comparative observational cohorts (Appendix 3; available online only).^{13,15–17,19,24,25,27,28,30,32,34,36,39,41,43,48} The majority of the studies did not provide adequate description of the study participants or how they were included in the study. The majority of studies utilized appropriate measurements and had complete outcome reporting.

Eight studies were assessed using the MMAT quality criteria for quantitative descriptive studies, which consisted of a single cohort

(Appendix 4; available online only).^{23,31,33,35,37,38,42,44} The majority of these studies reported a clear research question, relevant sampling strategy, and had low risk for nonresponse bias. Several studies, however, did not report how their statistical analysis was completed and several provided insufficient information on the study participants.

Finally, two studies were assessed using the MMAT quality criteria for mixed methods studies (Appendix 5; available online only).^{26,40} Overall, the quality of the two studies varied with one study judged as not effectively integrating the mixed methods approach to answer their research question and having low quality for the quantitative and qualitative components.⁴⁰

Primary Outcome

Overall Triage Accuracy—Twenty-four studies included the absolute numbers for overall correct triage and form the analysis group for the main study outcome (Figure 2). The reported proportion for correct triage in the included studies ranged from 0.27 to 0.99. Using the random effects model, the pooled estimate for proportion of correct triage was 0.73 (95% CI, 0.67 to 0.78). The pooled estimate was 0.70 (95% CI, 0.70 to 0.71) using the fixed effect model. There was significant heterogeneity among the included studies ($P < .0001$).

Five studies contained adequate information to be included in the assessment of accuracy of the triage category **black** (Figure 3). The proportion of cases correctly triaged ranged from 0.58 to 0.98. The pooled estimate for the proportion of victims correctly triaged was 0.85 (95% CI, 0.63 to 1.0) using the random effects model. Using the fixed effect model, the accuracy was 0.82 (95% CI, 0.80 to 0.84). There was significant heterogeneity among studies ($P < .0001$).

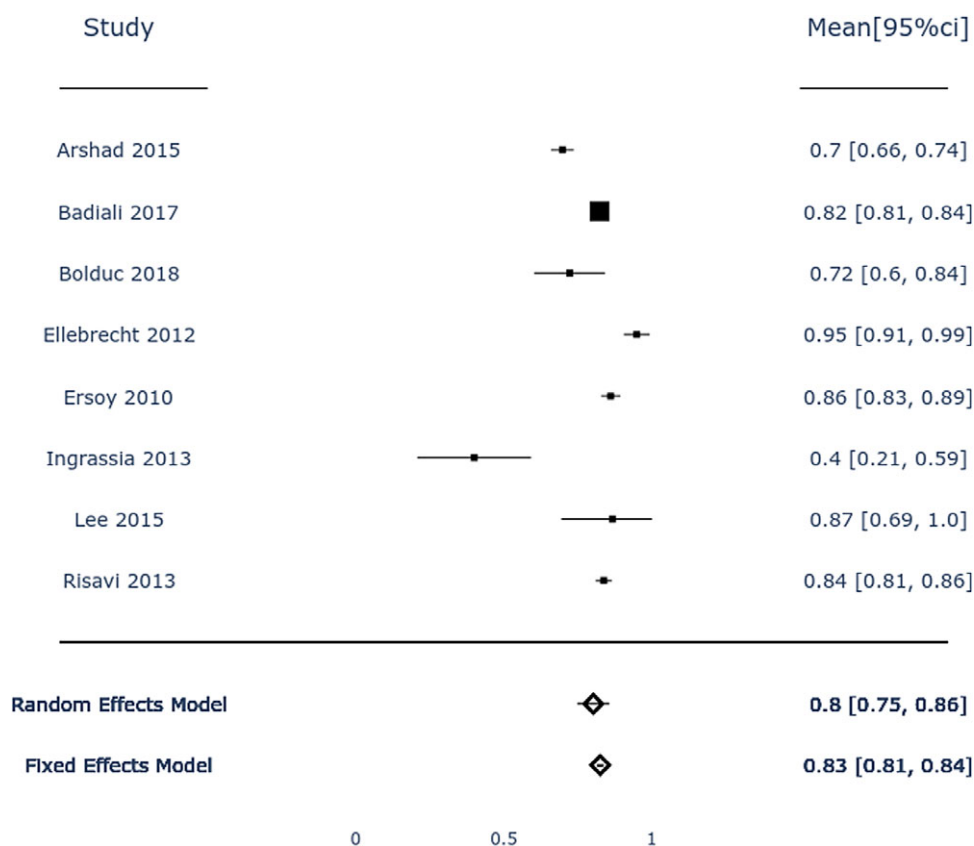


Figure 4. Category RED Triage Correct.

