DANISH MEDICAL JOURNAL

Systematic Review

Dan Med J 2023;70(11):A09220516

Accuracy of triage systems for mass casualty incidents in live simulations – a systematic review

Christian Elleby Marcussen, Karoline Bendix Bräuner, Henrik Alstrøm & Ann Merete Møller

Department of Anaesthesia, Copenhagen University Hospital – Herlev and Gentofte Hospital, Denmark

Dan Med J 2023;70(11):A09220516

ABSTRACT

INTRODUCTION. In mass casualty incidents, insufficient triage may lead to increased morbidity and mortality due to delayed evacuation and treatment of the most critically injured patients. We report current litterature findings on accuracy of primary prehospital triage systems for mass casualty incidents in full-scale live simulations and map the challenges that lie ahead for finding the most accurate triage system.

METHODS. This study was registered with PROSPERO ID: CRD42018091889. We searched the databases EMBASE, MEDLINE, Central, Web of Science, Reference lists, Scopus, ClinicalTrials.gov and Google Scholar. We included primary triage systems, studies reporting accuracy or outcomes convertible to accuracy and studies performed in full-scale live simulations. We excluded studies using paediatric, chemical, biological, radiological or nuclear populations or triage systems. Bias rating was based on a modified version of the QUADAS-2.

RESULTS. A total of 15 studies were included. Six of 41 existing triage methods were tested. The studies showed substantial heterogeneity in both study characteristics and findings. Different reference standards were applied and most were based on author-defined triage categories. All studies carried an unclear-to-high risk of bias. Therefore, no quantitative comparisons were made.

CONCLUSION. In general, the studies suffered from substantial heterogeneity and risk of bias. A standardised protocol for future live simulations is needed to encourage consistent and comparable data collection. We identified some of the most important topics to address in such a protocol.

KEY POINTS

- Numerous pre-hospital triage systems exist; however, the presented data are insufficient to determine which system is superior.
- A standardised protocol is needed to obtain comparable and conclusive results.
- We provide essential elements for such a protocol.

As mass casualty incidents (MCI) cause great strain on both pre- and in-hospital resources, prioritising patients before they arrive at the hospital is necessary [1, 2].

Minimising time from injury to treatment is vital for the most severely injured casualties as shown in several studies on trauma patients [3-6]. One way to speed up that process is quick and correct assessments of the

casualties and subsequent prioritisation based on the need for lifesaving interventions.

Triage systems are developed exactly for this reason. They identify the most severely injured people and give them priority for evacuation and treatment at all levels throughout the evacuation chain [7].

The fact that many triage systems exist indicates a lack of direction as to the best way to sort casualties. Most of these systems have never been examined quantitatively. Before we can design high-quality studies to determine the most accurate system, a status on what is currently known about triage systems is needed. Existing reviews have not been conducted systematically [8], used a narrow search strategy [9] or were made more than seven years ago [9, 10], making it likely that new results have emerged since then. Furthermore, the most recent review did not report accuracy [8]. Thus, an up-to-date and systematic review is needed.

Ideally, randomised controlled trials should be included, but no such trials exist as they are both ethically and practically unfeasible in MCI. However, many other types of studies have examined triage systems for MCIs. Because of methodological heterogeneity, we conducted a series of systematic reviews to obtain comparable results. In the first review, we focused on trauma register studies [11]. In this second review, we are examining the accuracy of prehospital triage systems in full-scale live simulations.

Unfortunately, the current literature describing the accuracy of primary prehospital triage systems for MCIs in full-scale live simulations has notable issues relating to methodology, reporting and heterogeneity (cf. this review). We highlight these gaps and aim for the present review to serve as an important tool to direct future research on this topic.

METHODS

A protocol was registered at PROSPERO, an international prospective register of systematic reviews, with registration ID: CRD42018091889 [12]. Where applicable, this review was reported according to the PRISMA-DTA guidelines [13]. However, typical DTA measures and approaches to synthesis were not applicable.

Eligibility criteria

Our eligibility criteria were as follows:

Population: We included trials that examined triage systems in full-scale live simulations. We excluded trials if the population was children, burn casualties or chemical, biological and nuclear (CBRN) casualties.

Intervention: Trials examining one or more primary triage systems for MCIs were included. Primary triage systems were defined as triage systems designed to be applied by first responders at the incident site. If the examined triage system was designed for children, burn or CBRN casualties, it was excluded.

Outcomes: To be included, trials needed to provide results as or convertible to accuracy in percentage.

Reasoning for inclusion and exclusion criteria are provided in the discussion.

We defined full-scale live simulations as simulations achieving a high level of realism by using actors in a MCI-imitating setting. The use of mannequins was allowed if the study also used actors. MCIs are defined by the WHO as an event requiring exceptional emergency arrangements and extraordinary assistance [14].