Eight studies were included in the **red** triage category with accuracy ranging from 0.40 to 0.95 (Figure 4). Using the random effects model, the pooled estimate for proportion correctly triaged was 0.80 (95% CI, 0.75 to 0.86). With the fixed effect model, pooled triage accuracy was 0.83 (95% CI, 0.81 to 0.84). Again, there was evidence of significant heterogeneity among studies ($P < .0001$).

The proportion of **yellow** triage category victims correctly triaged was reported in eight studies with accuracy ranging from 0.49 to 0.87 (Figure 5). The pooled estimate for proportion correctly triaged was 0.66 (95% CI, 0.55 to 0.77) with significant heterogeneity among studies ($P < .0001$) using a random effects model. Using a fixed effects model, the pooled estimate was 0.70 (95% CI, 0.69 to 0.71).

In the **green** triage category, eight studies reported an accuracy between 0.70 and 0.94 (Figure 6). The estimated pooled proportion correct was 0.87 (95% CI, 0.82 to 0.92) with significant heterogeneity among studies ($P < .0001$) using the random effects model. Using a fixed effects model, pooled accuracy was 0.86 (95% CI, 0.86 to 0.87).

Sensitivity Analysis—In four studies that did not explicitly report the absolute number of simulated patients triaged correctly, the results could be calculated by imputation.^{15,26,34,39} Addition of these four studies resulted in minimal change to the pooled estimates. When added to the fully reported studies, 28 studies could be included for overall assessment of START accuracy. Among these studies, the proportion of cases triaged correctly ranged from 0.27 to 1.0. The pooled effect estimate for accuracy was 0.74 (95% CI, 0.70 to 0.79) using a random effects model.

The pooled triage accuracy was 0.72 (95% CI, 0.71 to 0.72) using a fixed effects model. There was significant heterogeneity among studies ($P < .0001$).

Four studies initially included in the analysis group did not contain sufficient information to estimate triage accuracy and could not be included in the analysis of triage accuracy.^{24,29,35,42}

Secondary Outcome

Over-Triage—The proportion of over-triage was assessed in 18 (56%) studies (Figure 7). Over-triage proportions ranged from 0.006 to 0.53. The pooled over-triage estimate was 0.14 (95% CI, 0.11 to 0.17) using the random effects model. Using the fixed effects model, the pooled over-triage proportion was 0.10 (95% CI, 0.09 to 0.10). There was significant heterogeneity among studies ($P < .0001$).

Under-Triage—Under-triage was assessed in 18 (56%) studies and the reported proportion of under-triage ranged from 0.0061 to 0.25 (Figure 8). Using the random effects model, the pooled proportion of under-triage was 0.10 (95% CI, 0.07 to 0.14). The pooled estimate was 0.06 (95% CI, 0.06 to 0.07) using the fixed effects model. There was evidence of significant heterogeneity among studies ($P < .0001$).

Discussion

Using a robust search of the literature and efforts to avoid publication and selection bias, this meta-analysis identified the available studies utilizing the START tool for the assessment of critically ill or injured individuals in a real or simulated disaster. With a triage accuracy of 73%, this meta-analysis suggests that the START

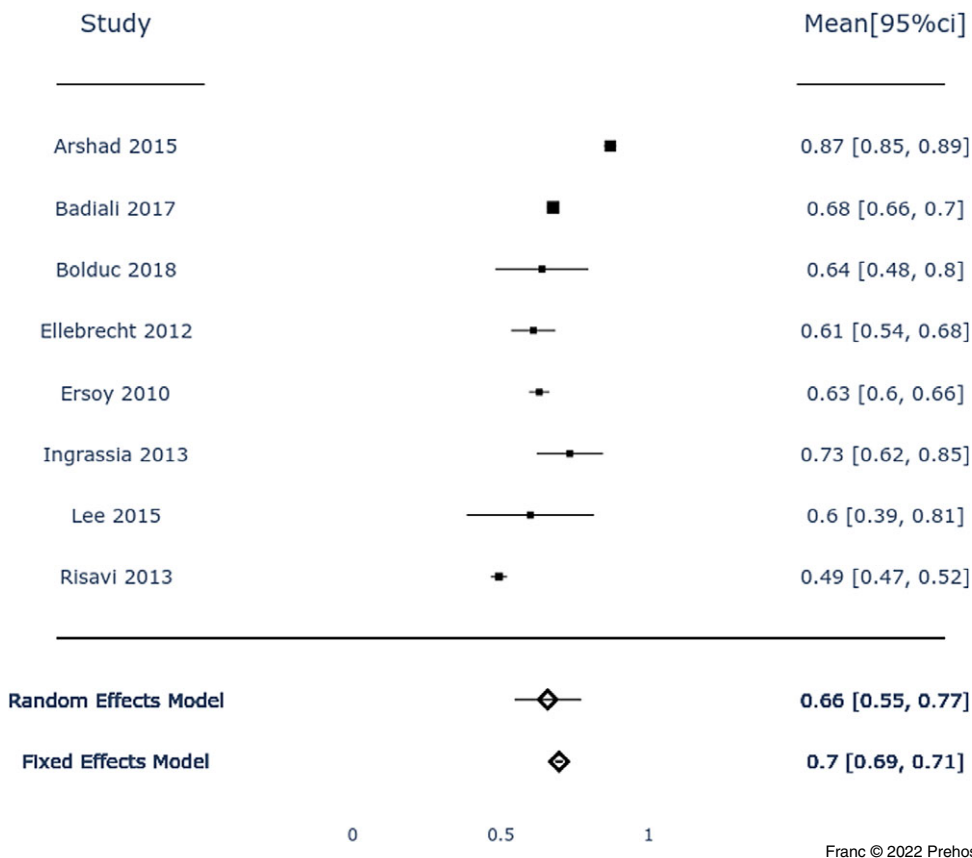


Figure 5. Category YELLOW Triage Correct.

method is not sufficiently accurate to serve as a reliable disaster triage tool.

In this meta-analysis, there was evidence of significant heterogeneity for each of the outcomes. As the included studies featured a number of different methodologies, disaster scenarios, and providers, this is not surprising. In fact, as disaster triage systems are expected to function across a wide range of disaster scenarios and provider experience, the heterogeneity of the studies represents one of the strengths of this meta-analysis: assessment of the true accuracy of START across a variety of scenarios. For a disaster triage system such as START to be useful, it must be reproducible. In addition, mass-casualty plans should contain a triage system that can be used by health care providers at a variety of levels.

Published studies of other disaster triage methods also show inconsistent and often low accuracy. For instance, recent studies of the Sort Assess Life-Saving Triage (SALT) method have found accuracy from 52% to 79%.^{47,49–51} Accuracy of Careflight triage among recently published articles varied from 39% to 94%, while triage sieve accuracy varied from 16% to 90%.^{51–54} The Sacco Triage Method (STM) may hold promise of a higher rate of accuracy, however, there are few published studies employing this method.^{27,55,56} By comparison, START was found to have a pooled accuracy of 73% with similar wide ranges (27% to 99%).

Traditionally, most disaster triage systems have been designed primarily to avoid under-triage as this protects disaster victims from a dangerous situation where they are denied early access to necessary medical care. Unfortunately, only a few studies (n = 18) included in this review reported under-triage as an outcome.

Analysis of those studies suggests that under-triage of START victims occurs in six percent to ten percent of victims.

Conversely, although over-triage represents little danger to the victim (victims are prioritized to get care quickly even if they do not need it), it represents a danger in overwhelming the health care response system. Again, only a small number of studies (n = 18) reported the outcome of over-triage, which was found to occur in seven percent to fourteen percent of victims.

This review also demonstrated that triage accuracy was lowest in the mid-acuity (yellow) patients at 66%. Accuracy was highest among the least acute (green) victims at 87%. Triage accuracy of black and red victims was 85% and 80%, respectively. Intuitively, having the highest accuracy among the most acute groups seems desirable as mistakes in triage among more acute victims are more likely to affect the victim's outcomes. The risk of the mid-acuity (yellow) victims being triaged incorrectly, however, does have the potential to result in maldistribution of resources that may be limited.

This review investigated the accuracy of the START method across all levels of providers. By definition, disaster scenarios exist because there is a lack of adequate resources to care for victims in the usual manner. Disaster plans commonly provide operating procedures to increase surge capacity by repurposing spaces, increasing supplies, and creating special operating protocols. In addition, responding to increased surge capacity often involves repurposing of staff. Thus, during a disaster, the task of prioritizing patients may be assigned to personnel not usually accustomed to performing triage. To be effective for use in disaster situations, a structured triage