Search strategy

Preliminary information was retrieved to find relevant medical subject headings (MeSH). Our search strategy was formed from the discovered terms with the assistance of an information specialist. We searched the EMBASE, MEDLINE, Central and Web of Science databases. For EMBASE and MEDLINE, we used the OVID

interface. No limitations on language, publication date or publishing status were applied. The final search was performed on 19 July 2022. Search strategies are provided in Supplementary materials page 1-2. Reference lists of the included articles were hand searched and a Scopus citation search was performed. We searched for unpublished literature through ClinicalTrials.gov and Google Scholar.
Titles and abstracts of retrieved articles were screened independently by two authors (CEM and KBB) followed by independent full-text screening of potentially eligible articles by the same two authors. Finally, the same two authors used a standardised and piloted form to extract data. Disagreements on study selection and data extraction were resolved by discussion. If disagreement persisted, a third author (AMM) was consulted.

Data extraction

Data were extracted for: type of triage system, type of MCI simulated, duration of pre-simulation triage course, distribution of cases into triage categories, whether or not a flowchart of the triage system was handed out, how vital parameters were obtained, what reference the results were compared to, occupation of triage performers, total number of triage decisions, number of cases played by actors, number of cases displayed with mannequins, accuracy, rate of undertriage, rate of overtriage, primary outcomes, secondary outcomes, conflicts of interest and funding sources.

Risk of bias

As there are no guidelines on how to rate the risk of bias in simulation studies, we predefined new criteria mainly based on the QUADAS-2 [15] – a bias rating tool developed by the Cochrane collaboration. The exact signalling questions are available in the QUADAS-2 guidelines and their modifications in our discussion. A piloted form was used to assess risk of bias by two of the authors (CEM and KBB) and disagreements were resolved by discussion or by involving a third author (AMM). The QUADAS-2 elements assessed were: patient selection, index test, reference standard and flow and timing. A selection of the reported results was assessed according to ROBINS-I [16] as this is not comprised by the QUADAS-2. Lastly, bias due to deviation from the intended triage category was assessed with the following signalling questions:

1. Was every patient triaged exactly as the triage system suggested?

2. Is it true that NO parameters were imputed from another vital characteristic?

3. If imputations were made, is it fair to assume that imputations did NOT bias the results

If the answer to question 1 or both question 2 and 3 was "no", the domain would be rated with a high risk of bias. If the answer to question 1 or 3 was "unclear", the domain would be rated as unclear. If the answer to question 1 and 3 was "yes", the domain was rated with a low risk of bias. Furthermore, as we did not believe that the assessment criteria used would identify every possible type of bias, we included additional observations under "other bias" as relevant. The studies were graded as proposed by the QUADAS-2 as either having a low, unclear or high risk of bias. Each domain was rated for bias risk on an outcome-specific level.

The overall study level bias rating was done according to the QUADAS-2 without modifications.