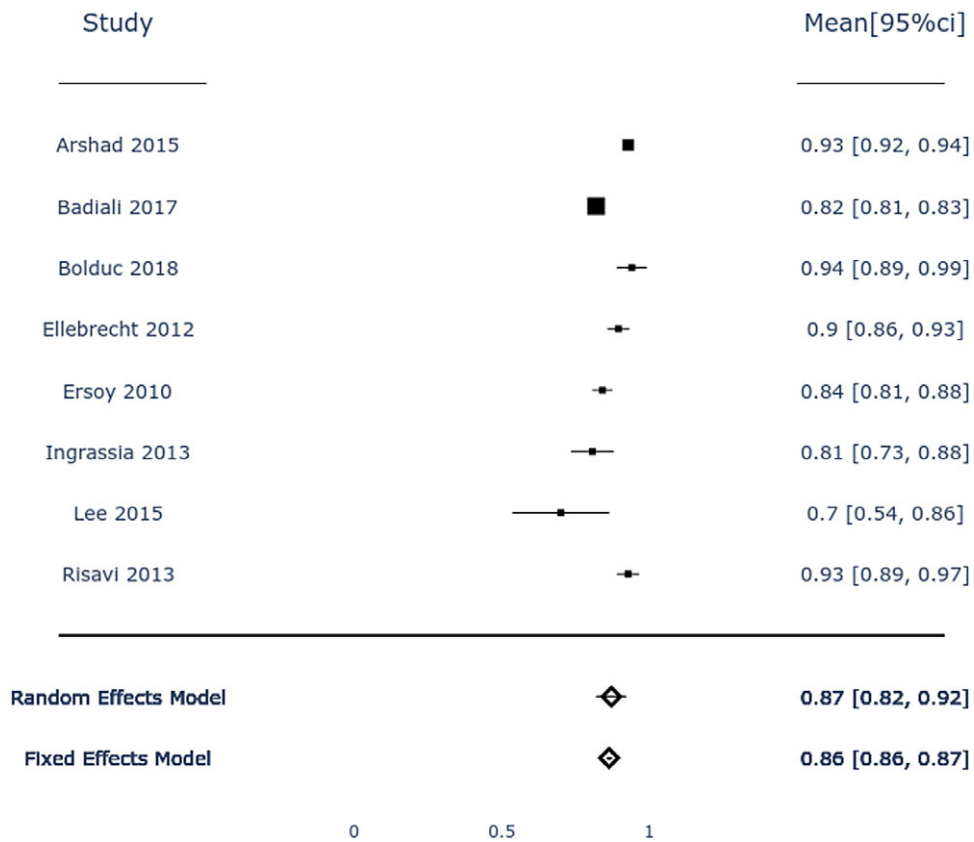


Figure 6. Category GREEN Triage Correct.