Outcomes

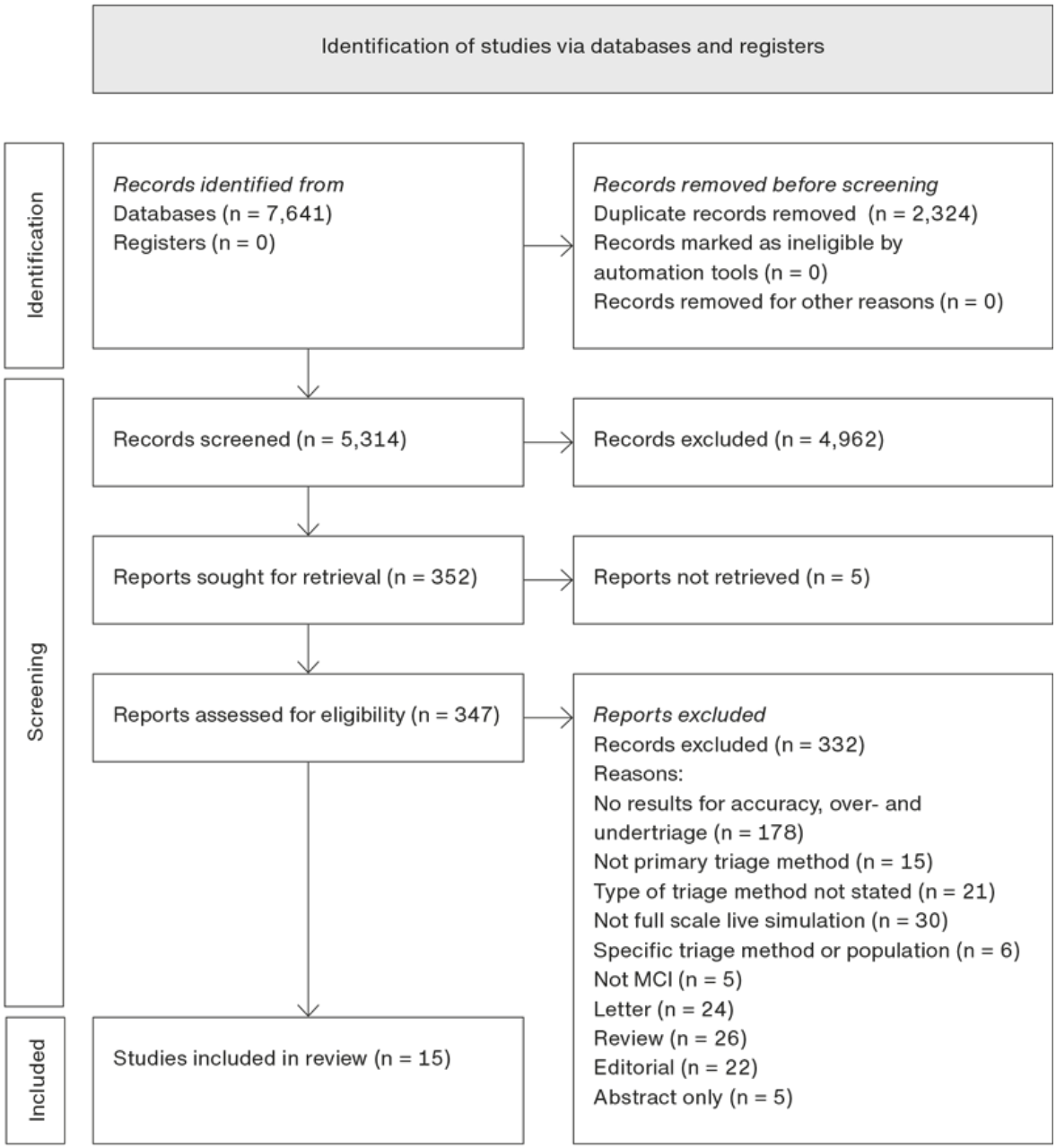
Initially, we had accuracy as our main outcome. However, as our work progressed, we realised that a comparison of accuracy would not yield any meaningful results due to heterogeneity between studies. Therefore, we changed our main outcome to be a description of the studies' differences in methodology, study characteristics and their potential risk of bias. We kept accuracy as a secondary outcome to illustrate the heterogeneity.

RESULTS

Selected studies

We found 7,641 records matching our search criteria. After removing duplicates, 5,314 records were screened. Among the 352 records that were full-text screened, 15 studies met our eligibility criteria [17-31] (**Figure 1**). No further studies were found in the citation search or the reference lists of the included studies.

FIGURE 1 PRISMA flow chart of study selection process [13].



MCI = mass casualty incidents.

Characteristics of included studies

The characteristics of the included studies showed variation in several categories. In 33 cases, study

characteristics were unclear or not reported in the studies (**Table 1**). (Download Table 1 as PDF – https://ugeskriftet.dk/sites/default/files/2023-10/A09220516_Table1.pdf)

TABLE 1 Study characteristics.

Reference	Triage system	Type of MCI	Development of reference standard	Presimulation triage course	Was a flow chart of the triage system handed out?	Experience of triage performers	Cases played by actors, n	Cases played by mannequins, n	Intended distribution of cases into triage categories, % (n)	Possible conflicts of interest	Funding sources
Silvestri et al, 2017 [19]	START[a]	Chlorine gas explosion and a live shooter at a university	Published consensus-based reference standard achieved in a Delphi process	-	-	EMT-P	47	4: only dead	Red: 33 (17) Yellow: 12 (6) Green: 47 (24) Black: 8 (4)	0 declared	-
Ingrassia et al, 2015 [20]	START	Motor vehicle accident	Compared to triage performed by a computer	2-hr. training on day 2: between 1st 2nd session	-	56 senior medical students	10	0	-	Authors are owners of VR software used in study	-
Schenker et al, 2006 [21]	START	Train collision with blast injury and chemical release	Predefined triage category	No	-	40 EMS personnel	99	31	Red: 23 (30) Yellow: 12 (16) Green: 40 (51) Black: 24 (33)	0 declared	-
Ellebrecht & Latasch, 2012 [22]	START	Airplane crash at airport	Unclear	1-hr. training in start during "the weeks up to the exercise"	Yes: PDA	EMT-P	520	0	Red: 19 (100) Yellow: 32 (170) Green: 49 (260) Black: 0	The Company which developed the PDA was involved in design of the study	-
Bolduc et al, 2018 [23]	START	Train derailment and chemical spill	Predefined triage category	30-min. training same day as exercise	Unclear	2 registered nurses	30	0	Red: 30 (9) Yellow: 20 (6) Green: 47 (14) Black: 3 (1)	0 declared	-
Price et al, 2018 [24]	START	-	Unclear	-	-	35 emergency and special care nursing master's degree students	20	0	Red: 40 (8) Yellow: 25 (5) Green: 10 (2) Black: 25 (5)	0 declared	Spanish government Seneca foundation
Fernandez-Pacheco et al, 2018 [25]	START	-	Compared to 3 emergency physicians' triage	4-hr. training prior to the exercise	-	68 nursing students	Unclear	Unclear	Red: 25 (10) Yellow: 42.5 (17) Green: 25 (10) Black: 7.5 (3)	0 declared	-
Jain et al, 2018 [26]	START	Motor vehicle accident	Based on medical records from trauma patients	15-min. training prior to the trial	Yes	20 PCP 20 advanced care PCP	10	0	Red: 30 (3) Yellow: 40 (4) Green: 20 (2) Black: 10 (1)	0 declared	Holland college paramedicine programme
Lee et al, 2016 [27]	SALT	4-car motor vehicle collision	Predefined triage category	30-min. training immediately before exercise	No	38 PCP 29 FS	8	0	Immediate: 25 (2) Urgent/delayed: 12.5 (1) Minimal: 37.5 (3) Expectant: 12.5 (1) Dead: 12.5 (1)	0 declared	-
Cone et al, 2009 [28]	SALT	Incident at airport	Predefined triage category	90-min. training session	Yes, though not used by participants	Experienced EMT-P	52	0	Immediate: 31 (16) Urgent/delayed: 23 (12) Minimal: 27 (14) Expectant: 0 Dead: 19 (10)	Cone participated in the development of SALT	Uppsala University
Lerner et al, 2010 [29]	SALT	Bomb blast at a community concert	Predefined triage category	30-min. training the day before the exercise	-	73 mixed healthcare personnel	19	10	Immediate: 26.7 (58) Delayed: 24.4 (53) Minimal: 35 (76) Expectant: 6.5 (14) Dead: 7.4 (16)	0 declared	-
Cicero et al, 2015 [30]	SMART	Airplane crash at airport	Predefined triage category	-	-	*CG* 2 EMT-P	*CG* 27 or 29	0	*CG* Red: 31 (9) Yellow: 34 (10) Green: 17 (5) Black: 17 (5)	0 declared	-
						IG With Google Glass: 2 EMT-P with 10 yrs extra experience	*IG* 20 or 21		*IG* Red: 38 (8) Yellow: 42 (9) Green: 5 (1) Black: 14 (3)		

To be continued >

TABLE 1 (continued) Study characteristics.

Reference	Triage system	Type of MCI	Development of reference standard	Presimulation triage course	Was a flow chart of the triage system handed out?	Experience of triage performers	Cases played by actors, n	Cases played by mannequins, n	Intended distribution of cases into triage categories, % (n)	Possible conflicts of interest	Funding sources
Navin et al, 2010 [17]	START	Building collapse	Predefined triage category	20-min. immediately before exercise	-	EMT-I EMT-P	20	79	*START* Red: 22 (22) Yellow: 25 (25) Green: 48 (47) Black: 5%(5)	Sacco invented STM and is among the authors	-
									STM RPM-score: 0: 5 (5) 1: 0 2: 0 3: 0 4: 0 5: 0 6: 2 (2) 7: 1 (1) 8: 5 (5) 9: 5 (5) 10: 4 (4) 11: 13 (13) 12: 65 (64)		
Offterdinger et al, 2014 [31]	MSTART	Scenario 1: crash landing of an airplane Scenario 2: airport shuttle bus collision Scenario 3: explosion of the airfield Scenario 4: patients injured by firearms	Compared to expected outcomes based on symptoms	4-hr. course the day before the exercise	Yes	12 EMT-P 12 emergency physicians	10	0	-	Connection to authors of MSTART [18, 33]	-
Rehn et al, 2010 [32]	TAS	Bus crash	Unclear	2-day course in TAS: MCI management system Exercise immediately after course	Yes	A mix of healthcare professionals, firefighters, police officers and others	20	0	Red: 25 Yellow: 25 Green: 25 Black: 25	Authors developed the examined system	Norwegian air ambulance foundation

CG = control group: EMS = emergency medical service; EMT-I = emergency medical technician – intermediate; EMT-P = paramedics; FS = fire science students; IG = intervention group; MCI = mass casualty incident; MSTART = modified START; PCP = paramedical students; PDA = personal digital assistants; RPM = respiration, pulse and motor function; SALT = Sort, Assess, Lifesaving interventions, Treatment/Transport; START = Simple Triage and Rapid Treatment; STM = Sacco Triage Method; TAS = Tverretatlig Akuttmedisinsk Samarbeid; VR = virtual reality.
a) Noted "field triage".

METHODS

Patients (cases) were defined with vital parameters corresponding to the examined triage system, and no baseline characteristics were reported. The distribution of cases corresponding to the correct triage category (reference standard) varied between studies (P1: 19-40%; P2: 12-43%; P3: from 5 to ≥ 64% (based on the Sacco Triage Method (STM) score 12); P4: 0-25%).