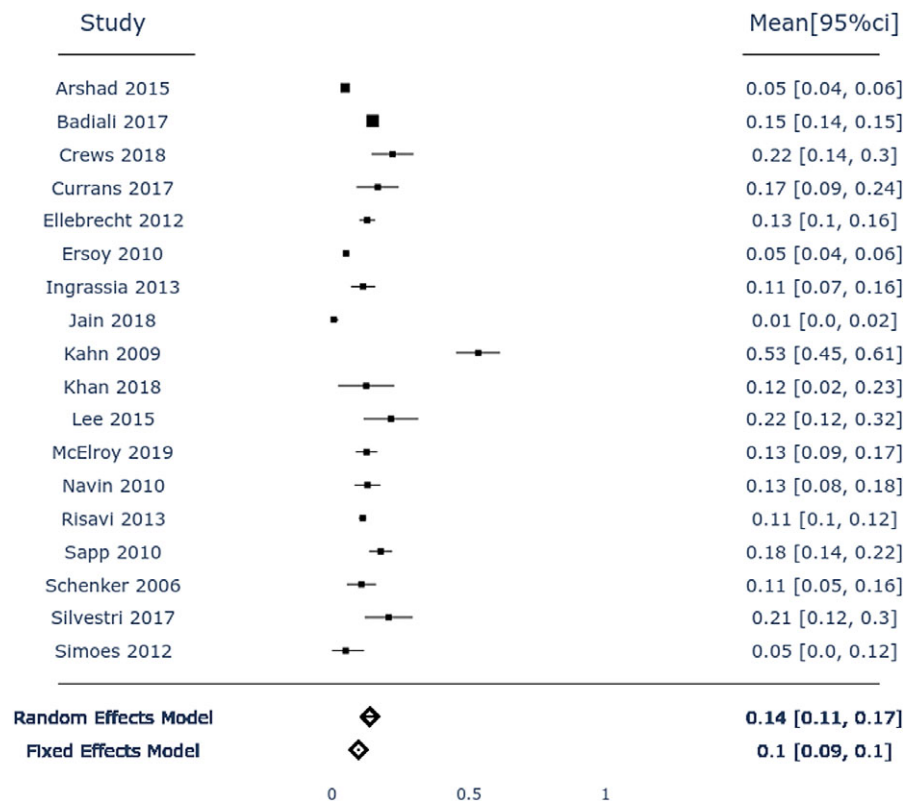
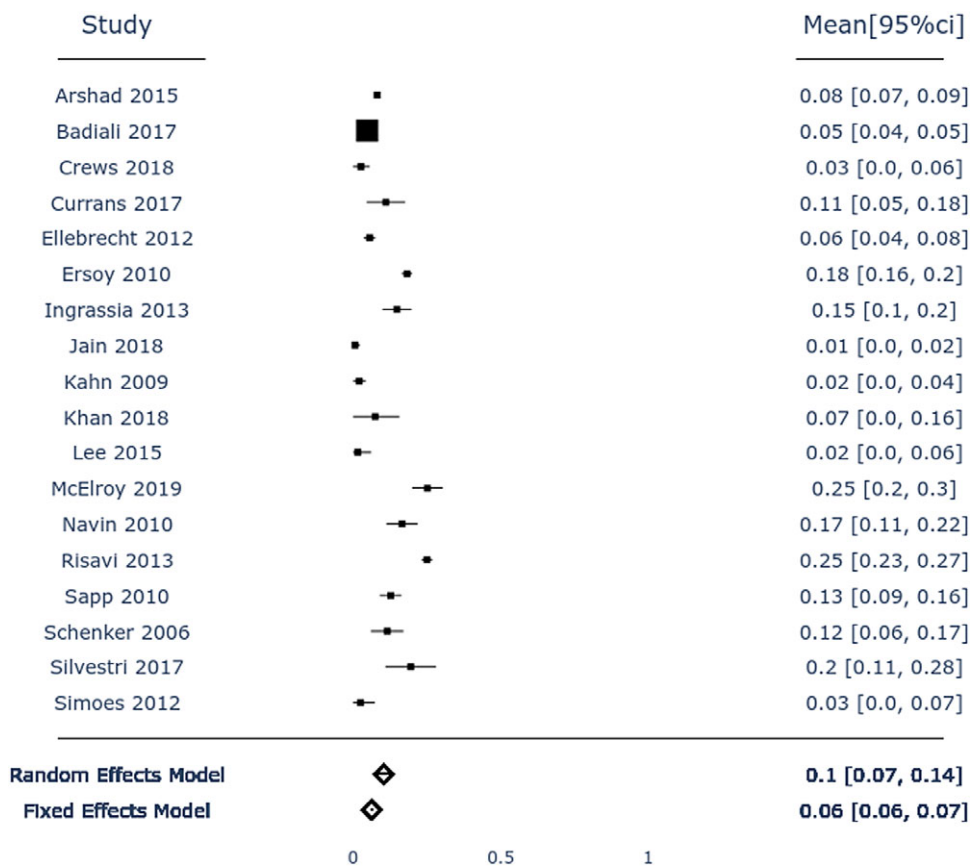


Figure 7. Over Triage.



Franc © 2022 Prehospital and Disaster Medicine

Figure 8. Under Triage.

protocol such as START should be adequately intuitive to be used across the entire range of providers: physicians, nurses, and prehospital providers. Triage systems that cannot be used by providers of all levels — while acceptable for normal operation of the emergency medical system — are suboptimal for disaster response.

Since the development of the START system in the 1980s, there has been a massive change in available technology. The ability to assess disaster victims using only a flowchart on a piece of paper and colored tags was clearly innovative forty years ago. As the use of hand-held electronic devices has become ubiquitous, electronic triage tools may hold promise. A recently published, large multi-center trial of 1,491 emergency department patients showed an increase in triage accuracy from 75.4% to 92.7% after introduction of the electronic Canadian Triage and Acuity Scale (eCATS) tool.⁵⁷ Conversely, another study comparing head-to-head electronic versus non-electronic methods failed to find a significant improvement.²⁵ Advancement in simulation training has also occurred over the years, from written summaries to live simulations with actors/mannequins with moulage to simulation exercises in VR. The impact of new technological advancements in triage assessment and simulation training on the accuracy of START is not well-understood and should be the subject of further investigation.

Limitations

The current study addresses the accuracy of START against self-reported reference standards. As many studies did not describe in detail how reference standards were developed or obtained, it is

possible that study investigators introduced bias in the development of these standards. This may be most concerning when the group studying the triage system and the group setting the reference standards included the same people.