Six different types of MCIs were used as settings for the simulations. Some of the studies were designed with a pre-simulation triage course, with a duration ranging from 15 minutes to two days. Likewise, some studies made a flowchart of the triage system available during the entire simulation; however, most studies did not report this aspect.

The participants had a wide mix of occupations, ranging from non-healthcare professionals to emergency physicians.

Triage systems: index test

Six different triage systems had been tested in eligible studies. The Simple Triage and Rapid Treatment (START) was included in nine studies, and the Sort, Assess, Lifesaving interventions, Treatment/Transport (SALT) was included in three studies. The last four triage systems – Smart, Modified START (MSTART) (more systems with this name exist – see referenced study for exact version [32]), STM unadjusted, and Tverretatlig Akuttmedisinsk Samarbeid (TAS) – were each employed in one study.

Reference standards

The reference standards were developed in six different ways. One was developed through a Delphi process [33]. In seven studies, the reference standards were developed by the authors alone. One studied compared results to triage categories assigned by a computer, another study compared results to the expected outcome based on the case symptoms, and a third study based the used reference standard on conveniently selected patient records from a trauma registry. Finally, one study had three emergency physicians triage each case and used their decisions as a reference standard. Three studies did not report how their reference standards were developed. Most studies did not specify their reference standard other than how they had developed it. As target condition, all studies used level of urgency.

Outcomes

The primary outcome was accuracy of the tested triage system in 12 out of 15 studies.

Conflicts of interest and funding sources

Six studies had possible conflicts of interest, even though this was not always declared by the study. Most of the possible conflicts were because one or more of the authors had developed the triage system. Most studies did not state how they were funded, but those that did, had not received funding from sources with possible conflicts of interest.

Risk of bias

On an overall study level, all studies were at risk of bias. By far the most often used rating of each domain was "unclear" (**Table 2**), which was due to a lack of reporting in almost every case.

TABLE 2 Results of individual studies.

Reference	Triage information	Triage decisions, n	Accuracy (95% CI)[a], %	Overall undertriage, %	Overall overtriage, %
START					
Silvestri et al, 2017 [19]		51	45	37	18
Ingrassia et al, 2015 [20]	D1	280[b]	58	-	-
Schenker et al, 2006 [21][c]		130	71	22	11
Navin et al, 2010 [17]		193	71	12[b]	18[b]
Ellebrecht & Latasch, 2012 [22]		520	81	6[b]	13[b]
Bolduc et al, 2018 [23]		60[b]	83	-	-
Ingrassia et al, 2015 [20]	D3	280[b]	84	-	-
Price et al, 2018 [25]		Unclear	88	-	-
Fernandez-Pacheco et al, 2018 [25]		Unclear	89	-	-
Jain et al, 2018 [26]		400[b]	100	0	0
SALT					
Lee et al, 2016 [27]	FS	-	72	13[b]	15
Cone et al, 2009 [28]		52	79	4	17[b, d]
Lee et al, 2016 [27]	PCP	-	80	10[b]	10
Lerner et al, 2010 [29]		217	83 (78-88)	10	6
SMART					
Cicero et al, 2015 [30]	CG	29	76[e]	-	-
Cicero et al, 2015 [30]	IG	21	86[e]	-	-
STM					
Navin et al, 2010 [17]		193	92	-	-
MSTART					
Offterdinger et al, 2014 [31]		480[b]	97	1	2
TAS					
Rehn et al, 2010 [32][f]		78	100	0	0

CG = control group; CI = confidence interval; D1 = day 1 (before training); D3 = day 3 (after training); FS = fire science students; IG = intervention group; MSTART = modified START; PCP = paramedic students; SALT = Sort, Assess, Lifesaving interventions, Treatment/Transport; START = Simple Triage and Rapid Treatment; STM = Sacco Triage Method; TAS = Tverretatlig Akuttmedisinsk Samarbeid.
a) 95% CI was only reported in 1 study.
b) Calculated from other results in the original study – for calculations see supplementary materials page 3-5.
c) Results from final treatment area.
d) Reported as 14% in the original paper, but the correct result is 17% (9/52).
e) Difference between CG and IG was insignificant with a p value of 0.39.
f) Results taken from course 1 as this course was the only course without paediatric patients (results were identical to the remaining courses).

Results of indidividual studies

The reported accuracy, undertriage and overtriage of the systems studied varied substantially (**Figure 2**). This variation was observed both between studies assessing the same triage system and between studies assessing different triage systems. Due to the low quality of evidence, we are highly uncertain about the overall utility of the systems examined in the included studies.

In some studies, the same case was triaged by several participants, which explains the discrepancy between the number of cases (Table 1) and the number of triage decisions.

FIGURE 2 Risk of bias within studies.

Reference	Triage system	Patient selection	Index test	Reference standard		Flow and timing	Selection of reported results	Deviation from intended triage algorithm	Other bias
				ROB	appl.[a]				
Silvestri et al, 2017 [19]	START	Unclear	Unclear	Unclear	Low	Low	Unclear	Unclear	Low
Ingrassia et al, 2015 [20]	START	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	High[b]
Schenker et al, 2006 [21]	START	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Low
Ellebrecht & Latasch, 2012 [22]	START	Unclear	Unclear	Unclear	Unclear	Low	Unclear	High	Low
Bolduc et al, 2018 [23]	START	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Low
Price et al, 2018 [24]	START	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Low
Fernandez-Pacheco et al, 2018 [25]	START	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Low
Jain et al, 2018 [26]	START	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Unclear	
Cone et al, 2009 [28]	SALT	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Low
Lee et al, 2016 [27]	SALT	Unclear	Low	Unclear	Unclear	Low	Unclear	Unclear	Low
Lerner et al, 2010 [29]	SALT	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Low
Cicero et al, 2015 [30]	SMART	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Low
Navin et al, 2010 [17]	STM & START	Unclear	Unclear	Unclear	Unclear	Low	Unclear	High	Low
Offterdinger et al, 2014 [31]	MSTART	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear	Low
Rehn et al, 2010 [32]	TAS	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Low

appl. = applicability; MSTART = modified START; ROB = risk of bias; SALT = Sort, Assess, Lifesaving interventions, Treatment/Transport; START = Simple Triage and Rapid Treatment; STM = Sacco Triage Method; TAS = Tverretatlig Akuttmedisinsk Samarbeid.
a) For applicability of Patient Selection and Index test see discussion.
b) The participants were tested twice in the same cases with two days interval. One in a live simulation and the other in virtual reality simulation. This may yield a falsely higher level of accuracy.

DISCUSSION

Overall, the evidence was too uncertain to determine which of the included primary triage systems for MCIs performed best in live simulations. Inconclusive results were also reached in previously published reviews examining triage systems in settings other than live simulation or adopting a different focus [8-10].

When examining aspects of MCIs, live simulations are much more ethically acceptable than RCTs, but this comes at a loss of study quality. Even simulations with a good apparent level of realism are not fully able to mimic the chaos of an actual MCI. Some examples of this are given below.

Firstly, as the actors could not simulate true vital signs expected in the emergency setting (e.g., pulse and respiratory rate), these data were generated by the study investigators and not measured within the simulation. Thus, these measurements would have required substantially less effort and time than anticipated in a real MCI in which the emergency responder would have to manually measure the vital parameter.

Secondly, professionals triaging during a real incident would be placed under an amount of pressure exceeding that of a simulation. We suspect that this compromise would likely cause them to make more errors during a real incident, and conceivably even more as the complexity of the triage system increases. Therefore, in the absence of anticipated real-life stress, live simulations risk overvaluing more complex triage systems, which may perform better in a controlled context.

Another issue is the fact that real patients have comorbidities and characteristics (e.g., age) that may affect their chance of survival. These parameters are not factored into the models of the included studies' cases.

As mentioned, no pre-existing bias-rating guidelines exist that are directly applicable to the included studies. After an evaluation of different risk of bias rating tools, we believe that the most fitting tool is a modified version of the Quadas-2 [15]. Thus, we modified the Quadas-2 by leaving our patient-selection in the individual assessment and rated it as unclear for all studies. We chose this because MCIs are very different in terms of composition of injuries and their severity. The distributions of casualties were defined by the authors before the simulation, and it remains unknown how this fact affects the results for risk of bias and applicability. Similarly, we did not rate the risk of bias for each study caused by threshold values determined *after* performing the index test. The thresholds are incorporated in the triage systems and were always defined before the studies began. Additionally, we did not make individual assessments of applicability for index tests as this aspect is addressed in

our inclusion criteria and in the assessment of bias caused by deviation of intended triage system. We removed the signalling question indicating if there was an appropriate time between index test(s) and the reference standard. We removed the question because the disease states of the fictional cases were constant, i.e. the index test and the reference standard triaged the patient in the same condition.
Data on study characteristics and risk of bias assessment were often scarce. Unclear risk of bias was due to insufficient reporting in almost all cases. Thus, the actual risk of bias may have been higher than we reported. An example of this is that most of the studies that used an author-made pre-defined reference standard did not further specify the elements of the reference standard. The authors may be biased in their opinion of how each case would turn out and might therefore have created a flawed reference material. Another example is that no studies cited a protocol. In the absence of study protocols, it was not possible to determine risk of bias due to the selection of reported results.

Another important consideration is the heterogeneity of study characteristics. As outlined in Table 1, no two studies were the same in all characteristics and some shared no commonalities. This heterogeneity may explain some of the variation observed in the results, with some factors having a larger influence than others.