The majority of included studies were based on a reference standard of expert opinion. As such, most studies evaluated the reliability of the START method to provide consistent assignment of the victims to the four groups. However, the prognostic capacity of the triage system was usually not assessed. A vital question that should be a topic for further studies would be to assess the prognostic ability of the test. Does the START method predict actual need for treatment? Furthermore, the final and most important outcome would be to evaluate the efficacy of the triage system in actually providing the best balance of individual patient care to optimize the disaster response. Although logistically and ethically challenging to perform, these studies would help delineate the true role of the START method. It should be considered, however, that if a test cannot be reliably applied by the group of providers using it, it is unlikely that the test will have good prognostic performance.

The majority of studies included in this review involved triage of simulated patients. The fidelity of the simulations varied from pen-and-paper to simulated patients to VR scenarios. While simulations are commonly used for training and evaluation of disaster medicine scenarios, all simulations have limitations. It is possible that true performance of START in real-life scenarios is different from that in simulated scenarios. Further studies to assess START performance in true disaster situations may be helpful in quantifying this difference. However, logistically performing this study

would likely be very difficult to coordinate in a safe and ethical manner given the unpredictable nature of disasters.

Unfortunately, there are several limitations due to the quality of the included studies. Only five of the included studies were randomized trials. Many studies suffered quality issues such as inadequate description of randomization, lack of blinding, and inadequate description of participant selection.

This study did not address pediatric disaster triage. As injury patterns and clinical presentations in pediatric disaster victims may vary greatly from adult victims, this study should not be used as a judgement of the accuracy of START among pediatric patients.

Finally, while there is a risk of selection and publication bias in this review, steps were taken to minimize these risks, including using two independent reviewers to identify eligible studies with a third-party mediator, as well as conducting an extensive search of the published and unpublished literature.

Conclusions

The results of this systematic review and meta-analysis suggest that with an overall accuracy of 73%, START triage may not

be sufficiently accurate to serve as a reliable disaster triage tool. Although meta-analysis revealed relatively low proportions of over- and under-triage, only approximately one-half of the included studies reported these outcomes. Victims in the mid-acuity (yellow) group appear to be the least accurately triaged. The vast majority of the included studies were simulation exercises and used expert opinion as the reference standard leading to potential biases in the results. In general, although the accuracy of START may be similar to other models of disaster triage, development of a more accurate triage method should be urgently pursued.

Acknowledgement

The authors would like to thank Jillian Meyer (JM) for her role in assisting with the review.