Two of these characteristics were of particular interest: Firstly, some studies handed out a card showing a flowchart of the triage system, making it easier to remember, whereas others did not hand out such a card. Among the studies that did use handouts, only two reported how much the participants used them. One of these two studies [29] reported that the participants who used STM were more consistent in following every step of the triage system than those who used START. This difference is very likely to cause bias when comparing the results. Secondly, the duration of the pre-simulation triage course varied greatly, giving some participants an advantage over others.

Standard protocol

The fact that the studies were considerably different with respect to study characteristics and methodology made it unreasonable to make a quantitative synthesis of the results. One of the main reasons for the heterogeneity and low reporting may be the lack of guidelines on how to report and conduct this type of study. A standardised protocol is one way to solve this problem. We believe that a standardised protocol is best developed through a Delphi process as some of the components described below are a matter of opinion and prioritising. Nonetheless, it is important to reach a consensus to get comparable results. Based on our findings and the experience gained through our work with this review, we believe that a standardised protocol should consider the components for study characteristics, reporting and results described in Supplementary materials page 6-9.

Limitations

This review was not without limitations. To focus its scope, we chose to focus on full-scale live simulations. This was done to exclude studies employing tabletop exercises, virtual reality and computer games as they conceivably have less validity than full-scale live simulations. Thus, we also excluded non-simulation studies such as registry studies. However, we have evaluated these types of studies in a second review [11].

We chose to report the findings according to the PRISMA-DTA as this seemed to be the best fit, though there are some limitations as no current guidance has been adapted to apply to simulation studies.

Furthermore, triage systems for a population of children, burn casualties or CBRN victims were not examined in this review as a different pathophysiology applies to these types of patients [34, 35]. Triage systems designed for children are very similar to those for adults, but have thresholds adapted to their physiology. Burn triage systems include Total Body Surface Area (TBSA) burnt and CBRN triage systems consider how contaminated or exposed the patient is.

The strength of our review is the adaptation of systematic review methodology as rigorously as possible to the simulation study context, where guidance for conduct is currently lacking. Additionally, we were the first to systematically apply an adapted QUADAS-2 rating to studies testing triage systems.

CONCLUSION

To summarise, this study found that the evidence is insufficient and too heterogenous to determine which of the included primary triage systems for MCIs is more accurate. To determine the triage system of highest quality, a standard protocol for future live simulation studies is needed to obtain comparable results. Our study shows that the main issues concern study characteristics, reporting and risk of bias. We provide specific elements that should be discussed in a future standardised protocol.

Supplementary https://content.ugeskriftet.dk/sites/default/files/2023-08/a09220516-supplementary.pdf

Correspondence *Christian Elleby Marcussen*. E-mail: Christianmarcussen@gmail.com

Accepted 10 August 2023

Conflicts of interest none. Disclosure forms provided by the authors are available with the article at ugeskriftet.dk/dmj

Acknowledgements We would like to thank *L. Christensen*, Department of Anaesthesia, Zealand University Hospital Koege, for assisting the initial article screening. We would also like to thank *J. Vendt*, Department of Anaesthesiology, Herlev and Gentofte Hospital, for helping us with the creation of our search strategy. Finally, we would also like to thank *C. Christensen*, *T. Quay* and *E. Jarnholt* for language corrections and grammatical assistance.

Cite this as Dan Med J 2023;70(11):A09220516

REFERENCES

1. Ball CG, Kirkpatrick AW, Mulloy RH et al. The impact of multiple casualty incidents on clinical outcomes. J Trauma. 2006;61(5):1036-9.
2. Abir M, Choi H, Cooke CR et al. Effect of a mass casualty incident: clinical outcomes and hospital Ccarges for casualty patients versus concurrent inpatients. Acad Emerg Med. 2012;19(3):280-6.
3. Meizoso JP, Ray JJ, Karcutskie CA 4th et al. Effect of time to operation on mortality for hypotensive patients with gunshot wounds to the torso: the golden 10 minutes. J Trauma Acute Care Surg. 2016;81(4):685-91.
4. Barbosa RR, Rowell SE, Fox EE et al. Increasing time to operation is associated with decreased survival in patients with a positive FAST examination requiring emergent laparotomy. J Trauma Acute Care Surg. 2013;75(1 suppl 1):S48-S52.
5. Alarhayem AQ, Myers JG, Dent D et al. Time is the enemy: mortality in trauma patients with hemorrhage from torso injury occurs long before the “golden hour.” Am J Surg. 2016;212(6):1101-5.
6. Rogers FB, Rittenhouse KJ, Gross BW. The golden hour in trauma: dogma or medical folklore? Injury. 2015;46(4):525-7.
7. Montan KL. Triage. In: Lennquist S, ed. Medical response to major incidents and disasters, a practical guide for all medical staff. 1st ed. Springer-Verlag Berlin Heidelberg, 2012:63.
8. Streckbein S, Kohlmann T, Luxen J et al. Sichtungskonzepte bei Massenanfallen von Verletzten und Erkrankten&;: ein Uberblick 30 Jahre nach START. Unfallchirurg. 2016;119(8):620-31.
9. Kilner TM, Brace SJ, Cooke MW et al. In “big bang” major incidents do triage tools accurately predict clinical priority?: a systematic review of the literature. Injury. 2011;42(5):460-8.
10. Timbie JW, Ringel JS, Fox DS et al. Systematic review of strategies to manage and allocate scarce resources during mass casualty events. Ann Emerg Med. 2013;61(6):677-689.e101.
11. Marcussen CE, Bräuner KB, Alstrøm H et al. Accuracy of prehospital triage systems for mass casualty incidents in trauma register studies – a systematic review and meta-analysis of diagnostic test accuracy studies. Injury. 2022;53(8):2725-33.