Supplementary Materials

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1049023X2100131X>

References

- Ghanbari V, Ardalan A, Zareiyan A, Nejadi A, Hanfling D, Bagheri A. Ethical prioritization of patients during disaster triage: a systematic review of current evidence. *Int Emerg Nurs*. 2019;43:126–132.
- Hogan DE, Burstein JL. *Disaster Medicine*. Philadelphia, Pennsylvania, USA: Lippincott Williams and Wilkins; 2002.
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLoS Med*. 2021;18(3):1–15.
- Fey MK, Gloe D, Mariani B. Assessing the quality of simulation-based research articles: a rating rubric. *Clin Simul Nurs*. 2015;11:496–504.
- Mariani B, Fey MK, Gloe D. The simulation research rubric: a pilot study evaluating published simulation studies. *Clin Simul Nurs*. 2018;22:1–4.
- Bond K, Ospina M, Blitz S, et al. *Interventions to Reduce Overcrowding in Emergency Departments*. Technology Report no 67.4. Ottawa, Ontario: Canadian Agency for Drugs and Technologies in Health; 2006.
- Higgins JP, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Hoboken, New Jersey USA: John Wiley and Sons, Inc.; 2019.
- Wells G, Shea B, O'Connell D, Peterson J. *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-Analyses*. Ottawa, Ontario: Ottawa Hospital Research Institute; 2015.
- Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Educ Inf*. 2018;34(4):285–291.
- Pace R, Pluye P, Bartlett G, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *Int J Nurs Stud*. 2012;49(1):47–53.
- Devore J. *Probability and Statistics for Engineering and the Sciences*. 7th ed. Belmont, California USA: Cengage Learning; 2011.
- Egger M, Smith GD, Altman DG. *Systematic Reviews in Health Care*. London, UK: BMJ Publishing Group; 2001.
- 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.
- Badiali S, Giugni A, Marcis L. Testing the START triage protocol: can it improve the ability of nonmedical personnel to better triage patients during disasters and mass casualty incidents? *Disaster Med Public Health Prep*. 2017;11(3):305–309.
- Ferrandini Price M, Escribano Tortosa D, Nieto Fernandez-Pacheco A, et al. Comparative study of a simulated incident with multiple victims and immersive virtual reality. *Nurse Educ Today*. 2018;71:48–53.
- Ingrassia PL, Colombo D, Barra FL, Carenzo L, Franc J, Della Corte F. Impact of training in medical disaster management: a pilot study using a new tool for live simulation. *Emergencias*. 2013;25(6):459–466.
- Ingrassia PL, Ragazzoni L, Tengattini M, Carenzo L, Corte FD. Nationwide program of education for undergraduates in the field of disaster medicine: development of a core curriculum centered on blended learning and simulation tools. *Prehosp Disaster Med*. 2014;29(5):508–515.
- Ingrassia PL, Ragazzoni L, Carenzo L, Colombo D, Gallardo AR, Corte FD. Virtual reality and live simulation: a comparison between two simulation tools for assessing mass casualty triage skills. *Eur J Emerg Med*. 2015;22(2):121–127.
- Izumida K, Kato R, Shigeno H. *A Triage Training System Considering Cooperation and Proficiency of Multiple Trainees*. Vol 10397 LNCS. (1) Graduate School of Science and Technology, Keio University; (2) Faculty of Science and Technology, Keio University: Springer Verlag; 2017.
- Jain T, Sibley A, Stryhn H, Hubloue I. Comparison of unmanned aerial vehicle technology-assisted triage versus standard practice in triaging casualties by paramedic students in a mass-casualty incident scenario. *Prehosp Disaster Med*. 2018;33(4):375–380.
- Khan K. Tabletop exercise on mass casualty incident triage, does it work? *Heal Sci J*. 2009;12(3):1–6.
- Lee JS, Franc JM, Lee J. Impact of a two-step emergency department triage model with START, then CTAS, on patient flow during a simulated mass-casualty incident. *Prehosp Disaster Med*. 2015;30(4):390–396.
- Lima DS, De-Vasc Oncelos IF, Queiroz EF, et al. Multiple victims incident simulation: training professionals and university teaching. *Rev Col Bras Cir*. 2019;46(3).
- Loth S, Cote AC, Shaafi Kabiri N, et al. Improving triage accuracy in first responders: measurement of short structured protocols to improve identification of salient triage features. *World Med Heal Policy*. 2019;11(2):163–176.
- Bolduc C, Maghraby N, Fok P, Luong TM, Homier V. Comparison of electronic versus manual mass-casualty incident triage. *Prehosp Disaster Med*. 2018;33(3):273–278.
- Mills B, Dykstra P, Hansen S, et al. Virtual reality triage training can provide comparable simulation efficacy for paramedicine students compared to live simulation-based scenarios. *Prehosp Emerg Care*. 2020;24(4):525–536.
- Navin DM, Sacco WJ, Waddell R. Operational comparison of the Simple Triage and Rapid Treatment Method and the Sacco Triage Method in mass casualty exercises. *J Trauma Inj Infect Crit Care*. 2010;69(1):215–225.
- Risavi BL, Lee W, Terrell MA, Holsten DL. Prehospital mass-casualty triage training-written versus moulage scenarios: how much do EMS providers retain? *Prehosp Disaster Med*. 2013;28(3):251–256.
- Riza'i A, Ade WRA, Albar I, Sulitio S, Muharris R. Teaching start triage: a comparison of lecture and simulation methods. *Adv Sci Lett*. 2018;24(9):6890–6892.
- Sapp RF, Brice JH, Myers JB, Hinchey P. Triage performance of first-year medical students using a multiple-casualty scenario, paper exercise. *Prehosp Disaster Med*. 2010;25(3):239–245.
- Schenker JD, Goldstein S, Braun J, et al. Triage accuracy at a multiple casualty incident disaster drill: the Emergency Medical Service, Fire Department of New York City experience. *J Burn Care Res*. 2006;27(5):570–575.
- Silvestri S, Field A, Mangalat N, et al. Comparison of START and SALT triage methodologies to reference standard definitions and to a field mass casualty simulation. *Am J Disaster Med*. 2017;12(1):27–33.
- Simões RL, Duarte Neto C, Maciel GSB, Furtado TP, Paulo DNS. Prehospital care to trauma victims with multiple simulated. *Rev Col Bras Cir*. 2012;39(3):230–237.