12. Marcussen CE, Møller A, Alstrøm H et al. Accuracy of primary triage systems for mass casualty incidents in full-scale live simulations: a systematic review. www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=91889 (Aug 2023).
13. McInnes M, Moher D, Thombs B et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The prisma-DTA Statement. JAMA. 2018;319(4):388-96.
14. WHO. Mass casualty management systems. Strategies and guidelines for building health sector capacity. WHO, 2007. https:// apps.who.int/iris/bitstream/handle/10665/43804/9789241596053_eng.pdf?sequence=1&isAllowed=y (Aug 2023).
15. Whiting PF, Rutjes AWS, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
16. Sterne JA, Hernán MA, Reeves BC et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
17. 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.
18. Ingrassia PL, Ragazzoni L, Carenzo L et al. Virtual reality and live simulation: a comparison between two simulation tools for assessing mass casualty triage skills. Eur J Emerg Med. 2015;22(2):121-7.
19. 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):570-5.
20. Ellebrecht N, Latasch L. Paramedic triage during a mass casualty incident exercise. Notfall Rettungsmed. 2012;15(1):58-64.
21. Bolduc C, Maghraby N, Fok P et al. Comparison of electronic versus manual mass-casualty incident triage. Prehospital Disaster Med. 2018;33(3):273-8.
22. Price MF, Tortosa DE, Fernandez-Pacheco AN et al. Comparative study of a simulated incident with multiple victims and immersive virtual reality. Nurse Educ Today. 2018;71:48-53.
23. Fernandez-Pacheco AN, Delgado RC, Gonzalez PA et al. Analysis of performance and stress caused by a simulation of a mass casualty incident. Nurse Educ Today. 2018;62:52-7.
24. Jain T, Sibley A, Stryhn H et al. Comparison of unmanned aerial vehicle technology-assisted triage versus standard practice in triaging casualties by paramedic students in a mass-casualty incident scenario. Prehospital Disaster Med. 2018;33(4):375-80.
25. Lee CWC, McLeod SL, Van Aarsen K et al. First responder accuracy using SALT during mass-casualty incident simulation. Prehosp Disaster Med. 2016;31(2):150-4.
26. Cone DC, Serra J, Burns K et al. Pilot test of the SALT mass casualty triage system. Prehosp Emerg Care. 2009;13(4):536-40.
27. Lerner EB, Schwartz RB, Coule PL et al. Use of SALT triage in a simulated mass-casualty incident. Prehospital Emerg Care. 2010;14(1):21-5.
28. Cicero MX, Walsh B, Solad Y et al. Do you see what I see? Insights from using google glass for disaster telemedicine triage. Prehosp Disaster Med. 2015;30(1):4-8.
29. 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-Injury Infect Crit Care. 2010;69(1):215-25.
30. Offterdinger M, Ladehof K, Paul AO et al. Using a simple checklist in pretriage with the mSTaRT algorithm. First experiences in simulation training. Notfall Rettungsmed. 2014;17(5):415-9.
31. Rehn M, Andersen JE, Vigerust T et al. A concept for major incident triage: Full-scaled simulation feasibility study. BMC Emerg Med. 2010;10:17.
32. Kanz KG, Hornburger P, Kay M V et al. mSTaRT-Algorithmus für Sichtung, Behandlung und Transport bei einem Massenanfall von Verletzten. Notfall Rettungsmed. 2006;9(3):264-70.
33. 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):267-71.
34. Kissoon N, Dreyer J, Walia M. Pediatric trauma: differences in pathophysiology, injury patterns and treatment compared with adult trauma. CMAJ. 1990;142(1):27-34.
35. Neal DJ, Barbera JA, Harrald JR. Prehospital mass-casualty triage: a strategy for addressing unusual injury mechanisms. Sci Med. 2019;29.