34. Wu Y-L, Shu C-C, Chung C-C. A simple method for pre-hospital dispatcher-aided consciousness assessment in trauma patients. *J Emerg Med Taiwan*. 1999;7(2): 69–77.
35. McCoy CE, Alrabah R, Weichmann W, et al. Feasibility of tele-simulation and Google glass for mass casualty triage education and training. *West J Emerg Med*. 2019;20(3):512–519.
36. Buono CJ, Lyon J, Huang R, et al. Comparison of mass casualty incident triage acuity status accuracy by traditional paper method, electronic tag, and provider PDA algorithm. *Ann Emerg Med*. 2007;50(3):s12–s13.
37. McElroy JA, Steinberg S, Keller J, Falcone RE. “Operation continued care: a large mass-casualty, full-scale exercise as a test of regional preparedness.” In: *Surgery (United States)*. Vol 166. Maryland Heights, Missouri USA: Mosby Inc.; 2019: 587–592.
38. Kahn CA, Schultz CH, Miller KT, Anderson CL. Does START triage work? An outcomes assessment after a disaster. *Ann Emerg Med*. 2009;54(3):424–430.
39. Challen K, Walter D. Major incident triage: comparative validation using data from 7th July bombings. *Injury*. 2013;44(5):629–633.
40. Crews CM. Disaster Response: Efficacy of Simple Triage and Rapid Treatment in Mass Casualty Incidents. Thesis. California USA: California State University, Long Beach; 2018.
41. Curran-Sills G, Franc JM. A pilot study examining the speed and accuracy of triage for simulated disaster patients in an emergency department setting: comparison of a computerized version of Canadian Triage Acuity Scale (CTAS) and Simple Triage and Rapid Treatment (START) methods. *Can J Emerg Med*. 2017;19(5).
42. Djalali A, Carenzo L, Ragazzoni L, et al. Does hospital disaster preparedness predict response performance during a full-scale exercise? A pilot study. *Prehosp Disaster Med*. 2014;29(5):441–447.
43. Ellebrecht N, Latasch L. Vorsichtung durch Rettungsassistenten auf der Großübung SOGRO MANV 500: Eine Vergleichende Analyse der Fehleinstufungen. *Paramed Triage Dur a Mass Casualty Incid Exerc a Comp Anal Ina Triage Lev Assignments*. 2012;15(1):58.
44. Ersoy N, Akpınar A. Triage decisions of emergency physicians in Kocaeli and the principle of justice. *Ulus Travma ve Acil Cerrahi Derg*. 2010;16(3):203–209.
45. Wisnesky UD, Kirkland SW, Rowe BH, Franc JM. A qualitative assessment of studies evaluating the accuracy of START in disaster triage: a scoping review. Press. 2021.
46. Alinier G. A typology of educationally focused medical simulation tools. *Med Teach*. 2007;29(8):e243–e250.
47. Fink BN, Rega PP, Sexton ME, Wishner C. START versus SALT triage: which is preferred by the 21st century health care student? *Prehosp Disaster Med*. 2018;33(4):381–386.
48. Waddell RK, Navin M. An objective comparison of the START triage protocols and the Sacco Triage Method. *Prehosp Disaster Med*. 2009;24(S1):s144.
49. Hartman EN, Daines B, Seto C, Shimshoni D, Feldman ME, LaBrunda M. Sort, Assess, Life-Saving Intervention, Triage with drone assistance in mass casualty simulation: analysis of educational efficacy. *Cureus*. 2020;12(9):e10572.
50. Celik DH, Mencl FR, Debacker M, Kurland L, Wilber ST, Frey JA. Triage performance of school personnel using the SALT system. *Prehosp Disaster Med*. 2019;34(4):401–406.
51. McKee CH, Heffernan RW, Willenbring BD, et al. Comparing the accuracy of mass casualty triage systems when used in an adult population. *Prehosp Emerg Care*. 2020;24(4):515–524.
52. Wallis LA, Carley S. Comparison of pediatric major incident primary triage tools. *Emerg Med J*. 2006;23(6):475–478.
53. Cuttance G, Dansie K, Rayner T. Paramedic application of a triage sieve: a paper-based exercise. *Prehosp Disaster Med*. 2017;32(1):3–13.
54. Vassallo J, Smith J. Major incident triage and the evaluation of the Triage Sort as a secondary triage method. *Emerg Med J*. 2019;36(5):281–286.
55. 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.
56. Cross KP, Cicero MX. Independent application of the Sacco disaster triage method to pediatric trauma patients. *Prehosp Disaster Med*. 2012;27(4):306–311.
57. McLeod SL, McCarron J, Ahmed T, et al. Interrater reliability, accuracy, and triage time pre- and post-implementation of a real-time electronic triage decision-support tool. *Ann Emerg Med*. 2020;75(4):524–531.